



HAL
open science

Applying transcriptomic, metabolomic and high-throughput screening approaches to unravel the underlying mechanisms of somatic embryogenesis in *Coffea arabica*

Rayan Awada

► **To cite this version:**

Rayan Awada. Applying transcriptomic, metabolomic and high-throughput screening approaches to unravel the underlying mechanisms of somatic embryogenesis in *Coffea arabica*. *Vegetal Biology*. Université Montpellier, 2020. English. NNT : 2020MONTG005 . tel-04218314

HAL Id: tel-04218314

<https://theses.hal.science/tel-04218314>

Submitted on 26 Sep 2023

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.

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Biologie du Développement

École doctorale GAIA – Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau

Unité de recherche IPME

APPLYING TRANSCRIPTOMIC, METABOLOMIC AND HIGH- THROUGHPUT SCREENING APPROACHES TO UNRAVEL THE UNDERLYING MECHANISMS OF SOMATIC EMBRYOGENESIS IN *Coffea arabica*

Présentée par Rayan AWADA

Le 11 juin 2020

Sous la direction de Hervé ETIENNE
et Benoît BERTRAND

Devant le jury composé de

Philippe LABEL, Directeur de Recherche (HDR), INRAE Clermont-Ferrand

Rapporteur

Pierre COUTOS-THEVENOT, Professeur (HDR), Université de Poitiers

Rapporteur

Laurent LEGENDRE, Professeur (HDR), Université de Lyon 1

Examineur

Soazig GUYOMARC'H, Maître de conférences (HDR), Université de Montpellier

Examineur

Pierre CZERNIC, Professeur (HDR), Université de Montpellier

Président du jury

Hervé ETIENNE, Chercheur (HDR), CIRAD Montpellier

Directeur de thèse

Benoît BERTRAND, Chercheur (HDR), CIRAD Montpellier

Invité, co-directeur

David BRETON, Chef de projet, Nestlé Tours

Invité, encadrant



UNIVERSITÉ
DE MONTPELLIER



Nestlé

Good food, Good life

“I was just guessing
At numbers and figures
Pulling your puzzles apart

Questions of science
Science and progress
Do speak as loud as my heart”

- *The Scientist*, Coldplay

Remerciements

Je tiens tout d'abord à remercier Nestlé Research et la direction du département Bios du CIRAD qui ont financé cette thèse.

Je souhaite remercier Fabrizio Arigoni, responsable de l'unité Plant Science au sein de Nestlé Research pour m'avoir accueilli au sein du laboratoire de Tours. Je remercie également Gilles Béna, directeur de l'unité IPME pour m'avoir accueilli au sein de cette unité de recherche. Je tiens aussi à remercier Simone de Faria Maraschin qui m'a accordé l'opportunité de faire cette thèse après mon stage de Master 2.

Je tiens à exprimer ma profonde gratitude envers Hervé Etienne et Benoit Bertrand, mes directeurs de thèse, ainsi que David Breton, mon encadrant de thèse, pour m'avoir fait découvrir le monde de la recherche fondamentale et appliquée. Merci pour la qualité de l'enseignement qu'ils m'ont apportée, pour leur encadrement déterminé, pour leur patience ainsi que pour leur disponibilité.

Je tiens à remercier Philippe Label et Pierre Coutos-Thévenot pour avoir accepté la tâche d'être rapporteurs de ma thèse. Je remercie également Soazig Guyomarc'h et Laurent Legendre pour avoir accepté d'être les examinateurs. Un merci particulier à Pierre Czernic, qui a répondu présent pour être président du jury.

Un grand merci à Dominique Crouzillat et Maud Lepelley qui ont toujours répondu présents pour s'assurer du succès de la partie transcriptomique. Un grand merci à l'équipe de Patrick Descombes au Nestlé Research, Lausanne, Suisse pour avoir mis à notre disposition leur plateforme de séquençage, et au consortium ACGC pour avoir donné accès à leur génome de référence. Un grand merci également à la plateforme PHIV où j'ai pu réaliser mes études d'imagerie cellulaire.

Un grand merci à Laurent Legendre qui m'a permis d'établir une collaboration avec le Max Planck Potsdam en Allemagne et où j'ai pu passer 2 mois très enrichissants durant ma thèse. Un grand merci à Joachim Kopka qui m'a accueilli au sein de son équipe.

Merci à l'école doctorale GAIA toujours présente pour répondre aux questions de ses doctorants, ainsi qu'au Collège Doctoral pour la qualité des formations proposées. Un merci particulier à Jean-Luc Verdeil, mon référent de thèse, qui a répondu présent aux différents comités de suivi de thèse où il m'a incité à poser les bonnes questions de recherche.

Je tiens à remercier tous les membres qui ont assisté à mes différents comités de suivi de thèse et qui ont permis de mener ce projet à bon port, à savoir : Marie-Anne Lelu-Walter, Jean-François Trontin, Laurent Legendre, Pierre Czernic, Claudine Campa, Sophie Léran, Maud Lepelley, David Breton, Benoit Bertrand et Hervé Etienne.

Un merci particulier à Jean-François Trontin qui m'a incité à participer à la 5^e édition du congrès IUFRO qui s'est tenu au Portugal et où j'ai pu rencontrer la communauté scientifique qui s'intéresse aux méthodes de micropropagation des arbres et plus particulièrement à l'embryogenèse somatique.

Evidemment un grand merci à tous les membres du personnel de l'unité Plant Science de Nestlé Research où j'ai passé ma première année de thèse. Un merci particulier à Jean-Paul Ducos, Audrey Fillodeau, Caroline Guillou, Dorothée Verdier, Eric Brulard, Solène Froger, Estelle Gibault, Mathieu Simon et Alexandre Talaron que j'ai côtoyés quotidiennement. Merci pour tout ce que vous m'avez appris.

Mes remerciements vont également à toute l'équipe CoffeeAdapt basée à l'IRD et qui m'a accueilli jusqu'à la fin de ma thèse. Un merci particulier à Claudine Campa et Cécile Abdallah qui m'ont beaucoup aidé sur la partie métabolisme secondaire, à Eveline Déchamp qui m'a toujours accompagné dans le labo de culture *in vitro*, à Isabelle Héroult qui a géré toute la logistique derrière ce projet de thèse. Je remercie également Fred Georget, Pierre Marraccini, Sophie Léran, Jean-Christophe Breitler, Luc Villain, Eric Lacombe et Delphine Mieulet pour tous leurs conseils scientifiques.

Je remercie aussi mes collègues de bureau : Adrian, Ngan, Doaa, Trixie, Sondo ainsi que tous les jeunes scientifiques de l'IRD qui étaient toujours là pour partager des conseils ou pendant les moments de détente. Un merci particulier à mon frère Jad et mes amis Ali, Zivile, Ahmad, Haitham et Nathalie pour les belles soirées estivales passées à Montpellier.

Et enfin, un grand MERCI à mes parents qui étaient toujours là pour m'apporter le réconfort dont j'avais besoin.

Résumé général

INTRODUCTION

L'amélioration génétique du caféier Arabica (allotétraploïde, autogame) est rendue très difficile en raison d'une diversité génétique très étroite et de la pérennité de la plante (durée des cycles biologiques supérieure à cinq ans). Traditionnellement, l'amélioration génétique est basée sur une sélection généalogique qui se traduit par la sélection de lignées pures (homozygotes) propagées par les semences. Les gains génétiques avec de tels programmes sont lents et à faible valeur ajoutée. Au cours des années 1990 en Amérique centrale, le CIRAD et ses partenaires du réseau de recherche centraméricain PROMECAFE ont développé des hybrides F1 d'Arabica caractérisés par leur productivité élevée, leur tolérance aux stress biotiques et abiotiques et leur excellente qualité de tasse. Compte tenu de la nature hétérozygote du matériel sélectionné, il a été nécessaire de développer une technique de multiplication végétative (clonage) par embryogenèse somatique (ES) pour permettre une dissémination rapide et à grande échelle de ces hybrides. Après 6-8 ans d'essais clonaux multi-locaux dans cinq pays d'Amérique centrale, certains clones d'hybrides résultant du croisement entre des variétés américaines cultivées et des individus « sauvages » d'Éthiopie ont été sélectionnés comme meilleurs candidats en vue d'une diffusion commerciale. Un accord de partenariat a été signé en 2002 entre une entreprise privée (ECOM) et le CIRAD et a permis la montée en puissance de la technologie, des productions pilotes et enfin le transfert industriel du procédé ES avec la création de deux laboratoires commerciaux au Nicaragua et au Mexique. Pendant ce temps, Nestlé, leader de l'industrie agroalimentaire et acteur majeur de la filière café, dirigeait des recherches sur l'ES en réalisant une propagation à grande échelle de clones Robusta en Thaïlande et aux Philippines, sélectionnés afin de sécuriser sa chaîne d'approvisionnement ; et débutait un programme de création variétale d'Arabica basé sur des hybrides F1 propagés par l'embryogenèse somatique.

Les équipes de Nestlé et celles du CIRAD partagent la même conviction : dans un secteur café peu innovant, il est nécessaire de proposer des nouvelles variétés pour augmenter à la fois la productivité, les résistances aux maladies mais également garantir des qualités physiques et sensorielles adaptées à l'industrie. Mais pour cela, il faut nécessairement lever les verrous liés à la micropropagation. En effet, la demande des producteurs pour les variétés élites n'est que de 2 à 4 millions de plantes par an du fait de prix de vente jugés trop élevés (une plante issue de l'*in vitro* coûte de 2 à 3 fois le prix d'une plante issue de graines). Comment fournir 50 à 100 millions de plantes par an ? L'une des clés de réussite est l'amélioration de l'ES. Ce processus de régénération qui a été mis en évidence il y a soixante ans chez la carotte, fait encore face à de nombreux défis technologiques et scientifiques lors de son utilisation à grande échelle. Quelle que soit l'espèce considérée, les études sur l'ES ont été principalement empiriques et à faible débit, axées sur l'optimisation de la production d'embryons somatiques bien plus que sur la compréhension des mécanismes sous-jacents à ce phénomène biologique qui reste encore bien mystérieux. Cela s'est traduit par un progrès technique très lent au cours des dernières années, qui continue d'empêcher le développement industriel de la technologie ES.

Bien que l'ES ait été décrite comme un outil très prometteur pour la diffusion rapide et à grande échelle de variétés élites dans de nombreuses espèces, ses coûts de production restent élevés principalement en raison de la faible efficacité biologique et des effets génotypiques élevés. Quelle que soit l'espèce végétale, un certain nombre de goulots d'étranglement communs ont été notés tout au long des trois principales étapes communes aux processus d'ES actuels : 1) l'induction du cal embryogène, 2) sa prolifération et 3) sa différenciation en embryons. Par exemple, un fort effet génotypique a souvent été remarqué lors de la première étape. Les explants de certains clones intéressants générés par des programmes de sélection ne répondent pas au stimulus auxinique (récalcitrance), ou répondent en produisant directement mais à faible fréquence des embryons sans passer par une étape de cals empêchant ainsi la possibilité d'une propagation à grande échelle. Pour la deuxième étape, les cals de certains clones perdent leur potentiel embryogène pendant la prolifération. Pour la troisième étape,

maîtriser la qualité des embryons régénérés et maximiser les rendements embryonnaires est une nécessité pour réduire les coûts de production.

Cette thèse représente un travail pionnier où les deux principaux groupes travaillant sur l'ES du caféier ont décidé d'unir leurs forces pour clarifier les principaux goulots d'étranglement du processus en utilisant des technologies de pointe, afin de comprendre les mécanismes moléculaires sous-jacents lors des étapes clés. La mise en œuvre de cette recherche fondamentale a été rendue possible par le fait que l'espèce de caféier Arabica bénéficie d'un des processus d'ES industriels les plus avancés et les plus fiables au monde. Un autre objectif, appliqué celui-ci, est visé : il s'agit de passer d'une recherche empirique et à bas-débit qui consiste à améliorer le procédé par tâtonnements successifs à une recherche rationnelle, holistique et à haut-débit, basée sur des indicateurs biologiques fiables. Les deux objectifs visés vont au-delà du caféier et devraient intéresser la communauté scientifique et industrielle travaillant sur l'ES.

Dans le premier chapitre du manuscrit, nous présentons une synthèse bibliographique visant à apporter des connaissances générales sur l'embryogenèse somatique, l'état de l'art de cette technologie sur le caféier, les enjeux de son optimisation et l'émergence des outils « omiques ». Cette partie nous permettra également de présenter nos principaux objectifs et hypothèses qui seront détaillés plus loin dans les trois chapitres.

Le chapitre II présentera comment nous avons développé, pour la première fois en micropropagation des plantes, un système de criblage miniaturisé et automatisé capable de répondre aux normes de haut débit afin d'optimiser rapidement les protocoles d'ES en testant, dans un délai limité, une large gamme de composés actifs (nutriments, régulateurs de croissance, etc.), multipliant les conditions expérimentales et réduisant les volumes ainsi que l'espace et la main-d'œuvre nécessaires.

Les chapitres III et IV montreront comment nous avons mis à profit les derniers outils métabolomiques et transcriptomiques et les avons appliqués au procédé d'ES du caféier, afin de démêler les mécanismes moléculaires globaux sous-jacents, de la dédifférenciation cellulaire jusqu'à la redifférenciation cellulaire, ce qui nous a permis d'identifier des gènes et des métabolites candidats capables de piloter l'optimisation de ce procédé complexe.

Enfin, le dernier chapitre fera l'objet d'une discussion sur l'ensemble des résultats obtenus et d'une conclusion générale ouvrant sur plusieurs perspectives à partir du présent travail de thèse.

RESULTATS

- Dans le premier chapitre du manuscrit, et plus particulièrement dans la revue que nous avons publiée dans le journal *Frontiers in Plant Science*, nous décrivons comment les recherches, menées jusque-là de façon empirique, ont conduit à des transferts industriels réussis lancés dans les années 2000 en Amérique latine, en Afrique et en Asie. L'ES du caféier peut être considérée comme un modèle pour d'autres cultures pérennes ligneuses pour les raisons suivantes : (i) une efficacité biologique acceptable a été démontrée pour les variétés multipliées à tous les stades de développement, et (ii) la variation somaclonale est maîtrisée grâce à des recherches intensives combinant des marqueurs moléculaires et des observations sur le terrain. L'ES du caféier est également un modèle utile étant donné les fortes contraintes économiques spécifiques à ces deux espèces. Brièvement, l'ES a rencontré quatre difficultés : (i) le coût élevé des plants dérivés de cette technologie par rapport au coût des plants propagés par semences, (ii) les problèmes logistiques liés à l'accès aux petits caféiculteurs, (iii) la nécessité d'une certification, et (iv) le manque de solvabilité des petits producteurs. Les activités en pépinière ont été professionnalisées en introduisant la certification variétale, le contrôle de la qualité en ce qui concerne les problèmes horticoles et les variations somaclonales, et le contrôle

sanitaire concernant *Xylella fastidiosa*. En outre, différents transferts technologiques ont été effectués pour assurer la diffusion mondiale des hybrides F1 Arabica améliorés et des clones Robusta. Les innovations ont été décisives pour une montée en puissance de production réussie, telles que le développement de bioréacteurs à immersion temporaire pour la production en masse d'embryons pré-germés, leur semis direct sur substrat horticole et la propagation de plantules issues d'ES et ainsi rejuvénilisées par la technique horticole de mini-boutures enracinées. Aujourd'hui, l'ES est un outil puissant qui est largement utilisé chez le caféier pour des applications biotechnologiques, y compris la propagation et la transformation génétique. Des recherches fondamentales ont tout récemment pu s'appliquer sur ces procédés d'ES industriels optimisés et bénéficiant d'une forte reproductibilité biologique. Basée sur des outils -omiques, les recherches visent à déchiffrer les événements moléculaires impliqués dans le succès des transitions entre les stades de développement clés de l'ES du caféier. En parallèle, le développement d'un outil de criblage à haut débit de molécules actives sur l'ES est plein de promesses pour permettre d'accélérer l'optimisation des protocoles.

- Le chapitre II correspond à un article de recherche méthodologique publié dans *Scientific Reports*. Nous montrons dans cet article comment une approche à haut débit pourrait être une solution pour optimiser les protocoles d'ES en testant une plus large gamme de composés actifs dans des systèmes de culture miniaturisés et robotisés. Nous présentons dans ce chapitre un système de criblage automatisé à haut débit capable d'accélérer les optimisations du procédé ES chez *C. arabica*. Notre objectif était de produire des embryons à partir de suspensions cellulaires embryogènes établies. Cette étape de l'ES étant classiquement effectuée dans des Erlenmeyers de 250 mL, une miniaturisation dans des plaques à 24-puits était donc nécessaire. La granulométrie des suspensions cellulaires a également été miniaturisée afin de l'adapter à la plateforme de pipetage. La plateforme automatisée a ensuite été programmée de manière à

valider l'homogénéité de la distribution de la suspension ainsi que la conversion cals-embryons. Après une mise en place réussie du système, une preuve de concept a été réalisée. Quatre composés (Oxamflatine, SAHA, Scriptaid et trichostatine A) appartenant à la famille des inhibiteurs d'histone désacétylase ont été testés à trois concentrations différentes : 0,1, 1 et 10 μM , en tant que composés potentiellement actifs sur la différenciation des embryons de *C. arabica*. Seuls les cals traités avec 1 μM de trichostatine A ont montré une augmentation significative (multiplication par 3) de la production d'embryons au torpille. En guise de validation, ce traitement a été re-testé dans des conditions standards (Erlenmeyers de 250 ml) où il a également permis une augmentation significative du nombre d'embryons obtenus (multiplication par 3,5), de leur taille (30%) et du poids frais (60 %). La trichostatine A, qui joue sur les modifications épigénétiques, n'a pas eu d'effet sur le phénotype des plantules lors de la conversion embryon-plantule. Nos résultats montrent que ce système peut être adapté pour cribler des milliers de composés dans une période de temps restreinte. Une analyse plus approfondie doit être effectuée pour comprendre l'effet moléculaire de la trichostatine A sur la stimulation de la production d'embryons.

- Même si l'approche développée dans le chapitre II est une approche qui peut conduire à une accélération de l'optimisation des milieux de culture, elle demeure toutefois empirique. En utilisant le caféier comme modèle, nous rapportons dans le chapitre III (article publié dans *International Journal of Molecular Sciences*) la première analyse globale de la dynamique du métabolome et des contenus hormonaux aux cours de l'embryogenèse somatique, avec l'objectif de mettre en œuvre une approche plus rationnelle de l'optimisation de l'ES en démêlant les mécanismes régulant le destin cellulaire et la totipotence. L'échantillonnage à partir de la dédifférenciation des explants de feuilles jusqu'au développement de l'embryon a couvert 15 étapes clés. Une analyse statistique approfondie réalisée sur 104 métabolites a révélé

qu'une reconfiguration massive des voies métaboliques induisait l'ES. Au cours de la dédifférenciation initiale, une forte diminution des composés phénoliques et des niveaux de caféine a également été observée tandis que les teneurs en auxine, cytokinine et éthylène étaient à leur maximum. La totipotence a atteint son expression la plus élevée au cours des stades de cal lorsqu'une inhibition des voies hormonales et métaboliques liées à l'hydrolyse du sucre et des substances énergétiques qui a été mise en évidence. L'Acide abscissique, la Leucine, le Maltotriose, le Myo-inositol, la Proline, les métabolites du cycle de l'acide tricarboxylique et la Zéatine sont apparus comme des marqueurs métaboliques clés du potentiel embryogène. De plus, la combinaison de la métabolomique avec la microscopie multiphotonique a permis d'identifier les acides chlorogéniques comme marqueurs de la redifférenciation embryonnaire. Cette étude montre que les profils métaboliques sont des signatures du destin cellulaire et représentent un point de départ pour optimiser les protocoles ES de manière rationnelle.

- Tout comme pour l'approche métabolomique, nous rapportons dans le chapitre IV (article prêt à soumettre à *BMC Genomics*) le premier profilage de transcriptome effectué sur l'ensemble du procédé d'ES chez le caféier, également pour démêler les mécanismes régulant le destin cellulaire et la totipotence. L'ARN-seq de 48 échantillons (12 stades de développement x 4 répétitions biologiques) a généré un total de 90 millions de lectures de haute qualité par échantillon, dont environ 74% étaient uniquement 'mappées' sur le génome d'*Arabica*. Premièrement, une approche statistique des données de transcription a clairement regroupé les stades de développement de l'ES en sept phases distinctes (feuille, dédifférenciation, cal primaire, cal embryonnaire, suspensions cellulaires embryogènes, redifférenciation et embryon) conduisant à l'identification de six transitions ('switches') de développement clés, qui sont stratégiques pour l'efficacité biologique globale de la régénération par ES. L'expression différentielle des gènes et l'analyse fonctionnelle ont montré que les transcrits des gènes codant pour les facteurs de transcription, ainsi que ceux des gènes liés au stress au

métabolisme et à la signalisation hormonale étaient considérablement enrichis. Deuxièmement, l'influence des facteurs environnementaux classiquement utilisés pour maîtriser l'ES, à savoir la lumière, les régulateurs de croissance et la densité cellulaire, est clairement perceptible au niveau moléculaire à différents stades de développement. Troisièmement, les profils d'expression des gènes liés à l'auxine, aux facteurs de transcription et au métabolisme secondaire ont été analysés au cours de l'ES. Des réseaux de co-expression de gènes ont également été déduits. Les gènes liés à l'auxine ont été surexprimés pendant la dédifférenciation et la redifférenciation tandis que les gènes liés aux facteurs de transcription ont été activés à partir du stade 'cal embryogène' et des stades ultérieurs. L'expression des gènes liés au métabolisme secondaire a été inhibée lors de la dédifférenciation cellulaire puis progressivement réactivée dès le début de l'étape de redifférenciation. Les teneurs en métabolites secondaires et en AIA endogène sont fortement liés à l'expression des gènes correspondants. Enfin, la comparaison des transcriptomes des cellules embryogènes et non embryogènes d'*Arabica* a permis d'identifier les processus biologiques impliqués dans l'acquisition de la capacité embryogène. La présente analyse a montré que les profils transcriptomiques sont également des signatures discriminantes du destin cellulaire tout comme les profils métabolomiques, et sont sous l'influence directe des facteurs environnementaux. Au total, 23 marqueurs moléculaires ont pu être identifiés au cours des 12 stades de développement jalonnant l'ES. Ils doivent maintenant être testés chez d'autres génotypes de caféier plus ou moins récalcitrants et chez d'autres espèces végétales, puisque les mécanismes mis en évidence dans lesquels ils interviennent pourraient être conservés pour optimiser les protocoles de manière rationnelle.

CONCLUSION

Nous avons réussi à apporter la preuve que les étapes de l'ES peuvent non seulement être caractérisées au niveau phénotypique (macroscopique et cellulaire) mais aussi au niveau moléculaire grâce aux profils métaboliques et transcriptomiques. Ces profils reflètent des changements dramatiques qui se manifestent lors des transitions de destin cellulaire et sont étroitement liés aux conditions environnementales (hormones exogènes, densité cellulaire, lumière). Ainsi, en recherchant des marqueurs moléculaires, le processus d'ES pourrait être piloté d'une façon rationnelle et nous ouvre la voie, pour la première fois, sur de nouvelles perspectives après de nombreuses années marquées par un très faible progrès scientifique. Un nombre de gènes et de métabolites candidats ont pu être identifiés et sont en attente de validation. L'étape suivante serait de valider les gènes candidats en les testant sur un panel de génotypes plus ou moins récalcitrants (dont on connaît les efficacités biologiques aux différentes étapes de l'ES) en réalisant des tests de RT-qPCR. Il sera ainsi possible de vérifier la corrélation existante entre le niveau d'expression des gènes marqueurs et l'efficacité biologique au niveau du switch de développement qu'ils doivent piloter. Quelques métabolites candidats identifiés comme marqueurs du potentiel embryogène (stade cal embryogène) entrent déjà dans la composition des milieux utilisés dans les protocoles d'ES industriels ; c'est le cas du myo-inositol et le maltose par exemple. Les concentrations de ces composés peuvent donc être directement modifiées dans la composition du milieu afin d'évaluer leur impact sur le potentiel embryogène, qui représente jusque-là l'étape la plus limitante de l'ES du caféier. Une fois les marqueurs transcriptomiques et métabolomiques validés, la démarche haut-débit peut être mise en œuvre afin d'accélérer l'optimisation des protocoles d'ES en venant tester simultanément des centaines de composés intervenant dans les mêmes voies métaboliques afin de sélectionner ceux qui sont capables de moduler les niveaux d'expression de ces marqueurs transcriptomiques ou les concentrations des marqueurs métaboliques. Nous avons réalisé la preuve de concept de la démarche haut-débit sur une échelle pilote, l'étape suivante sera de réaliser cette démarche sur une échelle industrielle qui

nécessitera certainement de nouveaux équipements et de nouvelles pratiques de laboratoire. L'optimisation des protocoles d'ES pourra ainsi venir accompagner les programmes de sélection variétale pour diffuser à grande échelle le progrès génétique afin de soutenir les caféiculteurs face aux défis issu du changement climatique. Nos données montrent aussi que les mécanismes régulant l'ES sont conservées entre espèces et pourraient donc être utiles à toute la communauté scientifique.

Published papers and communications

Published papers

Awada R, Verdier D, Froger S, Brulard E, de Faria Maraschin S, Etienne H and Breton D (2020) An innovative automated active compound screening system allows high-throughput optimization of somatic embryogenesis in *Coffea arabica*. *Sci Rep* 10:810. doi: 10.1038/s41598-020-57800-6

<https://www.nature.com/articles/s41598-020-57800-6>

Awada R, Campa C, Gibault E, Déchamp E, Georget F, Lepelley M, Abdallah C, Erban A, Martinez-Seidel F, Kopka J, Legendre L, Lérant S, Conéjéro G, Verdeil J-L, Crouzillat D, Breton D, Bertrand B and Etienne H (2019) Unravelling the Metabolic and Hormonal Machinery During Key Steps of Somatic Embryogenesis: A Case Study in Coffee. *Int. J. Mol. Sci.* 20:4665. doi: 10.3390/ijms20194665.

<https://www.mdpi.com/1422-0067/20/19/4665/htm>

Etienne H, Breton D, Breitler J-C, Bertrand B, Déchamp E, **Awada R**, Marraccini P, Lérant S, Alpizar E, Campa C, Courtel P, Georget F and Ducos J-P (2018) Coffee Somatic Embryogenesis: How Did Research, Experience Gained and Innovations Promote the Commercial Propagation of Elite Clones From the Two Cultivated Species? *Front. Plant Sci.* 9:1630. doi: 10.3389/fpls.2018.01630

<https://www.frontiersin.org/articles/10.3389/fpls.2018.01630/full>

To be submitted

Awada R, Lepelley M, Charpagne A, Berry V, Campa C, Breitler J-C, Lérant S, Breton D, Crouzillat D, Bertrand B and Etienne H (2020) Global transcriptome profiling reveals differential regulatory, metabolic and hormonal networks during somatic embryogenesis in *Coffea arabica*. *BMC Genomics*.

Oral communications

Awada R, Campa C, Gibault E, Déchamp E, Georget F, Lepelley M, Abdallah C, Erban A, Martinez-Seidel F, Kopka J, Legendre L, Lérant S, Conéjéro G, Verdeil J-L, Crouzillat D, Breton D, Bertrand B and Etienne H (2020) Unravelling the Metabolic and Hormonal Machinery During Key Steps of Somatic Embryogenesis: A Case Study in Coffee (oral communication). *In*: 28th International Conference of ASIC, Montpellier, Jun. 22-25.

Awada R, Verdier D, Brulard E, Froger S, de Faria Maraschin S, Etienne H and Breton D (2018) Automated active compounds screening system allows high-throughput optimization of somatic embryogenesis in *Coffea arabica* (oral communication). *In*: 5th International Conference of the IUFRO Working Party 2.09.02, Coimbra, Sept. 10-15.

Table of Contents

Chapter I. Literature Review	25
Part I. Generalities about coffee	26
1. Coffee cultivation, consumption and economic importance in producing countries	26
2. Coffee taxonomy	28
3. Botanic description of <i>C. arabica</i>	29
4. Arabica genome characteristics	30
5. Genetic diversity of <i>C. arabica</i>	31
Part II. Coffee breeding	33
1. <i>C. arabica</i> breeding programs prior to 1990s	33
2. <i>C. arabica</i> breeding programs from 1990s until today	35
Part III. Vegetative propagation techniques	38
1. Seed propagation vs. vegetative propagation	38
2. Traditional vegetative propagation techniques	38
3. Micropropagation	40
4. Somatic embryogenesis and totipotency	43
5. Somatic embryogenesis vs. zygotic embryogenesis	44
6. Direct and indirect somatic embryogenesis	45
7. Factors influencing the success of SE	46
Part IV. Coffee somatic embryogenesis	50
1. Sexual and vegetative techniques for coffee propagation	50
2. Coffee somatic embryogenesis: How did research, experience gained and innovations promote the commercial propagation of elite clones from the two cultivated species?	50
Part V. New tools to understand SE underlying mechanisms	51
1. SE description based on morphological studies in the 1990s	51
2. Development of “omics” techniques in the 2000s-2010s	51
3. Molecular insights gained in SE	53
4. Achievements in coffee SE	57
5. High-throughput chemical screening	58
Part VI. Challenges limiting large-scale dissemination of the genetic progress by somatic embryogenesis in the cultivated coffee species	60
Part VII. Thesis general objectives	62

Chapter II. An innovative automated active compound screening system allows high-throughput optimization of somatic embryogenesis in <i>Coffea arabica</i>	63
Background	64
Article	64
Chapter III. Unravelling the Metabolic and Hormonal Machinery During Key Steps of Somatic Embryogenesis: A Case Study in Coffee	65
Background	66
Article	66
Chapter IV. Global transcriptome profiling reveals differential regulatory, metabolic and hormonal networks during somatic embryogenesis in <i>Coffea arabica</i>	67
Background	68
Article	68
Chapter V. General discussion, conclusion and perspectives	69
General discussion	70
1. The key strengths of this thesis	71
2. The key limitations of this thesis	73
3. How did the “omics” tools contribute to the characterization of cell fate?	74
4. Identification of transcriptomic and metabolic markers during the key developmental phases of somatic embryogenesis	77
5. Is the high-throughput optimization of somatic embryogenesis still a dream for biotechnologists?	80
6. How can cell totipotency be defined?	81
7. An urgency to optimize somatic embryogenesis in order to disseminate the genetic progress of varieties resulting from breeding programs	82
Conclusion	85
Perspectives	87
References	91
Annexes	103
Résumé - Abstract	104

List of figures

Figure 1. Coffee cultivation areas in the bean belt	27
Figure 2. Coffee per-capita consumption across markets in 2018	27
Figure 3. Botanic description of <i>C. arabica</i>	29
Figure 4. Coffee berry and bean morphology	30
Figure 5. Genetic origin of allotetraploid <i>C. arabica</i>	31
Figure 6. Centers of genetic diversity of <i>C. arabica</i>	32
Figure 7. <i>C. arabica</i> breeding programs prior to 1990s and until today.....	34
Figure 8. Symptoms caused by leaf rust disease on <i>C. arabica</i>	35
Figure 9. Comparison between <i>C. arabica</i> cv Caturra and a clone of F1 hybrids after 3 years since germination	37
Figure 10. Difference between coffee cultivated in full-sun and in agroforestry systems	37
Figure 11. The process of horticultural cutting	39
Figure 12. The process of horticultural grafting	40
Figure 13. The process of micro-cuttings	41
Figure 14. The process of organogenesis	42
Figure 15. The indirect somatic embryogenesis (SE) process as first described in <i>Daucus carota</i> ..	43
Figure 16. Comparison between somatic and zygotic embryogenesis	45
Figure 17. Comparison of forward and reverse chemical screening.....	59
Figure 18. Schematic representation of the research pipeline leading to the scaling-up of the Arabica somatic embryogenesis process	80
Figure 19. Representation of the objectives set to reach a rational and accelerated optimization of SE protocols.....	86
Figure 20. A representation of the stem cell differentiation process	90

List of tables

Table 1. Composition of the different media used for the different stages of somatic embryogenesis in coffee in mg/l and μM (growth regulators) (from Etienne et al., 2005)	48
Table 2. Up-regulated transcriptomic and over-accumulated metabolic markers characterizing the different developmental phases of coffee somatic embryogenesis.....	79

Abbreviations

2,4-D	2,4-Dichlorophenoxyacetic acid
2-iP	N ⁶ -(2-Isopentenyl) adenine
4CL	4-CINNAMOYL COA LIGASE
6-BA	6-Benzylaminopurine
ABA	<i>cis</i> -Abscisic acid
ABAGE	ABA glucose ester
ABI	ABSCISIC ACID INSENSITIVE
ACC	1-Aminocyclopropane-1-carboxylic acid
ACS	1-AMINOACYCLOPROPANE 1-CARBOXYLATE SYNTHASE
AGL	AGAMOUS-LIKE
ARF	AUXIN RESPONSE FACTOR
ARR	ARABIDOPSIS RESPONSE REGULATOR
BBM	BABY BOOM
BSTFA	N,O-Bis(trimethylsilyl)trifluoroacetamide
CHS	CHALCONE SYNTHASE
CLV	CLAVATA
COMT	CAFFEIC ACID O-METHYLTRANSFERASE
CQA	Caffeoylquinic acid
DAD	Photodiode array detector
DEG	Differentially expressed genes
DW	Dry weight
DXMT	CAFFEINE SYNTHASE
EC	Embryogenic callus
ERF	ETHYLENE RESPONSE FACTOR
FQA	Feruloylquinic acid
FUS	FUSCA
GC-EI-TOF/MS	Gas chromatography-electron impact ionization-time of flight/mass spectrometry
GC-MS	Gas chromatography-mass spectrometry
HPLC	High-performance liquid chromatography
HQT	HYDROXYCINNAMOYL-COA QUINATE HYDROXYCINNAMOYL TRANSFERASE
IAA	Indole-3-acetic acid
IBA	Indole 3-butyric acid
iP	Isopentenyladenine
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC-MS	Liquid chromatography-mass spectrometry
LEC	LEAFY COTYLEDON
MeOH	Methanol
MS	Murashige and Skoog
N/C ratio	Nucleocytoplasmic ratio
NEC	Non-embryogenic callus
PAL	PHENYLALANINE AMMONIA LYASE
PCA	Principal Component Analysis
PEMs	Pro-embryogenic masses
PIN	PIN-FORMED

<i>PYR</i>	PYRABACTIN RESISTANCE
<i>SE</i>	Somatic embryogenesis
<i>SERK</i>	SOMATIC EMBRYOGENESIS RECEPTOR KINASE
<i>TAA</i>	TRYPTOPHAN AMINOTRANSFERASE
<i>TCA</i>	Tricarboxylic acid
<i>TF</i>	Transcription factor
<i>UPLC-ESI-MS/MS</i>	Ultra-performance liquid chromatography-electrospray tandem mass spectrometry
<i>W/L ratio</i>	Width-to-length ratio
<i>WOX</i>	WUSCHEL-RELATED HOMEODOMAIN PROTEIN
<i>WUS</i>	WUSCHEL
<i>XMT</i>	XANTHOSINE METHYLTRANSFERASE
<i>Z</i>	Zeatin (<i>cis</i> - and <i>trans</i> -)
<i>ZR</i>	Zeatin riboside (<i>cis</i> - and <i>trans</i> -)

Preface

The genetic improvement of the Arabica coffee tree (allotetraploid, autogamous) is made very difficult due to a very narrow genetic diversity and the perennial nature of the plant (duration of biological cycles greater than five years). Traditionally, genetic improvement is based on genealogical selection which results in the selection of pure (homozygous) lines propagated by seeds. Genetic gains with such programs are slow and of low added value. During the 1990s in Central America, CIRAD and its partners in the Central American research network PROMECAFE developed Arabica F1 hybrids characterized by their high productivity, their tolerance to biotic and abiotic stresses and their excellent cup quality. Given the heterozygous nature of the material selected, it was necessary to develop a technique of vegetative multiplication (cloning) by somatic embryogenesis (SE) to allow rapid and large-scale dissemination of these hybrids. After 6-8 years of multi-local clonal trials in five Central American countries, some hybrid clones resulting from the cross between cultivated American varieties and "wild" individuals from Ethiopia have been selected as the best candidates for commercial distribution. A partnership agreement was signed in 2002 between a private company (ECOM) and CIRAD and enabled the scale-up of the SE technology, pilot productions and finally the industrial transfer of the SE process with the establishment of two commercial laboratories in Nicaragua and Mexico. Meanwhile, Nestlé, leader in the food industry and a major stakeholder in the coffee sector, was leading the research on SE through achieving large-scale propagation of Robusta clones in Thailand and the Philippines in order to secure its supply chain; and was starting an Arabica breeding program based on F1 hybrids propagated by somatic embryogenesis.

The Nestlé and CIRAD teams share the same conviction: in a lowly innovative coffee sector, it is necessary to offer new varieties to increase both productivity and resistance to disease, but also to guarantee physical and sensory qualities adapted to the industry. But for this, it is necessary to resolve the bottlenecks related to micropropagation. In fact, producers demand for elite varieties is only 2 to 4 million plants per year due to selling prices deemed too high (a plant produced *in vitro* costs 2 to 3 times the price of a seedling). How to provide 50 to 100 million plants per year? One of the keys to success is improving SE. This regeneration process that was highlighted 60 years ago in carrot, still faces many technological and scientific challenges when used on a large scale. Whatever the species considered, studies on SE have been mainly empirical and at a low-throughput level, focused on optimizing the regeneration of somatic embryos much more than on understanding the mechanisms underlying this biological phenomenon which is still very mysterious. This has resulted in very slow

technical progress in recent years, which continues to prevent the industrial development of SE technology.

Although SE has been described as a very promising tool for the rapid and large-scale dissemination of elite varieties in many species, its production costs remain high mainly due to low biological efficiency and high genotypic effects. Whatever the plant species, a number of common bottlenecks have been noted throughout the three main steps common to current SE processes: 1) induction of the embryogenic callus, 2) its proliferation and 3) its differentiation into embryos. For example, a strong genotypic effect was often noticed during the first step. The explants of certain interesting clones generated by breeding programs do not respond to the auxin stimulus (recalcitrance), or respond by producing embryos directly at low frequency without going through a callus stage, thus preventing the possibility of large-scale propagation. For the second step, calli of some clones lose their embryogenic potential during proliferation. For the third step, controlling the quality of the regenerated embryos and maximizing the embryonic yields is a necessity to reduce production costs.

This thesis represents a pioneering work where the two main groups working on coffee SE decided to join forces to clarify the main bottlenecks of the process using the most advanced technologies, in order to understand the molecular mechanisms underlying the key developmental stages. The establishment of this fundamental research was made possible by the fact that the Arabica coffee species benefits from one of the most advanced and reliable industrial SE processes in the world. One applied objective is moving SE research from an empirical and low-throughput approach i.e. consisting in improving the process by successive trial-and-error approaches, towards a more rational, holistic and high-throughput research, based on reliable biological indicators. The two targeted objectives go beyond coffee and should interest the scientific and industrial community working on SE.

In the first chapter of the manuscript, we present a literature review aiming to provide general knowledge on somatic embryogenesis, the state of the art of this technology on coffee, the challenges facing its optimization and the emergence of the “omics” tools. This chapter will also allow us to present our main objectives and hypotheses which will be detailed later in the three chapters.

Chapter II will present how we developed, for the first time in micropropagation of plants, a miniaturized and automated screening system able to meet high-throughput standards in order to rapidly optimize SE protocols by testing, within a limited time, a wide range of active compounds

(nutrients, growth regulators, etc.), multiplying experimental conditions and reducing volumes as well as the space and labor required.

Chapters III and IV will show how we have taken advantage of the latest metabolomic and transcriptomic tools and applied them to the coffee SE process, in order to unravel the underlying global molecular mechanisms, from cell dedifferentiation to redifferentiation, which allowed us to identify a number of candidate genes and metabolites able to pilot the optimization of this complex process.

Finally, the last chapter will be the subject of a discussion on all the results obtained and a general conclusion opening on several perspectives based on the present thesis work.

Chapter I

LITERATURE REVIEW

PART I

Generalities about coffee

I.1. Coffee cultivation, consumption and economic importance in producing countries

Coffee is the most popular beverage. According to the International Coffee Organization (www.ico.org), about 2.2 billion cups of coffee are consumed every day worldwide. Coffee beverage is consumed either hot or cold by about one-third of the people in the world. Its popularity can be attributed to its stimulant effect, which is produced by caffeine, an alkaloid found in green coffee in amounts between 0.8 - 1.4% for the Arabica varieties and 1.7 - 4.0% for the Robusta (*Coffea canephora*) varieties (www.ico.org). Coffee is cultivated between the Tropics of Cancer and Capricorn forming a “bean belt” and occupies a total of 11 million hectares worldwide, 7 million for Arabica and 4 million for Robusta (**Figure 1**). The Arabica coffee tree typically grows at 600-2200 m altitude, requires a temperature between 15°-24°C and an annual rainfall of about 1200-2200 mm/year. Robusta tree grows at 0-800 m altitude, requires a warmer temperatures of 18°-36°C and slightly more rainfall (2200-3000 mm/year) than Arabica trees.

Coffee production in 2019 totaled 168.7 million 60-kg bags while consumption worldwide totaled 169.3 million 60-kg bags (www.ico.org). Brazil is responsible for about a third of all coffee production (62.9 million 60-kg bags), making it by far the most important coffee-producing country, followed by Vietnam, Colombia and Indonesia with 31.1, 13.9 and 9.4 million 60-kg bags respectively. USA, European Union nations and Japan are the traditional biggest coffee consumers but other markets are also emerging particularly in East Asia (China, South Korea) (**Figure 2**). Even though coffee is not considered as a food crop, the incomes of millions of people from producing countries, especially in Latin America, depend on coffee exports to consuming countries (Bertrand et al., 2012). Most of them (approx. 25 million) are considered small producers having less than 10 hectares of culture area. With the increasing harmful impact of climate change, the living conditions of these producers are at stake.

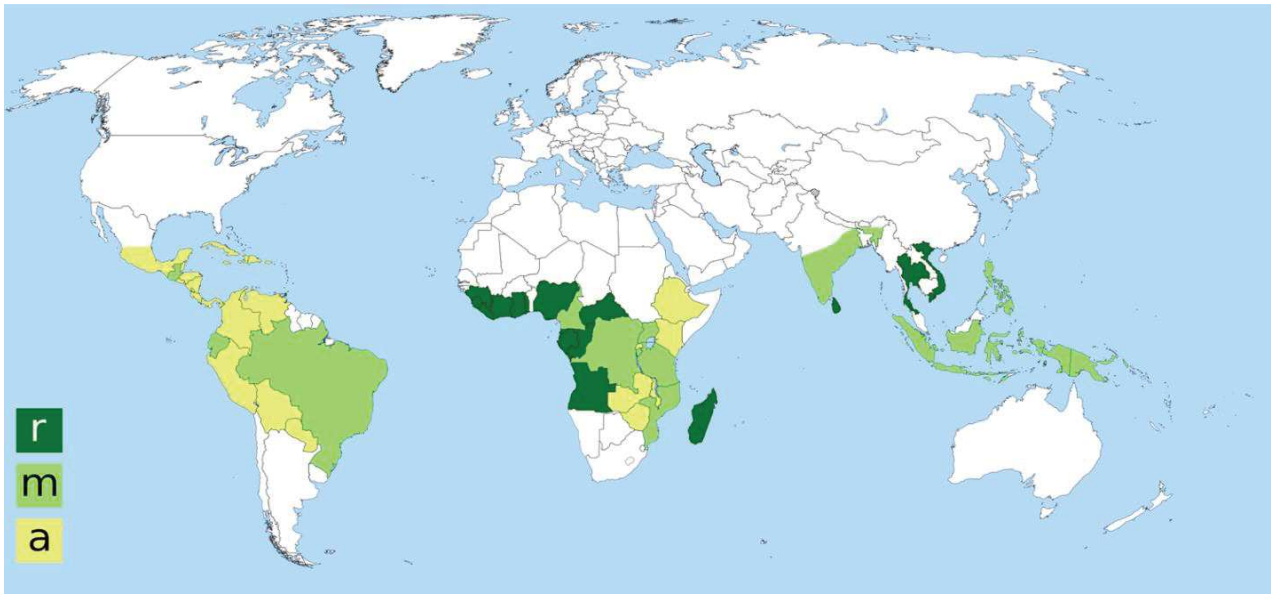


Figure 1. Coffee cultivation areas in the bean belt. (r) *Coffea canephora* cv Robusta, (m) *Coffea canephora* cv Robusta & *Coffea arabica*, (a) *Coffea arabica* (wikipedia).

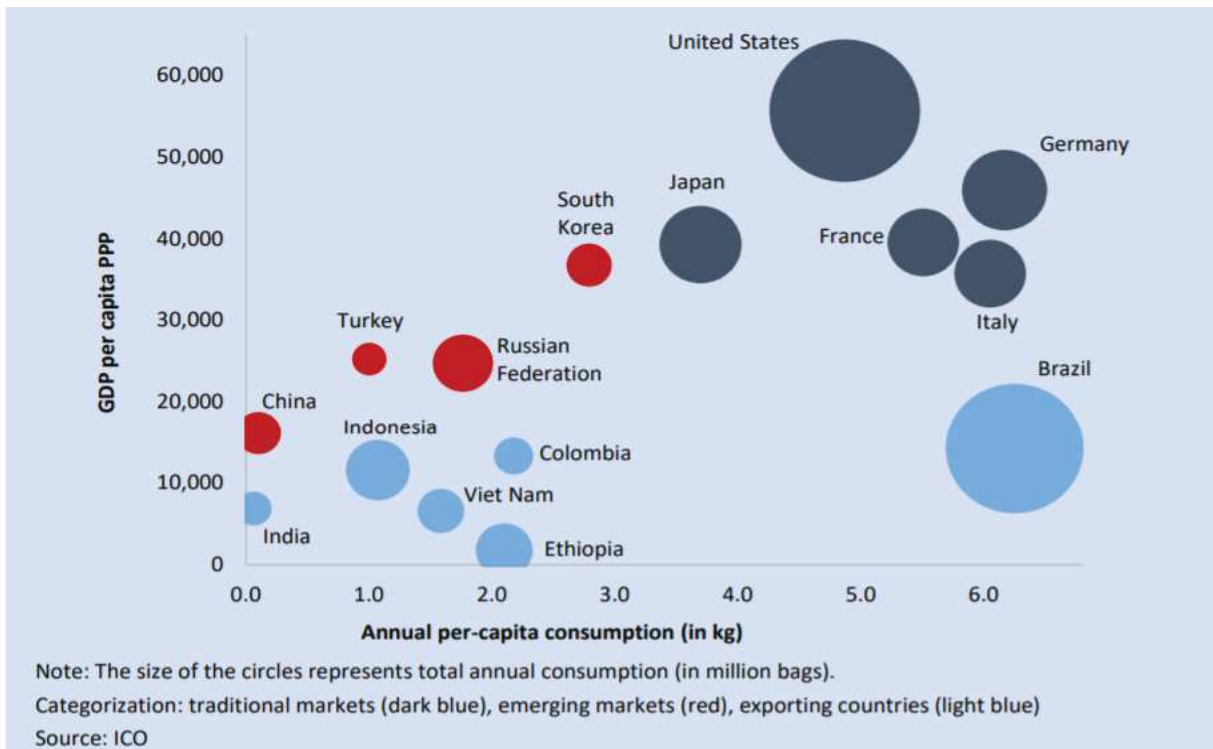


Figure 2. Coffee per-capita consumption across markets in 2018 (www.ico.org). GDP = Gross domestic product.

I.2. Coffee taxonomy

Coffee belongs to Rubiaceae family, a family of flowering plants, with ~ 13,000 species classified in 620 genera, more than 40 tribes and three subfamilies, which makes it the fourth-largest angiosperm family (Goevarts et al., 2006). Rubiaceae are found on all continents but most taxa are in tropical or subtropical areas (Goevarts et al., 2006). Most of them are woody trees and shrubs. Plants of economic importance in this family include: Quinine (*Cinchona*) which possesses medicinal properties, *Ixora* and *Gardenia* cultivated as ornamentals plants, madder (*Rubia* sp) used as dyes and *Coffea* by far the most agronomical important species of Rubiaceae family. Historically, based on morphological data, the species belonging to the genus *Coffea* have been divided by Chevalier (1947) into four groups: Eucoffea (true coffee tree), Mascarocoffea (low caffeine content), Argocoffea Pierre and Paracoffea Miq. (represented by the genus *Psilanthus* Hook). The first three groups contain coffee trees exclusively from Africa while most of the accessions in the fourth group are from India, Indochina, Ceylon and Malaysia. Since the 1980s, progress has been made in taxonomy and this classification has been reviewed by combining molecular and morphological data. The Argocoffea section has been excluded from the genus *Coffea*. Davis et al. (2006) proposed a new classification which separated the genus *Coffea* into two subgenera: the subgenus *Coffea* containing 95 species and the subgenus *Baracoffea* containing 8 species. The subgenus *Baracoffea* is restricted to seasonally dry deciduous forests and bushes located in western Madagascar, northeastern Kenya and southeastern Somalia. The *Coffea* subgenus is present in the evergreen rainforests where the three cultivated species for coffee production are found, *C. arabica*, *C. canephora* and *C. liberica* (Maurin et al., 2007). Davis (2011) included the species classified in the genus *Psilanthus* to the genus *Coffea*, making *Psilanthus* a subgenus of *Coffea*, bringing the number of species belonging to *Coffea* to 124. The flowers in *Coffea* subgenus have long style, medium corolla tube, bare anthers while in *Psilanthus* subgenus the flowers have short style, long corolla tube and encased anthers (**Figure 3a, 3b**). The subgenus *Coffea* share the typical coffee bean morphology, i.e. a groove on the flat side of the seed. The ‘husk’ or ‘parchment’ (endocarp) also has a deep ventral groove, which follows the invagination of the outer layer of the seed (Davis et al., 2006). Species included into the subgenus *Coffea* are self-incompatible with the exception of the allotetraploid species *C. arabica* (Carvalho et al., 1991) and the diploid species *C. heterocalyx* Stoff (Coulibaly et al., 2002) and *C. anthonyi* found out in Cameroon and in Republic of Congo (Stoffelen et al., 2009), respectively. Only two species are commercialized worldwide: *C. arabica* and *C. canephora* cv Robusta. They represent respectively 60% and 40% of world production (Wintgens, 2012). The superior quality beverage is produced from *C. arabica* which is the most important traded species.

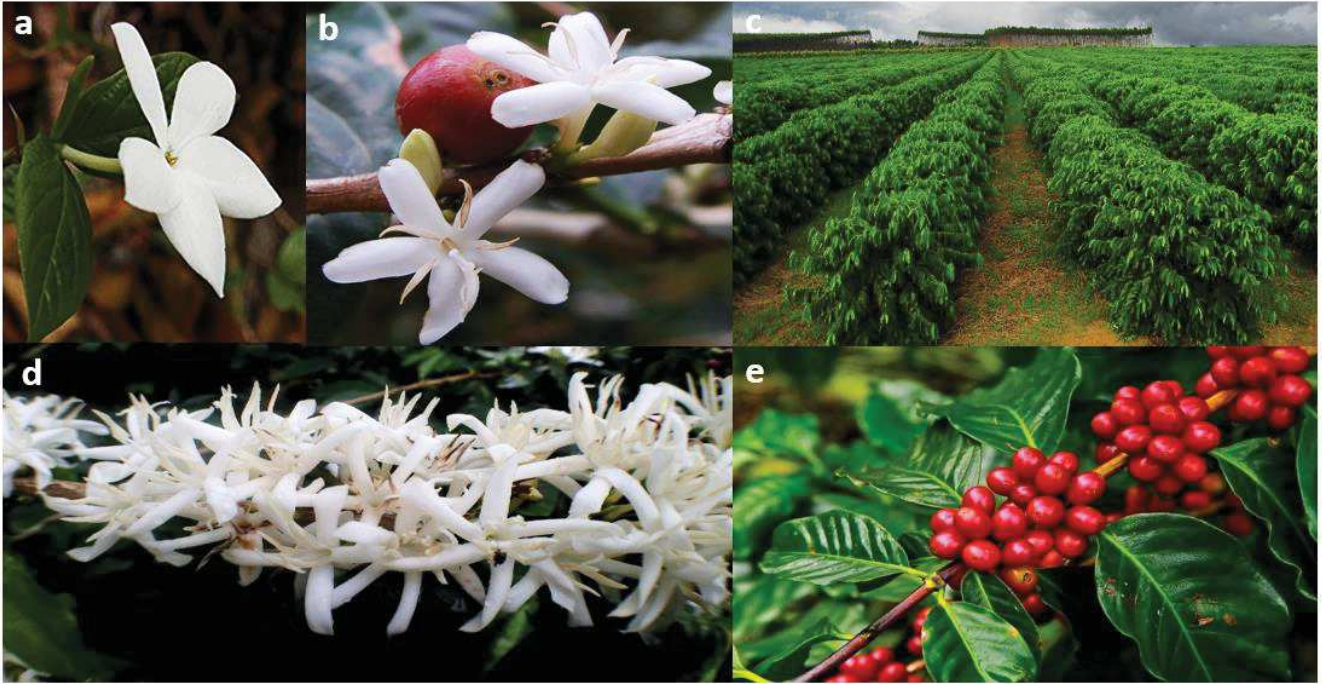


Figure 3. Botanic description of *C. arabica*. Difference between *Psilanthus bengalensis* (a) and *C. arabica* (b) flowers; morphology of *C. arabica* trees (c); flowers are produced in dense clusters (d) and their fruits have a cherry form (e).

I.3. Botanic description of *C. arabica*

Coffea arabica is an evergreen, shrub or small tree, up to 5 m tall when unpruned, glabrous, with small glossy leaves (Figure 3c). It has a main vertical trunk (orthotropic) and primary, secondary and tertiary horizontal branches (plagiotropic). Axillary and sub-axillary buds often develop into reproductive lateral branches. Leaves are simple, opposite, thin, dark-green, shiny-surfaced and fairly stiff (Dengler, 1999; Matsunaga et al., 2016). Leaves have a petiole that connects them to the stem. Flowers are produced in dense clusters (4 to 18 flowers) along reproductive branches in the axils of the leaves (Figure 3d). Flowers are white, sweet-scented, star-shaped (5 sepals and petals) and carried on stout but short peduncles. The fruits have a cherry form (Figure 3e) which contain two seeds, 8.5-12.5 mm long, ellipsoidal in shape that are joined by their “Coffean” invagination on their flat side (Figure 4). Their other side is convex. A thin, silvery skin follows the outline of the endosperm, so fragments are often found in ventral groove after preparation. Seeds consist mainly of a green corneous endosperm, folded in a peculiar manner, and a small embryo near the base (Figure 4). Immature berries are dull green and on ripening the skin color changes through yellow to bright crimson. Dried seeds, after removal of the silvery skin, provide the commercial coffee beans (green coffee).

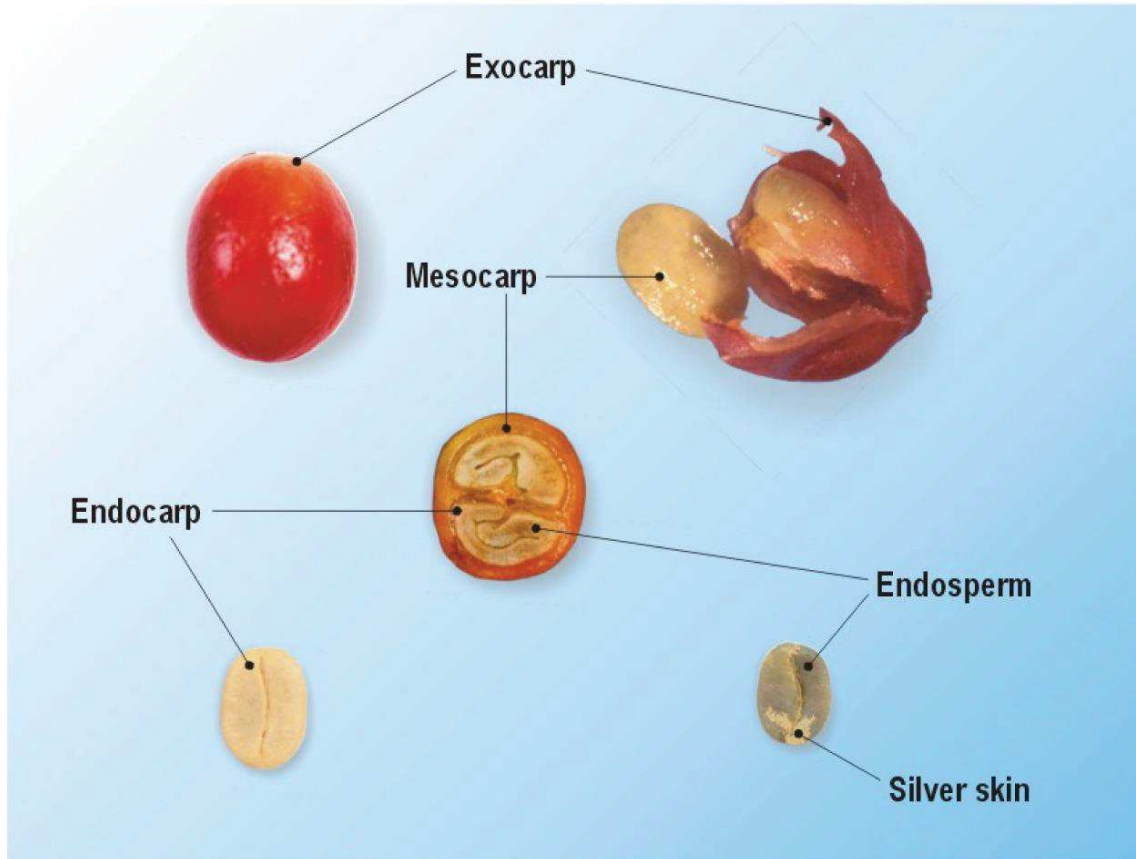


Figure 4. Coffee berry and bean morphology

I.4. Arabica genome characteristics

Coffea arabica is an allopolyploid species ($2n=4x=44$) resulting from a spontaneous hybridization between two wild diploid species most closely related to *C. eugenioides* and *C. canephora* (Lashermes et al., 1999) (**Figure 5**). The allopolyploid speciation of *C. arabica* has a broad time interval estimate spanning between 10,000 and 665,000 years before present day (Cenci et al., 2012). In *C. arabica*, a disomic type inheritance has been observed by analysis of segregation in an F₂ population. *C. arabica* has a meiotic behavior comparable to that of diploid species. At meiosis, the homologous chromosomes would systematically pair to form bivalents but none of the homeologous chromosomes of *C. arabica* pair together, and this is likely due to the functioning of genetic factors that regulate pairing (Lashermes et al., 1999). Flow cytometry has been used to estimate the amount of nuclear DNA in fifteen *Coffea* species native to Africa (Cros et al., 1995). In diploid species, the amounts of 2C of nuclear DNA are very variable with values between 0.95 and 1.78 pg per nucleus. These observed variations would be due to variations in the number of copies of repeated DNA sequences (Lashermes et al., 2008). Regarding the tetraploid *C. arabica*, its quantity of DNA was estimated to be twice as high (2.61 pg) as the quantities of the diploid species (Cros et al., 1998). Compared to other

plant species, the genome of coffee trees appears relatively small (Robusta = 710 million DNA base pairs (Mbp), Arabica = 1,300 Mbp, maize = 2,500 Mbp, wheat = 15,000 Mbp) (Kochko et al., 2010).

Arabica cultivars and landraces are typically propagated by seed. The mating system is primarily based on self-fertilization, although pollinator-mediated outcrossing may occasionally occur. The predominant autogamy leads to high levels of inbreeding.

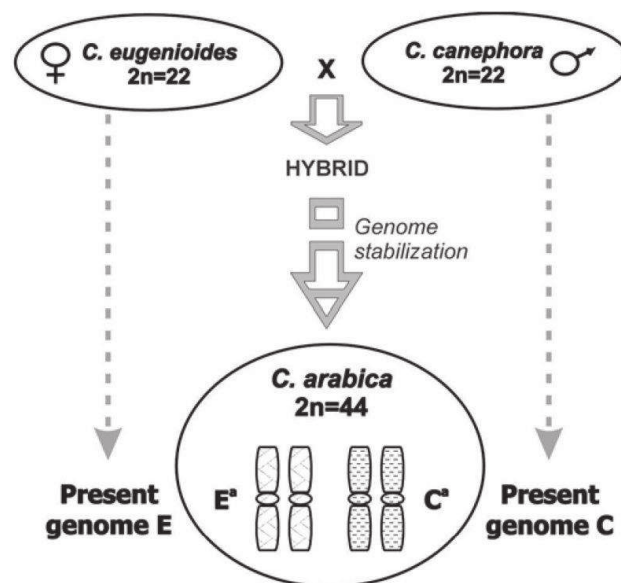


Figure 5. Genetic origin of allotetraploid *C. arabica*. A natural hybridization between *C. eugenioides* as female parent and *C. canephora* as male parent originated the *C. arabica* species.

1.5. Genetic diversity of *C. arabica*

C. arabica is indigenous to Ethiopia and South Sudan which represents its primary center of diversity (Sylvain, 1955, FAO-1964, ORSTOM-1966) (**Figure 6**). Yemen is a secondary dispersal center (Fernie et al., 1968). Several accounts of the early history of *C. arabica* germplasm usage and movements are available in the literature, but the most complete and best documented publication is the one of Haarer (1958). During the 14th century, coffee seeds were brought out of the forests of South Western Ethiopia to Yemen, where coffee cultivation expanded to satisfy the demand of a growing number of coffee houses in Mokka (Yemen) and Cairo (Egypt) at the end of the 15th century. The center of diversity in South Western Ethiopia corresponded to the area that has been for centuries under the control of the Kingdom of Kafa. This area was described as an impenetrable “citadelle” until Menelik II conquered the Kingdom in 1897. Hence, early escapes of coffee seeds from South Western Ethiopia were likely occasional and the coffee cultivation in Yemen has started from a narrow genetic basis. *C. arabica* was then spread worldwide from Yemen rather than from its primary center of origin in Ethiopia. Significant out-of-Yemen movements were documented i) in 1670, from Yemen to India

with some seeds smuggled by Baba Budan and ii) in 1715, from Yemen to Bourbon Island (today Ile de la Réunion) with very few seeds. The former gave rise to the Typica variety after the Dutch brought some seeds from India to today's Indonesia in 1696 and 1699. From Indonesia through Europe, Typica reached the Americas in 1723. The latter gave rise to the Bourbon variety that reached the Americas and East Africa in the mid-19th century. In India, the early seeds introduced in 1670 were grown locally for centuries. The history of Arabica has resulted in extremely low genetic diversity as revealed in the very recent work of Scalabrin et al. (2020) increasing predispositions to biotic and abiotic stress. This study showed a low diversity on the 699 accessions collected in Ethiopia and the 95 accessions collected in Yemen. The very low genetic diversity of cultivated varieties used today, their high susceptibility to diseases and global warming represent a major challenge for coffee production in the coming years (Ghini et al., 2011; Lin, 2011).

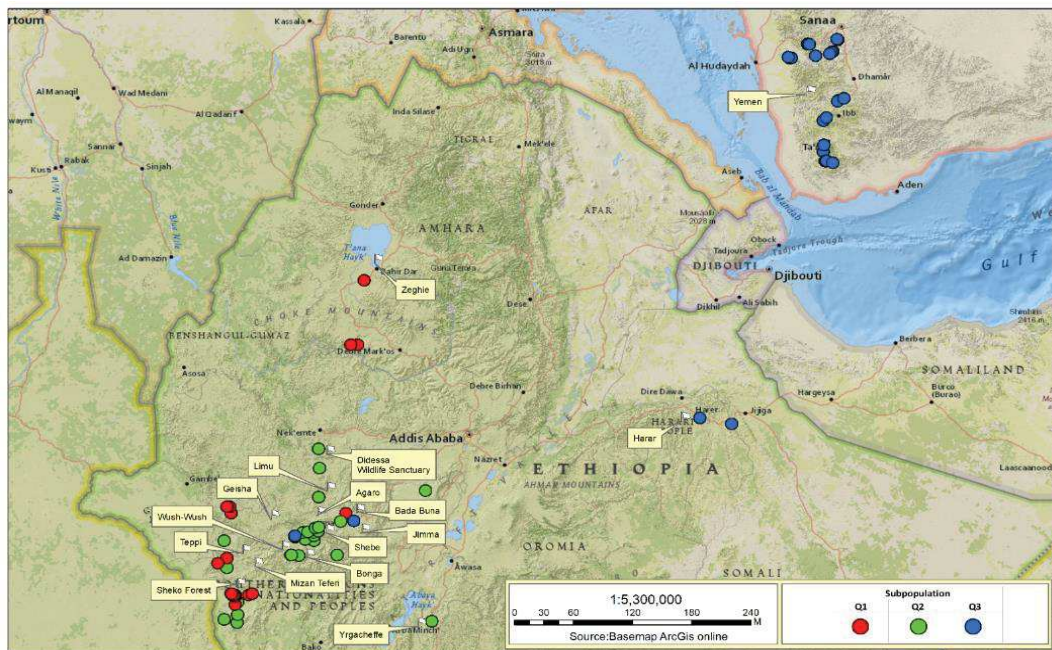


Figure 6. Centers of genetic diversity of *C. arabica*. Arabica is indigenous to Ethiopia and South Sudan which represents its primary center of diversity while Yemen is a secondary dispersal center. The Q1 subpopulation represents forest-adapted Ethiopian landraces, the Q3 subpopulation represents full-sun adapted Yemeni species and the Q2 subpopulation is an intermediate Ethiopian subpopulation between Q1 and Q3 (Scalabrin et al., 2020)

PART II

Coffee breeding

II.1. *C. arabica* breeding programs prior to 1990s

For a long time, agronomists have distinguished two cultivated forms: the Typica cultivar and the Bourbon cultivar (Van Der Vossen et al., 2015) (**Figure 7**). Typica coffee plants have a conical shape with a main vertical trunk and secondary verticals that grow at a slight slant. Typica plants can be 4-5 m tall. Bourbon cultivar characteristic is to be of a smaller size than Typica but with a denser vegetation and generally has a better productivity as well as a better quality by the cup. The natural cross between Bourbon and Typica gave the vigorous and productive cultivar "Mundo novo" in Brazil. Spontaneous mutants showing desirable characteristics were exploited for cross-breeding purposes. The Caturra variety originated from a mutation in the Bourbon variety. The dwarf-sized Caturra trees (based on a single dwarfing gene) allows intensification and mechanization of the harvest and to adopt the new cultural practices of the Green Revolution which began in coffee culture in the 1950-60s (Bertrand et al., 2012). A cross between Mundo novo and Caturra makes it possible to obtain another dwarf-sized variety particularly well-adapted to Brazilian full-sun culture conditions: the Catuai variety. Today, Caturra and the derived varieties (Mundo novo and Catuai) represent 80% of coffee plantations and represent a benchmark in terms of cup quality (**Figure 7**). These varieties are propagated by seeds. They have the inconvenience to be highly susceptible to the leaf rust disease which is caused by a fungal pathogen, *Hemileia vastatrix* (**Figure 8**). *H. vastatrix* is present in almost all coffee producing countries and is considered as the most devastating disease for this crop with losses of up to 30% for susceptible varieties if the disease is not treated (Rozo et al., 2012).

In the 1940s, the 'Timor Hybrid' was discovered on the island of Timor (Indonesia). It is a natural allotetraploid hybrid resulting from a cross between an Arabica and a Robusta variety, and showed better resistance to leaf rust. This hybrid gave rise to the Introgressed varieties (**Figure 7**). A cross between either a Typica or a Bourbon and cultivars derived from 'Timor Hybrid' gave rise to two groups:

- The "Catimor" group (Caturra or Catuai x Timor Hybrid) e.g. 'Costa Rica 95' variety;
- The "Sarchimor" group (Villa Sarchi (Bourbon) x Timor Hybrid) e.g. 'Obata', 'IAPAR59' and 'Marsellesa' varieties.

A better productivity of approximately 15% was obtained in Central America, Brazil and Colombia with introgressed varieties. However, the gene introgression was usually accompanied with a

substantial decrease in cup quality (Bertrand et al., 2003). On average, 1 point was lost on a scale ranging from 1 to 10. Hence, most national coffee organizations in charge of disseminating varieties in Central America have now suspended the distribution of these introgressed varieties. After 5-6 years of mass dissemination of introgressed varieties, the loss of resistance to rust raised many concerns for coffee growers since this resistance was the main advantage of these varieties.

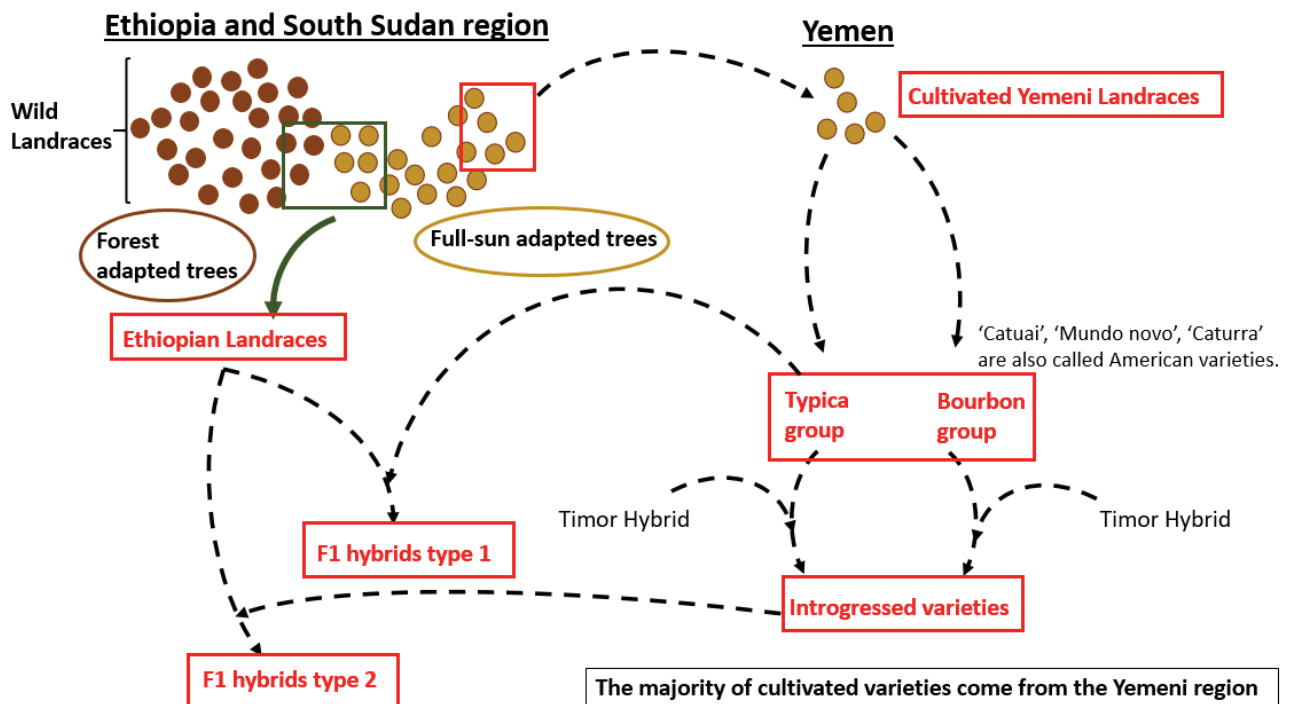


Figure 7. *C. arabica* breeding programs prior to 1990s and until today. Arabica is characterized by an extremely low genetic diversity. The F1 hybrids resulting from a cross between Ethiopian landraces and American varieties can broaden the Arabica genetic diversity and be more tolerant to biotic and abiotic stresses (section II.2.).



Figure 8. Symptoms caused by leaf rust disease on *C. arabica*. (a) Significant defoliation of a susceptible tree. (b)(c) Development of spores on the abaxial side of the leaf (from Rozo et al., 2012).

II.2. *C. arabica* breeding programs from 1990s until today

As explained in Part I.5., *C. arabica* has an extremely narrow genetic diversity. Since their domestication in Yemen during the 14th century, Arabica cultivars were adapted to full-sun culture systems. Until the 1990s, breeding programs in Latin America have always privileged productivity of cultivars over sustainability, enhancing full-sun cultivation and pesticide use.

The context of climate change came as a turning point for coffee breeding programs objectives. A number of harmful effects of climate change were observed for the first time in the history of the intertropical belt: highest temperatures scored; longer drought periods in Brazil, Central America, East Africa; excessive rain in Colombia, Peru and Central America; and crop frosts in Vietnam.

The impact on coffee was devastating:

- Fewer culture areas were available; coffee was now cultivated in higher altitudes.
- Higher temperatures and loss of resistance to leaf rust resulting in more leaf rust epidemics in Latin America.
- Growing pest problems (coffee berry borer (*Colletotrichum kahawae*) and root nematodes (*Meloidogyne* spp.)).
- Declining productivity and cup quality.
- Imbalances between supply and demand.
- The living conditions of 25 million producing families have dramatically degraded.

In this context of climate change, the objective of breeding programs was to select a more complex set of traits, combining productivity, better quality and tolerance of biotic and abiotic stresses (Bertrand et al., 2012). In order to achieve this objective, a collaboration between CIRAD, PROMECAFE and CATIE was proposed to create interspecific F1 hybrids, notably by crossing wild Ethiopian material with the cultivated American varieties (Caturra, Catuai, Mundo novo) (Van der Vossen, 2001) (**Figure 7**). The main objective of F1 hybrids was to broaden the genetic diversity of Arabica cultivars since the wild Ethiopian varieties could add tolerance to stresses and organoleptic quality to the American varieties. Creating F1 hybrid varieties was a faster and more convenient mean than the creation of pure line varieties in order to accumulate, in a same genotype, the maximum of favorable dominant genes (7 years instead of 25 years). The advantage of F1 hybrids for the autogamous Arabica species was demonstrated for the first time in the Kenyan breeding program (Van Der Vossen et al., 2015). Bertrand et al. (2005) evaluated the heterosis of F1 hybrids from crosses between wild Sudanese-Ethiopian cultivars and American cultivars between 22 and 47%. They conclude in a later study that heterosis gives F1 hybrids an advantage over American varieties in terms of vigor, productivity and cup quality but also in terms of yield stability due to their homeostasis (Bertrand et al., 2011) (**Figure 9**).

Since 2003 and until today, the CIRAD/ECOM alliance as well as Nestlé Research continue to develop new F1 hybrids as the consumer demand has shifted towards a coffee labelled “organic”, “high-quality”, “fair-trade” and “coming from a specific terroir”. Producers are aware of the need to crop their coffee in agroforestry systems (AFS) —where small coffee trees are grown among tall forest trees— since this system can sustainably dilute the harmful effects of climate change (lower temperatures in shade, better adaptability to drought and rain, less pesticide input) (**Figure 10**). However, American varieties showed a 20–40% drop in productivity in AFS (Bertrand et al., 2005). For this economic reason, AFS have been progressively abandoned by producers that preferred full-sun conditions. F1 hybrids have proven to be suitable for agroforestry-based ecologically intensive cropping systems since American varieties are crossed with wild Ethiopian varieties that were historically grown in the deep forest. In AFS, the mean yield of hybrids was 58% higher than that of the American cultivars (Caturra, Catuai and Catimor cultivars) while the mean yield of hybrids in the full-sun system was also 34% higher (Bertrand et al., 2011).

In order to select the outstanding individuals, F1 hybrids are subject to individual selection. F1 hybrids can't be propagated by seeds as they need to be disseminated in a vegetative way (asexual reproduction) in order to create clones having the same traits as the selected hybrid.

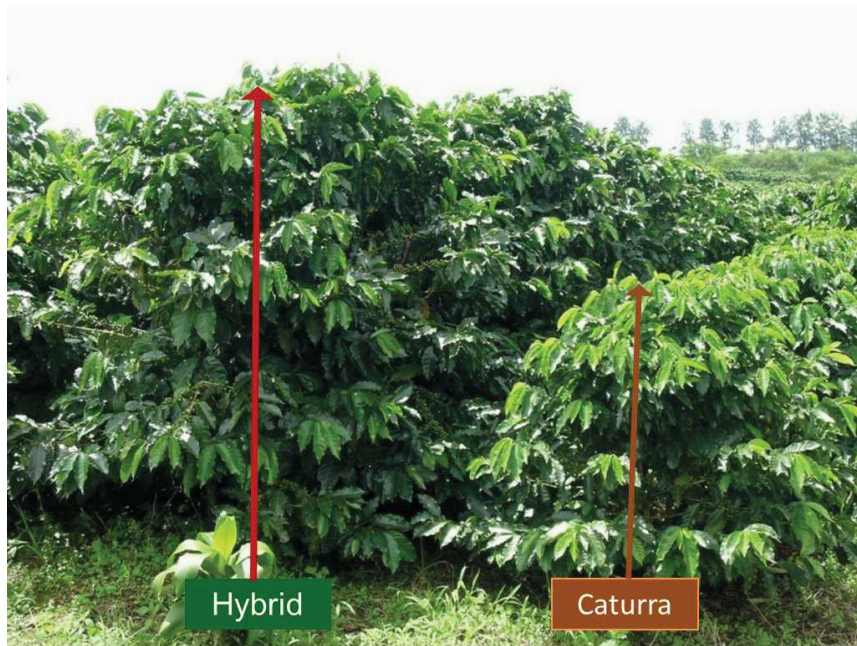


Figure 9. Comparison between *C. arabica* cv Caturra and a clone of F1 hybrids after 3 years since germination (photo E. Alpizar (ECOM)).



Figure 10. Difference between coffee cultivated in full-sun (a) and in agroforestry systems (b) (wikipedia).

PART III

Vegetative propagation techniques

III.1. Seed propagation vs. vegetative propagation

Plant propagation is the process of plant reproduction. There are two types of propagation: sexual and asexual. Sexual propagation involves the floral parts of a plant and consists of the union of the pollen and ovum to produce a seed. The seed is made up of three parts: the outer seed coat, which protects the seed; the endosperm, which is a nutrients reserve; and the embryo, which contains the genes of the two parents and will give rise to a new and original, third individual. When a seed is mature and placed in a favorable environment, it will germinate. Vegetative propagation is a form of asexual reproduction of a plant. Only one plant is involved and the offspring is the result of one parent. New plants are genetically identical to their parent. Asexual propagation involves vegetative parts of a plant (e.g. stems, roots, leaves) or sexual ones (e.g. floral explants). While the main advantage of sexual propagation is its low costs, several advantages of vegetative reproduction have been reported, mainly the fact that the produced offspring are clones of their parent plants, which avoids the difficulties of cross-pollination. If a plant has favorable traits, it can continue to pass down its advantageous genetic information to its offspring. It can be economically beneficial for commercial growers to clone a certain plant to ensure consistency throughout their crops. Vegetative propagation is also well-suited for breeding programs by allowing the use of clonal tests for the genetic, phenotypic and environmental evaluation of individuals in the frame of bio-assays in controlled environmental conditions. Another advantage of vegetative propagation is that it allows to avoid the complex process of producing flowers, seeds and fruits, especially in plants that have a long life cycle (woody species).

III.2. Traditional vegetative propagation techniques

For perennial plants, vegetative propagation in natural conditions is rare and limited to individuals placed in extreme environmental conditions. Different horticultural techniques exist to induce vegetative dissemination of plants. The two major methods are cuttings and grafting.

a) Horticultural cuttings

A cutting is a vegetative plant part, which is severed from the parent plant and rooted in order to regenerate itself, thereby forming a whole new plant identical to the parent plant (**Figure 11**). Cuttings technique is carried out either on tree branches or on pruned shoots. Rhizogenesis can be spontaneous but is generally initiated by immersing the base of the cuttings in rooting solution containing for

instance indole butyric acid (IBA). The roots appear after 10 weeks in *C. arabica* (Georget et al., 2017). The success of this technique will depend mainly on the quality and the genetic origin of the material, and the control of the cuttings environment. Horticultural cuttings in woody plants are quite easy to carry out as long as the biological material is juvenile. However, the cuttings technique has two important limits: it is genotype-dependent and has a low multiplication rate allowing only a limited number of copies per individual (maximum 65 in *Hevea brasiliensis*) (Masson et al., 2013). In addition, the rooting of the cuttings and the quality of the root system depend both on the clone and on the age of the parent plant (Lelu-Walter et al., 2013).

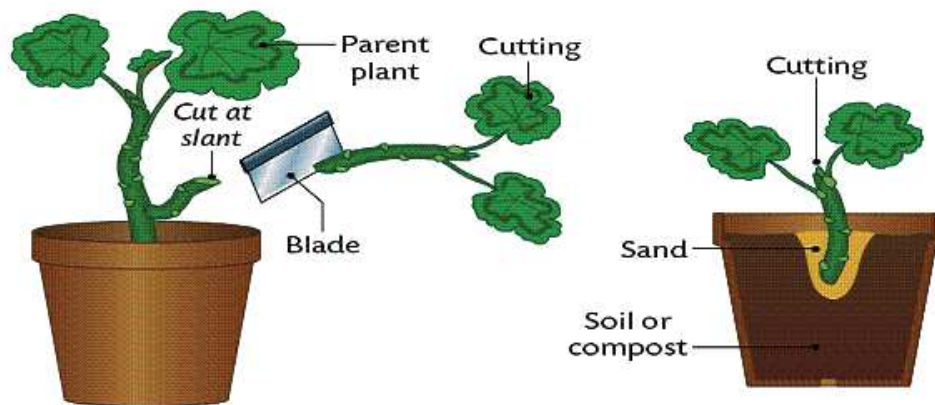


Figure 11. The process of horticultural cutting. A cutting is a vegetative plant part, which is severed from the parent plant and rooted in order to regenerate itself, thereby forming a whole new plant identical to the parent plant.

b) Horticultural grafting

The grafting technique consists of joining two plant parts so they will grow as one plant (**Figure 12**). The portion of the cultivar that is to be propagated is called the scion. It consists of a piece of the shoot with dormant buds that will produce the stem and branches. The rootstock provides the new plant root system. This technique is used to propagate cultivars that will not root as well as cuttings or whose own root systems are inadequate. The scion begins to develop after 1 month in woody plants, and the recovery rate is usually very high. However, a number of challenges faces this technique mainly high production costs due to the fact that it needs a skilled and meticulous manpower, in addition to the problems of incompatibility between scions and rootstocks.

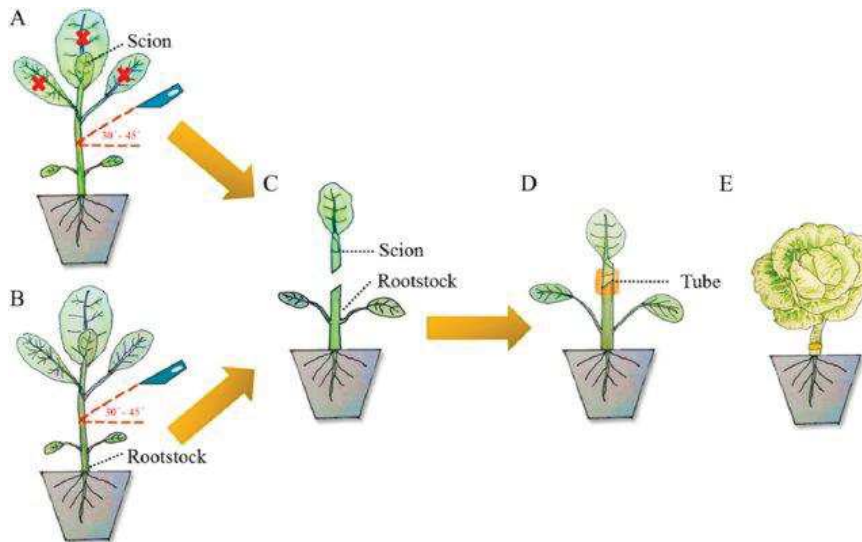


Figure 12. The process of horticultural grafting. Grafting consists of joining two plant parts so they will grow as one plant. The portion of the cultivar that is to be propagated is called the scion. The portion that provides the new plant root system is called the rootstock. The zone unifying the two plant portions is called “union”.

III.3. Micropropagation

Micropropagation is the rapid vegetative propagation of plants under *in vitro* conditions i.e. controlled environment and a defined nutrient medium. The technique has been applied to a substantial number of plant species (Maximova et al., 2002; Yang et al., 2012; Lelu-Walter et al., 2013). Micropropagation involves the culture of cells, tissues or explants (meristems, leaf explants, roots or shoots) under defined media and sterile culture conditions; established proliferation cycles for a number of cell generations and the subsequent regeneration and acclimatization of plants (Larkin & Scowcroft, 1981). Micropropagation technology offers the potential for efficiently propagating disease-free, genetically uniform and massive amounts of plants since its conception (Honda et al., 2001). The establishment of stable, efficient *in vitro* regeneration systems is not only important as a method for the mass propagation of economically important crops, it is also a prerequisite for biotechnology and molecular breeding applications (Neelakandan & Wang, 2012). The three major techniques of micropropagation are: micro-cuttings, organogenesis and somatic embryogenesis.

a) Micro-cuttings

Micro-cuttings technique is a method which utilizes tissue culture-based technology to propagate plants by using axillary buds of seedlings as explants (**Figure 13**). Rubber (*Hevea brasiliensis*) and coffee species (*C. arabica* and *C. canephora*) are two examples where micro-cuttings were successfully implemented (Carron et al., 2003). The multiplication rate varies from 1.3 to 2.3 per

month in woody plants depending on the genotypes and culture age. In field trial, cloned plants had a similar architecture to that of plants obtained from seeds. However, there are some constraints to implement micro-cuttings technique in plant mass propagation mainly the high level of contamination, low proliferation rate, poor rooting, and difficulty in acclimatization process. Moreover, considerable manpower is needed by this technology in order to realize monthly cuttings and transfers on fresh media resulting in high plant production costs, which have limited this technology to trees mainly with low planting density and high plus-value. Micro-cuttings technique is essentially used for maintaining a collection of genotypes *in vitro*.

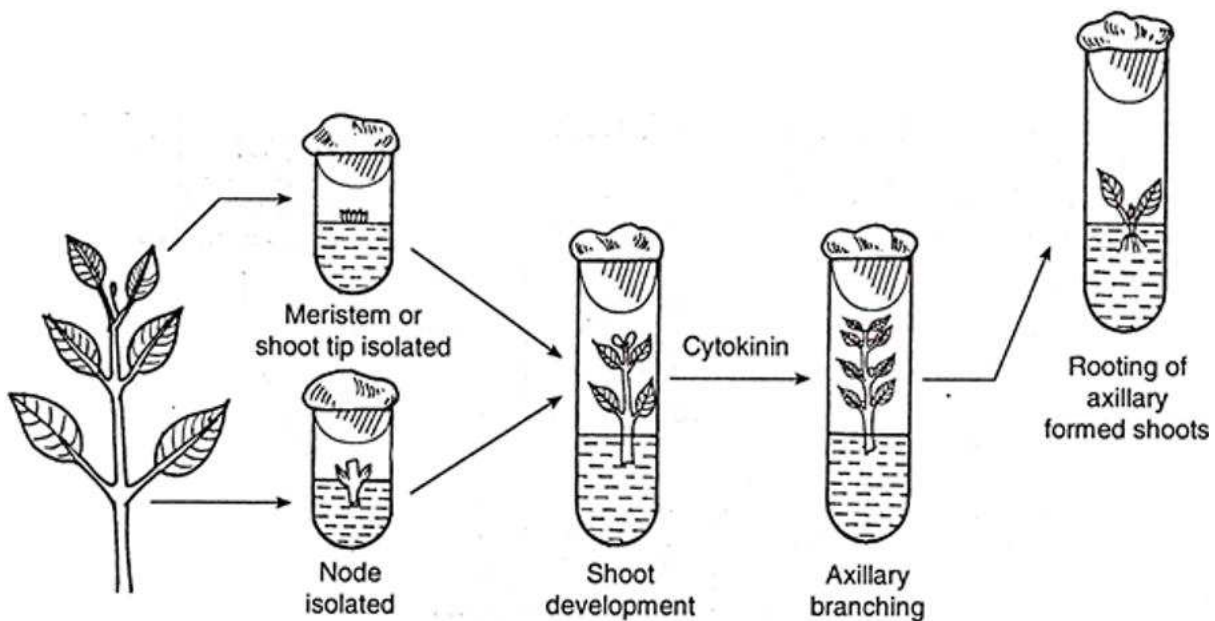


Figure 13. The process of micro-cuttings. First, axillary buds (nodes) or shoot tips are cultured *in vitro* until shoot development. Newly formed axillary buds are severed from the parent plant and rooted in order to regenerate themselves, thereby forming a whole new plant identical to the parent plant.

b) Organogenesis

Organogenesis is the regeneration pathway in which shoots or roots are induced directly from plant explants (e.g. leaf, roots, buds) or indirectly after callus induction (proliferation of meristematic cells) (Duclerq et al., 2011) (**Figure 14**). In organogenesis, plant growth regulators (PGRs) play a major role as a high cytokinin to auxin ratio induces shoot morphogenesis, whereas an opposite low ratio results in root development. However, organogenesis methods are often labor-intensive and require many steps of manipulations resulting in high labor costs (Brown & Thorpe, 1995). For example, in order to regenerate a whole new plant, shoots are first selected and manually separated before inducing plant rooting in separated stages of the process. The main disadvantage in implementing organogenesis in

mass propagation programs is related to the low multiplication rates offered by this technique (Honda et al., 2001). Until now, this technology has never been used for coffee.

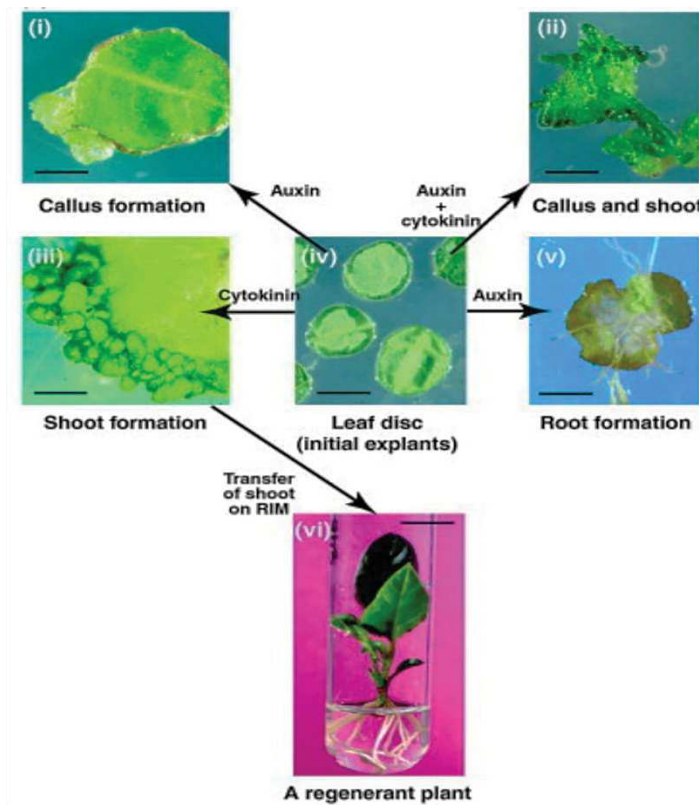


Figure 14. The process of organogenesis. Organogenesis is the regeneration pathway in which shoots or roots are induced directly from plant explants (e.g. leaf, roots, buds) or indirectly after callus induction (from Duclercq et al., 2011). RIM: root-inducing medium.

c) Somatic embryogenesis

Somatic embryogenesis is considered as the developmental process by which somatic cells (i.e. not involved in the sexual propagation process), under suitable induction conditions undergo restructuring through the embryogenic pathway to generate embryonic cells (Yang & Zhang, 2010). This developmental pathway was first described in carrot (*Daucus carota*) cell suspensions by the teams of Steward et al. (1958) and Reinert (1958) (**Figure 15**). Since these works, SE technology has particularly progressed in species that have a long life cycle and are difficult to propagate by conventional horticultural methods (woody plants). Today, SE is an important *in vitro* plant regeneration approach that has demonstrated significant benefits when applied to forest tree species, bringing important advantages including rapid clonal mass propagation, cryopreservation of valuable germplasm and genetic transformation (Corredoira et al., 2006).

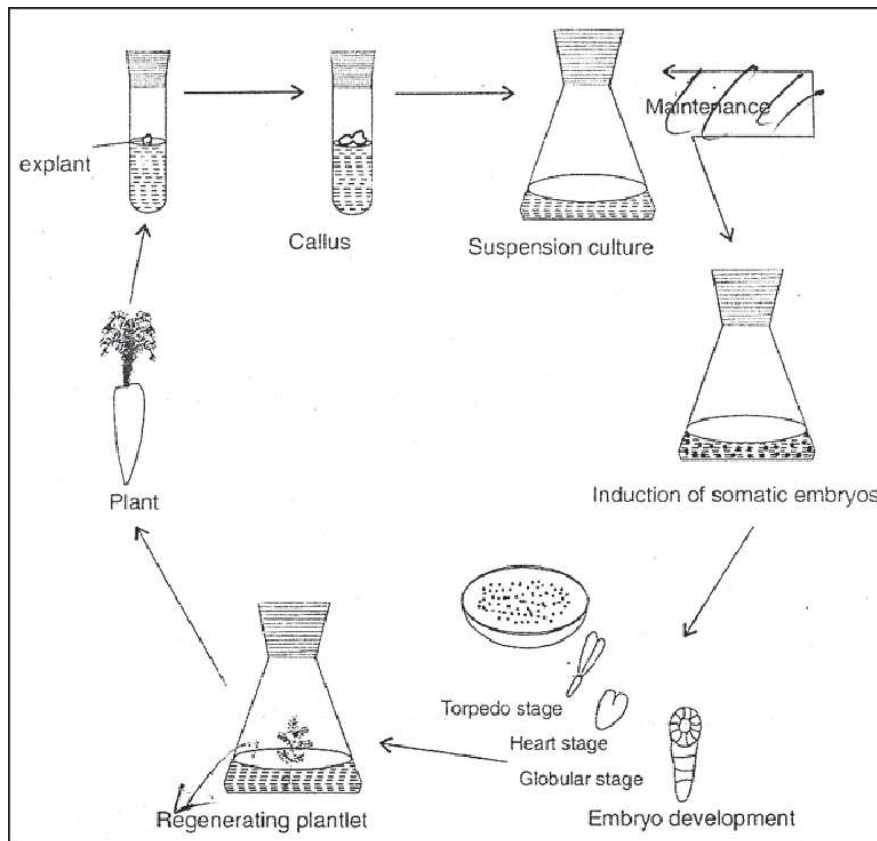


Figure 15. The indirect somatic embryogenesis (SE) process as first described in *Daucus carota* (Steward et al., 1958). Plant regeneration *via* indirect SE includes five common steps after explant selection: (1) induction of embryogenic cultures by culturing the primary explant on medium supplemented with plant growth regulators (PGRs), mainly auxin and cytokinin; (2) the proliferation of embryogenic cultures (maintenance and multiplication) on solidified or liquid medium supplemented with PGRs; (3) somatic embryo development in medium lacking PGRs; this inhibits proliferation and stimulates embryo formation and early development; (4) embryo maturation by culture on medium supplemented with ABA or with reduced osmotic potential and (5) the development of plants on medium lacking PGRs, until plants are rooted and able to be transferred to sterilized soil for hardening in greenhouse.

III.4. Somatic embryogenesis and totipotency

Somatic cells within the plant contain all the genetic information necessary to create a complete and functional plant (Von Arnold, 2008). This cell property is known as totipotency (i.e. the potential or inherent capacity to regenerate *in vitro* an entire plant). Somatic cells have the unique ability to dedifferentiate, proliferate, and regenerate into mature plants under appropriate culture conditions (Neelakandan & Wang, 2012). SE provides a good model system to study the mechanisms of cell totipotency (Komamine et al., 1992). One particular aspect has been taken as evidence of cell totipotency, the stage of “callus formation” (Sugimoto et al., 2018) (**Figure 15**). Nevertheless, not all somatic cells are capable of embryogenesis (Dodeman et al., 1997). The factors influencing the *in vitro* adaptability and regeneration are varied, ranging from genotype, origin of explant and culture

conditions (Neelakandan & Wang, 2012). Theoretically, all plant cells are totipotent; however, it seems that the initiation of the embryogenic pathway is restricted to certain responsive cells owning the potential to activate genes involved in the generation of an embryogenic response (Feher, 2008). The cells which are more competent for SE are generally those coming from juvenile tissues found in young plants and immature zygotic embryos among others (Gutierrez-Mora et al., 2012). It is believed that the induction of SE consists of the termination of a current gene expression pattern in the explant tissue and its replacement with embryogenic cell expression program (Quiroz-Figueroa et al., 2006; Von Arnold, 2008). In this respect, PGRs and stress response to wounding and osmotic medium could play a central role mediating the signal transduction cascade leading to the reprogramming of gene expression (Verdeil et al., 2007; Von Arnold, 2008).

III.5. Somatic embryogenesis vs. zygotic embryogenesis

Somatic embryos are defined as bipolar structures containing both shoot and root poles that are from anatomical and physiological perspectives comparable to zygotic embryos (Zimmerman et al., 1993; Phillips, 2004; Verdeil et al., 2007). Thus, SE systems can be used as models for studying molecular, regulatory and morphogenetic events in plant embryogenesis (Quiroz-Figueroa et al., 2006; Deo et al., 2010). Zygotic embryogenesis can be divided in two main steps: embryogenesis *sensu stricto* beginning with the sporophytic generation of the zygote and finishing in the embryo at the cotyledonary stage; and the maturation of a seed followed by germination (Dodeman et al., 1997). The sporophytic generation is initiated by a double fertilization event, a process unique in angiosperms in which both male gametes (spermatic cells) participate in the act of fusion, one unites with the egg cell (female gamete) to form the diploid zygote from which the embryo develops, while the other gamete fuses with the central cell of the embryo sac and then develops into the triploid endosperm that provides nutrients for embryo development and later for germination (Quiroz-Figueroa et al., 2006). The embryogenesis *sensu stricto* can be divided into the following stages: globular-shaped, heart-shaped, torpedo-shaped, and cotyledonary embryo stages in dicots; globular, scutellar and coleoptilar stages in monocots; and globular, early cotyledonary and late cotyledonary embryos in gymnosperms (Yang & Zhang, 2010). Somatic embryos pass through characteristic morphological stages of zygotic embryogenesis, which are: globular-shaped, heart-shaped, torpedo-shaped and cotyledonary embryo (Zimmerman, 1993; Dodeman et al., 1997; Phillips, 2004; Quiroz-Figueroa et al., 2006). At the cotyledon stage, instead of entering desiccation or dormancy as observed in zygotic embryos, the somatic embryos continue to grow into fully differentiated plantlets (Zimmerman, 1993; Yang & Zhang, 2010). These stages are called the 'late stages of somatic embryogenesis'. Nevertheless, dormancy can also be induced in somatic embryos, hence long-term storage or even the creation of

artificial seeds is possible (Deo et al., 2010; Corredoira et al., 2015). The completion of all the stages of embryogenesis results in the regeneration of a new plant (Zimmerman, 1993; Quiroz-Figueroa et al., 2006). **Figure 16** shows the characteristic developmental stages in (a) the zygotic and (b) the somatic embryogenesis and their main differences. Similar processes are found in coffee, however, somatic embryos are normally more irregular than zygotic embryos due to a different cell elongation and different storage components between both processes (Etienne et al., 2013).

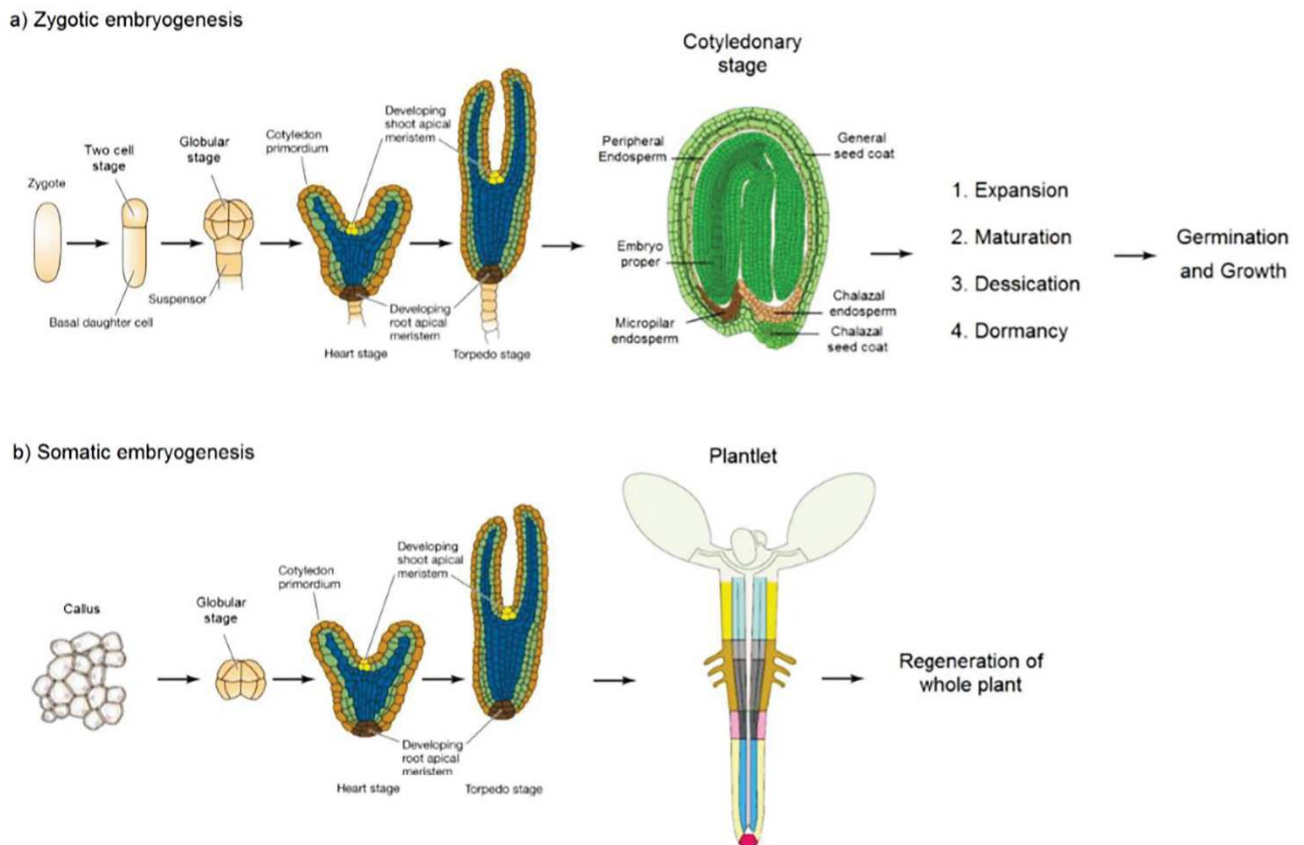


Figure 16. Comparison between somatic and zygotic embryogenesis (adapted from Zimmerman, 1993).

III.6. Direct and indirect somatic embryogenesis

Similar to organogenesis, somatic embryos can differentiate either directly from the explant without an intervening callus phase (direct SE) or indirectly after a callus phase involving the proliferation of embryogenic callus (indirect SE) (Dodeman et al., 1997; Von Arnold et al., 2002; Quiroz-Figueroa et al. 2006, Namasivayam, 2007). Direct SE is normally faster than indirect methods and could represent an advantage for the rapid release of cloned material (e.g. 70 days in coffee). However, direct SE is characterized by sharply lower yields than indirect SE (Campos et al., 2017). In indirect SE, a friable embryogenic tissue is formed either in semi solid media (embryogenic callus) or in liquid media (cell

clusters) from which the somatic embryos arise (Etienne, 2005) (**Figure 15**). Plant regeneration via indirect SE includes five common steps after explant selection: (1) induction of embryogenic cultures by culturing the primary explant on medium supplemented with PGRs, mainly auxin and cytokinin (Von Arnold et al., 2002); (2) the proliferation of embryogenic cultures (multiplication) on solidified or liquid medium supplemented with PGRs (Dodeman et al., 1997; Von Arnold et al., 2002); (3) somatic embryo development in medium lacking PGRs; this inhibits proliferation and stimulates embryo formation and early development (Dodeman et al., 1997; Von Arnold et al., 2002); (4) embryo maturation by culture on medium supplemented with ABA or with reduced osmotic potential and (5) the development of plants on medium lacking PGRs, until plants are rooted and able to be transferred to sterilized soil for hardening in greenhouse. Direct and indirect SE were largely described for the cultivated species of coffee (Nic-Can et al., 2015, Etienne et al., 2018)

III.7. Factors influencing the success of SE

a) Type of explant

Somatic embryogenesis can be induced using different kinds of explants (leaves, immature zygotic embryos, anthers, petals) but generally young leaves, which are available throughout the year, easy to harvest and to sterilize, are the most commonly used. To obtain somatic embryos, many parameters need to be taken into account such as the physiological state of the mother plants. For example, in coffee, the leaf age (young, mature, completely expanded) is determinant as well as the environmental growth conditions of the explant donor (mother plant) and even the month of collection since it can affect the physiology of the mother plant (Molina et al., 2002; Santana et al., 2004). In most species including coffee, the most determining factor for efficient SE remains the genotype (Molina et al., 2002; Etienne, 2005; Rezende et al., 2011). Molina et al. (2002) and Rezende et al. (2011) studied SE in different *C. arabica* genotypes and showed that some genotypes are completely recalcitrant (no callus production or production of non-embryogenic callus) while in others 73% of formed callus was able to generate embryos.

b) Nutrients

The nutrient medium for SE may have different compositions depending on the species considered and the culture step. In coffee, the most commonly used mineral medium is based on MS (Murashige & Skoog, 1962) and used in half-strength (Etienne, 2005; Gatica-Arias et al., 2008; Simões-costa et al., 2009). In order to maintain the cells in an undifferentiated callus state, cell suspensions maintained at high inoculum density, in a medium enriched with auxin and cultured in the dark, are commonly

used. Stably established embryogenic cell suspensions are obtained generally 2–3 months after their initiation in liquid medium. These cell suspensions are sub-cultured monthly by renewing the culture media completely until the transfer to differentiation medium with low inoculum density and further plantlet regeneration. The cell suspension method has been applied with success for multiplication and short-term maintenance (< 6 months) of undifferentiated embryogenic calli for many species, including coffee (Etienne, 2005; Bobadilla Landey et al., 2015) in order to control somaclonal variation occurrence which is highly related to exogenous auxin levels. The SE technology is applied at the industrial level since 2006.

c) Plant Growth Regulators

Plant growth regulators (PGRs) play an essential role in SE of plant species. The right balance between the kind of growth regulator and concentration is quite specific for each cultivar which makes SE an empirical process (Santana-Buzzy et al., 2007). Dedifferentiation of explant tissues is classically obtained by cultivating explants on a medium with high concentration of PGRs mainly auxin and cytokinins. Traditionally used PGRs are auxins and cytokinins, but other groups of growth regulators are being tested in established protocols, such as ethylene, oligosaccharides, jasmonates, polyamines, and brassinosteroids, in order to clarify their roles in SE in many species (Jiménez, 2005; Ruduś et al., 2006; Yang & Zhang, 2010; Mira et al., 2016). For coffee SE, most of the protocols are based and adapted from the indirect protocol developed by Berthouly & Michaux-Ferriere (1996; van Boxtel & Berthouly, 1996; Etienne & Bertrand, 2003; Etienne, 2005; Gatica-Arias et al., 2008; Menéndez-Yuffá et al., 2010; Bobadilla Landey et al., 2015). **Table 1** shows the composition of the different media used for the different stages of coffee somatic embryogenesis described by Etienne et al. (2005). For *C. arabica*, only auxins and cytokinins are sufficient to induce SE.

Table 1. Composition of the different media used for the different stages of somatic embryogenesis in coffee in mg/l and μM (growth regulators) (from Etienne et al., 2005).

	Callus induction medium (C) (mg/l)	Embryogenic callus production medium (ECP) (mg/l) (except when otherwise stated)	Callus maintenance medium (CP) (mg/l)	Regeneration medium (R) (mg/l)	Embryo maturation medium (M) (mg/l)
Basic medium	MS/2	MS/2	MS/2	MS/2	MS
Casein hydrolysate	100	200	100	400	-
Malt extract	400	800	200	400	-
Thiamine	10	20	5	10	10
Nicotinic acid	1	-	0.5	1	-
Pyridoxine	1	-	0.5	1	-
Myoinositol	100	200	50	200	100
Glycine	1	20	-	2	-
Cysteine	-	40	10	10	-
Adenine sulfate	-	60	-	40	-
2,4-D	2.26 μM	4.52 μM	4.52 μM	-	-
IBA	4.92 μM	-	-	-	-
2iP	9.84 μM	-	-	-	-
BAP	-	17.76 μM	-	17.76 μM	1.33 μM
KIN	-	-	4.65 μM	-	-
Sucrose	88 mM	88 mM	44 mM	88 mM	117 mM
PH	5.6	5.6	5.6	5.6	5.6
Phytigel	2.5 g/l	2.5 g/l	-	-	-

d) Growing Environment

Another very important factor that influences SE is the culture environment, for example, gaseous concentrations in the culture flasks, pH, temperature, hygrometry or light intensity. The concentration of dissolved CO_2 or oxygen influences the development of somatic embryos (Fisichella & Morini, 2003). A theory about how CO_2 influences somatic embryo formation is through the pH alteration of the medium. In general, the pH used to start a cell suspension is 5.7 ± 0.1 . In the first days of culture there is a continuous decrease, reaching to pH 4, followed by an increase until almost the initial level. The moment when the pH is increasing again was reported to coincide with the first embryo formation (de Feria et al., 2003; Barbón et al., 2008). Changes in pH seems to be an important factor in the regulation of SE and can be controlled naturally by the growth and normal substances release from the cells (Chung et al., 2006). The cell density also influences SE. A high density of cells leads to multiplication of embryogenic cells instead of development of embryos. Embryo differentiation thus benefits from a low density. This fact is observed in many species and happens probably due to a production of some conditioning substances such as 4-hydroxybenzyl alcohol that inhibits the embryo

formation (Kobayashi et al., 2003). During *in vitro* cultivation all these parameters act together, and influence each other. The way how these SE environmental drivers interact can help to improve or predict the development of SE.

PART IV

Coffee somatic embryogenesis

IV.1. Sexual and vegetative techniques for coffee propagation

Coffee varieties in both cultivated species are traditionally propagated by seeds (Juma et al., 1994; Etienne & Bertrand, 2003). The main constraints for coffee seed propagation are associated with the uncontrolled genetic variation found in heterozygous cultivars (e.g. F1 hybrids), the short span of seed viability (Dussert et al., 2006) and the slow rates of seed production i.e. starting only after 4 years of plant development (Juma et al., 1994). Arabica pure line varieties are sold in seed form after a relatively lengthy pedigree selection process of at least 20 years (Etienne et al., 2005).

Coffee clonal vegetative propagation is possible by traditional horticultural methods like grafting and cuttings, and by more recent and advanced vegetative techniques e.g. micro-cuttings and organogenesis, which implementation can only be done in an *in vitro* culture laboratory. However, many problems have limited the use of these techniques:

- (1) The risk of disease propagation which limits the interchange of traditional techniques-derived plants across country borders (Etienne et al., 2005).
- (2) The limited numbers of orthotropic cuttings that a coffee tree can produce (6-10 per plant/year).
- (3) Rooting induction on vegetative cuttings can be very difficult, particularly in Arabica (less than 10% efficiency), normally requiring the application of hormones (Juma et al., 1994).
- (4) Finally, the most important limiting factor is associated to the low multiplication rates resulting from organogenesis and micro-cuttings (Juma et al., 1994; Etienne, 2005). A typical Arabica plantation normally requires a high planting density (4,000-6,000 plants/ha) meaning that production rates using these micropropagation techniques are economically incompatible to satisfy the market demand.

IV.2. Coffee somatic embryogenesis: How did research, experience gained and innovations promote the commercial propagation of elite clones from the two cultivated species?

Published review in *Frontiers in Plant Science* on the 12th of November, 2018.



Coffee Somatic Embryogenesis: How Did Research, Experience Gained and Innovations Promote the Commercial Propagation of Elite Clones From the Two Cultivated Species?

Hervé Etienne^{1,2*}, David Breton³, Jean-Christophe Breittler^{1,2}, Benoît Bertrand^{1,2}, Eveline Déchamp^{1,2}, Rayan Awada^{1,2,3}, Pierre Marraccini^{1,2}, Sophie Lérant^{1,2}, Edgardo Alpizar⁴, Claudine Campa⁵, Philippe Courtel⁴, Frédéric Georget^{1,2} and Jean-Paul Ducos³

OPEN ACCESS

Edited by:

Sandra Isabel Correia,
Universidade de Coimbra, Portugal

Reviewed by:

Elena Corredoira,
Consejo Superior de Investigaciones
Científicas (CSIC), Spain
Sebastien Christian Carpentier,
Bioversity International, Belgium

*Correspondence:

Hervé Etienne
herve.etienne@cirad.fr

Specialty section:

This article was submitted to
Plant Evolution and Development,
a section of the journal
Frontiers in Plant Science

Received: 23 July 2018

Accepted: 19 October 2018

Published: 12 November 2018

Citation:

Etienne H, Breton D, Breittler J-C,
Bertrand B, Déchamp E, Awada R,
Marraccini P, Lérant S, Alpizar E,
Campa C, Courtel P, Georget F and
Ducos J-P (2018) Coffee Somatic
Embryogenesis: How Did Research,
Experience Gained and Innovations
Promote the Commercial Propagation
of Elite Clones From the Two
Cultivated Species?
Front. Plant Sci. 9:1630.
doi: 10.3389/fpls.2018.01630

¹ CIRAD, UMR IPME, Montpellier, France, ² IPME, Université de Montpellier, IRD, CIRAD, Montpellier, France, ³ Nestlé R&D Center Tours – Plant Science Research Unit, Tours, France, ⁴ ECOM, Exportadora Atlantic, Managua, Nicaragua, ⁵ IRD, CIRAD, Université de Montpellier, IPME, Montpellier, France

Since the 1990s, somatic embryogenesis (SE) has enabled the propagation of selected varieties, Arabica F1 hybrid and Robusta clones, originating from the two cultivated coffee species, *Coffea arabica* and *Coffea canephora*, respectively. This paper shows how mostly empirical research has led to successful industrial transfers launched in the 2000s in Latin America, Africa, and Asia. Coffee SE can be considered as a model for other woody perennial crops for the following reasons: (i) a high biological efficiency has been demonstrated for propagated varieties at all developmental stages, and (ii) somaclonal variation is understood and mastered thanks to intensive research combining molecular markers and field observations. Coffee SE is also a useful model given the strong economic constraints that are specific to this species. In brief, SE faced four difficulties: (i) the high cost of SE derived plants compared to the cost of seedlings of conventional varieties, (ii) the logistic problems involved in reaching small-scale coffee growers, (iii) the need for certification, and (iv) the lack of solvency among small-scale producers. Nursery activities were professionalized by introducing varietal certification, quality control with regard to horticultural problems and somaclonal variation, and sanitary control for *Xylella fastidiosa*. In addition, different technology transfers were made to ensure worldwide dissemination of improved F1 Arabica hybrids and Robusta clones. Innovations have been decisive for successful scaling-up and reduction of production costs, such as the development of temporary immersion bioreactors for the mass production of pre-germinated embryos, their direct sowing on horticultural soil, and the propagation of rejuvenated SE plants by rooted mini-cuttings. Today, SE is a powerful tool that is widely used in coffee for biotechnological applications including propagation and genetic transformation. Basic research has recently started taking

advantage of optimized SE protocols. Based on -omics methodologies, research aims to decipher the molecular events involved in the key developmental switches of coffee SE. In parallel, a high-throughput screening of active molecules on SE appears to be a promising tool to speed-up the optimization of SE protocols.

Keywords: technology transfer, innovation, somaclonal variation, cuttings, bioreactors

INTRODUCTION: FROM THE PROOF OF CONCEPT OF COFFEE SOMATIC EMBRYOGENESIS TO THE MASS PROPAGATION OF SELECTED ELITE CLONES OF ARABICA AND ROBUSTA

Since pioneer studies by Staritsky (1970), Söndhal and Sharp (1977), Pierson et al. (1983), and Dublin (1984), somatic embryogenesis (SE) has been proposed as an alternative way to propagate elite coffee cultivars from the two cultivated coffee species *Coffea arabica* and *C. canephora*. Tissue cultures of these two coffee species exhibited a high embryogenic potential: the first development of embryoid (direct SE) was reported by Staritsky (1970) and high-frequency embryogenesis from leaf explant-derived friable calli (indirect SE) by Söndhal and Sharp (1977). The morphological and histological characterization of both embryogenic and non-embryogenic calli as well as the histological events leading to the embryo regeneration in direct and indirect SE technologies have been described by many authors in detail until recently (Berthouly and Michaux-Ferriere, 1996; Quiroz-Figueroa et al., 2006; Gatica-Arias et al., 2008; Padua et al., 2014; Silva et al., 2014). These pioneer studies in the seventies provided the proof of concept of coffee SE, but at that time, no selected varieties were available to be propagated. In addition, the studies showed that several well-known bottlenecks hindering the development of SE in other crops such as the difficulty to regenerate embryogenic tissues and a strong genotypic effect were very limited in coffee. Hence most research efforts focused very early on the scaling-up of embryo production in liquid medium (Zamarripa et al., 1991a,b).

The allotetraploid *C. arabica* ($2n = 4x = 44$) autogamous species is mainly grown in Latin America (>80% of world production). The species was introduced using a small number of plants, which resulted in very low genetic diversity in the field (Lashermes et al., 2009). These varieties are lineages reproduced via seeds since the species is self-pollinating. From 1990, CIRAD and its public and private partners [CATIE (Tropical Agricultural Research and Teaching Center, Turrialba, Costa Rica), ICAFEs (Central American Coffee Research Institutes), ECOM-trading (ECOM-trading is a leading global commodity merchant and sustainable supply chain management company)] created F1 hybrid varieties of *C. arabica*, based on the crossing of genetically distant American lineages and wild individuals from Ethiopia and Sudan (van der Vossen et al., 2015) to introduce more genetic diversity. These hybrids enabled a 30–60% increase in production (Bertrand et al., 2011) and also improved the aromatic quality (Bertrand et al., 2006). The selection time of these hybrids, which is significantly shorter than that of

conventional lines (8 vs. 25 years), is another argument in their favor. After their phenotyping in field conditions, their initial selection is made by family and then by plant. The product of the selection is a hybrid (heterozygous genetic structure) that has to be vegetatively propagated (cloning) to conserve all the characteristics of the selected mother plant. The annual seed requirements for orchard renewal in Latin America are estimated at several hundred million coffee plants. Given the lack of efficient protocols for horticultural propagation by grafting and cuttings, SE was considered to be the most promising tool to insure the large scale dissemination of hybrids.

Robusta (*C. canephora* var. Robusta) is a strictly allogamous species that is traditionally propagated by seeds. Because of uncontrolled pollination, the planting material generated from seeds is highly heterogeneous, and the yields of the majority of trees are poor. The strong competition between industrial crops for profitable land is putting pressure on the cultivation of Robusta. One way to increase the profitability of the crop is to distribute new clones with higher agronomic performances that can be propagated using vegetative methods like grafting or rooted cuttings. In the 1990s, Nestlé R&D-Tours created their “Core Collection” of 55 clones representative of Robusta diversity, and planted them in the field in Ecuador and Thailand (Pétiard et al., 2004). The biomolecular aspects, agricultural performances, biochemical composition, processing performances and sensory (organoleptic) characteristics of the collection were extensively studied, leading to the selection of the most advantageous clones. From this collection, those with added genetic value were propagated via SE to set up multilocation field trials in several coffee producing countries. With the collaboration of National Institutes, the genotypes were validated under local conditions before being mass propagated by SE and distributed to farmers.

In this paper, we show how three major breakthroughs led to scaling-up of SE in coffee for the mass propagation of selected Robusta clones and Arabica hybrids: (1) in the early 1990s, liquid culture media were used for the multiplication of embryogenic cells and for the production of embryos up to the torpedo-stage (Zamarripa et al., 1991a,b; Ducos et al., 1993; Zamarripa, 1993); (2) in 1995, it became possible to pre-germinate embryos up to the cotyledonary-stage using temporary immersion culture in liquid media (Berthouly et al., 1995; Etienne-Barry et al., 1999); (3) in the late 1990s, it was possible to achieve *ex vitro* germination by direct sowing of cotyledonary embryos in the greenhouse (Ducos et al., 1999; Etienne-Barry et al., 1999).

Starting from this progress, pilot processes for large scale propagation were launched for Arabica in 2002 (CIRAD/ECOM alliance, laboratory of Sebaco, Nicaragua) and for Robusta in 2003 (Nestlé R&D Tours, France). In this paper, we describe the

background and main steps in coffee SE from research up to industrial applications and summarize the key research results obtained by these two leading groups in coffee SE. Finally, we present current technological and basic research on coffee SE.

ARABICA MASS PROPAGATION THROUGH SOMATIC EMBRYOGENESIS: THE BIRTH OF A COMMERCIAL ACTIVITY

C. arabica SE: A Model for Perennial Plants Under Strong Economic and Social Constraints

Once a variety of coffee has been created by breeders, it needs to be propagated at a large scale, respecting conformity, to be subsequently distributed to growers at a fair price (Bertrand et al., 2012). In Costa Rica, Honduras, Colombia and Brazil, for example, state or cooperative organizations distribute the seeds of varieties with high germination quality and varietal purity ranging between 90 and 95% for a subsidized cost of US \$8–15/kg. In other producing countries (Guatemala, El Salvador, Nicaragua, Peru, Mexico, Panama, Bolivia, Jamaica, and Dominican Republic), basic services for variety improvement and/or seed propagation and certification, do not exist. In these countries, coffee growers have no choice but to turn to non-certified seed producers, which increases the risks of losing varietal purity (due to seed mixtures and cross-pollination), germination capacity (highly variable depending on the batch), and possibly subsequent coffee productivity and quality. Given this situation, and also taking into account that the lifespan of a coffee plant is around 20 to 30 years, it is urgent to structure the coffee seed production chain in producing countries since we know that the choice of the variety to be planted by most producers is based on either the arguments of their neighbor or on (often) outdated information provided by coffee institutes, which increasingly lack the means to fulfill their role as advisors to producers.

The propagation of F1 hybrids creates new problems, both commercial and logistical. Indeed, as these varieties are propagated vegetatively (and not by seeds), they must be distributed to small producers in the form of plantlets grown at the lowest possible cost. Four main difficulties limit the massive distribution of such hybrids: (i) the high cost of F1 plantlets, (ii) the fact that small producers have no capital, (iii) fluctuations in the price of coffee, and (iv) logistic problems in reaching small producers, who often live in remote and mountainous areas. Possible ways to overcome these difficulties are presented below.

Horticultural and Agronomic Solutions to Reduce the Cost of F1-Plantlets

Several hundred thousand F1 hybrid plants are produced each year using the SE technique (Etienne et al., 2012, 2016; Bobadilla Landey et al., 2013) by a laboratory located in Sébaco (Nicaragua) and managed by the Agritech Company. The price at the

laboratory gate is high (US \$0.5–0.6), representing the production cost of a 2 month old acclimatized plantlet derived from a somatic embryo, and the plant material is too heterogeneous.

The recent development of propagation by rooted mini-cuttings (Georget et al., 2017) significantly increased both the number (multiplication rate > 10–30) and the quality of plantlets derived from SE and subsequently divided the production costs by half (from a mean of US \$0.50–0.60 to 0.25–0.30 per plant).

Considering the Solvency of Small Producers

When leaving the nursery, a 6 month-old F1 plant propagated by SE and ready to be planted in the field has a set sales price of around US \$0.8–1.0 versus US \$0.3–0.5 for a seed-derived plantlet. The cost is still too high for the majority of coffee growers. Knowing that between 4,000 and 6,000 plantlets are required per ha, planting F1 clones represents an additional cost of US \$2,000–3,600 per ha compared to the cost of setting up a conventional hectare (between US \$4,000 and 7,000 per ha).

In fact, due to the improved productivity of these hybrids and their better resilience, it is easy to show that this extra cost can be refunded after three to four harvests (out of 15–20 harvests that represent the life span of a normal coffee plantation).

This problem could also easily be overcome if the banks considered that small coffee growers (who represent the majority of producers) are solvent. However, because coffee prices are extremely volatile, banks often do not trust small producers and enforce usurious agricultural credit rates (between 8 and 14%). Thus, only a limited number of coffee growers (the biggest) are able to take the financial risk of planting F1 hybrids (or even of modernizing their plantations).

Fluctuating Coffee Prices and Direct Trading

One way to improve the living conditions of coffee growers and their production tools is to not consider produced coffee as a raw material whose price is fixed on the stock market by speculators, but to consider it as a product with a price that can be fixed between the grower and a buyer (roaster) who will guarantee a fair price and limited fluctuation over a long period (direct trading). More and more traders, notably ECOM are turning to this type of trading. Once the price is fairly established between them, it will be easier for the coffee grower to obtain credit at a reasonable rate.

Logistic Problems Involved in Trying to Reach Small Producers and Setting Up Cooperatives

Since the remoteness of many of the coffee growers made industrialization impossible, we developed a concept of rooted mini-cuttings in a rural environment. Inspired by farmers' seed systems, which are mainly suitable for seed reproduction, networks called "cooperatives" were created. The horticultural technique of rooted mini-cuttings derived from somatic embryo-derived plants (somatic seedlings) was transferred to these women's cooperatives in order to (i) reduce the production costs of mini-cuttings, (ii) reduce gender inequalities, and (iii) facilitate access to F1 hybrids and their use. This experiment is currently being tested in three very different countries (Vietnam, Cameroon, and Nicaragua) in the framework of the BREEDCAFS

H2020 project (BREEDing Coffee for Agroforestry Systems¹) supported by the European Commission.

From a small number of initial SE-derived plantlets provided by a certified plantlet producer and renewed every year, women cooperatives commit to produce rooted mini-cuttings they then sell to coffee growers. The cooperatives also commit to pay a royalty recognizing the breeder's right each year.

Breakthrough Innovations and Research to Scale Up the Technology and Reduce Production Costs

Mass Production of Pre-germinated Embryos in Temporary Immersion Bioreactors

CIRAD was a pioneer in the development of temporary immersion bioreactors, whose advantages for micropropagation were clearly demonstrated in the coffee SE model through high yielding and reliable processes (Figure 1). Twenty clones of *C. arabica* F1 hybrids successfully reproduced using a 1 L-RITA[®] temporary immersion bioreactor (Etienne et al., 1997). Depending on the genotype, yields ranged from 15,000 to 50,000 somatic embryos per gram of embryogenic suspension. The most spectacular effect obtained by temporary immersion was a dramatic improvement in the quality of the resulting somatic embryos. While the observed proportion of normal torpedo-shaped embryos is usually 20–30% in a bioreactor or in Erlenmeyer flasks (Zamarripa et al., 1991a; Noriega and Söndhal, 1993; De Fera et al., 2003), with temporary immersion, the proportion reached over 90%. This improvement in quality is also reflected in higher plant conversion rates on agar medium (80–90%) and, more particularly, in the highly successful regeneration of plants after direct sowing on horticultural substrate (Barry-Etienne et al., 1999). The adjustment of immersion time is critical with this type of bioreactor. According to Teisson and Alvard (1995) with 1-L bioreactors, 15 min of immersion at 6 h intervals led to repetitive embryogenesis and germination of coffee embryos, whereas in an identical culture medium, 1 min immersions at 24 h intervals halted embryo development and stimulated the production of adventive embryos. The use of bigger temporary immersion bioreactors, like the new 5 liter (5L)-MATIS[®] (area 355 cm²) bioreactor designed horizontally specifically to favor embryo dispersion and light transmittance to somatic embryos, further improved (91%) embryo-to-plantlet conversion, for the first time, enabling the regeneration of large numbers of Arabica plantlets in a bioreactor (Etienne et al., 2013). Similarly, protocols for bulk mass production of pre-germinated embryos in temporary immersion bioreactors were also developed with the other crop species *C. canephora* (Ducos et al., 2007a,b).

Direct Sowing in a Nursery of Pre-germinated Embryos

To further reduce production costs, the plant conversion steps on gel media in the laboratory had to be eliminated because they require a lot of labor for transplanting, many culture containers

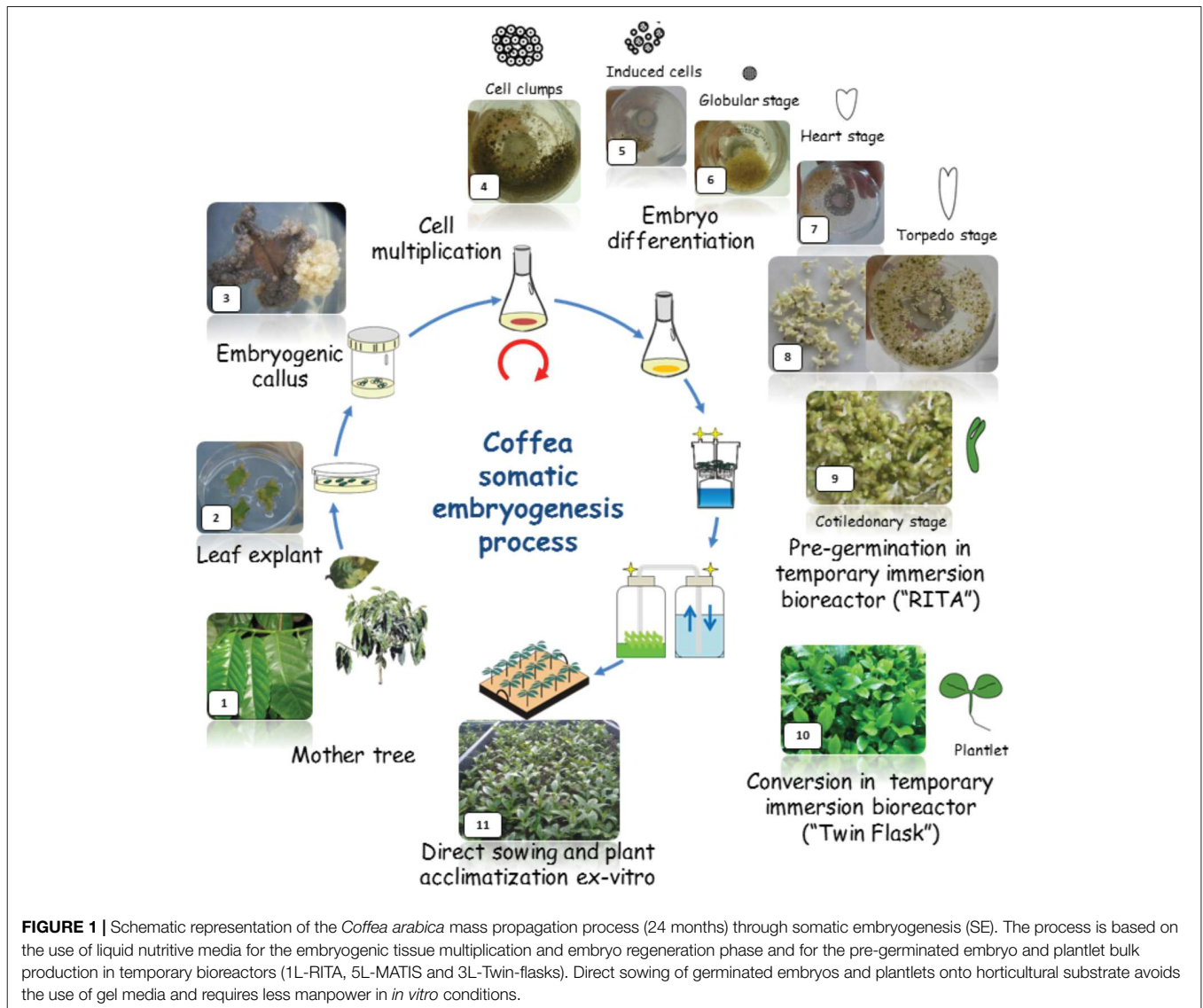
to be used and washed and a lot of shelf space. The effect of the germination conditions on the morphology of *C. arabica* somatic embryos mass-produced in a 1 L bioreactor was shown to play a determining role in successful plant conversion in the nursery after direct sowing in soil (Figure 1). Using pre-germinated embryos, direct sowing resulted in a highly successful conversion of embryos into plantlets. A culture density of more than 1,600 embryos per 1 L-bioreactor affected embryo morphology positively as it caused higher embryonic axis elongation (+4 to 5 mm) (Barry-Etienne et al., 1999). Furthermore, direct sowing reduced handling time to 13% and the required shelf space to 6.3% of those required by conventional acclimatization of plants developed on gel media.

In this way, approximately 800 Arabica somatic embryos were obtained in a 1L-RITA[®] bioreactor, 86% of which reached the pre-germinated stage but with morphological heterogeneity (Barry-Etienne et al., 2002a). For example, the pre-germinated embryos can be sub-divided into three categories according to cotyledon area which is proportional to embryo age: “small” (32%), “medium” (36%), and “large” (4.5%), the remaining material comprising non-germinated embryos. Embryos with an intermediate (medium) cotyledon area (0.86 cm²) were shown to not only have the best plant conversion rates *ex vitro* (63%), but also resulted in vigorous plantlets. Compared to *in vitro* germination conditions, mortality was higher in nursery conditions, but better plant development was obtained (Barry-Etienne et al., 2002b). The quality of plantlets produced in *ex vitro* conditions was reflected in the better growth of the aerial and root systems, but also by similar morphological, mineral and water status characteristics to those of seedlings. Direct sowing of pre-germinated somatic embryos therefore rapidly resulted in vigorous plantlets in *ex vitro* conditions, whilst eliminating the need for tricky and costly conventional acclimatization procedures. However, it was also clearly demonstrated that the heterogeneity of the embryos in the bioreactor resulted in variability in both the plant conversion efficiency in soil and plant growth in the nursery, delaying growth, mainly in plantlets derived from somatic embryos with small cotyledons (Barry-Etienne et al., 2002a).

Rooted Mini-Cuttings From Somatic Seedlings

Based on these facts, work was done to reduce the nursery plant heterogeneity and production costs. Mature tissues of *C. arabica* are known to be recalcitrant to rooting (around 8–12% rooting efficiency). However, we observed that between 10 and 25 weeks of development in the nursery, SE plantlets derived from *C. arabica* hybrids were temporarily able to root with a high success rate (up to 90%) whatever the genotype tested, before gradually losing that capacity. We took advantage of this transient rooting capacity, which was probably due to the rejuvenation process occurring during SE (resulting from the successive cell dedifferentiation/re-differentiation mechanisms), to establish a new propagation system in nursery conditions based on the serial rooting of cuttings obtained from an initial SE-derived plant batch, known as horticultural rooted mini-cutting (HRMC) (Georget et al., 2017). The excessively low embryo-to-plantlet conversion rates

¹<http://www.breedcafs.eu/>



(around 50%) during acclimatization and hardening steps in nursery that limit SE profitability can be more or less offset by the much higher HRMC multiplication rate (14 in 6 months) and better overall quality (Figure 2). Fifteen week-old rooted mini-cuttings were more uniform (2–4.5 cm vs. 1–5.5 cm for plant height distribution) and vigorous (1.41 vs. 0.81 mm for stem diameter) than SE-derived plantlets of the same age. We showed that this effect persists for several years after planting. Moreover, SV were not observed in the plants propagated by HRMC.

Long-Term Studies Led to the Control of Somaclonal Variation

To be able to use industrial scale procedures, developing embryogenic cell suspensions is the best way to ensure synchronous and massive somatic embryo production, and this process can be easily scaled up in bioreactors. Although embryogenic cell suspensions have been developed for some

major crops, this has not been the case for commercial uses. The use of embryogenic cell suspensions has frequently been associated with the increased likelihood of genetic and epigenetic instability and somaclonal variation (SV) in the regenerated plants (Etienne et al., 2016). Preliminary studies on coffee showed that seven types of phenotypic variants [Color of juvenile leaves, Giant, Dwarf, Thick leaf (Bullata), Variegata, Angustifolia, and Multi-stem] were systematically observed in all the regenerated plant batches (Etienne and Bertrand, 2001, 2003; Figure 3). For all genotypes, the degree of SV was always low (around 1.3%) for plants directly produced from embryogenic callus or 3 month old embryogenic cell suspensions. Thereafter, the frequency of variants increased significantly, reaching 6, 10, and 25% in plants produced from suspensions aged 6, 9, and 12 months, respectively. It was also observed that the severity of SV increased with the age of the embryogenic suspension with the successive appearance of mainly of “Dwarf,” “Angustifolia,” and “Multi-stem” phenotypic variants.



FIGURE 2 | Somatic embryogenesis of *C. arabica*: model of industrialization and professionalization. **(A)** Embryogenic cell suspension in liquid medium; **(B)** embryo differentiation in liquid medium (*torpedo stage*); **(C)** embryo development phase in TIS (temporary immersion system); **(D)** germinated embryos in Twin-Flask Bioreactor (above: cotyledonary embryos obtained in 1L-RITA); **(E)** Conversion into plantlets of pregerminated embryos onto horticultural substrate; **(F)** rooted cuttings in plug trays; **(G)** industrial hybrid nurseries with plastic pots and peat moss substrate. **(H)** Rooted cutting ready for field planting.

We conducted further research into industrial culture conditions, which were expected to be weakly mutagenic thanks to the combined use of a short proliferation period (6 months) and low auxin supply (0–1.4 μM 2,4-D) and compared two proliferation systems: secondary embryogenesis and embryogenic cell suspensions (Bobadilla Landey et al., 2013). Independently of the proliferation system studied, amplified fragment length polymorphism (AFLP) and methylation-sensitive amplified polymorphism (MSAP) molecular analyses on 145 SE derived-plants showed that genetic and epigenetic polymorphisms between mother plants and SE derived-plants were always low but present, ranging from 0 to 0.003% and 0.07 to 0.18%, respectively. Based on MSAP analyses, no plant was found to cumulate more than three methylation polymorphisms. It is expected that more powerful molecular tools could evidence more variations. Moreover, *C. arabica* is a young allopolyploid still having the most of its genes in duplicated copies (Cenci et al., 2012). It could be hypothesized that the impact of genetic or epigenetic variations on phenotype was restricted because of the buffer effect due to polyploidy.

Massive phenotypic observations in the nursery and field plots revealed very low levels of SV (0.9% in 800,000 plants), probably due to the short duration of the proliferation period applied as cell culture aging has clearly proved to strongly enhance SV (Etienne and Bertrand, 2003; Bobadilla Landey et al., 2015). Cytological analyses showed that most coffee somaclonal variants presented abnormal chromosome numbers (41–43, 45) instead of 44 as in phenotypically normal plants of *C. arabica*. The presence of mitotic aberrations, including double prophase, lagging chromosomes, aneuploid and polyploid cells, has previously been described in leaves (Ménendez Yuffá et al., 2000; Zorinians et al., 2003) and embryogenic calli

(Ménendez Yuffá et al., 2000) of *C. arabica* but not in the later steps of SE, and without establishing any relation with SV.

Stressful experimental conditions were also applied by using extended proliferation periods (4, 12, and 27 months) for three independent embryogenic cell lines established in the presence of high concentrations of growth regulators (4.5 μM 2,4-D, 17.8 μM 6-benzylaminopurine [6-BA]) to understand the mechanisms of culture aging and their relations with SV (Bobadilla Landey et al., 2015). This showed that the proliferation time strongly affected the SV frequency in the regenerated plants ($n = 180$) and in a very similar way in the three embryogenic cell lines. No variants were found after 4 months of proliferation whereas 30 and 94% phenotypic variants were observed in plants derived from 12 and 27 month cultures, respectively. Like the plants produced in industrial conditions, phenotypic variants systematically showed abnormal chromosome numbers.

Our research thus clearly demonstrated that SE based on embryogenic cell suspensions cultured for less than 6 months with low auxin supply was efficient and reliable for true-to-type propagation of selected *C. arabica* varieties with over 99% of regenerated coffee trees in full conformity with the mother plant morphologically and conserved their original physiological and agronomic traits.

Arabica Somatic Embryogenesis: From Pilot Scale to Commercial Production Scaling-Up Arabica Somatic Embryogenesis

Between 1995 and 2001, CIRAD, in partnership with the coffee institutes of Central America (ICAFE-Costa Rica, IHCAFE-Honduras, PROCAFE-Salvador, ANACAFE-Guatemala) and CATIE (Costa Rica), conducted laboratory research that led to

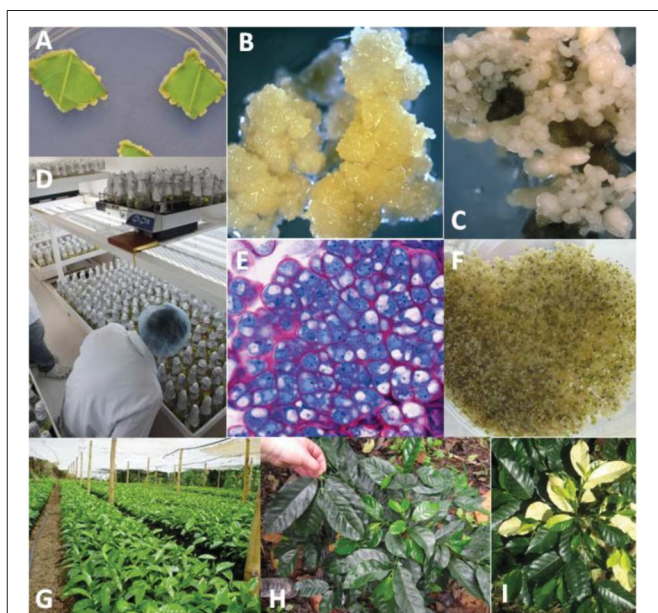


FIGURE 3 | Illustration of the production, multiplication, and plant regeneration from embryogenic cells in coffee (*C. arabica*). **(A)** Formation of a primary callus made of meristematic cells at the edge of leaf explants 3 weeks following *in vitro* introduction; **(B)** embryogenic callus obtained 6 months after *in vitro* introduction; **(C)** regeneration of somatic embryos from the embryogenic callus; **(D)** proliferation of embryogenic cell suspensions and subsequent regeneration of somatic embryos in Erlenmeyer flasks agitated on orbital shakers in an industrial coffee micropropagation laboratory (CIRAD/ECOM unit, Sebaco, Nicaragua); **(E)** histological appearance of embryogenic cell clusters in suspension cultures; embryogenic cells have a high nucleus cytoplasm ratio and dense cytoplasm rich in soluble and reserve proteins; **(F)** mass regeneration of torpedo-stage somatic embryos from embryogenic cell suspensions; **(G)** coffee SE-derived plants in the nursery (Cartago, Costa Rica); **(H)** “Dwarf” phenotypic variant (on the right) compared to a normal plant on the left; **(I)** “Variegata” phenotypic variant (from Etienne et al., 2016).

mass production in bioreactors and the possibility of sowing in the nursery. Over this period, tens of thousands of plants were produced every year. In the meantime, reassuring data were obtained concerning the low rate of SV and the better performances of F1 hybrids compared to pure line varieties, especially their remarkable adaptation to both full sun and agroforestry conditions (Bertrand et al., 2011).

In 2002, sufficient information had been collected to test F1 hybrids and SE on a larger scale. In partnership with a private company (ECOM trading group), CIRAD decided to apply the SE method to four *C. arabica* F1 clones first at the pilot scale, and subsequently at industrial scale in Nicaragua. The first commercial production of vitroplants took place in 2008. Between 2008 and 2010, more than 5 million pre-germinated embryos were produced in the laboratory and nearly 2 M hybrid plants were distributed to coffee growers mainly located in Nicaragua (Etienne et al., 2012). Over this period, several technical advances in the SE process helped support this change in scale. The first was the mastery of embryogenic suspensions and the production of somatic embryos in liquid medium in Erlenmeyer flasks. This innovation resulted in a

dramatic reduction in production costs associated with the cost of labor, and also contributed to the synchronization and homogenization of plant material development compared to the original protocol – asynchronous repetitive SE in RITA® bioreactors. Two distinct developmental stages were characterized, “cell aggregates” in the multiplication phase (Figure 2A) and torpedo-shaped embryos at the end of the differentiation phase (Figure 2B). The highly embryogenic potential of Arabica cell suspensions was also reported (e.g., 250,000–500,000 embryos/mL of cell suspension).

In 2010, *C. arabica* SE was relatively well mastered at the semi-industrial level (Etienne et al., 2012; Bobadilla Landey et al., 2013), but some technical constraints remained that prevented its scaling-up: a low embryo-to-plantlet conversion rate obtained in RITA (50%) associated with a long acclimatization stage (between 20 and 30 weeks) and obvious asynchronicity (2 to 3 harvests were needed to collect all the plantlets derived from embryo sowing). However, major optimizations were achieved over the period 2011 to 2018, which allowed the production of hybrids to be scaled up. Among these optimizations was the use of large Twin-Flask bioreactors (Figure 2C) leading to an increased plant conversion rate after sowing (Figures 2D,E) (70%) and the development of rooted mini-cuttings (Figure 2F). These mini-cuttings made it possible to use the same horticultural methodologies for coffee plants as those already available for industrial production of other crops (automated filling of the containers, watering and ferti-irrigation, automatic climate management, manufacturing of customized horticultural containers, etc.) (Figures 2G,H). The result was a gain in reproducibility enabling the production of high quality SE derived-plants at a lower cost.

In the laboratory full traceability of production was also achieved with the systematic verification of mother plants using the SNP microsatellite markers previously used in our laboratory (Combes et al., 2000). Similarly, to prevent the risk of a health threat linked to the presence of *Xylella fastidiosa*, mother plants, as well as some progeny plants, are systematically indexed using a commercial detection kit² to prevent the risk to plant health linked to the presence of *X. fastidiosa* reported to be detected in coffee plants elsewhere (Jacques et al., 2016).

Regarding SV, even though it was demonstrated that the new SE industrial protocol generated very few variants (see above) a strategy for risk dilution was also created (Etienne et al., 2016). It consists in (i) introducing full production traceability through the creation of lots and production lines (plants derived from the same embryogenic calli), (ii) limiting the number of plants produced per line (100,000 plants) and the cell suspension age to 6 months, and (iii) early screening of somaclonal variants in nurseries, given that most coffee mutants are phenotypically detectable at this stage. These different measures allowed us to routinely obtain mutation rates of less than 0.5%. Today, the CIRAD/ECOM alliance produces between 2 and 3 M hybrid plants per year in Nicaragua, and regularly exports several hundred thousand to neighboring countries (El Salvador, Mexico, Guatemala, Costa Rica, and Peru).

²<http://www.agdia-biofords.com/en/prod>

The Professionalization of the Production of Arabica Hybrid Vitroplants

In the absence of private seed companies in the coffee business, one of our main objectives was to professionalize the production of hybrid vitroplants of *C. arabica* to ensure their distribution on a very large scale (several hundred million), with full traceability, excellent horticultural and phytosanitary quality (i.e., disease and pest free) and following a well-defined and innovative technical itinerary. In recent years, the CIRAD/ECOM alliance has initiated this process in Nicaragua with the help of the World Coffee Research (WCR), which supported the verification and certification program for varieties and producing nurseries. WCR Verified? is the first global standard to confirm that producers of SE-derived F1 hybrids and associated nurseries produce healthy and genetically pure plants³. A catalog of WCR-recommended varieties was published in 2017, including the new F1 hybrid varieties jointly developed by CIRAD/ECOM, which represents a first step in the restructuring of the coffee seed sector⁴. Finally, ECOM is currently registering the new hybrid varieties in the UPOV (International Union for the Protection of New Varieties of Plants) in order to recognize breeders' selection work, limit free access and consequently better regulate their propagation.

Technology Transfers and Somatic Embryogenesis Democratization for the Propagation of Selected Arabica Hybrids Industrial Prototype in Nicaragua

The CIRAD-ECOM alliance promotes participatory research and assists farmers with this new hybrid plant material (Figure 4), which is a complementary activity to that of SE technology transfer. This collaboration led to the construction of an *in vitro* production unit in Sebacco (Nicaragua) and associated industrial nurseries in Matagalpa (Nicaragua) and Cartago (Costa Rica) that are still active today. In the last 10 years, more than 20 million Arabica hybrid plants have been planted in Central America.

Democratization of Arabica SE and Rooted Mini-Cuttings (Cameroon, Vietnam, Nicaragua)

Dissemination procedures must include an economic perspective and prevent monopolistic practices and exclusive control of new genetic material. The Mesoamerican experience paves the way for mass production and easy dissemination to farmers of millions of rooted mini-cuttings derived from SE (Georget et al., 2017). The objective of the BREEDCAFS European project (2017–2021) is to popularize these methods in Asia, Africa, and Latin America to allow the mass propagation of hybrids and the diversification of the actors involved in this propagation. In this context, procedure handbooks were distributed to partners, and theoretical and practical courses in technology transfer of SE started in 2017 for personnel from IRAD (Institute of Agricultural Research for Development, Cameroon) and AGI (Agricultural Genetics Institute, Vietnam) (Figures 5A,B).

The role of women in coffee production is often considered minor, despite the fact that they are traditionally responsible for

³<https://worldcoffeeresearch.org/work/seed-and-nursery-verification-program/>

⁴<https://worldcoffeeresearch.org/work/variety-intelligence>



FIGURE 4 | Promoting participatory research and assistance to farmers with new hybrid plant material. **(A)** Farmer showing “his” Arabica hybrid plants in a farm located in Naranjo at 1300 m a.s.l., Alajuela province, Costa Rica; **(B,C)** visit of a farmer in San Ramon (1050 m a.s.l.), Alajuela province, Costa Rica; **(D)** planting of Arabica hybrids in “on farm” evaluation trials in the Són La province, Vietnam.



FIGURE 5 | Technology transfer of *C. arabica* SE and democratization of the rooted mini-cuttings companion method. **(A,B)** Training in SE at CIRAD (France) for researchers from Cameroon and Vietnam. **(C,D)** Women's mini-cuttings cooperative in Nicaragua at the “La Cumplida” farm, Matagalpa district.

seed production in micropropagation laboratories and micro-cuttings/micro-grafting nurseries. Our objective is to promote the setting-up of women's mini-cuttings cooperatives. In Cameroon, transfer of the technology is underway through COPAFERLOS, a cooperative of women from the coastal, west and south-west regions, which plays an important role in the education and promotion of women in agriculture. Women's mini-cutting cooperatives have been created in Nicaragua in the Matagalpa district (“La Cumplida” farm, Figures 5C,D) and in the north western provinces of Són La and Đìền Biên Phủ in Vietnam. The goal of the project is to assess the impact of these cooperatives

on women, particularly in improving access to, and control of, benefits derived from project participation. Innovation platforms have also been set up to promote dialog among stakeholders of the coffee value chain with the goal of facilitating the acceptance and efficient insertion of both propagation by SE/mini-cuttings and Arabica F1 hybrids in local economies.

MASS PROPAGATION OF ROBUSTA BY SOMATIC EMBRYOGENESIS

Since three decades, the Nestlé R&D Center Tours – Plant Science Research Unit has developed intensive research on SE and technological transfers to ensure the large scale dissemination of improved Robusta clones.

Robusta SE in Liquid Medium: Highlighting a Remarkable Mass Production Potential

Multiplication of the Embryogenic Callus

For the two Arabica and Canephora (cv. Robusta) cultivated species, both direct and indirect SE processes have been established showing better performances in the latter species. Direct SE allows the quick embryo regeneration (4 months after *in vitro* introduction) directly on the leaf explants without visible embryogenic callus production. However, as this process is highly asynchronous and leads to the production of low embryo quantities, its use was very early limited to basic studies on cell dedifferentiation/redifferentiation through histological approaches (Berthouly and Michaux-Ferriere, 1996; Quiroz-Figueroa et al., 2006), gene expression (Arroyo-Herrera et al., 2008), epigenetic regulations (Nic-Can et al., 2013), proteomics (Mukul-López et al., 2012), or to the establishment of *in vitro* germplasm core collections. Conversely, indirect SE was investigated in order to establish mass propagation protocols based on the addition of an embryogenic callus multiplication step, starting with *C. canephora*, the less recalcitrant species. These protocols had to be further adapted for Arabica.

The multiplication of embryogenic cells in liquid medium is a key step because it greatly and rapidly scales up the number of potential embryos that can be produced. The first establishment and growth kinetics of a coffee cell line were reported in the early 1990s for a model Robusta clone (R2) of *C. canephora* (Zamarripa, 1993). R2 embryogenic callus were induced and multiplied in Yasuda et al. (1985) medium containing 6-benzylaminopurine (BA) at 1.0 mg L⁻¹ as the only growth hormone. The R2 cell line was cultured in 0.25-L Erlenmeyer flasks starting with an initial cell density of 10 g FW L⁻¹. Growth occurred without an exponential phase and a maximum cell density of 60 g FW L⁻¹ was reached after 45 days when the carbohydrates were consumed. This cell line consisted of clumps ranging from 500 to 2,000 μm in size, composed of two cell types (Figures 6A,B). In the center of the clumps, the cells were vacuolated and highly loaded with voluminous starch grains. At the periphery, cells presented a high nucleoplasmic ratio, a slightly vacuolated cytoplasm containing small starch

granules and frequent mitotic figures (Figure 6C). These cells, also characterized in Arabica embryogenic cell suspensions, which can be defined as “embryogenic potential” cells, were clustered in peripheral nodules and gave rise to new nodules via their mitotic activity.

Expression Step

The production of R2 embryos in 0.25 L flasks was induced by transferring embryogenic clumps to a fresh Dublin (1984) medium consisting of macro and micro MS salts supplemented with 5.0 mg L⁻¹ BA, 40 mg L⁻¹ adenine, and 400 mg L⁻¹ malt extract. The initial cell density was identified as the key factor in causing the cell suspension to produce embryos, showing that low densities of 0.5–1.0 g FW L⁻¹ were optimal. After 6 weeks of culture, the total embryo concentration reached a plateau as high as 240,000 L⁻¹ with 40% of the embryos at the torpedo-stage (Figure 6D; Zamarripa et al., 1991a,b). On the contrary, high inoculation densities strongly inhibited SE. The same observation is true for the development of SE, expressed by the percentage of torpedo-stage embryos: the lower the initial density, the better their development. Similarly, inoculation density has shown to be a determinant factor for the regeneration of Arabica suspensions. Weekly renewal of the medium partly eliminated this inhibition. The inhibiting activity of the medium conditioned by cells cultured under high cell densities was related to low molecular weight factor(s), below 1 kDa, having a hydrophobic characteristic (Ducos et al., 2007c). After 6 to 8 weeks, the embryos were collected and cultured on solid medium containing 0.2 mg L⁻¹ BA for 4 weeks during which they turned green and acquired well-developed cotyledons. The germination steps were conducted *in vitro* by transferring the somatic embryos onto a similar medium, but without BA. With this two-media sequence, up to 70% of the embryos developed into plantlets bearing two to three pairs of leaves. When these plantlets were acclimatized in the greenhouse, the survival rate was more than 90%.

In a bioreactor, critical parameters for the success of mass regeneration of coffee embryos were investigated using a mechanically stirred fermentation apparatus operating at a 3 L working volume (Setric SGI, model SET4CV) (Ducos et al., 1993, 2007c). Although the agitation was provided by a 4-blade “cell-lift” propeller to minimize the shear stress, an initial agitation of 100 rpm had a detrimental effect on growth. Consequently, agitation was kept at the lowest level (50 rpm) until day 21 and then gradually increased to 100–120 rpm. The optimal initial oxygen transfer rate (kLa) was 5 h⁻¹. The bioreactor was charged with R2 embryogenic cells at a rate of 0.5 g FW L⁻¹. A maximum production of 180,000 embryos L⁻¹ was reached on day 58. The first torpedo-shaped embryos appeared on day 28 and 70% of the embryos had reached this stage at the end of the culture.

Mass production of *C. arabica* SE in a 5-L stirred bioreactor were reported by Noriega and Söndhal (1993). A total yield of 45,000 embryos were recovered within 3 months. De Feria et al. (2003) produced *C. arabica* SE in 2-L bioreactors with a rate of 70,000 embryo per g FW of inoculated calli. Beside the facility of the scaling-up of the volume which reduces the labor workforce, another advantage of the bioreactors is the possibility they offer to control various parameters as the aeration rate and

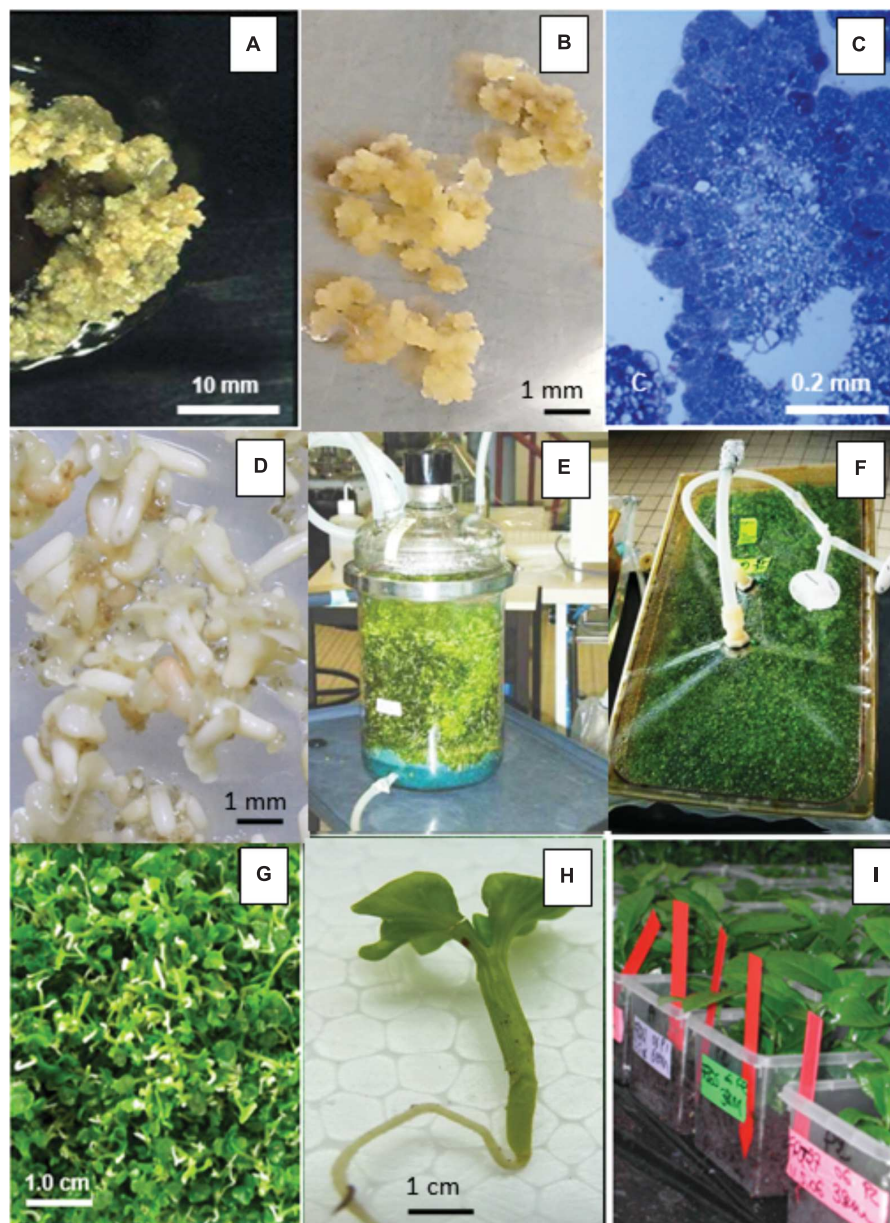


FIGURE 6 | Robusta somatic embryogenic tissues at different stages. **(A)** Primary embryogenic callus on a leaf explant, **(B)** embryogenic calli, **(C)** histological view of an embryogenic clump, **(D)** expression: torpedo-stage embryos at the end of the phase, **(E)** pre-germination: 10 L glass temporary immersion bioreactor (upper recipient), **(F)** Pre-germination: 10 L Box-in-Bag disposable bioreactor (upper recipient), **(G)** pre-germination: aspect of the cotyledonary-stage embryos, **(H)** selected cotyledonary embryos (pre-germinated embryos), **(I)** *ex vitro* germination test in the Nestlé R&D Tours greenhouse.

the gas composition in CO_2 , ethylene (C_2H_4) and particularly in oxygen dissolved concentration (DO_2). Various works reported the importance of aeration in coffee SE in stirred bioreactor. De Feria et al. (2003) showed that high DO_2 induced globular SE differentiation in Arabica, but for production of torpedo-shaped SE lower concentrations DO_2 (<50%) are needed. In the same species, Barbón et al. (2008) reported that the ventilation with an air mixture enriched with CO_2 at 5% led to more embryos than the control without CO_2 enrichment. In cyclamen, the regeneration ability of cell suspensions after being cultured in

bioreactors with CO_2 accumulation was better than those after culture in bioreactors without CO_2 accumulation (Hohe et al., 1999).

Application to Selected Clones From the Core Collection

It was confirmed that auxins were not required to obtain embryogenic calli and to multiply them in the majority of Robusta genotypes. Contrary to Arabica, these steps could be managed using a medium containing only a cytokinin as plant

growth regulator. However, for some genotypes, even if direct SE occurs during induction, the quantity of embryogenic callus is insufficient and, for the first 5 months, the foliar explants need to be cultured on Pierson medium (Pierson et al., 1983), containing indole butyric acid (IBA) at 5 mg L^{-1} .

At each subculture, the callus should be carefully selected to maintain a callus stock available on solid medium. This is a repetitive job that requires well trained operators and whose difficulty can be illustrated by the notion of “balance” (Figure 7). The ideal situation is an “equilibrium” between the undifferentiated status of the callus and their embryogenic potential. Typically, some genotypes “lean” to the left, others to the right. Agar, darkness, auxin, and high densities can help to limit precocious differentiation. Inversely, Gelrite, light and low densities facilitate differentiation.

To establish the cell line, a protocol was validated according to which the culture volumes are progressively increased by maintaining high concentrations of biomass, to avoid precocious differentiation, and by keeping the ratio of the volume medium to the volume of recipient identical throughout (Ducos et al., 2007b). At T0, 0.1 g friable callus is transferred into 10 mL of liquid medium in 25 mL flasks. After 2 weeks, the medium is discarded and the tissues transferred into 20 mL of fresh medium in 50 mL flasks. This is repeated every 2 weeks: 50 mL in 100 mL flasks at 4 week, 100 mL in 250 mL flasks at 6 week. In this way, each cell line is started with 0.1 g of selected callus and will produce 1.0 to 2.0 g of callus after 8 weeks of multiplication in liquid Yasuda medium.

True-to-Type Status of the Regenerated Plants

In order to validate Robusta propagation via SE in liquid medium, the true-to-type quality of the regenerated trees was checked in field conditions (Ducos et al., 2003). Field trials were run in five coffee-producing countries: Philippines, Thailand, Mexico, Nigeria, and Brazil from 1996 to 2000. These trials represented a total of 12,000 somatic seedlings from 10 clones. In the Philippines and in Thailand, a total of 5,000 trees of five clones (FRT01, 03, 06, 07, and 103) originating from 5 to 7 month old embryogenic cell lines were compared with control trees derived from *in vitro* axillary budding (microcuttings). No phenotypic variant trees were detected. No significant differences in the morphological traits studied and the characteristic yield between trees regenerated were observed using the two methods. Similar to what has been observed in Arabica, these studies in the Robusta species thus confirmed that this method of propagation preserves the agronomic traits to be applied for large-scale applications.

Pilot Scale Process for the Production of Pre-germinated Embryos

Description of the Process

In 2003, a standard procedure was established to implement the process schematized in Figure 8 at pilot scale (Ducos et al., 2007b). Typically, a batch lasts 6 months, from the initiation of the cell lines up to harvesting the embryos from the bioreactors (steps 3+4+5). Over a period of 3 years, starting from 0.1 g of selected callus from solid medium, we obtained 1.0 kg of callus after multiplication in liquid medium, then 4.4 M pre-germinated

embryos. To produce 1.0 M pre-germinated embryos per year, a full time employee was required in the laboratory. Depending on the clone, the embryo-to-plantlet conversion rate varied from 25 to 67%, an average of 41% for 17 clones.

Pre-germination in a 10 L Temporary Immersion Bioreactor

To scale up the pre-germination step, we developed a temporary immersion system comprising two glass bottles (a 10 L jar containing the embryos and a 5 L jar containing the culture medium) (Figure 6E). The collected embryo populations differed notably in size, from the early (1 mm) to fully expanded cotyledon stage (20 mm). The only embryos scored as “pre-germinated” are those with an hypocotyl longer than 5 mm which applied to only half the embryos. Light was obviously a limiting factor during culture, as it can only penetrate the first few centimeters of the biomass. We thus hypothesized that a horizontal design would be better than a vertical one as it would provide more light to the embryos thanks to a greater surface-to-volume ratio. The simplest solution consisted in placing a rigid polycarbonate box inside a disposable plastic bag (Figure 6F; Ducos et al., 2007a, 2008). This so-called “Box-in-Bag” bioreactor almost doubled the torpedo-to-pre-germinated-stage conversion rate: for identical inoculum, 16,000 pre-germinated embryos (Figures 6G,H) were collected per bioreactor instead of 8,600 per 10-L glass bioreactor (Ducos et al., 2010).

Ex vitro Germination in the Greenhouse in a Micro-Environment

Typically, the Robusta embryos are grown on peat or coconut fibers under a 0.5–1 m high plastic tunnel to maintain >95% relative humidity (Figure 6I). However, we observed that simply placing a transparent cover over the containers at a height of 2 to 3 cm above the embryos, significantly increased the *ex vitro* germination rates (Ducos et al., 2009). Due to the micro-environment, the germination rates increased from 42 to 66%. We showed that this positive effect was related to the CO₂ released by the horticultural media because enrichment with exogenous CO₂ promoted germination when the embryos are grown on media that emit no CO₂ such as rockwool cubes. Elevating CO₂ might facilitate the transition from heterotrophic embryos to autotrophic plants. That confirms the works of Kosai et al. (1997) who demonstrated the feasibility of photoautotrophic micropropagation using a sugar-free medium on different plant species. This method was applied to grow coffee plantlets from the interspecific hybrid *Coffea arabusta* from cotyledonary embryos in *in vitro* conditions. Cotyledonary-stage is the earliest stage capable of photosynthesizing due to a high chlorophyll contents (Afreen et al., 2002). To date, attempts to stimulate the transition to photo autotrophy in Arabica species using high concentrations of CO₂ combined with high irradiation remain to be achieved. They could improve the rates of success in acclimation and conversion of embryos into seedlings.

Quality Control/System of Quality Assurance

In 2003, a quality assurance procedure was designed to guarantee the quality of the regenerated plants. Since this date, the mother

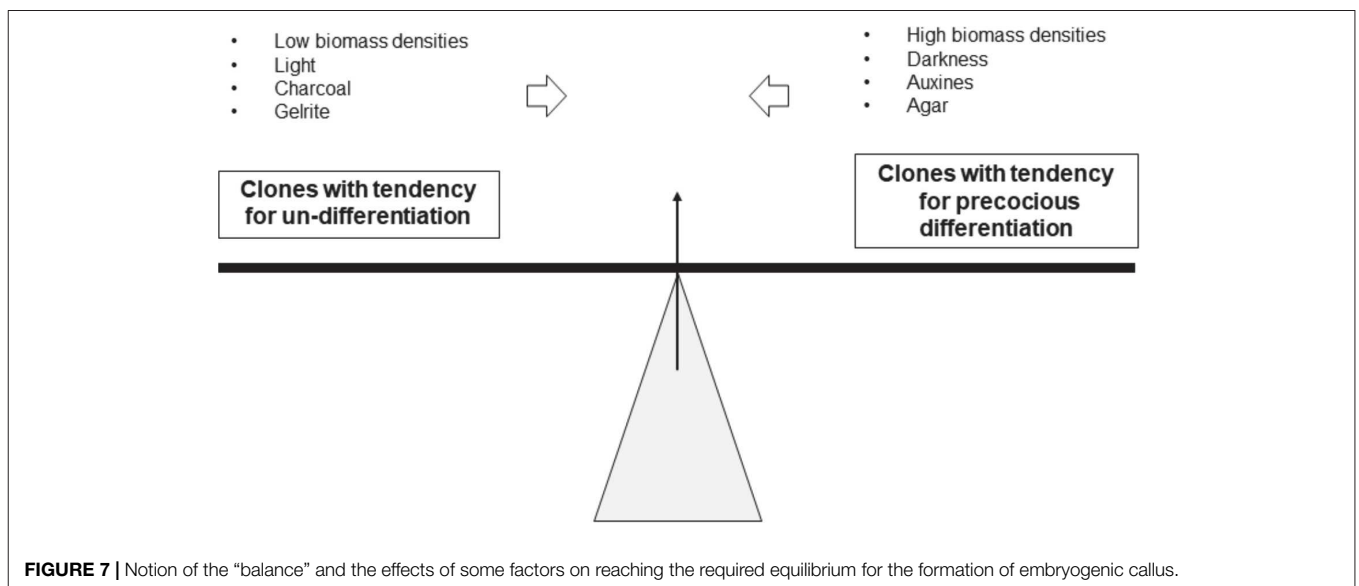


FIGURE 7 | Notion of the “balance” and the effects of some factors on reaching the required equilibrium for the formation of embryogenic callus.

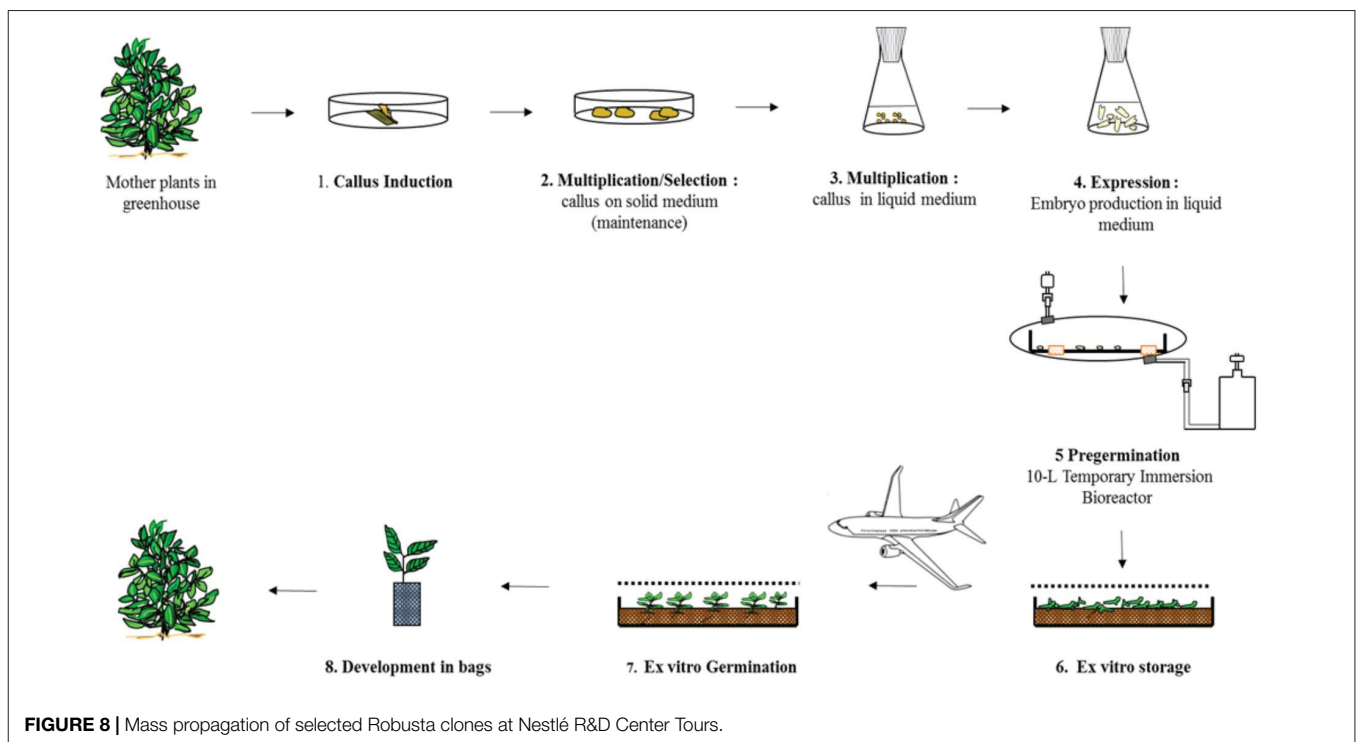


FIGURE 8 | Mass propagation of selected Robusta clones at Nestlé R&D Center Tours.

plants have been regularly checked for the absence of *X. fastidiosa* using Elisa tests, since 2011 using micro-satellites according to Legendre et al. (2013). The absence of phytopathogenic microorganisms was also checked through regular and systematic axenic tests. The clonal identities of the tissues were checked by sampling the embryogenic callus and analyzing it using specific polymorphic microsatellites. Concerning the true-to-type quality of the regenerated plants, it is well known that the frequency of SV is related to the age of the *in vitro* cultures. Consequently, to limit this risk, it was decided to limit the period between the disinfection date of the leaves and the

beginning of the production of an SE batch to 2 years. If auxin is used (IBA 5 mg L⁻¹ or 2,4-D 1 mg L⁻¹), its use is strictly limited to the first 5 months of the callus induction step. The multiplication in liquid medium is in all cases limited to 6 months.

Technology Transfers for the Propagation of Selected Robusta Clones

In the mid-2000s, Nestlé decided to strengthen its green coffee supply chain with a global initiative launched in 2010: the

Nescafé Plan⁵. For Robusta, the goals were to promote modern agricultural practices in producing countries as well as renovating old plantations with clones selected for yield, cup quality and disease resistance. The production potential of strategic markets was identified, including Mexico, Thailand, and the Philippines. To address the lack of a professional coffee plant propagation sector, the company invested in SE technology, not only to speed up the distribution of valuable Robusta plants. According to the countries, SE was either implemented in Nestlé R&D facilities (laboratory and nurseries) or locally transferred to public or private plant production units. When required, protocols (handbooks), laboratory or nursery layouts and training courses were provided.

Thailand

In 1998, the Department of Agriculture (DOA) and Nestlé Thailand decided to improve the yield of this crop by replacing the low yield trees with selected clones. The technique of SE was transferred to the DOA by means of a 1 year training of two scientists from the Chumphon Horticultural Research Centre (CHRC). The CHRC team produced and sold 0.55 M plants directly to the farmers and/or to Nestlé Thailand (Sanpote et al., 2006). A second propagation project started in 2005 with greenhouse facilities set up in the Nestlé's factory. Between 2005 and 2010, 1.8 M somatic seedlings were distributed to the farmers (Kunasol et al., 2010).

Mexico

To fill the gap between the lack of coffee beans and increasing consumer demand, Nestlé decided to transfer the SE process to a local company and to implement it there to boost the propagation of Robusta plantlets, an attractive alternative in the context of severe attacks of leaf rust disease in the Arabica species. The technology was transferred from Nestlé Mexico, the Mexican institute of Agronomic Research INIFAP to NSIP (formerly Agromod SA de CV), a private company with well-known expertise in plant micropropagation.

Due to the allogamous feature of *C. canephora*, a set of four Robusta varieties (clones) was designed to start the mass propagation. The objectives of the technology transfer were to (i) insure reliable production of embryogenic callus on solid and in liquid medium, (ii) insure efficient differentiation of the embryogenic callus into somatic embryos, and (iii) implement an effective acclimation protocol that would allow a 50% embryo-to-plantlet conversion rate. Quality control was also implemented throughout the production process. In particular, the SE-derived plant root system was evaluated prior to distribution.

As a result, the process was fully transferred between 2012 and 2014 and adapted for efficient plantlet production (Breton et al., 2017). A laboratory with a production capacity of 10 M somatic embryos per year was fully operational in 2014. Dedicated nurseries were designed to achieve an embryo-to-plant conversion mean rate of 50%.

Philippines – A Case of Technical Assistance

With the same objective, i.e., to be able to purchase coffee beans locally, Nestlé promoted the production and delivery of SE plantlets to farmers. This project has been underway since 2009. In their French laboratory, the Nestlé R&D team produces somatic embryos from existing embryogenic calli obtained from elite varieties using *in vitro* processes. The embryos are then pre-germinated in bioreactors to make them compatible with the greenhouse processes. Batches of embryos (around 1 M a year) are regularly shipped by plane from Tours to the city of Cabuyao, close to Manila. From there, they are transferred to Dalwangan (island of Mindanao) in the south of the archipelago. After this 5 day trip, the somatic embryos are sown, germinated, acclimated and grown until they reach the required size for transplantation in the field (**Figure 9**). Agronomists proceeded with the distribution of the coffee trees and provided technical assistance to farmers to ensure the plants express their full agronomic potential. Concomitantly, scientists from the Filipino Department of Agriculture were trained in the use of SE procedures at the Nestlé facility to promote the diffusion and use of the propagation technique. Between 2009 and 2017, around 1.9 M of Robusta SE trees were distributed to farmers (De Smet W., pers. comm.). This strategy enabled the start of mass propagation of elite clones of Robusta, juvenile SE plants being used as donor plants for classic rooted cutting techniques. Agronomists reported that a total number of 17 M trees were distributed thanks to this project.



FIGURE 9 | Robusta SE plants produced in Nestlé's French facility and distributed to farmers in the Philippines (**Top**); young trees in the field (Mindanao Island, **Bottom Left**); locally produced SE tree bearing fruits in a Mexican plantation (Chiapas, **Bottom Right**).

⁵<https://www.nestle.com/csv/raw-materials/coffee>

Lessons Learned

From those technology transfers, lessons were learned mainly concerning the time required to master the acclimation step of SE plants. Managing such soft and fragile plants requires strict control of the climatic and phytosanitary environment. Also, logistics and plant transport require particular care, especially in countries where the nurseries are located in remote places. Communication with the farmers themselves is also required on how to grow young SE plants, especially when local farmers are used to working with seedlings. Moreover, despite the fact scientists from governmental institutions were trained, few of them actually implemented the SE production locally, and so the initial investment for the production unit remained prohibitive. Significant SE production was only maintained when a private company was involved, highlighting the crucial role of investors in the promotion of SE technology. Impacts were observed at the national scale, as illustrated in Mexico where 10.2 million Robusta SE-derived plants were distributed to farmers from 2010 to 2017. When all these trees are fully productive and assuming a yield of 2 tons per hectare, it can be expected that around 18.6 kilotons of green coffee will be harvested in Mexico, corresponding to around 40% of local demand (Garcia Alvarez O., pers. comm.).

SOMATIC EMBRYOGENESIS, A POWERFUL TOOL FOR COFFEE FUNCTIONAL GENOMICS

Genetic Transformation and Functional Genomics

Genetically modified coffee plants have been designed in the past 15 years by different research groups worldwide with three main objectives: to develop efficient transformation and regeneration methods, to decipher gene function and/or to improve coffee genotypes. However, as neither consumers nor coffee stakeholders want GMO coffee, CIRAD has explored genetic transformation and functional genomics as powerful tools to understand gene function.

Although direct and indirect DNA delivery systems have been successfully developed to transform coffee cells, most published studies used only reporter genes and described methodological improvements (Ribas et al., 2005b; Mishra and Slater, 2012). Despite these significant methodological advances in the transformation process itself, there are very few examples of the introduction of genes of agronomic interest in coffee, reflecting the difficulty faced by many laboratories to achieve regeneration of transgenic plants. Essential progress in establishing reliable and efficient transformation protocols in coffee has been achieved by optimizing the production conditions of embryogenic cultures used as target tissues for transformation. The availability of in depth SE knowledge and well mastered procedures played key roles in removing the remaining major bottlenecks. Hence, proembryogenic masses (PEMs) were shown to be the competent target tissue (**Figure 10**) and improved transformation efficiency is also systematically

associated with increased quantities of PEMs depending on the adequate auxin/cytokinin balance, number of preliminary proliferation cycles on auxin rich medium and low mineral salt strength (Ribas et al., 2011).

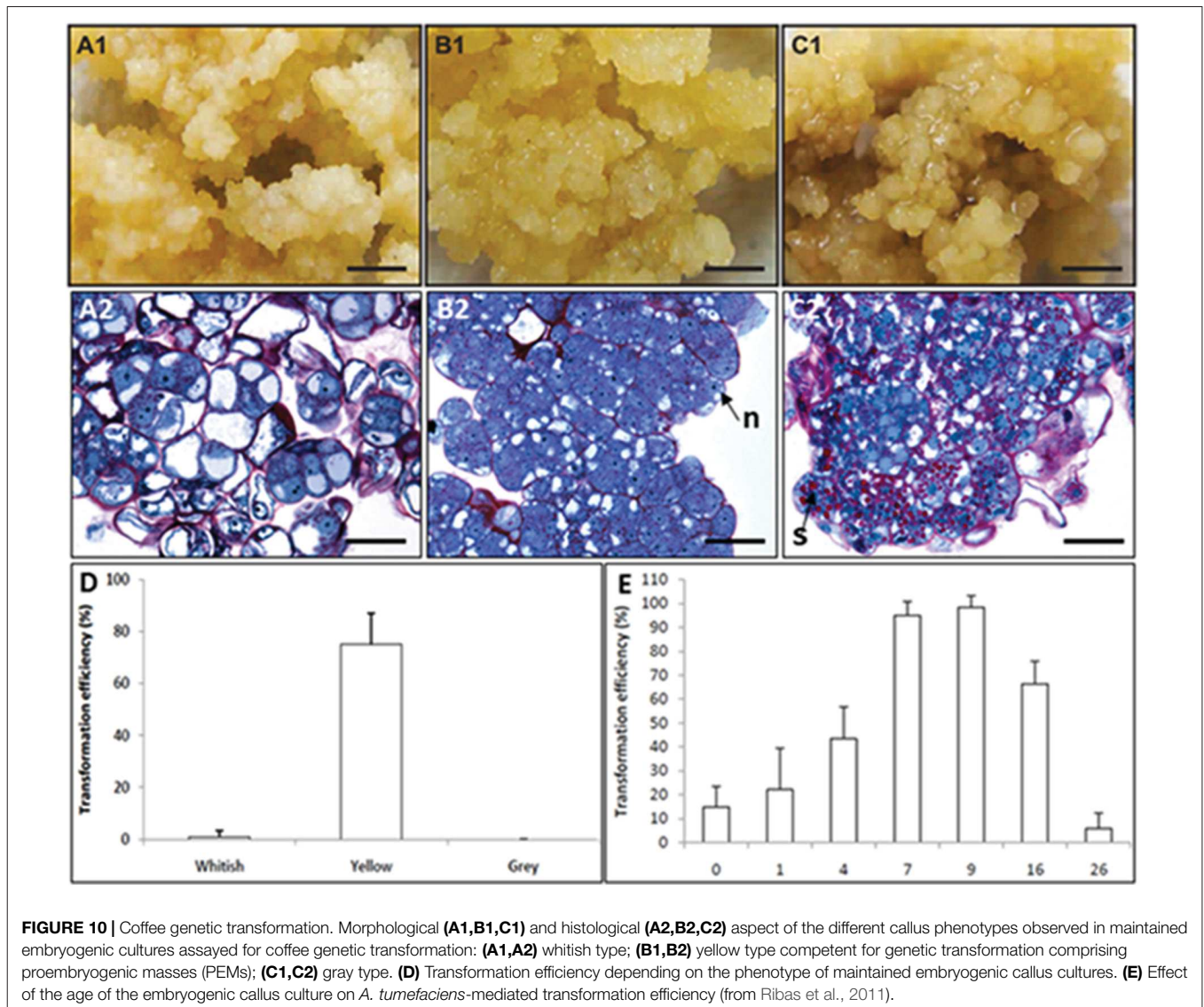
Some of the genes isolated from coffee were analyzed using a functional genomic approach comprising a theobromine synthase gene (*CaMXMTI*) to suppress caffeine biosynthesis (Ogita et al., 2003, 2004), an *ACC oxidase* gene involved in ethylene biosynthesis (Ribas et al., 2005a) and a *CaHSFA9* and *CaDREB2* gene respectively involved in the transcriptional activation of heat shock proteins (HSPs) and tolerance to heat and drought stress (Dussert et al., 2018). The promoter of *CcDREB1B* (Alves et al., 2017) and *CaWRKY1a* and *CaWRKY1b* (Petitot et al., 2013) were analyzed in transgenic coffee plants. A few studies conducted by other teams aimed to improve coffee genotypes with desirable traits (Leroy et al., 2000; Cruz et al., 2004; Ribas et al., 2005a, 2006; Arroyo-Herrera et al., 2008).

Genome Editing by CRISPR/Cas9

On July 2018, the Court of Justice of the European Union (ECJ) in Luxembourg ruled that gene-edited crops should be subject to the same stringent regulations as conventional genetically modified (GM) organisms (2001 directive)⁶. Moreover, organisms made using mutagenesis techniques developed after 2001 are not exempt from the 2001 directive relative to GMO regulation. This very rigid judgment goes against legislations in other countries such as the United States where the law exempts organisms whose genomes were modified using mutagenesis techniques, such as irradiation or genome editing, which introduce changes to an organism's DNA but does not add foreign genetic material. The coffee sector does not want GMO, but how is it going to consider the edited plants?

Genome editing technologies, which are an unprecedented technological breakthrough, provide novel opportunities for functional genomic studies and molecular breeding but rely on highly efficient genetic transformation and plant regeneration methods, which could be bottlenecks in the process (Altpeter et al., 2016; Buell and Voytas, 2017). One of them, the CRISPR/Cas9 technology is simple, versatile and cheap and provides a valuable means of creating targeted mutations and sequence replacements in metazoan and plant genomes with high specificity (Ding et al., 2016; Quétier, 2016). Although reports of genome editing in tree species are still limited (Jia and Wang, 2014; Fan et al., 2015; Zhou et al., 2015; Nishitani et al., 2016; Wang et al., 2018), this technology was recently successfully used to introduce mutations in the coffee genome (Breitler et al., 2018). Nevertheless, the use of SE as the method for regenerating genome-edited plants could restrict the choice of targeted genes to those that are not essential to the somatic embryo development and germination steps. For other genes, an alternative is to set up for coffee the shoot regeneration method used for other woody species. Indeed, poplar and apple tree albino plants were

⁶<https://www.nature.com/articles/d41586-018-05814-6>



obtained using the same method of genetic transformation and further regeneration. Leaf disks were dipped in *Agrobacterium* culture and cultured first on callus-induction medium and then on screening medium, to induce adventitious buds and finally shoot formation (Fan et al., 2015; Nishitani et al., 2016).

The recent advances in coffee genomics, including the improvement of coffee genetic transformation techniques and genome editing (Alpizar et al., 2008; Ribas et al., 2011; Déchamp et al., 2015; Breidler et al., 2018), now open the way for studies of the molecular and genetic determinism of many important agronomic traits, leading to the identification of molecular markers and candidate genes that can now be routinely functionally validated thanks to efficient genetic transformation pipelines, use of cryopreserved competent embryogenic calli and well established bioassay conditions (Marraccini et al., 2011, 2012; Vieira et al., 2013; Mofatto et al., 2016).

CURRENT RESEARCHES ON COFFEE SOMATIC EMBRYOGENESIS

Unraveling the Molecular Mechanisms Underlying the SE Process

As discussed above, research on SE has led to the successful industrialization of the process for the large-scale dissemination of improved coffee varieties of both *C. arabica* and *C. canephora* species. Over the last 12 years, intensive research has enabled the design of mastered and reproducible SE protocols, well-suited for all commercially propagated genotypes at acceptable production costs.

However, like for the other crops, coffee SE research remains mainly empirical, characterized by a low-throughput and hypothesis-driven research, leading to an overall slow technical progress. For example, the development of culture conditions for the establishment at the industrial level of embryogenic cell

suspensions in *C. arabica* and the mass regeneration of somatic embryos have been very long (10 years) and laborious, carried out only by an empirical approach. The lack of knowledge about the cellular and molecular events taking place during these developmental stages makes these stages real black boxes. While a scale-up is needed to meet increasing market demand – estimated at 50–100 million vitroplants per year – current overall production cannot meet this huge demand, mainly due to the lack of efficiency of certain steps in the SE process. We believe that the lack of understanding of the molecular mechanisms underlying the reprogramming of somatic cells is the main limitation to improving coffee SE protocols.

Recently, Campos et al. (2017) provided an overview of which achievements and molecular insights have been gained in coffee SE and encourage researchers to invest further in the *in vitro* technology and combine it with the latest omics techniques. Since the last 16 years, several works have evidenced the potential of omics technologies to optimize coffee SE. For example, Silva et al. (2015) showed higher expression levels of two BBM homologous sequences in embryogenic calli and cell suspensions of *C. arabica* when compared to non-embryogenic calli. They also identified SERK orthologs in embryogenic cell suspensions (Silva et al., 2014). Arroyo-Herrera et al. (2008) showed that an overexpression of a heterologous gene of WUS led to a huge increase in the production of somatic embryos. Quiroz-Figueroa et al. (2002) evidenced differential gene expression in embryogenic and non-embryogenic clusters from cell suspensions in *C. arabica*. Tonietto et al. (2012) reported the proteomic analysis of developing somatic embryos of *C. arabica* leading to the identification of 14 proteins linked to different developmental stages during SE as enolase and globulin S11 for torpedo stage. The proteome profile of embryogenic calli was recently published (Campos et al., 2016). It was shown that epigenetic regulation of genes could play a role in coffee SE. Nic-Can et al. (2013) proved that the genes LEC1, BBM1, and WUS related gene WOX4 are under epigenetic control in relation to the embryogenic capacity in *C. canephora*. However, these studies were carried out on several different protocols, in research laboratories and using non-optimized and often asynchronous SE regeneration processes giving rise to the co-existence of several stages of development in a same sample.

Today, a reliable, synchronized and efficient industrialized SE process, added to the recent boom in -omics technologies represents a real opportunity to understand the molecular events involved in the key developmental switches of coffee SE and those associated with the main phase changes, for example, at the transcriptomic (gene expression) level, associated with totipotency.

To better understand SE, it is important to look at the process as a sequence of events or developmental stages. These stages were chosen and characterized based on well described morphological and anatomical changes that occur when switching from one stage to another (Jayasankar et al., 2003). Nine developmental stages have been characterized in the indirect SE of dicots: explant, primary callus, embryogenic callus, PEMs, globular, heart-shaped, torpedo-shaped and cotyledonary embryos, and plantlet (**Figure 1**). During direct SE, globular

embryos appear directly on the explant with no intermediate callus phase and as this process is generally asynchronous it is not suitable for this kind of basic study in which the synchronous regeneration of well identified developmental stages is a prerequisite.

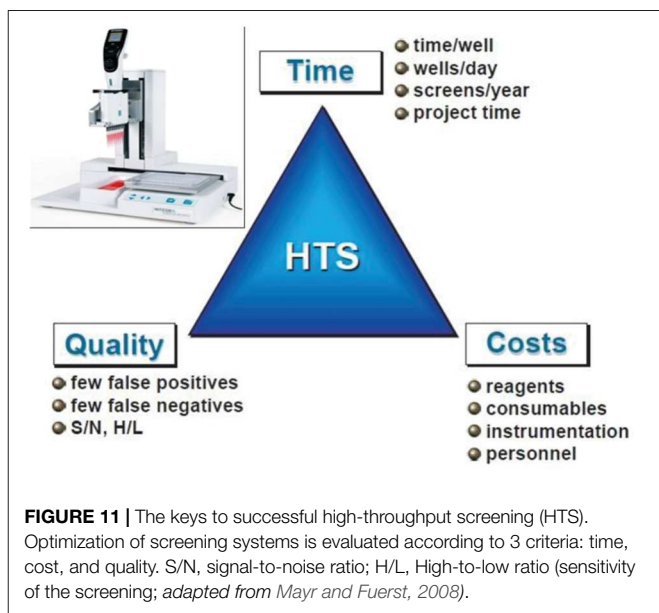
To undertake a comprehensive analysis of the indirect SE process in *C. arabica*, CIRAD and Nestlé R&D launched a significant joint research project, firstly to study the transcriptional changes that occur at the key developmental switches. With the very shortly expected delivery of the *C. arabica* genome sequence (Mueller et al., 2015), applying combined NGS techniques will make it possible to create a gene expression profile blueprint of coffee SE. The availability of the *C. arabica* genome follows the publication of the *C. canephora* genome (Denoëud et al., 2014).

In parallel to transcriptomic and detailed phenotyping approaches, full metabolomic profiles and the hormone contents at each of the developmental stages are being drawn. Visualizing the evolution of the primary and secondary metabolism at each stage and its interaction with hormone content and levels will help characterize cell functioning at each stage.

Integrating these results in the transcriptomic profiles obtained will definitely help to build the first regulatory networks of the Arabica SE process as a whole – an illustration of totipotency. As well as the huge basic knowledge accumulated, these networks will subsequently help identify candidate molecular and biochemical markers known to be key factors in the success of the different developmental stages. These markers will then enable further optimizations of the SE process (Campos et al., 2017), ending 40 years of empirical research.

High-Throughput Screening of Active Compounds to Optimize SE

Testing the effect of new compounds on SE efficiency is generally a laborious, intensive and tricky operation, which prevents considering comparing the effects of a large number of molecules and by that, greatly prevents accelerating culture media optimization. Nestlé R&D is working on a high-throughput screening (HTS) of active compounds to optimize SE protocols. A proof of concept of the high-throughput approach has been successfully validated in coffee with a first set of molecules and already allows the multiplication of experimental conditions, reduces culture volumes by a factor of a hundred as well as the space and labor required compared to the conventional research approach used until now. The next step is to scale up this technology through the screening of chemical libraries on robotic platforms. **Figure 11** shows the key factors in the success of HTS (Mayr and Fuerst, 2008). Once molecular knowledge about the SE process is gained thanks to the cutting edge -omics technologies described above, target-based active compound screening can be applied on a reasoned and scientific basis to optimize SE protocols. Adding active compounds to the culture medium, that have a potential effect on SE targets, and are capable of modulating the physiological state of cells, is a possible way to maximize embryo production, quality, and remove genetic (genotypic effects) or stress-induced recalcitrance. To effectively



select active substances, it is now possible to screen hundreds of libraries (Rodriguez-Furlan et al., 2016). Nestlé is currently reproducing some SE phases at a miniaturized level compatible with an automated screening platform, which appears to be unavoidable to reach the high-throughput standards.

CONCLUSION: TOMORROW'S CHALLENGES FOR LARGE-SCALE COFFEE DISSEMINATION BY SOMATIC EMBRYOGENESIS

Coffee SE is now one of the most advanced technologies. It has been industrialized (Robusta) or commercialized (Arabica) for 13–15 years for the two cultivated species. Very high yields and biological efficiency characterize both processes, with low genotypic effects and controlled SV. To date, very few examples of commercial applications of SE are available for any species. The industrial application of SE in hardwoods remains limited to a few species such as the hybrid yellow-poplar, cocoa, and coffee. Thirty five M improved coffee clones derived from SE have already been planted in the field.

Research by the two leading groups – the CIRAD/ECOM alliance and Nestlé – has often been way ahead and the resulting innovations have benefited the entire SE working community. This is particularly true for the temporary immersion bioreactors that are now extensively used for the micropropagation of many different species. It is interesting to observe the complementarities of the research done by these two teams and how they have facilitated the rapid development of micropropagation of both Arabica and Robusta.

The objective of the Nescafé Plan is to reach a total of 220 M coffee plants (Arabica and Robusta) distributed to farmers worldwide by 2020, using any method of propagation (seedling, rooted cuttings from SE-derived plants, grafting and SE). The

price of Robusta plants derived from SE was a limiting factor in the Philippines, which explains why rooted cuttings from SE-derived plants was privileged. Today, a total of c.a. 15 M of Robusta plants are reported to have been distributed to farmers using the SE method. To our knowledge, this is one of the biggest commercial SE-based propagation achieved in woody plants so far, however, it is still far behind the 74.5 M SE cocoa trees produced by ICCRI (Indonesian Coffee and Cocoa Research Institute) between 2009 and 2011 (Maximova and Guiltinan, 2015). The ECOM/CIRAD alliance has evaluated the demand for Arabica hybrid vitroplants at several dozen million per year for the next 5 years. The interest of these hybrid clones for agroforestry systems, to whose environmental conditions they are particularly well suited because of their high photosynthetic efficiency under low light, is boosting demand.

Based on the previous situations described for Thailand, Nicaragua, Costa Rica, Mexico, and the Philippines, SE is now a must-have in the toolbox of the Arabica and Robusta coffee plant producers. This *in vitro* method, already validated on more than 20 Robusta clones and 40 Arabica F1 varieties, allows significant time and space savings, and offers greater versatility for tree propagation. Indeed, breeding programs are now delivering new varieties (i.e., with improved cup quality as well as new disease resistant or drought tolerant clones) faster that can be rapidly deployed in the field. The SE process can also supply axenic planting material, which may be a prerequisite for successful international shipments. In a context of increasing parasite pressure including the major risk of the spread of *X. fastidiosa* or soil nematodes (*Meloidogyne* or *Pratylenchus* sp.), micropropagation has the great advantage of guaranteeing plants that are healthy and free from pests and diseases. This advantage will become increasingly important in the future.

Obviously, this system of propagation requires significant investments (laboratory and nursery) and skilled technicians to be effective. Training staff, who may need a long time to acquire sufficient experience, also has to be taken into account. Further improvements, especially focused on the automation of the acclimation step *ex vitro* are expected to dramatically reduce labor costs due to the need to handle each individual plant. In another connection, new technologies have recently been developed by the plant micropropagation industry for cost savings in the laboratory (LED lights, CO₂-enrichment, etc.). These developments may result in reconsidering the original strategies, particularly by investigating how embryo sowing could be automatized in the laboratory, using for example the Vitro Plus process implemented for the production of fern vitroplants⁷.

Contrary to what might have been expected a few years ago, SV has not been a bottleneck for the SE development for both cultivated coffee species. For each cultivated coffee species, a limited genotypic effect allows at the industrial scale the utilization of a common protocol whatever the cultivar propagated. The costs of production are more limiting. The use of *in vitro* cultivation to propagate new selected varieties represents in all cases an additional cost compared to traditional varieties that are agronomically less efficient but distributed at

⁷<http://www.vivi.nu/>

low cost in the form of seedlings, sometimes free of charge by governments. In spite of the very low selling prices reached with coffee vitroplants, they are often still too high for small farmers, but acceptable for larger producers who are aware of the agronomic added value. Line losses during the acclimatization and nursery stages (50% losses) are costly and significantly increase production costs. This problem is more and more minimized by the generalization of the use of mini-cuttings from SE-derived plants with which the losses are almost zero during these stages and hence the industrialization possible.

The coffee community working on SE chose to conduct in parallel the following strategies: to increase our knowledge about the molecular mechanisms controlling SE, to optimize SE by using innovative methods based on piloting with molecular markers and screening of active compounds using a miniaturized and automated system, to automate the acclimation step, to industrialize and generalize the rooted mini-cuttings method from somatic seedlings, to boost the market for Arabica hybrids and Robusta clones by bringing roasters and farmers closer together around innovation platforms so that the added value of improved varieties is better known, to democratize methods of cloning coffee and attract large micropropagation companies with the double aim of scaling up the coffee vitroplant production and being present in a maximum of producing countries.

This paper clearly shows that the success of coffee SE is tightly linked to the availability of selected varieties and clones to propagate. Up to now, SE research and development has been able to support the dissemination of new clones. However, a new huge up scaling appears to be necessary to be able to rapidly provide much larger quantities. In the case of Arabica, the reason is the recent loss of leaf rust resistance in all the cultivated conventional resistant varieties and the increasing impact of the harmful effects of climate change. In the case of Robusta, there is also an urgent need to replace unselected clones with improved ones, particularly in Vietnam – the world's

main coffee producer – where coffee orchards need to be totally renewed.

The absence of a functioning seed industry is a major deterrent for perennial tropical crops in many developing countries. Given the long lifespan of a tree plantation and hence the low renewal rate compared with annual crops, the propagation of perennial species has not been attractive so far as it is deemed uneconomic by investors. The investment of large micropropagation companies in the coffee planting business would guarantee the success of this future change of scale. With the availability of new improved clones and their economic added value, relevant business cases for the production of coffee plants should be disseminated to draw the attention of investors to the use of SE. All past and present efforts toward the democratization of coffee SE should generate huge interest in new varieties and the associated propagation techniques.

AUTHOR CONTRIBUTIONS

HE organized the plan and wrote the whole manuscript. DB and J-PD wrote all the parts concerning the Robusta SE. J-CB, SL, ED, CC, and RA wrote the current research part. BB, PM, EA, and HE were in charge of the analysis of social and economic constraints affecting *C. arabica*. EA, HE, FG, and PC wrote the paragraphs related to Arabica SE professionalization, technology transfer, and rooted mini-cuttings.

FUNDING

The BREEDCAFS project was supported by the European Commission under the Horizon 2020 – Research and Innovation Programme, H2020-SFS-2016-2, grant agreement number: 727934. RA benefits from a Ph.D. grant and funding from Nestlé and CIRAD.

REFERENCES

- Afreen, F., Zobayed, S. M. A., and Kozai, T. (2002). Photoautotrophic culture of *Coffea arabusta* somatic embryos: development of a bioreactor for the large scale plantlet conversion from cotyledonary embryos. *Ann. Bot.* 90, 20–29. doi: 10.1093/aob/mcf151
- Alpizar, E., Déchamp, E., Guilhaumon, C., Bertrand, B., Jourdan, C., Lashermes, P., et al. (2008). Agrobacterium rhizogenes-transformed roots of coffee: conditions for long-term proliferation, morphological and molecular characterization. *Ann. Bot.* 101, 929–940. doi: 10.1093/aob/mcn027
- Altpeter, F., Springer, N., Bartley, L., Blechl, A., Bruttner, T., Citovsky, V., et al. (2016). Advancing crop transformation in the era of genome editing. *Plant Cell* 28, 1510–1520. doi: 10.1105/tpc.16.00196
- Alves, G. S. C., Ferrera Torres, L., Dechamp, E., Breitler, J. C., Joët, T., Gatineau, F., et al. (2017). *Coffea canephora* DREB1D promoter regulates gene expression in mesophyll and stomatal guard cells under drought stress. *J. Exp. Bot.* 68, 3017–3031. doi: 10.1093/jxb/erx166
- Arroyo-Herrera, A., Ku Gonzalez, A., Canche Moo, R., Quiroz-Figueroa, F. R., Loyola-Vargas, V. M., Rodriguez-Zapata, L. C., et al. (2008). Expression of WUSCHEL in *Coffea canephora* causes ectopic morphogenesis and increases somatic embryogenesis. *Plant Cell Tissue Organ Cult.* 94, 171–180. doi: 10.1007/s11240-008-9401-1
- Barbón, R., Jiménez, E., and Preil, W. (2008). Influence of in vitro environment on somatic embryogenesis of *Coffea arabica* L. cv. caturra rojo: the effects of carbon dioxide on embryogenic cell suspensions. *Plant Cell Tissue Organ Cult.* 95, 155–161. doi: 10.1007/s11240-008-9427-4
- Barry-Etienne, D., Bertrand, B., Schlongvoigt, A., and Etienne, H. (2002a). The morphological variability within a population of coffee somatic embryos produced in a bioreactor affects the regeneration and the development of plants in the nursery. *Plant Cell Tissue Organ Cult.* 68, 153–162. doi: 10.1023/A:1013874221569
- Barry-Etienne, D., Bertrand, B., Vásquez, N., and Etienne, H. (2002b). Comparison of somatic embryogenesis derived coffee (*Coffea arabica* L.) plantlets regenerated in vitro or ex vitro conditions: morphological, mineral and water characteristics. *Ann. Bot.* 90, 77–85. doi: 10.1093/aob/mcf149
- Barry-Etienne, D., Bertrand, B., Vásquez, N., and Etienne, H. (1999). Direct sowing of *Coffea arabica* somatic embryos mass-produced in a bioreactor and regeneration of plants. *Plant Cell Rep.* 19, 111–117. doi: 10.1007/s002990050720
- Berthouly, M., Dufour, M., Alvard, D., Carasco, C., Alemanno, L., and Teisson, C. (1995). “Coffee micropropagation in a liquid medium using the temporary immersion technique,” in *Proceedings of the 16th Colloquium of International Coffee Science Association*, ed. ASIC (Vevey: ASIC), 514–519.

- Berthouly, M., and Michaux-Ferriere, N. (1996). High frequency somatic embryogenesis in *Coffea canephora*. *Plant Cell Tissue Organ Cult.* 44, 169–176. doi: 10.1007/BF00048196
- Bertrand, B., Alpizar, E., Etienne, H., and Charmetant, P. (2006). Comparison of bean biochemical composition and beverage quality of Arabica hybrids involving sudanese-ethiopian origins with traditional varieties at various elevations in Central America. *Tree Physiol.* 26, 1239–1248. doi: 10.1093/treephys/26.9.1239
- Bertrand, B., Alpizar, E., Lara, L., Santa-Creo, R., Hidalgo, R., Quijano, J. M., et al. (2011). Performance of Arabica F1 hybrids in agroforestry and full-sun cropping systems in comparison with pure lines varieties. *Euphytica* 181, 147–158. doi: 10.1007/s10681-011r-r0372-7
- Bertrand, B., Montagnon, C., Georget, F., Charmetant, P., and Etienne, H. (2012). Création et diffusion de variétés de caféiers Arabica: quelles innovations variétales? *Cah. Agric.* 21, 77–88. doi: 10.1684/agr.2012.0547
- Bobadilla Landey, R., Cenci, A., Georget, F., Bertrand, B., Camayo, G., Dechamp, E., et al. (2013). High genetic and epigenetic stability in *Coffea arabica* plants derived from embryogenic suspensions and secondary embryogenesis as revealed by AFLP, MSAP and the phenotypic variation rate. *PLoS One* 8:e56372. doi: 10.1371/journal.pone.0056372
- Bobadilla Landey, R., Cenci, A., Guyot, R., Bertrand, B., Georget, F., Dechamp, E., et al. (2015). Assessment of genetic and epigenetic changes during cell culture ageing in coffee and relations with somaclonal variation. *Plant Cell Tissue Organ Cult.* 3, 517–531. doi: 10.1007/s11240-015-0772-9
- Breitler J. C., Déchamp E., Campa C., Rodrigues L. A. Z., Guyot R., Marraccini P., et al. (2018). CRISPR/Cas9-mediated efficient targeted mutagenesis has the potential to accelerate the domestication of *Coffea canephora*. *Plant Cell Tissue Organ Cult.* 134, 383–394. doi: 10.1007/s11240-018-1429-2
- Breton, D., Garcia Martinez, C., Ducos, J. P., De Faria Maraschin, S., Navarro Mastache, L.C., and Broun, P. (2017). “Industrial implementation of somatic embryogenesis for the production of *Coffea canephora* P. plantlets,” in *Proceedings of the 4th International Conference of the IUFRO Unit 2.09.02*, eds J. M. Bonga et al. (La Plata: IUFRO).
- Buell, C., and Voytas, D. (2017). Technology turbocharges functional genomics. *Plant Cell* 29, 1179–1180. doi: 10.1105/tpc.17.00443
- Campos, N. A., Paiva, L. V., Panis, B., and Carpentier, S. C. (2016). The proteome profile of embryogenic cell suspensions of *Coffea arabica* L. *Proteomics* 16, 1001–1005. doi: 10.1002/pmic.201500399
- Campos, N. A., Panis, B., and Carpentier, S. C. (2017). Somatic embryogenesis in coffee: the evolution of biotechnology and the integration of omics technologies offer great opportunities. *Front. Plant Sci.* 8:1460. doi: 10.3389/fpls.2017.01460
- Cenci, A., Combes, M. C., and Lashermes, P. (2012). Genome evolution in diploid and tetraploid *Coffea* species as revealed by comparative analysis of orthologous genome segments. *Plant Mol. Biol.* 78, 135–145. doi: 10.1007/s11103-011-9852-3
- Combes, M. C., Andrzejewski, S., Anthony, F., Bertrand, B., Rovelli, P., Graziosi, G., et al. (2000). Characterization of microsatellites loci in *Coffea arabica* and related coffee species. *Mol. Ecol.* 9, 1178–1180. doi: 10.1046/j.1365-294x.2000.00954-5.x
- Cruz, A. R. R., Paixão, A. L. D., Machado, F. R., Barbosa, M. F., Junqueira, C. S., Cabral, G. B., et al. (2004). “Metodologia para obtenção de plantas transformadas de *Coffea canephora* por co-cultivo de calos embriogênicos com *Agrobacterium tumefaciens*,” in *Boletim de Pesquisa e Desenvolvimento*, Vol. 58, ed. M. S. P. Negrão (Brasília: Embrapa Recursos Genéticos e Biotecnologia), 15.
- De Faria, M., Jimenez, E., Barbon, R., Capote, A., Chavez, M., and Quiala, E. (2003). Effect of dissolved oxygen concentration on differentiation of somatic embryos of *Coffea arabica* cv. Catimor 9722. *Plant Cell Tissue Organ Cult.* 72, 1–6. doi: 10.1023/A:1021202305692
- Déchamp, E., Breitler, J. C., Leroy, T., and Etienne, H. (2015). “*Coffea arabica* (L.)” in *Agrobacterium Protocols*, 3rd Edn, ed. K. Wang (Totowa, NJ: Humana Press Inc.).
- Denoeud, F., Carretero-Paulet, L., Dereeper, A., Droc, G., Guyot, R., Pietrella, M., et al. (2014). The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* 345, 1181–1184. doi: 10.1126/science.1255274
- Ding, Y., Li, H., Chen, L., and Xie, K. (2016). Recent advances in genome editing using CRISPR/Cas9. *Front. Plant Sci.* 7:703. doi: 10.3389/fpls.2016.00703
- Dublin, P. (1984). Techniques de reproduction végétative in vitro et amélioration génétique chez les caféiers cultivés. *Café Cacao* 28, 231–244.
- Ducos, J. P., Alenton, R., Reano, J. F., Kanchanomai, C., Deshayes, A., and Pétiard, V. (2003). Agronomic performance of *Coffea canephora* P. trees derived from large-scale somatic embryo production in liquid medium. *Euphytica* 131, 215–223. doi: 10.1023/A:1023915613158
- Ducos, J. P., Gianforcaro, M., Florin, B., Pétiard, V., and Deshayes, A. (1999). “A technically and economically attractive way to propagate elite *Coffea canephora* (Robusta) clones: in vitro somatic embryogenesis,” in *Proceedings of the 18th Colloquium of International Coffee Science Association*, ed. ASIC (Vevey: ASIC), 295–301.
- Ducos, J. P., Gibault, E., Broun, P., and Lambot, C. (2010). “Coffee propagation by somatic embryogenesis at Nestlé R&D Center-Tours,” in *Proceedings of the International Conference of the IUFRO Unit 2.09.02*, eds Y. S. Park et al. (Suwon: IUFRO).
- Ducos, J. P., Chantanut, P., Vuong, P., Lambot, C., and Pétiard, V. (2007a). Mass propagation of Robusta clones: disposable plastic bags for pre-germination of somatic embryos by temporary immersion. *Acta Hort.* 764, 33–40. doi: 10.17660/ActaHortic.2007.764.3
- Ducos, J. P., Labbé, G., Lambot, C., and Pétiard, V. (2007b). Pilot scale process for the production of pre-germinated somatic embryos of selected Robusta clones. *In Vitro Cell. Dev. Biol. Plant* 43, 652–659. doi: 10.1007/s11627-007-9075-0
- Ducos, J. P., Lambot, C., and Pétiard, V. (2007c). Bioreactors for coffee mass propagation by somatic embryogenesis. *Int. J. Dev. Biol.* 1, 1–12.
- Ducos, J. P., Prévot, A., Lambot, C., and Pétiard, V. (2009). Positive effect of the CO₂ released by commercial substrates on the ex vitro germination of coffee somatic embryos. *Acta Hort.* 812, 329–336. doi: 10.17660/ActaHortic.2009.812.46
- Ducos, J. P., Terrier, B., Courtois, D., and Pétiard, V. (2008). Improvement of plastic-based disposable bioreactors for plant science needs. *Phytochem. Rev.* 7, 607–613. doi: 10.1007/s11101-008-9089-1
- Ducos, J. P., Zamarripa, A., Eskès, A., and Pétiard, V. (1993). “Production of somatic embryos of coffee in a bioreactor,” in *Proceedings of the 15th Colloquium of International Coffee Science Association*, ed. ASIC (Vevey: ASIC), 89–96.
- Dusseret, S., Serret, J., Bastos-Siqueira, A., Morcillo, F., Déchamp, E., Rofidal, V., et al. (2018). Integrative analysis of the late maturation program and desiccation tolerance mechanisms in intermediate coffee seeds. *J. Exp. Bot.* 69, 1583–1597. doi: 10.1093/jxb/erx492
- Etienne, H., and Bertrand, B. (2001). Trueness-to-type and agronomic characteristics of *Coffea arabica* trees micropropagated by the embryogenic cell suspension technique. *Tree Physiol.* 21, 1031–1038. doi: 10.1093/treephys/21.14.1031
- Etienne, H., and Bertrand, B. (2003). Somaclonal variation in *Coffea arabica*: effects of genotype and embryogenic cell suspension age on frequency and phenotype of variants. *Tree Physiol.* 23, 419–426. doi: 10.1093/treephys/23.6.419
- Etienne, H., Bertrand, B., Anthony, F., Côte, F., and Berthouly, M. (1997). “L’embryogénèse somatique: un outil pour l’amélioration génétique du caféier,” in *Proceedings of the 17th Colloquium of International Coffee Science Association*, ed. ASIC (Nairobi: ASIC), 457–465.
- Etienne, H., Bertrand, B., Déchamp, E., Maurel, P., Georget, F., Guyot, R., et al. (2016). Are genetics and epigenetic instabilities of plant embryogenic cells a fatality? The experience of coffee somatic embryogenesis. *Hum. Genet. Embryol.* 61, 1–5. doi: 10.4172/2161-0436.1000136
- Etienne, H., Bertrand, B., Georget, F., Lartaud, M., Montes, F., Déchamp, E., et al. (2013). Development of coffee somatic and zygotic embryos to plants differs in the morphological, histochemical and hydration aspects. *Tree Physiol.* 33, 640–653. doi: 10.1093/treephys/tpt034
- Etienne, H., Bertrand, B., Montagnon, C., Déchamp, E., Jourdan, I., Alpizar, E., et al. (2012). Un exemple de transfert technologique réussi en micropropagation : la multiplication de *Coffea arabica* par embryogénèse somatique. *Cah. Agric.* 21, 115–124. doi: 10.1687/agr.2012.0553
- Etienne-Barry, D., Bertrand, B., Vasquez, N., and Etienne, H. (1999). Direct sowing of *Coffea arabica* somatic embryos mass-produced in a bioreactor and regeneration of plants. *Plant Cell Rep.* 19, 111–117. doi: 10.1007/s002990050720
- Fan, D., Liu, T., Li, C., Jiao, B., Li, S., Hou, Y., et al. (2015). Efficient CRISPR/Cas9-mediated targeted mutagenesis in *Populus* in the first generation. *Sci. Rep.* 5:12217. doi: 10.1038/srep12217
- Gatica-Arias, A. M., Arrieta-Espinoza, G., and Espinoza Esquivel, A. M. (2008). Plant regeneration via indirect somatic embryogenesis and optimization of

- genetic transformation in *Coffea arabica* L. cvs. caturra and catuai. *Electron. J. Biotechnol.* 11, 1–12. doi: 10.2225/vol11-issue1-fulltext-9
- Georget, F., Courtel, P., Garcia, E. M., Hidalgo, M., Alpizar, E., Breitler, J. C., et al. (2017). Somatic embryogenesis-derived coffee plantlets can be efficiently propagated by horticultural rooted mini-cuttings: a boost for somatic embryogenesis. *Sci. Hortic.* 216, 177–185. doi: 10.1016/j.scienta.2016.12.017
- Hohe, A., Winkelmann, T., and Schwenkel, H. (1999). CO₂ accumulation in bioreactor suspension cultures of *Cyclamen persicum* Mill. and its effect on cell growth and regeneration of somatic embryos. *Plant Cell Rep.* 18:863. doi: 10.1007/s002990050675
- Jacques, M.-A., Denancé, N., Legendre, B., Morel, E., Briand, M., Mississippi, S., et al. (2016). New coffee plant-infecting xylella fastidiosa variants derived via homologous recombination. *Appl. Environ. Microbiol.* 82, 1556–1568. doi: 10.1128/AEM.03299-15
- Jayasankar, S., Bondada, B. R., Li, Z., and Gray, D. J. (2003). Comparative anatomy and morphology of *Vitis vinifera* (Vitaceae) somatic embryos from solid and liquid culture derived proembryogenic masses. *Am. J. Bot.* 90, 973–979. doi: 10.3732/ajb.90.7.973
- Jia, H., and Wang, N. (2014). Targeted genome editing of sweet orange using Cas9/sgrRNA. *PLoS One* 9:e93806. doi: 10.1371/journal.pone.0093806
- Kosai, T., Kubota, C., and Jeong, B. R. (1997). Environmental control for the large-scale production of plants through in vitro techniques. *Plant Cell Tissue Organ Cult.* 51, 49–56. doi: 10.1023/A:1005809518371
- Kunason, T., Ducos, J. P., Lambot, C., Kasinkasempong, Y., Chantanumat, P., and Broun, P. (2010). “Support to small coffee farmer in Thailand through large scale propagation of *Coffea canephora*,” in *Proceedings of the 23rd Colloquium of International Coffee Science Association*, ed. ASIC (Vevey: ASIC), 1012–1016.
- Lashermes, P., Bertrand, B., and Etienne, H. (2009). “Breeding coffee (*Coffea arabica*) for sustainable production,” in *Breeding Plantation Tree Crops: Tropical Species*, eds J. S. Mohan and P. M. Priyadarshan (New York, NY: Springer [Etats-Unis]), 525–543. doi: 10.1007/978-0-387-71201-7_14
- Legendre, B., Mississippi, S., Olivier, O., Morel, E., Crouzillat, D., Jacques, M. A., et al. (2013). “Identification and characterization of *Xylella fastidiosa* isolated from coffee plants in France,” in *Proceedings of the Pierce’s Disease Research Progress Reports*, ed. T. Esser (Sacramento, CA: California Department of Food and Agriculture).
- Leroy, T., Henry, A., Royer, M., et al. (2000). Genetically modified coffee plants expressing the *Bacillus thuringiensis* cry1Ac gene for resistance to leaf miner. *Plant Cell Rep.* 19, 382–389. doi: 10.1007/s002990050744
- Marraccini, P., Freire, L. P., Alves, G. S. C., Vieira, N. G., Vinecky, F., Elbelt, S., et al. (2011). RBCS1 expression in coffee: *Coffea orthologs*, *Coffea arabica* homeologs, and expression variability between genotypes and under drought stress. *BMC Plant Biol.* 11:85. doi: 10.1186/1471-2229-11-85
- Marraccini, P., Vinecky, F., Alves, G. S. C., Ramos, H. J. O., Elbelt, S., Vieira, N. G., et al. (2012). Differentially expressed genes and proteins upon drought acclimation in tolerant and sensitive genotypes of *Coffea canephora*. *J. Exp. Bot.* 6, 4191–4212. doi: 10.1093/jxb/ers103
- Maximova, S., and Gultinan, M. J. (2015). “Tissue culture,” in *Supplying New Cocoa Planting Material to Farmers: A Review of Propagation Methodologies*, eds B. Laliberte and M. End (Rome: Bioversity International), 67–85.
- Mayr, L. M., and Fuerst, P. (2008). The future of high-throughput screening. *J. Biomol. Screen.* 13, 443–448. doi: 10.1177/1087057108319644
- Ménendez Yuffá, A., Da Silva, R., Rios, L., and Xena de Enrech, N. (2000). Mitotic aberrations in coffee (*Coffea arabica* cv. ‘Catimor’) leaf explants and their derived embryogenic calli. *Electron. J. Biotechnol.* 3, 1–5. doi: 10.2225/vol3-issue2-fulltext-1
- Mishra, M. K., and Slater, A. (2012). Recent advances in the genetic transformation of coffee. *Biotechnol. Res. Int.* 10, 1–17. doi: 10.1155/2012/580857
- Mofatto, L. S., Carneiro, F. A., Vieira, N. G., Duarte, K. E., Vidal, R. O., Alekcevetch, J. C., et al. (2016). Identification of candidate genes for drought tolerance in coffee by high-throughput sequencing in the shoot apex of different *Coffea arabica* cultivars. *BMC Plant Biol.* 16:94. doi: 10.1186/s12870-016-0777-5
- Mueller, L., Strickler, S. R., Domingues, D. S., Pereira, L. F. P., Andrade, A. C., Marraccini, P., et al. (2015). “Towards a better understanding of the *Coffea arabica* genome structure,” in *Proceedings of the 25th Colloquium of International Coffee Science Association*, ed. ASIC (Armenia: ASIC), 8–12.
- Mukul-López, H. G., De-la-peña, C., Galaz-ávalos, R. M., and Loyola-Vargas, V. M. (2012). Evaluation of the extracellular proteome profile during the somatic embryogenesis process of *Coffea* spp. *J. Mex. Chem. Soc.* 56, 72–79.
- Nic-Can, G. I., López-Torres, A., Barredo-Pool, F., Wrobel, K., Loyola-Vargas, V. M., Rojas-Herrera, R., et al. (2013). New insights into somatic embryogenesis: LEAFY COTYLEDON1, BABY BOOM1 and WUSCHELRELATED HOMEBOX4 are epigenetically regulated in *Coffea canephora*. *PLoS One* 8:e72160. doi: 10.1371/journal.pone.0072160
- Nishitani, C., Hirai, N., Komori, S., Wada, M., Okada, K., Osakabe, K., et al. (2016). Efficient genome editing in apple using a CRISPR/Cas9 system. *Sci. Rep.* 6:31481. doi: 10.1038/srep31481
- Noriega, C., and Söndhal, M. R. (1993). “Arabica coffee micropropagation through somatic embryogenesis via bioreactors,” in *Proceedings of the 15th Colloquium of International Coffee Science Association*, ed. ASIC (Vevey: ASIC), 73–81.
- Ogita, S., Uefuji, H., Morimoto, M., and Sano, H. (2004). Application of RNAi to confirm theobromine as the major intermediate for caffeine biosynthesis in coffee plants with potential for construction of decaffeinated varieties. *Plant Mol. Biol.* 54, 931–941. doi: 10.1007/s11103-004-0393-x
- Ogita, S., Uefuji, H., Yamaguchi, Y., Koizumi, N., and Sano, H. (2003). RNA interference: producing decaffeinated coffee plants. *Nature* 423, 823–825. doi: 10.1038/423823a
- Padua, M. S., Paiva, L. V., Silva, L. C., Livramento, K. G., Alves, E., and Castro, A. H. F. (2014). Morphological characteristics and cell viability of coffee plants calli. *Ciênc. Rural* 44, 660–665. doi: 10.1590/S0103-84782014000400014
- Pétiard, V., Lepage, B., Lambot, C., Crouzillat, D., Florin, B., Brulard, E., et al. (2004). “Establishment and preliminary agronomic and qualitative evaluation of collections of *Coffea arabica* and *Coffea canephora* cultivated varieties,” in *Proceedings of the 20th Colloquium of International Coffee Science Association*, ed. ASIC (Vevey: ASIC), 834–841.
- Petitot, A. S., Barsalobres-Cavallari, C., Ramiro, D., Albuquerque, F. E., Etienne, H., and Fernandez, D. (2013). Promoter analysis of the WRKY transcription factors CaWRKY1a and CaWRKY1b homoeologous genes in coffee (*Coffea arabica*). *Plant Cell Rep.* 32, 1263–1276. doi: 10.1007/s00299-013-1440-3
- Pierson, E. S., Van Lammern, A. A. M., Schel, H. N., and Staritsky, G. (1983). In vitro development of embryoids from punched leaf dishes of *Coffea canephora*. *Protoplasma* 115, 208–216. doi: 10.1007/BF01279811
- Quétiér, F. (2016). The CRISPR-Cas9 technology: closer to the ultimate toolkit for targeted genome editing. *Plant Sci.* 242, 65–76. doi: 10.1016/j.plantsci.2015.09.003
- Quiroz-Figueroa, F., Mendez-Zeel, M., Sanchez-Teyer, F., Rojas-Herrera, R., and Loyola-Vargas, V. M. (2002). Differential gene expression in embryogenic and non-embryogenic clusters from cell suspension cultures of *Coffea arabica*. *J. Plant Physiol.* 159, 1267–1270. doi: 10.1078/0176-1617-00878
- Quiroz-Figueroa, F. R., Rojas-Herrera, R., Galaz-Avalos, R. M., and Loyola-Vargas, V. M. (2006). Embryo production through somatic embryogenesis can be used to study cell differentiation in plants. *Plant Cell Tissue Organ Cult.* 86, 285–301. doi: 10.1007/s11240-006-9139-6
- Ribas, A., Déchamp, E., Champion, A., Bertrand, B., Combes, M.-C., Verdeil, J.-L., et al. (2011). Agrobacterium-mediated genetic transformation of *Coffea arabica*(L.) is greatly enhanced by using established embryogenic callus cultures. *BMC Plant Biol.* 11:92. doi: 10.1186/1471-2229-11-92
- Ribas, A. F., Galvao, R. M., Pereira, L. F. P., and Vieira, L. G. E. (2005a). “Transformacao de *Coffea arabica* com o gene da ACC oxidase em orientacao antinverso,” in *Proceedings of the 50th Congresso Brasileiro de Genetica*, Sao Pulo, 492.
- Ribas, A. F., Kobayashi, A. K., Pereira, L. F. P., and Vieira, L. G. E. (2005b). Genetic transformation of *Coffea canephora* by particle bombardment. *Biol. Plant.* 49, 493–497. doi: 10.1007/s10535-005-0038-1
- Ribas, A. F., Kobayashi, A. K., Pereira, L. F. P., and Vieira, L. G. E. (2006). Production of herbicide-resistant coffee plants (*Coffea canephora* P.) via Agrobacterium tumefaciens-mediated transformation. *Brazil. Arch. Biol. Technol.* 49, 11–19. doi: 10.1590/S1516-89132006000100002
- Rodriguez-Furlan, C., Miranda, G., Reggiardo, M., Hicks, G. R., and Norambuena, N. (2016). High throughput selection of novel plant growth regulators: assessing the translatability of small bioactive molecules from *Arabidopsis* to crops. *Plant Sci.* 245, 50–60. doi: 10.1016/j.plantsci.2016.01.001
- Sanpote, S., Ducos, J. P., Lambot, C., Kunason, T., Kasinkasempong, Y., Chantanumat, P., et al. (2006). “First industrial massive propagation of *Coffea canephora* through somatic embryogenesis in Thailand,” in *Proceedings of the*

- 21th Colloquium of International Coffee Science Association, ed. ASIC (Vevey: ASIC), 89–96.
- Silva, A. T., Barduche, D., do Livramento, K. G., Ligterink, W., and Paiva, L. V. (2014). Characterization of a putative serk-like ortholog in embryogenic cell suspension cultures of *Coffea arabica* L. *Plant Mol. Biol. Rep.* 32, 176–184. doi: 10.1007/s11105-013-0632-x
- Silva, A. T., Barduche, D., do Livramento, K. G., and Paiva, L. V. (2015). A putative BABY BOOM-like gene (CaBBM) is expressed in embryogenic calli and embryogenic cell suspension culture of *Coffea arabica* L. *In Vitro Cell. Dev. Biol. Plant* 51, 93–101. doi: 10.1007/s11627-014-9643-z
- Söndhal, M. R., and Sharp, W. R. (1977). High frequency induction of somatic embryos in cultured leaf explants of *Coffea arabica* L. *Z. Pflanzenphysiol.* 81, 395–408. doi: 10.1016/S0044-328X(79)80123-3
- Staritsky, G. (1970). Embryoid formation in callus tissues of coffee. *Acta Bot. Neerl.* 19, 509–514. doi: 10.1111/j.1438-8677.1970.tb00677.x
- Teisson, C., and Alvard, D. (1995). “A new concept of plant in vitro cultivation liquid medium: temporary immersion,” in *Current Issues in Plant Molecular and Cellular Biology*, eds M. Terzi et al. (Dordrecht: Springer), 105–110.
- Tonietto, A., Sato, J. H., Teixeira, J. B., de Souza, E. M., Pedrosa, F. O., Franco, O. L., et al. (2012). Proteomic analysis of developing somatic embryos of *Coffea arabica*. *Plant Mol. Biol. Rep.* 30, 1393–1399. doi: 10.1007/s11105-012-0425-7
- van der Vossen, H., Bertrand, B., and Charrier, A. (2015). Next generation variety development for sustainable production of arabica coffee (*Coffea arabica* L.): a review. *Euphytica* 204, 243–256. doi: 10.1007/s10681-015-1398-z
- Vieira, N. G., Carneiro, F. A., Sujii, P. S., Alekcevetch, J. C., Freire, L. P., Vinecky, F., et al. (2013). Different molecular mechanisms account for drought tolerance in *Coffea canephora* var. *Conilon*. *Trop. Plant Biol.* 6, 181–190. doi: 10.1007/s12042-013-9126-0
- Wang, Z., Wang, S., Li, D., Zhang, Q., Li, L., Zhong, C., et al. (2018). Optimized paired-sgRNA/Cas9 cloning and expression cassette triggers high efficiency multiplex genome editing in kiwifruit. *Plant Biotechnol. J.* 16, 1424–1433. doi: 10.1111/pbi.12884
- Yasuda, T., Fujii, Y., and Yamaguchi, T. (1985). Embryogenic callus induction from *Coffea arabica* leaf explants by benzyladenine. *Plant Cell Physiol.* 26, 595–597. doi: 10.1093/oxfordjournals.pcp.a076946
- Zamarripa, A. (1993). *Study and Development of Somatic Embryogenesis in Liquid Medium of Coffee*. PhD thesis, Ecole Nationale Supérieure Agronomique, Rennes, 191.
- Zamarripa, A., Ducos, J. P., Bollon, H., Dufour, M., and Pétiard, V. (1991a). Production of coffee somatic embryos in liquid medium: effects of inoculation density and renewal of the medium. *Café Cacao* 35, 233–244.
- Zamarripa, A., Ducos, J. P., Tessereau, H., Bollon, H., Eskes, A. B., and Pétiard, V. (1991b). “Développement d’un procédé de multiplication en masse du caféier par embryogenèse somatique en milieu liquide,” in *Proceedings of the 14th Colloquium of International Coffee Science Association*, ed. ASIC (Vevey: ASIC), 392–402.
- Zhou, X., Jacobs, T. B., Xue, L. J., Harding, S. A., and Tsai, C. J. (2015). Exploiting SNPs for bi-allelic CRISPR mutations in the outcrossing woody perennial *Populus* reveals 4-coumarate: CoA ligase specificity and redundancy. *New Phytol.* 208, 298–301. doi: 10.1111/nph.13470
- Zorinians, S. E., Nosov, A. V., Monforte Gonzalez, M., Mendes Zeel, M., and Loyola-Vargas, V. M. (2003). Variation of nuclear DNA content during somatic embryogenesis and plant regeneration of *Coffea arabica* L. using cytophotometry. *Plant Sci.* 164, 141–146. doi: 10.1016/S0168-9452(02)00322-9

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Etienne, Breton, Breitler, Bertrand, Déchamp, Awada, Marraccini, Lérain, Alpizar, Campa, Courtel, Georget and Ducos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Part V

New tools to understand somatic embryogenesis underlying mechanisms

V.1. SE description based on ultrastructural studies in the 1990s

During this period, studies about SE in different species were mainly empirical and focused on establishing and optimizing somatic embryo production, much more than to understand the mechanisms behind this event. Studies were mainly focused on using light and electron microscopy to provide detailed descriptions of SE developmental stages. Nine developmental stages have been characterized in the indirect SE of dicots: explant, primary callus, embryogenic callus, embryogenic cell clusters, globular, heart-shaped, torpedo-shaped, cotyledonary embryos, and plantlet. These stages were chosen and characterized based on well-described morphological and anatomical changes that occur when switching from one stage to another (Jayasankar et al., 2003). Furthermore, literature shows that many authors have distinguished embryogenic calli from non-embryogenic calli based on morphological and histological studies (Verdeil et al., 2001, 2007; Na et al., 2007). In general, embryogenic calli are yellow and friable and composed of small, isodiametric cells, arranged in clusters, with a dense cytoplasm, a big nucleus with salient nucleoli (high nucleo-cytoplasmic ratio) and rich in small amyloplasts. Non-embryogenic calli are spongy and translucent showing cells that are more elongated, with the vacuole occupying a big volume of the cytoplasm, higher number of vesicles and absence of cytoplasmic organelles (Verdeil et al., 2001; Gatica-Arias et al., 2008).

V.2. Development of “omics” techniques in the 2000s-2010s

The latest “omics” technologies (genomics, transcriptomics, proteomics, epigenomics, metabolomics and phenomics) are gaining increasing attention. Transcriptomics and metabolomics will be further discussed. Many authors believe that combining “omics” technologies with the *in vitro* process can have a tremendous impact on the knowledge of the molecular mechanisms underlying SE (Yang et al., 2012; Campos et al., 2017, Pais, 2019, Mendez-Hernandez et al., 2019).

a) Transcriptomics

Probably the easiest way to study changes on a genome-wide scale is through transcriptomics. The structure of RNA is homogenous and relatively simple and therefore the analysis is the most straightforward. Zivy et al. (2015) provided an overview on the evolution of transcriptomics techniques. Briefly, Velculescu et al. (1995) developed serial analysis of gene expression (SAGE)

technique which is based on the generation of 15 bp tags from a defined position in each transcript, which are then concatenated, cloned into a plasmid vector and ultimately sequenced. In 2003, microarrays, which were significantly cheaper and was a more high-throughput technology, were also developed for transcriptomic studies. Microarrays use known probes that will hybridize with the labeled sample and based on the intensity of these dyes, transcript levels are estimated. This however, implies that sequence information exists before generation of the microarray and this is a serious limitation when applied to non-model crops. The limited sequence availability in non-models can be overcome by the use of microarrays of closely related species or by the generation of a species-specific microarray based on known expressed sequence tag (EST) data for instance, but these analyses will be less informative (Davey et al., 2009). Microarray analysis is also hampered by a high background noise due to cross-hybridization as well as saturation of signals. Microarrays therefore have a limited sensitivity. With the availability of next generation sequencing (NGS) technologies, the possibility to directly sequence mRNA at relatively reduced costs became available. This technique, termed RNA-seq, has clear advantages over the other transcriptomics methods: a higher sensitivity can be achieved (Wang et al., 2009) and no previous sequence knowledge is *per se* required. Reads can be mapped to a known reference genome or *de novo* assembled. *De novo* assembly of reads into contigs increases the use of this technique for crops whose genome has not been sequenced, as was demonstrated for wheat, agave and horse gram (Bhardwaj et al., 2013; Gross et al., 2013; Oono et al., 2013). However, a reference genome is highly recommended to assure the correct assembly of the reads and to deal with paralogs and allelic variants. Moreover, most read-mapping software has been written to analyze diploid genomes (Langmead & Salzberg, 2012) and is unsuited for polyploid organisms (Page et al., 2013). Categorizing DNA sequence reads into their respective genomes enables current methods to analyze polyploid genomes like the one of *C. arabica* as if they were diploid. This makes RNA-seq in combination with reference genomes an excellent tool to differentiate between isoforms of a gene family, which are a widespread phenomenon in complex crop genomes.

With the introduction of NGS, RNA-seq appears to be the transcriptomic tool for the future, especially in crops. At the moment, the costs associated with RNA-seq prevent large scale analysis of many varieties in different conditions and multiple biological replicates. However, as the NGS technique keeps evolving, costs are likely to drop and may no longer be a limiting factor in the future. As more and more genomes are sequenced, alignments to reference genomes should become standard practice, which will also significantly reduce the analysis time required for *de novo* assembly.

b) Metabolomics

In the “omics” field, metabolomics generates large datasets for the identification and quantification of small molecules. Usually, the approach is undertaken for the high-throughput detection of secondary (flavonoids, sugar-phosphates, phytohormones, phytoalexins, etc.) and primary metabolites (sugars, organic and amino acids, etc.). Complementary Mass Spectrometry (MS)-based Liquid Chromatography (LC) and Gas Chromatography (GC) approaches are required to adequately profile this metabolic diversity (Scherling et al., 2010). Although it is possible to gather tens of thousands of metabolic variables, their accurate identification remains the major bottleneck in metabolomics studies to date (Zivy et al., 2015). In contrast to proteomics, for metabolomics analyses, functional identification is not dependent on the availability of genome sequence information. However, the availability of purified standard substances and/or spectral libraries/databases is necessary. Of the various MS methods, two profiling strategies can be typically distinguished: a targeted and a non-targeted approach (Wienkoop et al., 2010). The non-targeted technique allows for the quantitative evaluation of unknown, yet unidentified, variables. Nevertheless, metabolomics databases and the availability of MS-derived metabolite fragment information are increasing. For example, the Golm Metabolome Database is now one of the most advanced databases for authenticated reference metabolites (Kopka et al., 2005; Dethloff et al., 2014). Furthermore, novel strategies for pathway and structural assignments of untargeted high-throughput metabolomics data are being developed. For example, metabolomics is also playing an increased role in stress marker detection and contributing to improved stress tolerance in crops (Martinez-Gomez et al., 2012; Rasmussen et al., 2012).

V.3. Molecular insights gained in somatic embryogenesis

a) Genes encoding transcription factors

Unlike histo-cytological studies which have been extensively driven, much less is known about the molecular mechanisms underlying the progression from somatic cells to embryo ontogenesis. Many model species (annual and perennial) were used, including carrot, Norway spruce, alfalfa, cotton, and *Arabidopsis* (Yang & Zhang, 2010, Horstman et al., 2017) to decipher the molecular mechanisms involved in SE. These works have resulted in the isolation of several key genes encoding transcription factors that are specifically activated or that exhibit differential expression patterns during SE (Rojas-Herrera et al., 2002; Feher et al., 2003), such as *SERK* (Schmidt et al., 1997), *LEAFY COTYLEDON* (*LEC*, Curaba et al., 2004; Gazzarrini et al., 2004; Gaj et al., 2005), *BABY BOOM* (*BBM*, Boutilier et al., 2002), and *WUSCHEL* (*WUS*, Zuo et al., 2002). Although many works have studied the expression

of these master regulators in SE, the spatio-temporal framework showing at which developmental stage these genes are switched on/off is still unclear and must be established.

a.1) *Somatic Embryogenesis Receptor Kinase (SERK) gene*

The gene *SERK* was first described in carrot (*dcSERK*) cell suspensions (Schmidt et al., 1997). Its expression is associated with the early stages of SE. Based on its expression pattern, Hecht et al. (2001) suggested to use *SERK* as a marker for embryogenic competence. The same expression pattern was found during early zygotic embryogenesis, proving the similarity of both types of embryogenic development (Schmidt et al., 1997). Homologs of *SERK* were identified in different species as maize, sunflower, rice and citrus (Baudino et al., 2001; Thomas et al., 2004; Shimada et al., 2005; Ito et al., 2005) and showed a role in the SE of these species. The *dcSERK* gene encodes a Leu-rich repeat (LRR) transmembrane receptor-like kinase (RLK). This transmembrane receptor plays a role in many biological processes as receptor of growth regulators, maintenance of balance between undifferentiated cells and cells designated to form new organs in meristematic regions (Schmidt et al., 1997; Hecht et al., 2001). This gene was overexpressed in *Arabidopsis* and its importance to acquire embryogenic competence has been confirmed (Hecht et al., 2001; Yang & Karlson, 2012; Ma et al., 2012).

a.2) *Leafy Cotyledon (LEC) gene*

The gene *Leafy Cotyledon* was first identified and characterized in *Arabidopsis* (Lotan et al., 1998). Its expression is crucial for the cotyledon identity formation and the completion of embryo maturation (Lotan et al., 1998). Ectopic expression of *LEC* in vegetative tissues suggests its importance for cellular differentiation and embryo morphogenesis through the induction of somatic embryo formation. *LECI* is expressed more profusely during the seed maturation and integrates different activities in both, zygotic embryo and SE to induce the embryogenic program and maturation (Lotan et al., 1998; Boutilier et al., 2002; Braybrook & Harada, 2008).

a.3) *Baby Boom (BBM) gene*

Baby boom is known to be expressed in developing embryos and is correlated to cell proliferation and morphogenesis (Boutilier et al., 2002; Nic-Can et al., 2013; Silva et al., 2015). Boutilier et al. (2002) studying *Brassica napus*, published one of the first reports linking *BBMI* to the induction of embryogenesis. They showed that the overexpression of *BBM* has a role in promoting cell proliferation and morphogenesis during embryogenesis. The *BBM* products are similar to the AP2/ERF family of proteins, that are specific plant proteins playing a role in many biological processes such as

determination of cell identity in leaf and floral organs, response to biotic and abiotic stress, embryonic processes and cellular proliferation in meristematic regions (Boutillier et al., 2002).

a.4) *Wuschel-Related Homeobox (WUS) gene*

Wuschel-related homeobox is a gene whose expression has been shown specifically during zygotic embryogenesis in *Arabidopsis* (Arroyo-Herrera et al., 2008). *WUS* is also known to keep cells in an undifferentiated state, plays a specialized function in the embryo development by maintaining the vascular procambium and is highly expressed during germination of zygotic embryos.

b) Cell cycle-related genes and hormone-responsive genes

Somatic embryogenesis is also a process where important changes in the cell cycle occur in order to initiate cell division and elongation. Essential alterations in the expression of a set of genes responsible for cell cycle and cell wall were reported (Yang & Zhang, 2010).

Plant growth regulators (PGRs) play an essential role in SE, mainly auxin and cytokinin. The exposure to auxin shock serves as a trigger, inducing cell division in epidermal cells and probably promoting their differentiation into somatic embryos. Studies showed that auxin-related genes were related to somatic calli induction (Gray, 2004; Quint & Gray, 2006) while cytokinin-related genes were related to meristematic cell maintenance (Su et al., 2011). Abscisic acid (ABA)-related genes were expressed during embryo development and maturation (von Aderkas et al., 2001; Kong et al., 2007; Vahdati et al., 2008). The ABA-regulated gene expression program includes an accumulation of storage proteins, which are considered markers of the maturation phase (Vondrakova et al., 2018) and is followed by the accumulation of late embryogenesis abundant (LEA) proteins. The processes underlying embryo maturation phase are much more documented than the calli induction or the early embryo development phases where studies were carried out on several different protocols in research laboratories, and using non-optimized and often asynchronous SE regeneration processes giving rise to the co-existence of several developmental stages in a same sample.

c) “Omics” studies applied on somatic embryogenesis

Even though different studies investigated the cellular and molecular changes during SE in many different plant species, the molecular basis of the factors involved in initiation of SE process are still not understood (Verdeil et al., 2007; Yang & Zhang, 2010; Mukul-lópez et al., 2012). Researchers are encouraged to invest further in the SE technology and combine it with the latest “omics” techniques

since they can provide a wealth of information in the elucidation of the molecular mechanisms underlying SE.

c.1) Transcriptome studies on SE

To date, very few authors have carried out an RNA-seq approach applied to SE. Very recent studies have shown that authors focused on the early events of SE induction in annual plants (*Arabidopsis* (Magnani et al., 2017), cotton (Yang et al., 2012), rice (Indoliya et al., 2016)). Their results showed that the enriched genes had a role in basic processes such as metabolic pathways and biosynthesis of secondary metabolites. These studies also highlighted a complex auxin signaling pathway involved in callus induction and early differentiation during SE. In woody plants, a number of studies mainly focused on highlighting differences between embryogenic and non-embryogenic calli. For example, Gautier et al. (2019) carried out a comparison between embryogenic callus (EC) and non-embryogenic callus (NEC) in Douglas-fir (*P. menziesii*), and more recently, Chen et al. (2020) compared EC, NEC and globular embryos (GE) in *Dimocarpus longan*. Even though these studies bring insights on the molecular pathways underlying embryogenesis, they exclude an understanding of the process as a whole by taking into account a limited number of developmental stages.

c.2) Metabolome studies on SE

Until recently, very few authors have carried out a metabolomics approach related to SE. All these studies were carried out on conifers. Using gas chromatography–mass spectrometry (GC-MS), metabolomic profiling has been successfully applied to identify metabolic processes regulating SE in five sampled stages in Norway spruce (*Picea abies*) covering cell line proliferation till embryo maturation (Businge et al., 2012). Dobrowolska et al. (2017) applied GC-MS to compare metabolic profiles of normal and aberrant SE germinants in Norway spruce. In the same species, using liquid chromatography-mass spectrometry (LC-MS), Vondrakova et al. (2018) drew endogenous phytohormone profiles in nine sampled stages covering proliferation, maturation, desiccation and germination. Using the same technique, Gautier et al. (2019) compared phytohormone profiles between embryogenic calli and non-embryogenic calli in Douglas-fir (*Pseudotsuga menziesii*). All these authors assumed that clear correlations may exist among the different metabolome or phytohormone profiles and some specific SE stages. Nonetheless, to date, no global analysis, covering explant dedifferentiation till embryo development, of both metabolome and hormone content has been performed on a complete SE process in order to provide a global and kinetic vision of metabolic changes.

V.4. Achievements in coffee somatic embryogenesis

As for other species, studies on SE in coffee is still mainly empirical and low-throughput, and focused on establishing and optimizing somatic embryo production, much more than to understand the mechanisms controlling the expression of this biological phenomenon. Today, research on coffee SE by two leading groups - the CIRAD/ECOM alliance and Nestlé - has led over the last 15 years to the successful industrial application of SE for the two cultivated species, large-scale dissemination of Arabica F1 hybrids and Robusta clones (Etienne et al., 2018). The reliability and efficiency of coffee SE as well as the availability of highly synchronized developmental stages over the coffee SE process make of this process a good model to be combined with “omics” technologies.

Recently, Campos et al. (2017) provided an overview of which achievements and molecular insights have been gained in coffee SE. Since the last 16 years, several works have evidenced the potential of “omics” technologies to optimize coffee SE. For example, Silva et al. (2015) showed higher expression levels of two *BBM* homologous sequences in embryogenic calli and cell suspensions of *C. arabica* when compared to non-embryogenic calli. They also identified *SERK* orthologs in embryogenic cell suspensions (Silva et al., 2014). Arroyo-Herrera et al. (2008) showed that an overexpression of a heterologous gene of *WUS* led to a huge increase in the production of somatic embryos. Quiroz-Figueroa et al. (2002) evidenced differential gene expression in embryogenic and non-embryogenic clusters from cell suspensions in *C. arabica*. The proteome profile of embryogenic calli was recently published (Campos et al., 2016). It was also shown that epigenetic regulation of genes could play a role in *C. canephora* SE as Nic-Can et al. (2013) proved that the genes *LECI*, *BBM1*, and *WUS*-related gene *WOX4* are under epigenetic control in relation to the embryogenic capacity in *C. canephora*.

The genome of *C. canephora* was published and became freely available (Denoëud et al., 2014), opening new possibilities to explore coffee diversity and study numerous biological processes by carrying out RNA-seq approaches (Bertrand et al., 2015). The *C. arabica* genome sequence is expected to be published soon (Mueller et al., 2015), enabling the establishment of a molecular blueprint of coffee SE by applying combined “omics” techniques.

V.5. High-throughput chemical screening

Another approach to understand the underlying mechanisms governing a biological process is the use of genetic mutants. However, this strategy is often impaired by redundancy, lethality or pleiotropy of gene functions, which prevent the isolation of viable mutants (Alves et al., 2017). The high-throughput chemical screening approach has been recognized as an alternative experimental strategy that consists in testing a large number of compounds in a miniaturized and automated biological system, to systematically screen for novel modifiers of a biological phenomenon. A successful chemical screening campaign comprises three equally important elements: (1) a reliable, robust, and quantitative bioassay, which allows to distinguish between potent and less potent compounds tested to modulate a particular biological activity, (2) a rigorous validation process for candidate compounds to establish their selectivity, and (3) an experimental strategy for elucidating a compound's mode of action and molecular target. Two types of chemical screening can be distinguished: forward and reverse (**Figure 17**, Serrano et al., 2015). The goal of forward or phenotypic chemical screening is to identify from an arrayed library of chemicals a bioactive compound causing a phenotypic alteration during a particular biological process, usually in a microplate format. Once a selective compound is found, the molecular target is identified, either by a genetic approach or some type of biochemical purification strategy. On the other hand, the goal of reverse or target-based chemical screening is to identify a compound that modulates the activity of a selected protein or pathway. Subsequently, the chemical is used to determine the phenotypic consequences when applied to growing cells or plants.

High-throughput strategies are gaining increasing attention, particularly in the fields of biofuel production and pharmaceuticals (Szymanski et al., 2011; Radzun et al., 2015). They combine large-scale chemistry and biology data, along with bioinformatics, which is required for data mining (Robert et al., 2009). The effectiveness of these strategies is enhanced by the fact that most active compounds are small molecules that modulate target proteins and/or pathways of a particular biological process (Hicks & Raikhel, 2012). In the last decade, several collections of bioactive compounds became available for the research community, enabling the screening of hundreds of libraries for novel activities (Connor et al., 2012). Nevertheless, this kind of approach has never been applied in plants to optimize culture media for micropropagation, particularly for SE.

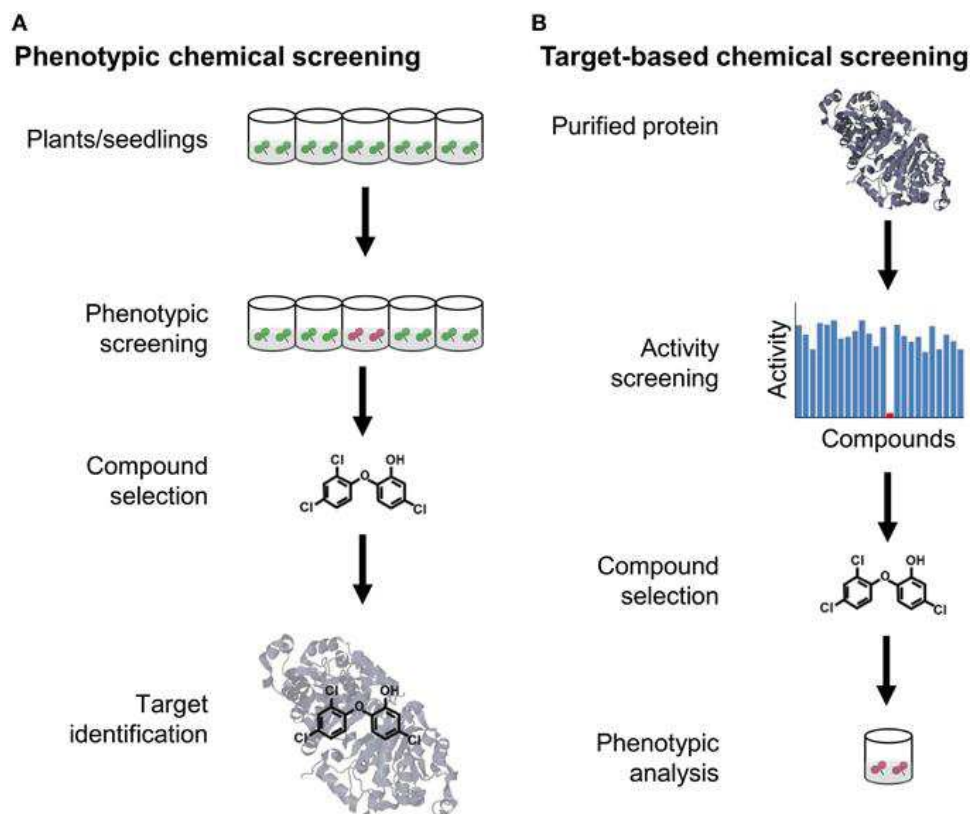


Figure 17. Comparison of forward and reverse chemical screening (from Serrano et al., 2015). **(A)** The goal of forward or phenotypic chemical screening is to identify from an arrayed library of chemicals a bioactive compound causing a phenotypic alteration, usually in a microplate format. Once a selective compound is found, the molecular target is identified, either by a genetic approach or some type of biochemical purification strategy. **(B)** The goal of reverse or target-based chemical screening is to identify a compound that modulates the activity of a selected protein or pathway. Subsequently, the chemical is used to determine the phenotypic consequences when applied to cells or plants.

Part VI

Challenges limiting large-scale dissemination of the genetic progress by somatic embryogenesis in the cultivated coffee species

To date, very few examples of commercial SE applications are available. The industrial application of SE in hardwoods remains limited to a few species, such as the hybrid yellow-poplar, cocoa, and coffee (Etienne et al., 2018). Although SE technology has been described as a very promising tool for rapid and large-scale dissemination of elite materials in many species, its production costs remain high mainly due to low biological efficiency and high genotypic effects (Campos et al., 2017; Guillou et al., 2018).

Whatever the plant species, a number of common bottlenecks have been noted throughout the main three steps of the current SE process –1) the induction of embryogenic callus, 2) its proliferation, and 3) its differentiation into embryos. For instance, a strong genotypic effect has often been reported in the first step (Park et al., 2006). Explants of some interesting clones generated from breeding programs either do not respond to auxin stimulus, or respond by directly producing embryos without passing through a proliferating callus step hence prohibiting the possibility of large-scale propagation enabled by the use of embryogenic cell cultures. For the second step, calli (i.e. unorganized cell proliferations) of some clones lose their embryogenic potential during proliferation (Bobadilla Landey et al., 2015). For the third step, maximizing embryo yields is mandatory to secure the supply chain and reduce production costs (Maximova et al., 2002).

Until now, the main objective of research programs on SE has been the optimization of the nutritive culture media through adding active compounds (nutrients, PGRs...) in the medium and testing their effect on the biological efficiency of the different developmental phases. The ability to use these compounds to improve tissue culture responses eliminates the need to create and market transgenic plants, especially in the food sector (Boutilier et al., 2016). However, to date, SE media optimization remains a long, tricky and tedious operation. Many authors reported that the lack of knowledge on the cellular and molecular mechanisms underlying the reprogramming of somatic cells represents the greatest limitation to the improvement of the SE process (Florez et al., 2015; Campos et al., 2017). As a result, research on SE remains mainly empirical, based on a trial-and-error approach and characterized by an overall slow technical progress.

Another factor contributing to the slow optimization of SE in recent years is that the optimization of nutritive media is being held on a low-throughput level. In fact, testing the effects of active compounds on SE in standard conditions (e.g. using Erlenmeyer flasks or baby-food jars as containers) is time-consuming, space-demanding and costly. To date, the literature on SE in several species has shown that only a limited number of compounds have been chosen to be tested, and when tested, only a very limited range of concentrations was chosen (Moon et al., 2013; Corredoira et al., 2015; Maciel, 2016).

Today, research on coffee SE by two leading groups - the CIRAD/ECOM alliance and Nestlé - has led over the last 15 years to the successful industrial application of SE for the two cultivated species, through the large-scale dissemination of Arabica F1 hybrids and Robusta clones (Etienne et al., 2018). Even though the processes established in the 2 species display low genotypic effects (Ducos et al., 2007; Etienne et al., 2012) and controlled somaclonal variation (Ducos et al., 2003; Bobadilla Landey, 2015), SE production costs established in laboratory conditions remain high and still cannot compete with the low production costs of conventional seedlings or cuttings obtained in horticultural conditions (Georget et al., 2017).

Indeed, as for other species, coffee SE research remains mainly empirical, also characterized by overall slow technical progress. For example, the development of culture conditions for the mass redifferentiation into somatic embryos have been very long (10 years) and laborious, carried out only by an empirical approach (Etienne et al., 2018).

While the micropropagation of coffee *in vitro* plantlets can today reach 2-3 M SE-derived plants per year, SE faces many challenges in fulfilling growing market demand for elite coffee varieties (Arabica F1 hybrids and Robusta clones), estimated at around 50-100 M plants per year (Etienne et al., 2018). This market demand for more resilient varieties comes from the recent loss of leaf rust resistance in all the conventional cultivated line varieties for Arabica, and the increasing impact of the harmful effects of climate change. For Robusta, there is also an urgent need to replace unselected clones with improved ones, particularly in Vietnam – the world’s main coffee producer – where aging coffee orchards need to be totally renewed.

Part VII

Thesis general objectives

Whatever the species considered, SE research remains essentially empirical and at a low-throughput level, resulting in many drawbacks for fulfilling market demand, especially due to overall slow technical progress over the last decades.

We are convinced that a high-throughput screening approach is the solution to quickly optimize SE protocols by testing, in a limited period of time, a wide range of active compounds (nutrients, growth regulators, etc.), multiplying experimental conditions, and reducing volumes as well as the space and manpower required. Despite high expectations in the plant biotechnology community, such a high-throughput approach has never been proposed for plant cells. We are convinced that it should have numerous impacting developments in micropropagation.

Using *C. arabica* as a model plant, the **first objective** of this thesis was to develop, for the first time in plant micropropagation, a miniaturized and automated screening system to optimize coffee somatic embryo formation protocols, able to meet high-throughput standards (elaborated in **Chapter II**).

Knowledge about the molecular events involved in the key steps of the SE process are also urgently needed to pilot the optimization of the protocols through the identification of key genes and metabolites. The **second objective** of this thesis was to take advantage of the latest metabolomics and transcriptomics technologies and apply them to one of the most advanced and reliable large-scale SE process —the one developed for coffee — in order to unravel underlying global mechanisms, from cell dedifferentiation until cell redifferentiation, and identify metabolic and molecular candidates with the final aim to be able to rationally pilot the optimization of the coffee SE protocols (elaborated in **Chapters III and IV**).

Together with the massive amount of information provided by “-omics” technologies allowing a description at the molecular level of the successive developmental pathways involved in key SE steps, high-throughput screening can be piloted in a rational way to identify specific key compounds and further use them to optimize SE industrial protocols, offering unprecedented perspectives in plant micropropagation.

Chapter II

An innovative automated active compound screening system allows high-throughput optimization of somatic embryogenesis in *Coffea arabica*

Published paper in *Scientific Reports* on the 21st of January, 2020.

Background

Of all the possible micropropagation techniques, vegetative propagation by SE is by far the most promising for rapid, large-scale dissemination of elite varieties. This technique is especially relevant for woody species with long life cycles that are difficult to propagate in a clonal form by conventional methods and for which there is an urgent need for a method allowing large-scale disseminating of new selected varieties. However, whatever the species considered, SE research performed either in research laboratory or directly in the industrial production units (in both situations, this research is expensive and difficult to quickly transfer onto commercial protocols) still remains essentially empirical and at a low-throughput level, resulting in many drawbacks for fulfilling market demand, especially due to overall slow technical progress for several decades, and the immense hopes associated with the initial boom in research into this regeneration process following the coffee SE discovery in the 1980s have not led to the many commercial applications announced at that time. We are convinced that a high-throughput approach is the solution to quickly optimize SE protocols by testing, in a limited time period, a wide range of active compounds (nutrients like sugars, organic compounds, mineral salts, growth regulators, etc.), multiplying experimental conditions, and reducing volumes as well as the space and manpower required; hence, reducing the overall cost of this optimization. Despite high expectations in the plant biotechnology community, such a high-throughput approach has never been proposed for plant micropropagation.

In this chapter, we will show how three years of research work on *C. arabica* SE has led us to develop, for the first time in plant micropropagation, a miniaturized and automated screening system able to meet high-throughput standards. Our results clearly will focus on a major step of the SE process, i.e. the embryo regeneration step. Our ambition was to successfully miniaturize and automatize this step without affecting regeneration capacity, allowing efficient, easy, quick, cheap, and reliable testing of numerous compounds at varied concentrations. We will also carry out pilot screening for potentially active compounds affecting SE as proof-of-concept for the screening method by testing one particular compound, Trichostatin A.

OPEN

An innovative automated active compound screening system allows high-throughput optimization of somatic embryogenesis in *Coffea arabica*

Rayan Awada^{1,2,3*}, Dorothée Verdier¹, Solène Froger¹, Eric Brulard¹, Simone de Faria Maraschin⁴, Hervé Etienne^{2,3} & David Breton¹

Somatic embryogenesis (SE) faces many challenges in fulfilling the growing demand for elite materials. A high-throughput approach is required to accelerate the optimization of SE protocols by multiplying experimental conditions within a limited time period. For the first time in plant micropropagation, we have developed a miniaturized and automated screening system to meet high-throughput standards. *Coffea arabica* embryo regeneration, classically achieved in 250-ml Erlenmeyer flasks, was successfully miniaturized in 24-well plates, allowing a volume downscaling factor of 100 and a space saving of 53 cm²/well. Cell clusters were ground and filtered to fit the automated pipetting platform, leading to fast, reproducible and uniform cluster distribution (23.0 ± 5.5 cell clusters/well) and successful regeneration (6.5 ± 2.2 embryos/well). Pilot screening of active compounds on SE was carried out. Compounds belonging to the histone deacetylase inhibitor family were tested for embryo regeneration efficiency. Cells treated with 1 μM Trichostatin A showed a marked 3-fold increase in the number of regenerated embryos. When re-tested in 250-ml flasks, the same enhancement was obtained, thereby validating the miniaturized and automated screening method. These results showed that our screening system is reliable and well suited to screening hundreds of compounds, offering unprecedented perspectives in plant micropropagation.

The rapid increase in environmental stresses, especially those associated with climate change, and growing demand from the food and timber industries, have put enormous pressure on trees¹. Developing new technologies for the clonal propagation, improvement and breeding of trees can help solve these problems². This has been achieved in part using biotechnology methodologies, such as *in vitro* propagation, genetic transformation, and marker-assisted breeding for the gradual genetic improvement of woody plants^{3,4}.

For *in vitro* propagation, shoot proliferation methods using adventitious shoots and axillary buds have been widely applied, though research is still largely focused on using somatic embryogenesis (SE) in woody plants⁵. Today, SE is considered as a very promising tool for the rapid and large-scale propagation of elite varieties in many species^{5,6}, and is particularly important for those that have a long life cycle and are difficult to propagate by conventional methods⁷.

However, to date, very few examples of commercial SE applications are available. The industrial application of SE in hardwoods remains limited to a few species, such as the hybrid yellow-poplar, cocoa, and coffee⁸. SE technology is still work-in-progress, mainly due to low biological efficiency and high genotypic effects resulting in high production costs^{6,9}.

To guarantee the security of the supply chain in response to market demand for elite materials, it is necessary to maximize SE-derived plant production and quality, i.e. to improve SE protocols. Whatever the plant species, a number of common bottlenecks have been noted throughout the main three steps of the current SE process –(1)

¹Nestlé R&D Center Tours, 101 avenue Gustave Eiffel, 37097 Cedex 2, Tours, France. ²CIRAD, UMR IPME, F-34398, Montpellier, France. ³UMR IPME, Université de Montpellier, CIRAD, IRD, F-34398, Montpellier, France.

⁴Bioalternatives SAS, 1 bis Rue des Plantes, 86160, Gençay, France. *email: rayan.awada@cirad.fr

the induction of embryogenic callus, (2) its proliferation, and (3) its differentiation into embryos. For instance, a strong genotypic effect has often been reported in the first step¹⁰. Explants of some interesting clones either do not respond to auxin stimulus, or respond by directly producing embryos without passing through a proliferating callus step. For the second step, calli of some clones lose their embryogenic potential during proliferation¹¹. For the third step, maximizing embryo yields is a necessity to secure the supply chain and reduce production costs¹².

Until now, one of the conventional methods for improving SE protocols has consisted in applying active compounds. Adding small (<500 Da) bioactive compounds (nutrients, growth regulators, epigenetic regulators, etc.) to the culture medium that have a potential effect on SE targets, and are capable of modulating the physiological state of cells, is a possible way of maximizing embryo production and quality, and of removing genetic or stress-induced recalcitrance¹³. The ability to use small compounds to improve tissue culture responses eliminates the need to create and market transgenic plants¹⁴. This is important in the food sector, where consumers are reluctant to consume transgenic products.

However, testing the effect of new compounds on SE efficiency under standard conditions is generally a laborious, intensive and tricky operation, because it is time-consuming, space-demanding and costly. To date, the literature on SE in several species has shown that only a limited number of compounds have been chosen to be tested, on the basis of prior hypotheses (targeted selection)^{15–17} and when tested, only a very limited range of concentrations was chosen. The above inconveniences also hampered the possibility of testing these compounds in combinations. The resulting low-throughput research strategy has led to overall slow technical progress on SE protocols in recent years.

At the same time, high-throughput strategies are gaining increasing attention, particularly in the fields of biofuel production and pharmaceuticals^{18,19}. These strategies consist in testing a large number of compounds at different concentrations (untargeted selection) in a miniaturized and automated system, to systematically screen for novel modifiers of a biological phenomenon. They combine large-scale chemistry and biology data, along with bioinformatics, which is required for data mining²⁰. The effectiveness of these strategies is enhanced by the fact that most active compounds are small molecules that modulate target proteins and/or pathways of a determinate biological process²¹. In the last decade, several collections of bioactive compounds became available for the research community, enabling the screening of hundreds of libraries for novel activities²². Nevertheless, to our knowledge, this kind of approach has never been applied to optimizing culture media for micropropagation, particularly for SE.

In plants and mammals, epigenetic regulation and histone modifications occur widely during cellular differentiation and development^{23,24}. Nic-Can *et al.*²⁵ showed that genes related to SE, such as *LEAFY COTYLEDON1*, *BABY BOOM1* and *WUSCHEL-RELATED HOMEODOMAIN 4*, are epigenetically regulated in *Coffea canephora*. Boutilier *et al.*¹⁴ showed that the switch to haploid embryogenesis is controlled by the activity of histone deacetylases (HDAC). Blocking HDAC activity with HDAC inhibitors (HDACi) in *Brassica napus*, *Brassica rapa*, *Arabidopsis thaliana* and *Capsicum annuum* male gametophytes resulted in a large increase in the proportion of cells undergoing embryogenic growth¹⁴.

Coffee (*C. arabica* & *C. canephora* cv Robusta) has a great economic impact in many producing countries, especially in South America²⁶. Today, coffee SE is one of the most advanced technologies. Research on SE by two leading groups - the CIRAD/ECOM alliance and Nestlé - has led over the last 13 to 15 years to the successful industrialization and commercial application of SE for the two cultivated species, through the large-scale dissemination of Arabica F1 hybrids and Robusta clones⁸. Even though the processes established in the 2 species display good biological efficiency, with low genotypic effects^{27,28} and controlled somaclonal variation^{11,29}, SE production costs remain high and still cannot compete with the production costs of conventional seedlings or cuttings³⁰.

While the micropropagation of coffee *in vitro* plantlets can today reach 2–3 M SE-derived plants per year, SE faces many challenges in fulfilling growing market demand for elite coffee varieties, estimated at around 50–100 M plants per year⁸. This market demand for more resilient varieties comes from the recent loss of leaf rust resistance in all the conventional cultivated line varieties for Arabica, and the increasing impact of the harmful effects of climate change. For Robusta, there is also an urgent need to replace unselected clones with improved ones, particularly in Vietnam – the world's main coffee producer – where aging coffee orchards need to be totally renewed⁸.

As for other woody species, optimizing coffee SE protocols is a long and laborious process. Yet, increasing production efficiency *in vitro* would make SE technology more cost-effective, hence the interest in finding new active compounds able to increase embryogenic callus formation and embryo yield. A high-throughput approach might be a solution, given the possibility of multiplying experimental conditions in a limited time period.

We focus in this paper on one of the main three SE steps in *C. arabica*: embryo regeneration from embryogenic cell clusters. This SE step is classically achieved in 250-ml Erlenmeyer flasks. Here, we describe how we miniaturized this step in 24-well plates and made it compatible with an automated pipetting platform. Cell cluster miniaturization was also required. Once miniaturization and automation were successful, we carried out pilot screening using compounds belonging to the HDACi family, for their potential role as active compounds to increase embryo yield. Lastly, treatments showing a significant positive effect on embryo yield were tested again under standard *in vitro* conditions to confirm treatment efficiency and validate the automated and miniaturized screening method.

Results

Grinding and filtration of embryogenic tissues to overcome the pipetting constraint.

Preliminary profiling of long-term cultured Arabica embryogenic cell clusters showed that their diameter fell within the range of 100–1000 µm and followed a statistical normal distribution, with a maximum peak around 500–600 µm (Fig. 1a). For automation purposes, uniformity and reproducibility are required. Hence, in order for the cell clusters to fit the pipetting tips, filtration at 700 µm was needed to exclude a diameter greater than 700 µm,

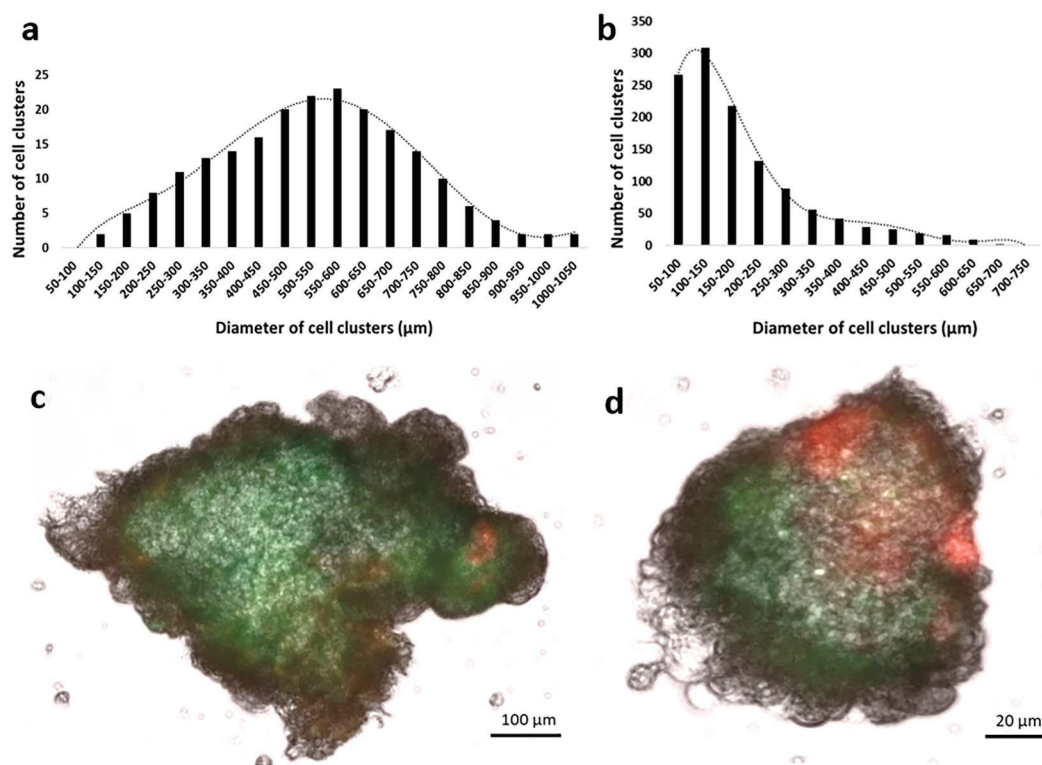


Figure 1. Successful miniaturization of Arabica embryogenic cell clusters. **(a)** Profiling of cell clusters according to their diameter before grinding and filtration. Data of 210 cell clusters were plotted. The black curves show that the plotted values are normally distributed. **(b)** Profiling of cell clusters according to their diameter after grinding and filtration at 700 μm . Data of 1210 cell clusters were plotted. Viability of cell clusters before **(c)** and after grinding **(d)** following double staining with FDA and PI. FDA fluoresces in green in living cells and PI fluoresces in red in dead cells.

which amounted to approximately 40% of the whole cell suspension. To benefit from all the available embryogenic material at our disposal, we decided to add a grinding step to reduce and homogenize cell cluster diameter. Figure 1b shows the effect of grinding on the particle diameter distribution. Cell cluster diameters maintained their statistical normal distribution, but with a shift of the maximum peak towards the range of 100–150 μm . Viability tests (Fig. 1c,d) showed similar staining profiles between ground and intact cell clusters, with a prevailing green light emission indicating that cells had high viability. Some peripheral cells emitted red light, indicating their dead status in both ground and intact clusters, though the red emission intensity was more noticeable in ground cells, due to the effect of the grinding blades.

The regeneration capacity of ground material had first to be tested under standard *in vitro* culture conditions (Erlenmeyer flasks) (Fig. 2). After 9 weeks of culturing, an average of $3,039 \pm 485$ normal torpedo-shaped embryos were obtained per gram of inoculated callus (embryos. g^{-1} callus) representing a biomass of 0.43 ± 0.12 g ($n = 6$ flasks/condition/replicate). These values were significantly lower compared to those obtained with intact cell clusters, i.e. $4,611 \pm 685$ embryos. g^{-1} callus and a biomass of 0.83 ± 0.18 g ($***P < 0.001$) (Fig. 2a,b). However, it should be noted that better synchronization of embryo development was observed. Embryos generated from ground cell clusters were morphologically more uniform, since very few proliferating clusters remained, limiting spatial competition (Fig. 2c,d).

Achieving uniform cell cluster distribution after automated pipetting in miniaturized culture containers.

In order to obtain uniform distribution of cell clusters at the start of culturing, two parameters were quantified to verify uniformity of distribution: the number of cell clusters in each well, and the total area they occupied. The values of 4 independent replicates ($n = 24$ wells/replicate) were uniform ($P > 0.05$) and similarly distributed between plates (1, 2, 3, 4) and between rows (A, B, C, D), having an average of 23.0 ± 5.6 clusters/well and occupying an area of 1.39 ± 0.12 mm^2 /well (Table 1, Fig. 3a).

The preliminary results showed medium evaporation in peripheral wells (loss of 40% of total volume after 4 weeks compared to central wells) causing non-reproducible embryo regeneration (data not shown). Parafilm-sealed plates were then placed in a culture chamber with 90% RH, which drastically limited the evaporation rate to less than 2% in all wells, and led to uniformity between tested replicates ($P > 0.05$). Similar distribution of biomass parameters was obtained between peripheral wells (22.5 ± 5.39 cell clusters/well; area occupied 1.39 ± 0.13 mm^2) and central wells (23.9 ± 5.63 clusters/well; area occupied 1.39 ± 0.11 mm^2) (Table 1).

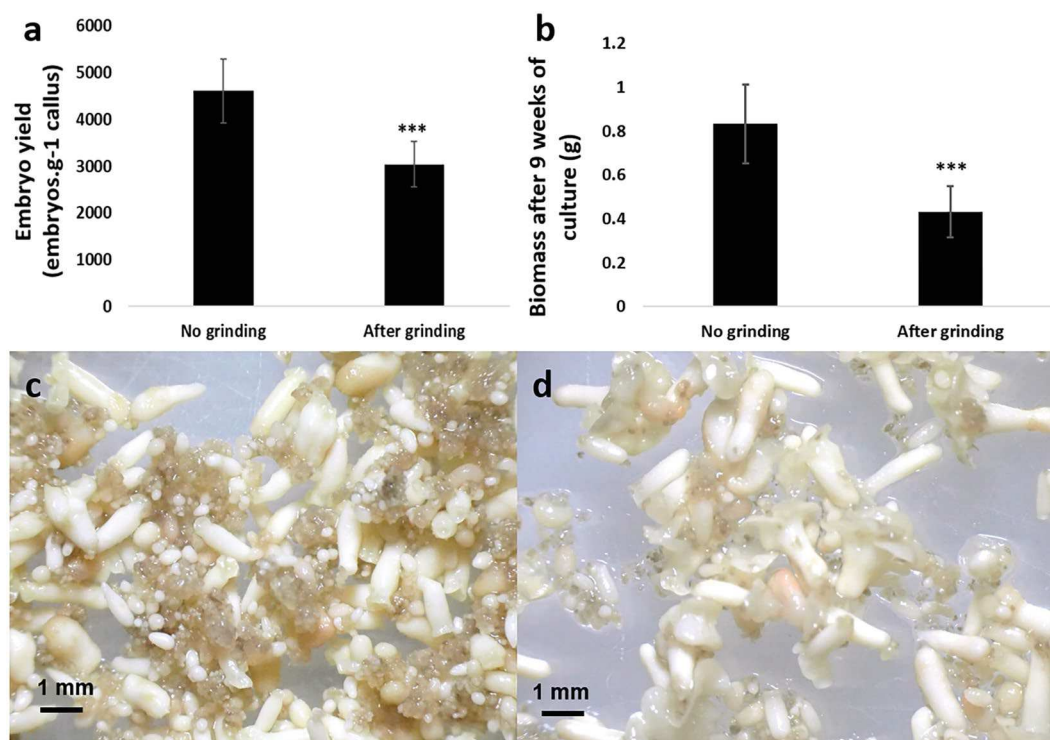


Figure 2. Assessing the effect of cell cluster grinding on embryo regeneration in Erlenmeyer flasks. **(a)** Embryo yield (embryos. g^{-1} callus) obtained after 9 weeks of embryo regeneration. **(b)** Biomass (g) obtained after 9 weeks of embryo regeneration. Bars show the mean of triplicate samples ($n = 6$ flasks/condition/replicate) and error bars represent the SD. Differences between both groups were analysed with t-tests ($***P < 0.001$). **(c)** Appearance of torpedo-shaped embryos (white structures) obtained after 9 weeks of embryo differentiation from intact cell clusters. Remaining embryogenic clusters can still be observed (brown structures). **(d)** Appearance of torpedo-shaped embryos regenerated after 9 weeks of differentiation from ground cell clusters. Better synchronization of embryo development is observed.

Maintenance of regeneration capacity and embryo yield reproducibility in a miniaturized environment.

The efficiency of embryogenic cell cluster conversion to embryos was validated in the 24-well plate culture system. The regenerated embryos were normally shaped, with a conversion rate of 6.5 ± 2.1 torpedo-shaped embryos/well after 9 weeks (Table 1, Fig. 3b). The values of 4 independent replicates ($n = 24$ wells/replicate) followed a statistical normal distribution. High uniformity in embryo yield rates was found between plates, rows and peripheral and central wells ($P > 0.05$), confirming the very limited impact of automated pipetting on cluster distribution, whatever the well position. However, it should be noted that expressing this embryo yield in wells by gram of inoculated callus resulted in a mean of $13,000 \pm 4,200$ embryos, which was higher than the embryo yield obtained in a flask culture system from ground or intact cell clusters, i.e. $3,039 \pm 485$ embryos and $4,611 \pm 685$ embryos, respectively. However, for the same inoculated cell cluster density, a delay in embryo development was found in the miniaturized system, where the first globular-shaped embryos emerged after 6 weeks (Fig. 4d–f), as opposed to 4 weeks in Erlenmeyer flasks.

Successful pilot screening for the effect of active compounds on somatic embryogenesis.

After satisfactory implementation of a miniaturized and automated embryo regeneration assay, pilot screening was carried out by testing histone deacetylase inhibitor (HDACi) treatments. Four compounds belonging to the HDACi family – Oxamflatin, SAHA, Scriptaid and Trichostatin A – were tested at 3 different concentrations (0.1, 1 and $10 \mu M$) as potential active compounds affecting *C. arabica* embryo differentiation. The regenerated embryo numbers of 3 independent replicates ($n = 18$ wells/HDACi/concentration/replicate) were uniform between plates and followed a statistical normal distribution between culture wells. Only clusters treated with $1 \mu M$ TSA showed a significant increase in the number of torpedo-shaped embryos obtained after 9 weeks (18.3 ± 3.9 embryos/well vs 6.0 ± 2.5 for untreated control, $***P < 0.001$) (Fig. 4a,b) and the appearance of the regenerated torpedo-shaped embryos was comparable between clusters treated or not with TSA (Fig. 4b). The area occupied by cell clusters in a well was recorded over a 6-week culture period up to the emergence of globular-shaped embryos (Fig. 4c,d). Comparing the resulting curves established with cell clusters treated with $1 \mu M$ TSA and untreated clusters showed that cluster growth started after 3 weeks and exponentially increased up to the sixth week of culturing ($R^2 = 0.93$), when the first globular-shaped embryos emerged (Fig. 4d). From 4 weeks under embryo regeneration conditions, the growth of cell clusters treated with $1 \mu M$ TSA was significantly better compared to the untreated control (total area of $35.0 \pm 3.9 \text{ mm}^2$ vs $21.9 \pm 3.4 \text{ mm}^2$ after 6 weeks, $***P < 0.001$), explaining the larger number of regenerated embryos obtained after 9 weeks (Fig. 4a).

Parameters	Assessment of distribution uniformity									
	Between plates				Between rows				Between central and peripheral wells	
	Plate 1	Plate 2	Plate 3	Plate 4	Row A	Row B	Row C	Row D	Central wells	Peripheral wells
Number of cell clusters per well	22.1 ± 5.7 _a	24.4 ± 6.0 _a	22.0 ± 5.5 _a	23.3 ± 5.0 _a	21.1 ± 4.7 _a	24.5 ± 5.4 _a	22.5 ± 5.0 _a	23.8 ± 6.6 _a	23.9 ± 5.6 _a	22.5 ± 5.4 _a
Total area occupied (mm ²)	1.40 ± 0.10 _a	1.41 ± 0.12 _a	1.37 ± 0.13 _a	1.38 ± 0.12 _a	1.39 ± 0.11 _a	1.38 ± 0.10 _a	1.41 ± 0.12 _a	1.38 ± 0.15 _a	1.39 ± 0.11 _a	1.39 ± 0.13 _a
Embryo yield per well	6.9 ± 2.5 _a	6.7 ± 2.4 _a	6.2 ± 1.7 _a	6.3 ± 2.2 _a	7.0 ± 2.9 _a	6.5 ± 1.8 _a	6.2 ± 2.2 _a	6.4 ± 1.8 _a	6.3 ± 2.0 _a	6.6 ± 2.3 _a

Table 1. Assessing the uniformity and reproducibility of automated distribution of Arabica cell clusters in 24-well plates. Uniformity of distribution was revealed between plates, between rows, and between central and peripheral wells. Three parameters were quantified: the number of clusters per well and the total area they occupied (mm²) at the beginning of the regeneration step, and the number of regenerated embryos per well after 9 weeks. Values represent the mean ± SD of 4 independent replicates (n = 24 wells/replicate). For each parameter, differences between plates and between rows were analysed with a one-way ANOVA test, while a t-test was used to analyse differences between central and peripheral wells. The letter “a” shows no significant difference between groups (P > 0.05).

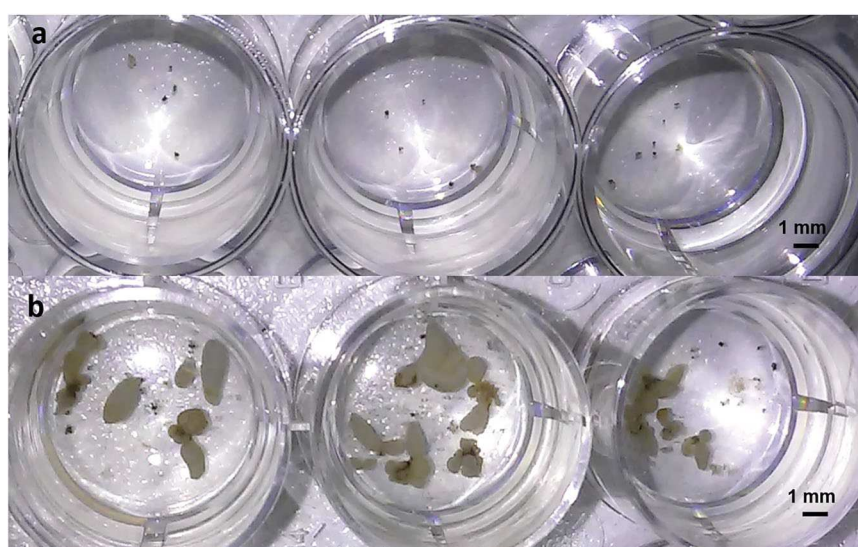


Figure 3. Successful uniform distribution of Arabica cell clusters resulting in a uniform number of regenerated embryos. Programming the automated pipetting platform resulted in uniform distribution of cell clusters between wells at the beginning of the regeneration step as well as a uniform number of embryos regenerated after 9 weeks. (a) Appearance of inoculated cell clusters in a 24-well plate after distribution using the automated pipetting platform after cell cluster grinding and filtration. (b) Appearance of embryos regenerated from ground clusters in a “24-well plate” miniaturized culture system.

To validate the screening carried out in the automated and miniaturized system, the best treatment obtained with 1 μM TSA, with an exposure time of 24 hours, was re-tested under standard *in vitro* conditions, i.e. 250-ml Erlenmeyer flasks. The TSA treatment enhanced embryo regeneration in the same way compared to the untreated embryogenic clusters (9,325 ± 2,240 embryos. g⁻¹ callus vs 4,673 ± 1,614 embryos. g⁻¹ callus, respectively; P < 0.001; Fig. 5a). The beneficial effects of TSA were also visible through the increased overall biomass (1.12 ± 0.26 g vs 0.69 ± 0.15 g, P < 0.001; Fig. 5b), and the increased length of the treated torpedo-shaped embryos (896 ± 463 μm vs 760 ± 429 μm, respectively, P < 0.001; Fig. 5c). Extended exposure to 1 μM TSA, which was also tested (7, 14 d), did not significantly improve the embryo regeneration yield compared to an exposure time of 24 h, but biomass and torpedo mean size significantly increased with an exposure of 14 d to 1 μM TSA.

Treating cell clusters with TSA does not affect plant regeneration. The phenotypic data recorded on 160 young plants ready for greenhouse transfer did not reveal any significant differences between plants regenerated from cell clusters treated, or not, with TSA. An ANOVA did not indicate any treatment effect on either main root or stem lengths (data not shown). All plants displayed orthotropic development, with uniform pigmentation and leaf shape (Fig. 6a). A Multiple Component Analysis (MCA) performed on the scores obtained for the main phenotypic categorical variables (number of pairs of leaves, number of roots, number of stems, leaf colour index, Table 2) provided 19 components reaching 3.8 inertia (data not shown). However, the first

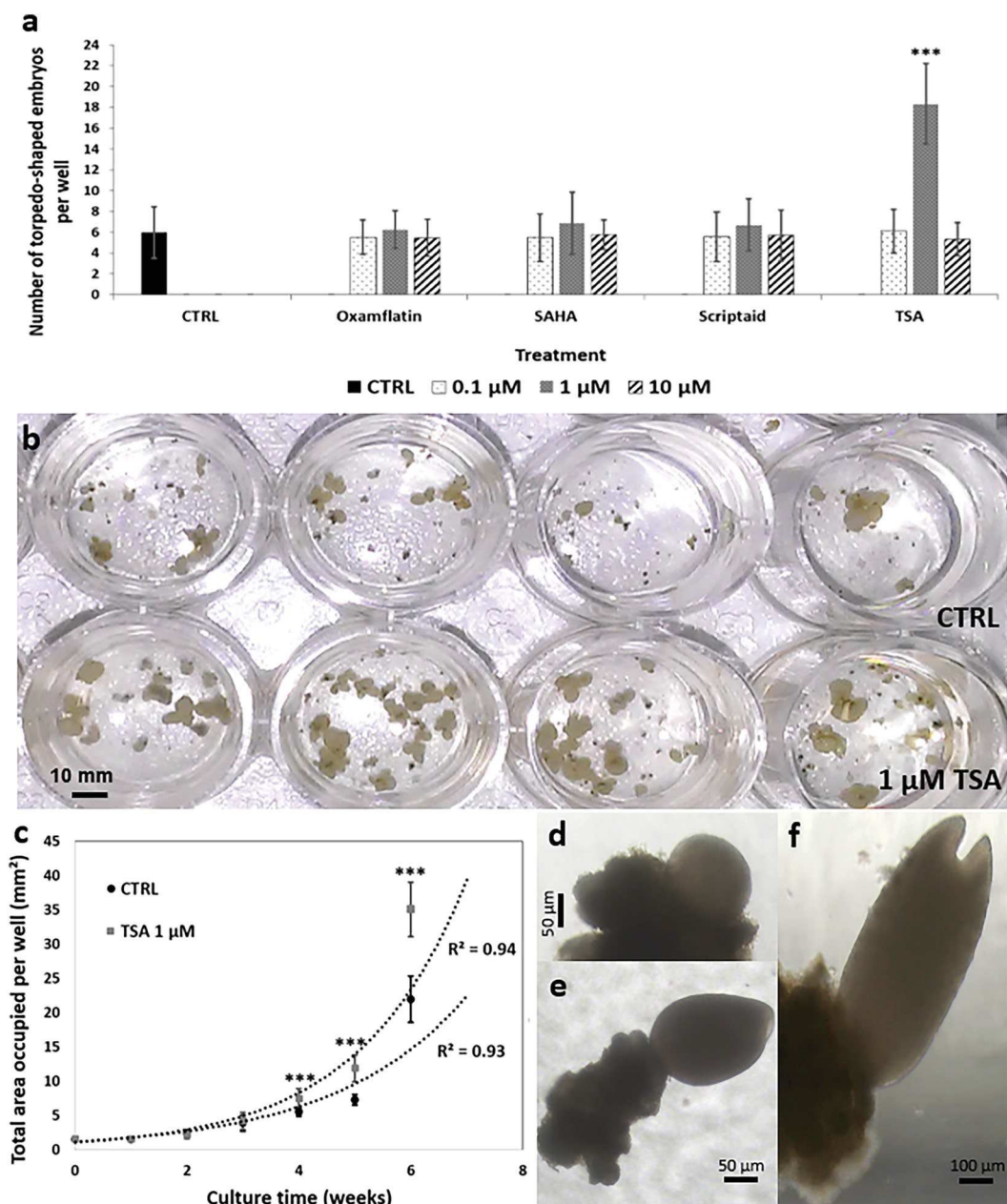


Figure 4. Successful pilot screening of active compounds affecting SE in the automated and miniaturized system. Four compounds belonging to the HDACi family (Oxamflatin, Scriptaid, SAHA and TSA) were tested in 24-well plates at three different concentrations (0.1, 1 and 10 μM), as potential active compounds affecting *C. arabica* embryo differentiation. Cell clusters were exposed for 24 h to treatment before washing. (a) Embryo regeneration was scored after 9 weeks. The “CTRL” group was untreated (control). Bars show the mean of triplicate samples ($n = 18$ wells/HDACi/concentration/replicate) and error bars represent the SD. Differences between treated clusters and the control were analysed with a one-way ANOVA test ($***P < 0.001$). (b) Picture showing the increased embryo yield in the wells treated with 1 μM TSA (bottom) vs control wells (top). (c) The total area occupied by cell clusters was monitored during the embryo regeneration period. Plotted values are the mean of triplicate samples ($n = 18$ wells/treatment/replicate) and error bars represent the SD. Differences between clusters treated with 1 μM TSA and the control were analysed with a t-test ($***P < 0.001$). Cell cluster growth followed an exponential trendline ($R^2 = 0.93$) up to the sixth week. (d–f) Pictures of an emerged globular-shaped, heart-shaped and torpedo-shaped embryo after 6, 7 and 8 weeks under embryo regeneration conditions, respectively.

two axes accounted for only 18.9% of total inertia. Plotting categorical data according to axes 1 and 2 revealed a strong contribution to global variability for 2 plants scored “376” and one scored “382” for their colour index (Supplementary Data). As a result, no distinction could be detected between TSA treatments so we decided to

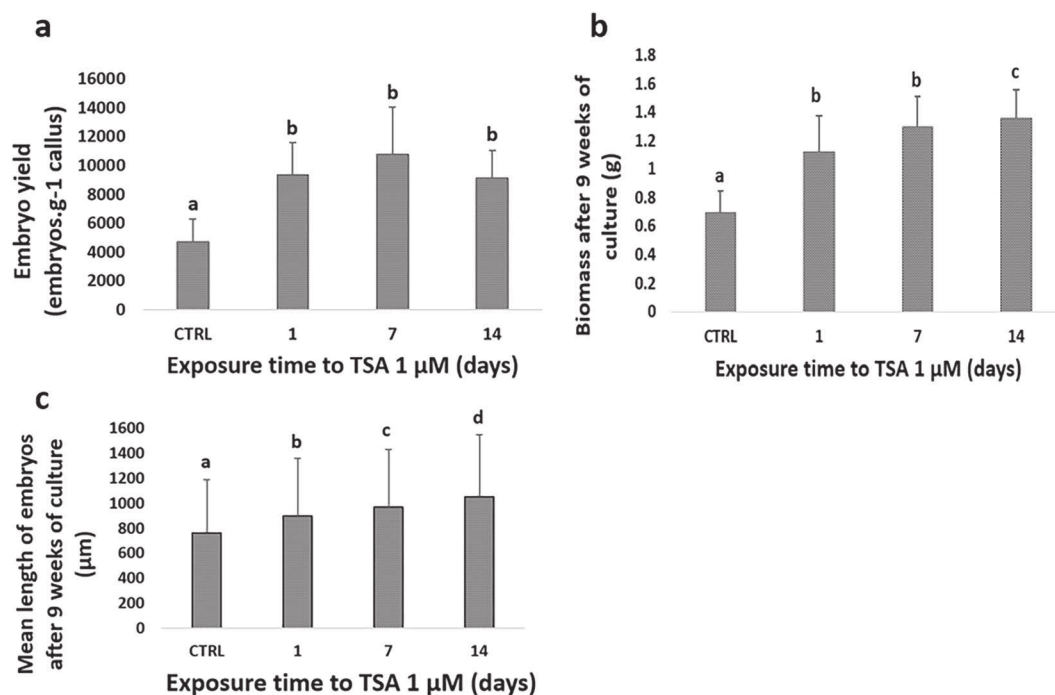


Figure 5. Validation in Erlenmeyer flasks of active compounds highlighted in the pilot screening. Treatment with 1 μM TSA was re-tested in standard 250-ml Erlenmeyer flasks with an exposure time varying between 1, 7 and 14 days. **(a)** Embryo yield (embryos.g⁻¹ callus) obtained after 9 weeks. **(b)** Biomass (g) obtained after 9 weeks. **(c)** Embryo length. Bars show the mean of triplicate samples ($n = 6$ flasks/exposure time/replicate for **(a,b)** and $n = 100$ embryos/exposure time/replicate for **(c)**) and error bars represent SD. The “CTRL” group was untreated (control). Differences between treated clusters and the control were analysed with a one-way ANOVA test for **(a,b)**, while a Kruskal-Wallis test was used for **(c)** followed by a Dunn test as a post hoc. Means with different letters are significantly different ($P < 0.001$).

discard these data and run another MCA. This time, 17 components were detected with a total level of 3.4 inertia (data not shown). The contribution of the first two axes to this inertia reached 19.6%. This time, a graphical analysis revealed 2 main groups of categorical data and a smaller third one including atypical, double stem and unrooted plants (Fig. 6e). This group did not include TSA treatment or control conditions, indicating a lack of correlation between variables. Considering the upper left of the graph, plants from untreated tissues and tissues treated for 7 days with TSA correlated with plants bearing 2 to 4 pairs of leaves, 1 to 2 roots and various leaf colour indexes (357, 364, 371). Conversely, on the right of the graph, plants arising from cell clusters treated for 1 and 14 days correlated with well-developed plants: 5 to 7 pairs of leaves, 3 to 4 roots, and a stable colour index (377). Thus, treating cell clusters with 1 μM TSA did not show any significant deleterious effect on regenerated plants.

Discussion

We proved in this study that a major step of SE processes, i.e. embryo regeneration from embryogenic cell clusters, can be successfully miniaturized and automated allowing easy, quick, cheap and reliable testing of numerous compounds at varied concentrations. Switching from conventional 250-ml flask culture containers filled with 100 ml of culture medium to 24-well plates with 1 ml of available nutritive medium allowed a miniaturization factor of 100. With this strong downscaling, consideration could be given to testing the effectiveness of a much larger number of molecules at different concentrations, while reducing the space required (2.5 cm² required per well compared to 55.4 cm² required per flask and the possibility of stacking multi-well plates). In addition, automation led to a considerable gain in speed and reduced manpower, where only 1 working full-time employee (FTE) was needed to supervise the pipetting platform. For instance, the pilot screening carried out using HDACi treatments required 36 plates, 864 wells, i.e. 864 independent culture environments. The same experiment in flasks was simply unmanageable, as 1 FTE can only inoculate 100 flasks a day, not to mention the space required, whereas inoculation of several hundred wells can be done within hours. Many authors have also shown the importance of miniaturization and automation in animal or plant cell screening systems^{31–33}. For instance, Rodriguez-Furlan *et al.*³³ also showed that the 24-well microplate format used to screen for compounds modulating *A. thaliana* seedling growth required a minimum of sample handling and allowed automatic acquisition of quantitative data and rapid monitoring of morphological traits (leaf and root growth). This container format, along with automation, allowed the screening of 10,000 structurally diverse compounds in a limited time period.

While animal cell screening systems are mainly derived from drug discovery programmes and cancer research^{18,34}, plant cell screening systems were only recently used to study compound effects on growth and development. To date, in plants, only *A. thaliana* seedlings have been used in such studies, because of their small size,

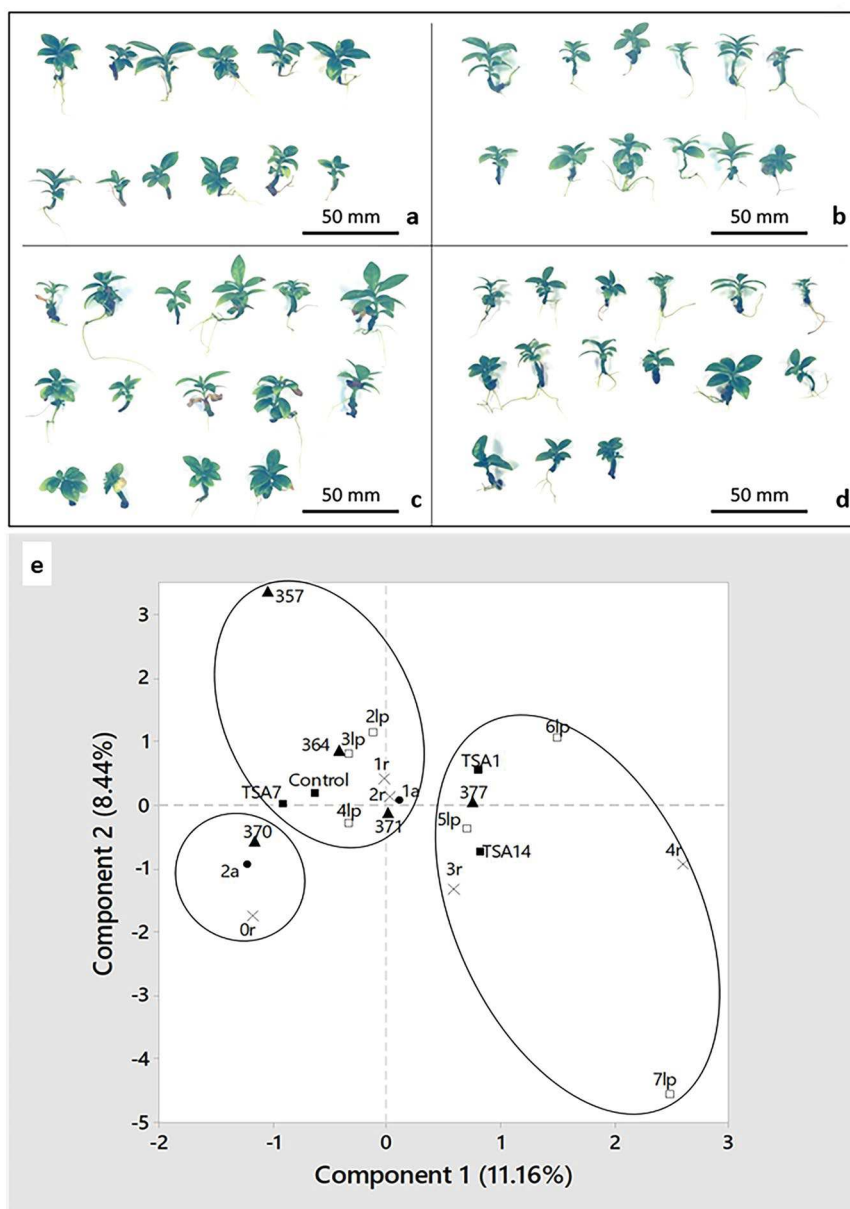


Figure 6. *In vitro* plant phenotypic characterization following 1 μM TSA treatment during the coffee SE process. (a) Plants regenerated from untreated embryogenic callus. (b–d) Plants regenerated from TSA-treated embryogenic clusters for 1, 7 and 14 days, respectively. (e) Graphic representation of a Multiple Component Analysis (MCA) performed on plant phenotypic variables. Data from 157 plants were plotted according to the first 2 axes of the MCA. The number of pairs of leaves are represented by white squares, the number of roots by crosses, the number of stems by black spots, the leaf colour index by triangles and the *in vitro* treatment prior to regeneration by black squares.

allowing development in multi-well containers^{33,35}. However, automated screening systems for active compounds have never been used in plant propagation procedures. Our system provided the first evidence in plants that miniaturized containers do not hamper embryogenic cell cluster growth and differentiation, and thus enable screening for the effectiveness of compounds in micropropagation, under real conditions and on economically important crops.

Our results showed that even after miniaturizing the embryo regeneration step, mechanisms occurring during conventional-scale SE were preserved, as reliable production of well-shaped embryos was achieved. SE miniaturization did not engender any developmental heterogeneity. However, the type of grinding chosen had significant effects by lowering embryo yields in flasks. Since this biological efficiency was reproducible during this step, we did not consider this difference to be a drawback, because the main objective of this miniaturized system was not large-scale embryo production, or to study molecular mechanisms underlying embryogenesis, but to use it on SE as a tool to screen for active compounds, with a view to speeding up protocol optimization.

Phenotypic scores								
Leaves	2lp	3lp	4lp	5lp	6lp	7lp		Total
Control		12	18	10				40
TSA1		13	14	11	2			40
TSA7		6	15	16	2	1		40
TSA14	1	17	16	6				40
Total	1	48	63	43	4	1		160
Roots	0r	1r	2r	3r	4r			Total
Control	5	20	13	2				40
TSA1	1	20	15	3	1			40
TSA7	3	17	14	5	1			40
TSA14	3	19	14	4				40
Total	12	76	56	14	2			160
Stems	1a	2a						Total
Control	36	4						40
TSA1	38	2						40
TSA7	37	3						40
TSA14	36	4						40
Total	147	13						160
Colour index	357	364	370	371	376	377	382	Total
Control	1	11	13	6	1	8		40
TSA1		2		5		33		40
TSA7		3	4	4	1	27	1	40
TSA14		9	20	4		7		40
Total	1	25	37	19	2	75	1	160

Table 2. Phenotypical distribution of the 160 acclimated plants as a function of the TSA treatments *in vitro*. Each plant was individually scored for their leaf pairs (lp) number, roots (r) number, stems (a) number and colour index. Treatments consisted in growing *in vitro* embryogenic clusters in contact with 1 μM TSA for 0 (control), 1, 7 and 14 days.

A difference in embryo development timing was also noted between the Erlenmeyer flask and 24-well plate culture systems, which could be explained by obviously different physico-chemical environments. While rotary shaking was set to the same speed for both systems (120 rpm), it showed different effects, mainly with fast spatial movement of the nutritive medium in flasks, compared to slow movement in wells, due to their smaller diameter. Such lower shaking is known to affect medium O_2/CO_2 uptake^{36,37} and probably explained the differences between the 2 culture systems.

Pilot screening was carried out as proof-of-concept to show that the miniaturized automated culture system we developed was efficient for testing the effect of putatively active compounds on SE, and was reliable and reproducible. Having uniformity of cell cluster distribution is essential before initiating regeneration, to avoid false positives and negatives when screening for active compounds. The use of an automated scanning microscope enabled a considerable gain in speed and quality in terms of image acquisition and analysis. Proof-of-concept was achieved by proving the stimulation of coffee SE by TSA. A 1 μM TSA supply promoted the proliferation of embryogenic cells and led to a 3-fold increase in embryo yield. When re-tested under standard conditions, treatment with 1 μM TSA confirmed its efficiency, hence validating the whole screening system pipeline. Other authors also showed positive effects of TSA on embryogenic structure formation. Li *et al.*³⁸ showed that 0.5 μM TSA promoted totipotency in male gametophytes of *B. napus* and *A. thaliana*. Male gametophytes, which are rather recalcitrant in haploid embryo development in cultures, formed embryogenic cell clusters after TSA treatment. In agreement with these authors, we hypothesize that the 0.1 μM TSA treatment used in our study was at suboptimal concentration to have any observable effect on SE expression while a 10 μM TSA treatment was toxic to cell clusters preventing embryogenesis to occur. Li *et al.*³⁸ associated the TSA mode of action with hyper-acetylation of histones H3 and H4 to block histone deacetylase (HDAC) activity. Wójcikowska *et al.*³⁹ recently showed that treatment of *A. thaliana* explants with TSA induces SE without the exogenous application of auxin and, thus, the TSA mode of action is related to the activation of auxin-responsive transcription factor genes. In conifers, TSA caused the de-repression of *LEAFY COTYLEDON* genes and stimulated the initiation of SE from cotyledonary embryos in *Picea abies*⁴⁰ and in *Pinus Sylvestris*⁴¹, but inhibited embryo germination. For *C. arabica*, 1 μM TSA did not have any significant deleterious effect on plants regenerated from treated cell clusters, and thus can now be considered as an active compound for optimizing SE protocols. As increasing the time of exposure to 1 μM TSA did not differ in embryo yield, we put forward the hypothesis that the mechanism by which TSA triggers its response, i.e. blocking histone deacetylase (HDAC) activity, had already occurred on day 1.

This screening system proof-of-concept is a key step towards high-throughput active compound screening platforms. The next step will be to scale-up this technology by screening hundreds of libraries on robotic platforms. Such platforms are widely used in drug discovery programmes^{18,36}. They combine robotics, data processing

and control software, liquid handling devices, and sensitive detectors to quickly conduct thousands of tests⁴², allowing high-throughput screening (HTS) for active compounds able to modulate a particular molecular pathway. Current generations of platforms are so robust and user-friendly⁴³ that HTS can be used to investigate new areas in biology, such as micropropagation. For coffee SE, achieving the intended change of scale, i.e. HTS of hundreds of libraries to optimize SE protocols, still requires three important upgrades to our screening system: (1) new equipment, e.g. high-throughput pipetting platforms; the platforms described above are coupled with a robotic arm and enable full automation of the screening process; (2) new laboratory strategies, e.g. preparation and maintenance of large amounts of embryogenic tissue; (3) new data acquisition strategies; as embryo development occurs in all 3 dimensions, embryo quantification is still difficult to analyse precisely with an automated 2-dimensional scanning microscope. To that end, early SE-markers would be very helpful.

Further high-throughput experiments will require a statistical methodology called “design of experiments”, which helps experimenters to predict how input variables (e.g. active compounds and nutritive compounds) interact to create output responses (e.g. embryo yield) in a process or system (e.g. regeneration process)⁴⁴. Full factorial and fractional factorial experiments are among the most useful multi-factor experiments for scientific investigations^{19,45}, in contradiction with a one factor-at-a-time approach used until now to optimize culture media in micropropagation, which requires many test runs and fails to consider any possible interaction between factors.

Thanks to masses of information obtained from “-omics” technologies making it possible to describe, in a global way, the pathways involved in the success of the key SE steps⁶, high-throughput screening can already be piloted in a rational way by selecting libraries able to alter target-molecules playing key roles within these pathways. In many cultivated species, including coffee, sequences of genomes, transcriptomes, proteomes and metabolomes are now available^{46–49}. This strategy combining these 2 approaches – “-omics” technologies and robotic screening of active compounds – is currently being implemented by our teams on both cultivated species, *C. arabica* and *C. canephora*, in order to speed up the optimization of SE cloning processes and the dissemination of improved varieties.

Our focus was turned to the SE differentiation phase, where research tends to maximize embryo yields, but other key phases of the coffee SE process could also be easily optimized as a result of the “automated screening of active compound” method, such as the explant-to-callus phase, or the proliferation phase.

These results clearly demonstrate the feasibility, in plants, of automated screening for active compounds in miniaturized SE processes. This proof-of-concept represents a key step towards high-throughput optimization of cell culture media in plant micropropagation. This new method is particularly adapted to SE processes carried out in liquid media and, more widely, to all cell culture-based propagation processes.

Material and Methods

Plant material and experimental design to set up a miniaturized and automated screening platform.

Embryogenic cell clusters of *C. arabica* cv. GPFA116 were initiated from leaf explants obtained from 1-year-old plants obtained by SE and grown in the greenhouse of the Nestlé Plant Science Research Unit (Tours, France). The embryogenic cultures were cultivated following the protocols described previously for *C. arabica*⁵⁰. Briefly, the embryogenic cultures proliferated in 250-ml Erlenmeyer flasks containing M ‘proliferation’ liquid nutritive medium, i.e. a medium supplemented with 0.3 mg/l 2,4-D (2,4-dichlorophenoxyacetic acid) and 1 mg/l BA (6-benzylaminopurine) (Fig. 7a). To stimulate differentiation of early somatic embryos, cultures are “classically” transferred to 250-ml Erlenmeyer flasks containing DIF ‘embryo regeneration’ liquid medium lacking the auxin 2,4-D at a rate of 0.1 g of embryogenic cell clusters per 100 ml of medium for 9 weeks. For miniaturization purposes, 24-well plates (ref. 734–2325, VWR International, Radnor – USA) were chosen as containers for optimum early embryo development. One ml of DIF medium was distributed per well. For automation purposes, the Viaflo Assist automated pipetting platform (Integra Bioscience Corp., Hudson – USA) equipped with a 6-channel pipette (Voyager, Integra Bioscience) was used. Disposable and sterile tips (Cat#4445, 1250 µl), as well as reservoirs (Cat#4322, 100 ml), were also purchased from Integra Bioscience. One g of Arabica cell clusters was first ground with a commercial blender (model: HGB2WT from Waring, Durrington – USA) with a sterilized MC-3 bowl (Waring Products, Durrington – USA) in 100 ml of differentiation medium for 30 sec at 18,000 rpm before filtration on a 700 µm diameter filter, in order to have particle diameters fitting the pipetting tips (Fig. 7b,c). The cell clusters were then inoculated in 24-well plates at an estimated rate of 1 mg per ml per well (Fig. 7d). The 6-channel pipette was programmed in order to have uniform distribution of cell clusters between plates, between rows, and between peripheral and central wells. This programme included a mixing step before pipetting, in order to avoid particle sedimentation. The plates were sealed with Parafilm (M from Bemis, Neenah – USA) and placed in the dark in an incubation shaker (CERTOMAT CT plus, Sartorius Stedim Biotech, Goettingen – Germany) set to 90% RH, 25 °C and 120 rpm for 9 weeks up to the regeneration of torpedo-shaped embryos (Fig. 7e). Uniformity of cluster distribution was assessed at the beginning of culturing, and embryo yield after 9 weeks. This experiment was carried out in 4 replicates with n = 24 wells/replicate.

At the same time, ground and intact cell clusters were cultured in DIF medium under standard *in vitro* conditions (n = 6 flasks/condition/replicate). Erlenmeyer flasks were placed on shakers (120 rpm) at 25 °C in the dark for 9 weeks. The effect of grinding on regeneration capacity was evaluated according to the number of torpedo-shaped embryos obtained per gram of inoculated callus. Total fresh weight was also reported. The experiment was carried out in triplicate.

Viability test on cell clusters before and after grinding. A viability test was carried out before and after grinding cell clusters. This test consisted of double staining with fluorescein diacetate (FDA) and propidium iodide (PI). In living cells, FDA is degraded by cytosolic esterases allowing it to fluoresce in green (emission at 517 nm) in living cells and PI fluoresces in red (emission at 636 nm) in dead cells. A staining solution was

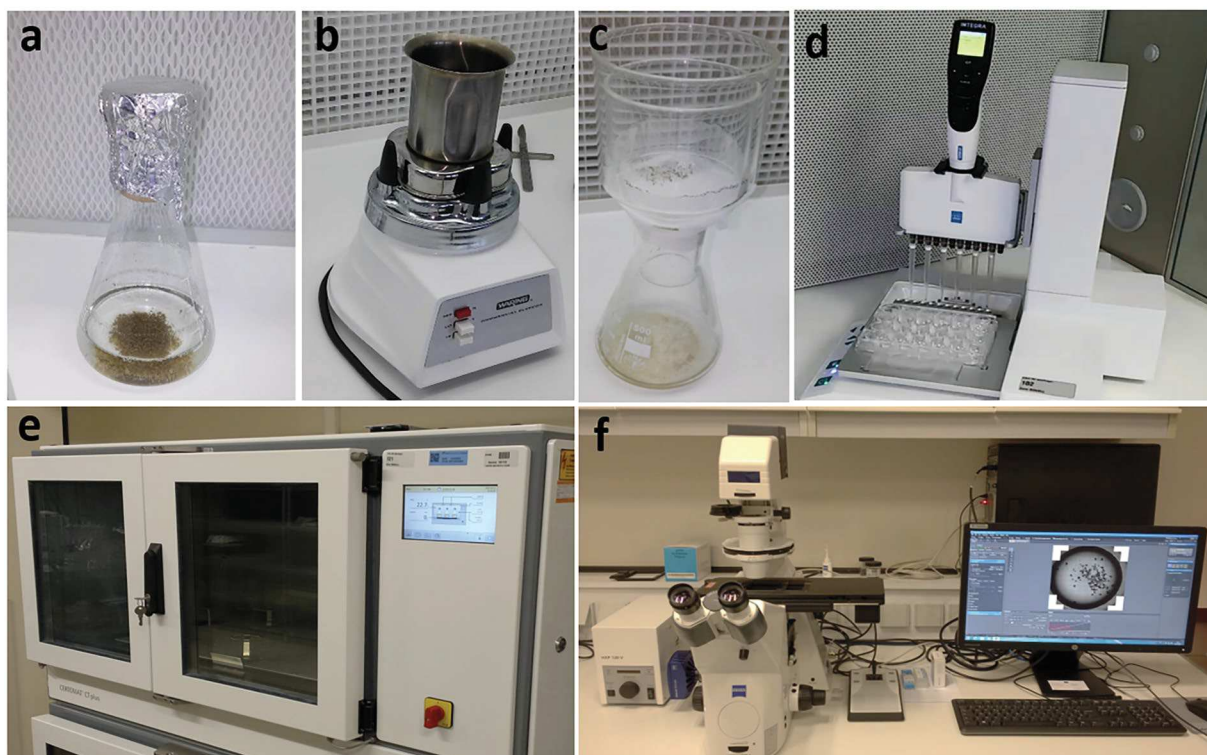


Figure 7. The pipeline established to screen for active compounds affecting coffee SE by miniaturizing and automating the embryo regeneration step. **(a)** Embryogenic cell clusters in M 'proliferation' medium. Cell clusters had a heterogeneous diameter range (50–1000 μm). **(b)** Grinding of cell clusters for 30 sec at 18,000 rpm in DIF 'regeneration' medium. **(c)** Cell cluster filtration on a 700- μm diameter filter to obtain a uniform diameter range (50–250 μm). **(d)** Automated distribution of ground cell clusters in miniaturized culture systems (24-well plates). The same automated platform was used to screen for potential active compounds affecting SE. **(e)** Plates were placed in the dark in incubation shakers (90% RH 25°C and 100 rpm) for 9 weeks up to embryo regeneration. **(f)** An automated inverted plate-reading microscope coupled with a digital camera and computer-assisted image analysis were used to count cell clusters, measure the total area they occupied in a well and monitor embryo development.

prepared with 5 ml of DIF medium, 8 μl of FDA at 5 g/l (F7378, Sigma Aldrich) and 50 μl of IP at 2 g/l (P4170, Sigma Aldrich). Cell clusters were incubated with 1 ml of staining solution for 10 minutes in the dark before washing 3 times with PBS pH 7.2. Clusters were then suspended in DIF medium and placed between 2 microscope slides for observation.

Image acquisition and analysis. Images from 24-well plates were automatically acquired using an inverted microscope equipped with a motorized scanning xy stage (Axio Observer D1, Zeiss, Jena – Germany) and a digital camera (Axiocam 503, Zeiss) (Fig. 7f). The software (Zen, Zeiss) was programmed to find the position of each well using auto-focusing (4 \times objective) and 8 horizontal images by 8 consecutive rows (8 \times 8) were taken in order to register each entire well. The resulting collection of 64 reconstructed images per well was later processed using ImageJ freeware (NIH, Maryland – USA) for cell cluster counting and measurement of their diameter, or the total area they occupied in a well.

For viability tests, images of cell clusters were acquired using the same microscope equipped with dichroic filter cubes for the detection of fluorochromes and connected to a UV light source (HXP 120 v, Zeiss).

Images of torpedo-shaped embryos were acquired using a digital camera (Moticam 5+, Motic, Hong-Kong) mounted on a trinocular stereomicroscope (SZH10, Olympus Optical, Hamburg – Germany) and controlled by Motic Images Plus (v.3.0, Motic) software for length measurements.

Screening of active compounds affecting SE by carrying out HDACi treatments in plates. A full 4 \times 4 factorial design of 18 wells replicated 3 times was used to implement the miniaturized bioassay. Four HDACi, Oxamflatin, SAHA, Scriptaid and Trichostatin A (O3139, SML0061, S7817 and T8552, respectively, from Sigma-Aldrich) were tested. Stock solutions of 10 mM were prepared by dissolving the HDACi powders in DMSO, then immediately sterilized by filtration (0.2 μm , Sartorius AG) and diluted in DIF differentiation medium to obtain final concentrations of 0, 0.1, 1 and 10 μM for each HDACi. Ground and filtered Arabica embryogenic clusters were prepared to inoculate 12 24-well plates at an inoculation density of 1 mg/ml of DIF differentiation medium using the Integra pipetting platform (Integra Bioscience Corp., Hudson – USA). For each HDACi, 3 24-well plates were filled with differentiation medium supplemented with 0 (6 wells of row A), 0.1

(wells of row B), 1 (wells of row C) and 10 μM HDACi (wells of row D) respectively ($n = 18$ wells/HDACi/concentration/replicate). The plates were sealed with Parafilm and placed in an incubation shaker (CERTOMAT CT plus, Sartorius AG) at 90% RH, 120 rpm, 25 °C, in the dark for 24 h. The medium was then gradually replaced with fresh DIF medium by replacing half of the well volume for 6 consecutive times using the pipetting platform (Integra Bioscience Corp.) in order to prevent loss of cell clusters. The plates were placed again in the CERTOMAT for 9 weeks under the same culture conditions. The response to treatment was evaluated according to the number of torpedo-shaped embryos obtained. This experiment was carried out in triplicate.

TSA re-testing assay under standard *in vitro* culture conditions. TSA at an optimum concentration of 1 μM was re-tested under standard conditions. To that end, Arabica embryogenic cell clusters were inoculated in 24 250-ml Erlenmeyer flasks at a rate of 0.1 g per 100 ml of DIF differentiation medium. Four exposure times to 1 μM TSA were tested (0, 1, 7 and 14 days), at the end of which the medium was replaced with fresh DIF differentiation medium ($n = 6$ flasks/exposure time/replicate). Flasks were placed on shakers (120 rpm) at 25 °C in the dark for 9 weeks. The response to treatment was evaluated according to the number of torpedo-shaped embryos obtained per gram of inoculated callus. Total fresh weight and mean length of embryos were also recorded. This experiment was carried out in triplicate.

Plant phenotyping assay post-TSA treatment. Torpedo-shaped embryos from cell clusters treated, or not, with TSA were placed on germination medium in 140 × 21 mm, vented Petri dishes (ref. 391–1500, VWR International) for further development. Dishes were maintained at 25 °C, under a 12/12 photoperiod of artificial light (24 $\mu\text{mol/s}$, Green Power TLED DR/W/MB 18 W, Philips Lighting, Zurich – Switzerland) delivering 65 $\mu\text{mol. m}^{-2}.\text{s}^{-1}$ for 2 to 3 months. After cotyledon development, green embryos were transplanted in 145 × 101 × 60 mm plastic boxes (ref: 017002, D. Dutscher S.A., Issy-Les-Moulineaux – France) containing growth medium for shoot and root development under the same conditions as for the Petri dishes. After 4 months, and for each TSA treatment, 40 SE-derived plantlets ready for greenhouse transfer from 3 boxes were assessed as follows: main root and stem lengths, number of pairs of leaves (lp, from 2 to 7) and roots (r, 0–4), number of stems (a, 1 or 2), growth type (orthotropic or plagiotropic), leaf pigmentation (uniform or variegated), leaf colour index (using a green colour chart, Pantone Carlstadt USA) and leaf shape (normal or abnormal).

Statistical analysis. All analyses were performed in R⁵¹ using the default functions. First, homogeneity between replicates was verified using a Kruskal-Wallis test ($P > 0.05$), then normality of data distribution and homogeneity of variances were assessed. If these conditions were verified, the ANOVA test was used to calculate any significant differences (at $P = 0.05$) between the groups (if groups > 2), or a t-test was used (if groups = 2). If data did not follow a normal distribution, a Kruskal-Wallis test was used to calculate any significant differences (at $P = 0.05$) between the groups, with a Dunn test as post hoc (dunn.test package).

Based on acclimation data, ANOVA were performed on numerical variables, such as main root and stem lengths, to detect any after-effect of the TSA treatment on plant development. For categorical variables, a multi-component analysis (MCA), performed with Minitab v. 18.1, was used to identify correlations between TSA treatments and phenotypes.

Received: 12 May 2019; Accepted: 7 January 2020;

Published online: 21 January 2020

References

1. Guan, Y., Li, S. -G., Fan, X. -F. & Su, Z. -H. Application of Somatic Embryogenesis in Woody Plants. *Front. Plant Sci.* **7** (2016).
2. Harfouche, A. *et al.* Accelerating the domestication of forest trees in a changing world. *Trends Plant Sci.* **17**, 64–72 (2012).
3. Merkle, S. A. & Dean, J. F. Forest tree biotechnology. *Curr. Opin. Biotechnol.* **11**, 298–302 (2000).
4. Lelu-Walter, M. -A. *et al.* Somatic Embryogenesis for More Effective Breeding and Deployment of Improved Varieties in Pinus spp.: Bottlenecks and Recent Advances. in *Somatic Embryogenesis: Fundamental Aspects and Applications* (eds. Loyola-Vargas, V. M. & Ochoa-Alejo, N.) 319–365 (Springer International Publishing, 2016).
5. Lelu-Walter, M. -A. *et al.* Somatic embryogenesis in forestry with a focus on Europe: state-of-the-art, benefits, challenges and future direction. *Tree Genet. Genomes* **9**, 883–899 (2013).
6. Campos, N. A., Panis, B. & Carpentier, S. C. Somatic Embryogenesis in Coffee: The Evolution of Biotechnology and the Integration of Omics Technologies Offer Great Opportunities. *Front. Plant Sci.* **8** (2017).
7. Isah, T. Induction of somatic embryogenesis in woody plants. *Acta Physiol. Plant.* **38**, 118 (2016).
8. Etienne, H. *et al.* Coffee Somatic Embryogenesis: How Did Research, Experience Gained and Innovations Promote the Commercial Propagation of Elite Clones From the Two Cultivated Species? *Front. Plant Sci.* **9** (2018).
9. Guillou, C. *et al.* Indirect somatic embryogenesis of Theobroma cacao L. in liquid medium and improvement of embryo-to-plantlet conversion rate. *Vitro Cell. Dev. Biol. - Plant* **54**, 377–391 (2018).
10. Park, Y. S. *et al.* Initiation of somatic embryogenesis in Pinus banksiana, P. strobus, P. pinaster, and P. sylvestris at three laboratories in Canada and France. *Plant Cell Tissue Organ Cult.* **86**, 87–101 (2006).
11. Bobadilla Landey, R. *et al.* Assessment of genetic and epigenetic changes during cell culture ageing and relations with somaclonal variation in Coffea arabica. *Plant Cell Tissue Organ Cult. PCTOC* **122**, 517–531 (2015).
12. Maximova, S. N. *et al.* Efficiency, genotypic variability, and cellular origin of primary and secondary somatic embryogenesis of Theobroma cacao L. *Vitro Cell. Dev. Biol. - Plant* **38**, 252–259 (2002).
13. Lamberth, C., Jeanmart, S., Luksch, T. & Plant, A. Current Challenges and Trends in the Discovery of Agrochemicals. *Science* **341**, 742–746 (2013).
14. Boutilier, K., Angenent, G. C., Castan, M. S. & Hui, L. Inventors; Stichting Wageningen Research, assignee. Haploid embryogenesis. United States patent US 20160212956A1. 2016 Jul 28th.
15. Corredoira, E., Ballester, A., Ibarra, M. & Vieitez, A. M. Induction of somatic embryogenesis in explants of shoot cultures established from adult Eucalyptus globulus and E. saligna × E. maidenii trees. *Tree Physiol.* **35**, 678–690 (2015).
16. Moon, H. -K., Kim, Y. -W., Hong, Y. -P. & Park, S. -Y. Improvement of somatic embryogenesis and plantlet conversion in Oplonax elatus, an endangered medicinal woody plant. *SpringerPlus* **2** (2013).

17. Maciel, A. L. R. *et al.* Large-scale, high-efficiency production of coffee somatic embryos. *Crop Breed. Appl. Biotechnol.* **16**, 102–107 (2016).
18. Szymański, P., Markowicz, M. & Mikiciuk-Olasik, E. Adaptation of High-Throughput Screening in Drug Discovery—Toxicological Screening Tests. *Int. J. Mol. Sci.* **13**, 427–452 (2011).
19. Radzun, K. A. *et al.* Automated nutrient screening system enables high-throughput optimisation of microalgae production conditions. *Biotechnol. Biofuels* **8**, 65 (2015).
20. Robert, S., Raikhel, N. V. & Hicks, G. R. Powerful Partners: Arabidopsis and Chemical Genomics. *Arab. Book* **7**, e0109 (2009).
21. Hicks, G. R. & Raikhel, N. V. Small Molecules Present Large Opportunities in Plant Biology. *Annu. Rev. Plant Biol.* **63**, 261–282 (2012).
22. Connor, C. J. O., Beckmann, H. S. G. & Spring, D. R. Diversity-oriented synthesis: producing chemical tools for dissecting biology. *Chem. Soc. Rev.* **41**, 4444–4456 (2012).
23. Feng, S., Jacobsen, S. E. & Reik, W. Epigenetic Reprogramming in Plant and Animal Development. *Science* **330**, 622–627 (2010).
24. He, X.-J., Chen, T. & Zhu, J.-K. Regulation and function of DNA methylation in plants and animals. *Cell Res.* **21**, 442–465 (2011).
25. Nic-Can, G. I. *et al.* New Insights into Somatic Embryogenesis: Leafy Cotyledon1, Baby Boom1 And Wuschel-Related Homeobox 4 Are Epigenetically Regulated in *Coffea canephora*. *PLOS ONE* **8**, e72160 (2013).
26. Bertrand, B., Montagnon, C., Georget, F., Charmetant, P. & Etienne, H. Création et diffusion de variétés de caféiers Arabica: quelles innovations variétales? *Cah. Agric.* **21**, 77–88 (2012).
27. Etienne, H. *et al.* Un exemple de transfert de technologie réussi dans le domaine de la micropropagation: la multiplication de *Coffea arabica* par embryogenèse somatique. *Cah. Agric.* **21**, 115–124 (2012).
28. Ducos, J. P., Labbe, G., Lambot, C. & Pétiard, V. Pilot scale process for the production of pre-germinated somatic embryos of selected robusta (*Coffea canephora*) clones. *Vitro Cell. Dev. Biol. - Plant* **43**, 652–659 (2007).
29. Ducos, J. P. *et al.* Agronomic performance of *Coffea canephora* P. trees derived from large-scale somatic embryo production in liquid medium. *Euphytica* **131**, 215–223 (2003).
30. Georget, F. *et al.* Somatic embryogenesis-derived coffee plantlets can be efficiently propagated by horticultural rooted mini-cuttings: A boost for somatic embryogenesis. *Sci. Hortic.* **216**, 177–185 (2017).
31. Hughes, J., Rees, S., Kalindjian, S. & Philpott, K. Principles of early drug discovery. *Br. J. Pharmacol.* **162**, 1239–1249 (2011).
32. Mayr, L. M. & Fuerst, P. The future of high-throughput screening. *J. Biomol. Screen.* **13**, 443–448 (2008).
33. Rodriguez-Furlán, C., Miranda, G., Reggiardo, M., Hicks, G. R. & Norambuena, L. High throughput selection of novel plant growth regulators: Assessing the translatability of small bioactive molecules from Arabidopsis to crops. *Plant Sci.* **245**, 50–60 (2016).
34. Gao, G., Chen, L. & Huang, C. Anti-cancer Drug Discovery: Update and Comparisons in Yeast, Drosophila, and Zebrafish. *Curr. Mol. Pharmacol.* **7**, 44–51 (2014).
35. Halder, V. & Kombrink, E. Facile high-throughput forward chemical genetic screening by *in situ* monitoring of glucuronidase-based reporter gene expression in Arabidopsis thaliana. *Front. Plant Sci.* **6**, 13 (2015).
36. Zhang, H., Lamping, S. R., Pickering, S. C. R., Lye, G. J. & Shamlou, P. A. Engineering characterisation of a single well from 24-well and 96-well microtitre plates. *Biochem. Eng. J.* **40**, 138–149 (2008).
37. Etienne, H., Dechamp, E., Barry-Etienne, D. & Bertrand, B. Bioreactors in coffee micropropagation. *Braz. J. Plant Physiol.* **18**, 45–54 (2006).
38. Li, H. *et al.* The Histone Deacetylase Inhibitor Trichostatin A Promotes Totipotency in the Male Gametophyte[W]. *Plant Cell* **26**, 195–209 (2014).
39. Wójcikowska, B. *et al.* Trichostatin A Triggers an Embryogenic Transition in Arabidopsis Explants via an Auxin-Related Pathway. *Front. Plant Sci.* **9** (2018).
40. Uddenberg, D. *et al.* Embryogenic potential and expression of embryogenesis-related genes in conifers are affected by treatment with a histone deacetylase inhibitor. *BMC Proc.* **5**, P151 (2011).
41. Abrahamsson, M., Valladares, S., Merino, I., Larsson, E. & von Arnold, S. Degeneration pattern in somatic embryos of *Pinus sylvestris* L. *Vitro Cell. Dev. Biol. - Plant* **53**, 86–96 (2017).
42. Kong, F., Yuan, L., Zheng, Y. F. & Chen, W. Automatic Liquid Handling for Life Science: A Critical Review of the Current State of the Art. *J. Lab. Autom.* **17**, 169–185 (2012).
43. Michael, S. *et al.* A Robotic Platform for Quantitative High-Throughput Screening. *Assay Drug Dev. Technol.* **6**, 637–657 (2008).
44. Ilzarbe, L., Álvarez, M. J., Viles, E. & Tanco, M. Practical applications of design of experiments in the field of engineering: a bibliographical review. *Qual. Reliab. Eng. Int.* **24**, 417–428 (2008).
45. Altekar, M. *et al.* Assay Optimization: A Statistical Design of Experiments Approach. *Clin. Lab. Med.* **27**, 139–154 (2007).
46. *Coffea arabica* genome. Available at, <https://worldcoffeeresearch.org/work/coffee-arabica-genome/> (Accessed: 6th March 2019).
47. Ivamoto, S. T. *et al.* Transcriptome Analysis of Leaves, Flowers and Fruits Perisperm of *Coffea arabica* L. Reveals the Differential Expression of Genes Involved in Raffinose Biosynthesis. *PLoS ONE* **12** (2017).
48. Campos, N. A., Paiva, L. V., Panis, B. & Carpentier, S. C. The proteome profile of embryogenic cell suspensions of *Coffea arabica* L. *Proteomics* **16**, 1001–1005 (2016).
49. Businge, E., Brackmann, K., Moritz, T. & Egertsdotter, U. Metabolite profiling reveals clear metabolic changes during somatic embryo development of Norway spruce (*Picea abies*). *Tree Physiol.* **32**, 232–244 (2012).
50. Etienne, H. Somatic Embryogenesis Protocol: Coffee (*Coffea arabica* L. and *C. canephora* P.). in *Protocols for Somatic Embryogenesis in Woody Plants* (eds Jain, S. M. & Gupta, P. K.) (Springer International Publishing, 2005).
51. R-Core-Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing (2018).

Acknowledgements

Rayan Awada benefits from a PhD grant and funding from Nestlé and CIRAD.

Author contributions

D.B. and S.F.M. designed the research. R.A., D.V., S.F. and E.B. performed the experiments. R.A. and D.B. analysed the data. R.A., D.B. and H.E. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-57800-6>.

Correspondence and requests for materials should be addressed to R.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

Chapter III

Unravelling the Metabolic and Hormonal Machinery During Key Steps of Somatic Embryogenesis: A Case Study in Coffee

Published paper in *International Journal of Molecular Sciences* on the 20th of September, 2019.

Background

In chapter II, we brought a proof of concept of an automated system able to reproduce the embryo regeneration step in miniaturized conditions and to screen for active compounds able to modulate SE in high-throughput standards. This proof of concept carried out on a pilot scale is a big step towards an industrial high-throughput approach for research on SE which was carried out until now on a low-throughput level. However, in the absence of identified molecular targets i.e. a list of specific biochemical compounds playing a key role in the efficiency of the SE process, research will remain empirical based on a trial-and-error approach. New analytic tools and knowledge about the molecular events involved in the SE key steps, are urgently needed to pilot the optimization of the protocols. In this chapter, we will show how we took advantage of the latest metabolomics technologies and applied them to one of the most reliable large-scale SE process, the one developed for coffee. To obtain solid and reliable information and massive data, five independent leaf introductions were carried out in February, April, June, October and December 2016, more than 4,000 leaf explants and a total of 25 independent cell lines. All obtained cell lines were high-yielding and time-synchronized during embryo regeneration enabling a successful sampling.

The objective of this study is to unravel the metabolic profiles and hormone dynamics during the whole SE process (14 sampled stages). A collaboration with the Max Planck Institute has been established specially to comply with this objective that allows us an access to one of the most up-to-date metabolite database, the Golm Metabolome Database. The data analysis will require the development of a specific pipeline for in-depth statistical analysis. We expect from this analysis to identify i) the developmental phases of coffee SE, ii) the key developmental switches that represent the transitions between the above development phases, iii) metabolite fingerprints of each phase and if possible of each developmental stage, iv) a list of key metabolites ('markers') associated with cell fate and/or with each developmental phase, or associated with the crucial embryogenic capacity. We hope that these markers can be used to control the optimization of the composition of culture media.

From a more general point of view, we hope that this metabolomic approach applied to the whole process of embryogenesis in coffee will provide us with a global knowledge of the massive changes involved in this process, but also with an intimate understanding of the fine mechanisms at key stages.



Article

Unravelling the Metabolic and Hormonal Machinery During Key Steps of Somatic Embryogenesis: A Case Study in Coffee

Rayan Awada ^{1,2,3,*} , Claudine Campa ^{3,4}, Estelle Gibault ¹, Eveline Déchamp ^{2,3}, Frédéric Georget ^{2,3}, Maud Lepelley ¹ , Cécile Abdallah ^{3,4}, Alexander Erban ⁵, Federico Martinez-Seidel ⁵ , Joachim Kopka ⁵, Laurent Legendre ⁶, Sophie Lérant ^{2,3}, Geneviève Conéjéro ⁷, Jean-Luc Verdeil ⁷, Dominique Cruzillat ¹, David Breton ¹, Benoît Bertrand ^{2,3} and Hervé Etienne ^{2,3}

¹ Nestlé Research—Plant Science Unit, 101 avenue Gustave Eiffel, F-37097 Tours CEDEX 2, France; estelle.gibault@rdto.nestle.com (E.G.); maud.lepelley@rdto.nestle.com (M.L.); dominique.cruzillat@rdto.nestle.com (D.C.); david.breton@rdto.nestle.com (D.B.)

² CIRAD (Centre de coopération internationale en recherche agronomique pour le développement), UMR IPME, F-34398 Montpellier, France; eveline.dechamp@ird.fr (E.D.); frederic.georget@cirad.fr (F.G.); sophie.leran@cirad.fr (S.L.); benoit.bertrand@cirad.fr (B.B.); herve.etienne@cirad.fr (H.E.)

³ UMR IPME (Interactions Plantes Microorganismes Environnement), University of Montpellier, CIRAD, IRD, F-34398 Montpellier, France; claudine.campa@ird.fr (C.C.); cecile.abdallah@ird.fr (C.A.)

⁴ IRD (Institut de recherche pour le développement), UMR IPME, F-34398 Montpellier, France

⁵ Max Planck Institute for Molecular Plant Physiology, Am Muehlenberg 1, D-14476 Golm, Germany; erban@mpimp-golm.mpg.de (A.E.); mseidel@mpimp-golm.mpg.de (F.M.-S.); kopka@mpimp-golm.mpg.de (J.K.)

⁶ Université de Lyon (Université Lyon 1, CNRS, UMR5557, Ecologie Microbienne, INRA, UMR1418), F-69622 Lyon, France; laurent.legendre@univ-lyon1.fr

⁷ Histo cytology and Plant Cell Imaging platform PHIV, UMR AGAP (CIRAD, INRA, SupAgro)-UMR B&PMP (INRA, CNRS, SupAgro, University of Montpellier), F-34095 Montpellier, France; genevieve.conejero@inra.fr (G.C.); jean-luc.verdeil@cirad.fr (J.-L.V.)

* Correspondence: rayan.awada@cirad.fr

Received: 20 August 2019; Accepted: 18 September 2019; Published: 20 September 2019



Abstract: Somatic embryogenesis (SE) is one of the most promising processes for large-scale dissemination of elite varieties. However, for many plant species, optimizing SE protocols still relies on a trial-and-error approach. Using coffee as a model plant, we report here the first global analysis of metabolome and hormone dynamics aiming to unravel mechanisms regulating cell fate and totipotency. Sampling from leaf explant dedifferentiation until embryo development covered 15 key stages. An in-depth statistical analysis performed on 104 metabolites revealed that massive re-configuration of metabolic pathways induced SE. During initial dedifferentiation, a sharp decrease in phenolic compounds and caffeine levels was also observed while auxins, cytokinins and ethylene levels were at their highest. Totipotency reached its highest expression during the callus stages when a shut-off in hormonal and metabolic pathways related to sugar and energetic substance hydrolysis was evidenced. Abscisic acid, leucine, maltotriose, myo-inositol, proline, tricarboxylic acid cycle metabolites and zeatin appeared as key metabolic markers of the embryogenic capacity. Combining metabolomics with multiphoton microscopy led to the identification of chlorogenic acids as markers of embryo redifferentiation. The present analysis shows that metabolite fingerprints are signatures of cell fate and represent a starting point for optimizing SE protocols in a rational way.

Keywords: cell imaging; cell fate; coffee; histology; hormone content; metabolomics; somatic embryogenesis; totipotency

1. Introduction

Somatic embryogenesis (SE) is a developmental process where a plant somatic cell can dedifferentiate to a totipotent embryogenic stem cell that has the ability to redifferentiate and give rise to an embryo under appropriate conditions [1–3]. This new embryo can further develop into a whole plant. Since its first description in carrot by Reinert [4] and Steward et al. [5], this process has been reported in a wide range of plant species, both annual [6–8] and perennial [9–13]. SE has demonstrated significant benefits when applied to forest tree species, bringing important advantages including clonal mass propagation, cryostorage of valuable germplasm and genetic transformation [14,15]. SE is particularly important for plants that have a long life cycle (woody species) and are difficult to propagate by conventional methods [16].

The ability of a somatic cell to undergo embryogenesis *in vitro* is both an inherent and an acquired characteristic that requires the right combination of explant and culture environment [17]. The most efficient treatments used to induce embryogenesis are diverse and range from application of exogenous growth regulators to abiotic stress. Under the appropriate conditions, the explant produces histo-differentiated embryos, either directly from the explant or indirectly from callus [18]. The direct SE is often described as a low yield method and the indirect SE as a high yield method [10]. The morphological and cellular changes that occur during *in vitro* embryogenesis have been well-described [1,19,20]. Briefly, nine developmental stages have been characterized in the indirect SE of dicots: explant, primary callus, embryogenic callus, embryogenic cell clusters, pro-embryogenic masses, globular-shaped embryos, heart-shaped embryos, torpedo-shaped embryos and cotyledonary embryos, before development into a new plant [1,19,20]. Unlike the detailed knowledge about morphological and histological events, little is known about the underlying mechanisms involved in the acquisition and expression of totipotency occurring during SE.

Although SE has already been widely described in a number of woody species [10–13] propagating adult woody plants remains a difficult, laborious, intensive and tricky operation. Many authors reported that the lack of knowledge on the mechanisms underlying the reprogramming of somatic cells represents the greatest limitation to the improvement of the SE process [21,22]. Research on SE remains mainly empirical, characterized by a low-throughput trial-and-error approach. A set of drawbacks have been reported especially a strong genotypic effect, a difficulty in obtaining embryogenic callus, a variability in the quality of regenerated embryos, and more generally a lack of efficiency of certain steps [9,23,24] leading to hitherto prohibitive production costs and an overall slow technical progress.

In parallel, the latest omics technologies (genomics, transcriptomics, proteomics, epigenomics, metabolomics and phenomics) are gaining increasing attention. Many authors believe that combining omics technologies with the *in vitro* process can have a tremendous impact on the knowledge of the molecular mechanisms underlying SE [22,25,26]. Metabolomics data can provide a wealth of information in the description and elucidation of physiological responses to environmental conditions in plants [27]. However, to date, very few authors have carried out a metabolomics approach related to SE. Using gas chromatography–mass spectrometry (GC-MS), metabolomic profiling has been successfully applied to identify metabolic processes regulating SE in five sampled stages in Norway spruce (*Picea abies*) covering cell line proliferation till embryo maturation [28]. Similarly, Robinson et al. [29] applied a GC-MS approach in loblolly pine (*Pinus taeda*) as a way to predict a model for embryo regeneration capacity. More recently, Dobrowolska et al. [30] applied GC-MS to compare metabolic profiles of normal and aberrant SE germinants in Norway spruce. In the same species, using liquid chromatography–mass spectrometry (LC-MS), Vondrakova et al. [31] drew endogenous phytohormone profiles in nine sampled stages covering proliferation, maturation, desiccation and germination.

Using the same technique, Gautier et al. [32] compared phytohormone profiles between embryogenic calli and non-embryogenic calli in Douglas-fir (*Pseudotsuga menziesii*). All these authors assumed that clear correlations may exist among the different metabolome or phytohormone profiles and some specific SE stages. Nonetheless, to date, no global analysis, covering leaf explant dedifferentiation till embryo development, of both metabolome and hormone content has been performed on a complete SE process.

Coffee is one of the world's favorite beverages. It has a great economic impact in many producing countries, especially in South America [33]. Today, SE applied to coffee is one of the most advanced technologies in plant mass vegetative propagation [24]. Thirty years of research on SE by two leading groups—the CIRAD/ECOM alliance and Nestlé—has led to a successful large-scale dissemination of *Coffea arabica* F1 hybrids and *C. canephora* cv. Robusta clones for 13 years now. Even though good biological efficiencies characterize the processes established in the two cultivated species, with low genotypic effects [34,35] and controlled somaclonal variation [36,37], SE production cost remains high and still cannot compete with production costs of seedlings or traditional cuttings [38].

Indeed, as for other species, coffee SE research remains mainly empirical, also characterized by overall slow technical progress. For example, the development of culture conditions for the mass redifferentiation into somatic embryos have been very long (10 years) and laborious, carried out only by an empirical approach [24]. The lack of knowledge about the cellular and molecular events taking place during these developmental stages makes these stages real black boxes. While a scale-up is needed to meet increasing market demand—estimated at 50–100 million coffee vitroplants per—current overall production cannot meet this huge demand [24]. However, optimization is possible for different stages if a rational optimization is developed on metabolic knowledge.

In this paper, we describe how we took advantage of the latest omics technologies combined with a reliable, synchronized and efficient large-scale coffee SE process to draw full metabolic profiles and study hormone dynamics over 15 key sampled stages. A solid statistical method was used to identify the metabolic pathway changes associated with the main developmental phases and phase switches. Histological analysis and cell imaging were also required to characterize sampled stages and associate metabolic profiles with cell structure organization. Lastly, comparing Arabica embryogenic and non-embryogenic calli enabled the identification of metabolic markers of the embryogenic capacity.

2. Results

2.1. Morphological and Histological Characterization of Arabica SE Key Sampled Stages

A characterization of the 15 key sampled stages of the Arabica SE process on a morphological level (Figure 1), as well as on a histocytological level is shown in Figure 2 and Table 1. SE was carried out from juvenile leaf explants (Figure 1L2) which could be characterized histologically by a well-structured anatomy i.e., from up to bottom, an upper-epidermis, palisade mesophyll cells, spongy mesophyll cells and a lower-epidermis (Figure 2A L1). Palisade mesophyll cells are elongated and cuboid (width-to-length ratio = 0.33 ± 0.04) (Figure 2B L1, Table 1) with a low nucleocytoplasmic (N/C) ratio (0.17 ± 0.02) and a small nucleus located at the periphery of the cell, pressed against the cell wall (Figure 2B L1). Leaf disinfection (Figure 2B L2) did not have a significant effect on cell morphology. After one week on culture medium, and even though no changes were noticed at a morphological level (Figure 1D1), a high mitotic activity was taking place in cells enduring dedifferentiation and evidenced by the number of cells undergoing cytokinesis (Figure 2B D1).

This mitotic activity was particularly observed in perivascular and spongy mesophyll cells of the leaf explant. Cell nuclei appeared intensely stained in blue by Naphthol Blue-Black and took a more or less central position, where they were maintained by spans of cytoplasm between which are found several vacuoles of small sizes. Cells were therefore described as star-shaped. These vacuolated cells had a large size ($647 \pm 71 \mu\text{m}^2$) and a high width-to-length ratio (0.67 ± 0.08) (Table 1). After two weeks on culture medium, whitish amorphous callus started to emerge from the wounded edge of leaf explants (Figure 1D2). On a histological level, clusters of actively proliferating meristematic cells appeared to be forming the emerging callus (Figure 2A D2). These cells were remarkably stained with Naphthol Blue Black. Dedifferentiated cells of the constituted callus had an irregular shape, but maintained their large size and high mitotic activity as well as a higher N/C ratio (2.7-fold) (Figure 2B D2, Table 1). After five weeks on culture medium, two types of calli could be observed: a whitish spongy callus which is non-embryogenic and a more compact and organized yellowish callus that emerged on the edge of leaf explants (Figure 1D3). The cells from the yellowish callus were also actively proliferating (numerous mitosis) with a larger size ($695 \pm 91 \mu\text{m}^2$) (Figure 2B D3, Table 1). Three months after the *in vitro* introduction, this callus was well-established and had a dark grey color, and named primary callus (Figures 1C1 and 2A C1). Primary callus cells were square-shaped, had also a large size ($784 \pm 56 \mu\text{m}^2$), a large nucleus ($114 \pm 25.3 \mu\text{m}^2$), and their vacuole occupied the main cell volume (Figure 2B C1, Table 1). A polysaccharide sugar-containing mucilaginous coating layer was present between cell walls (Figure 2B C1). Less cell divisions were observed in this type of slow proliferating callus. Seven months after explant introduction, embryogenic calli emerged (Figures 1C2 and 2A C2). Embryogenic calli were yellowish and friable, constituted by small cells ($214 \pm 18 \mu\text{m}^2$), isodiametric, surrounded by a thickened outer cell wall, arranged in clusters, with a dense cytoplasm (rich protein-staining) and a central nucleus with a prominent nucleolus stained in blue-black (N/C ratio = 0.33 ± 0.02) (Figure 2B C2, Table 1). The nucleus is surrounded by starch granules. These cells display the characteristic cytological features generally found in embryogenic calli [39]. A morphological and histological comparison between embryogenic and non-embryogenic calli (NEC) is described below. Embryogenic cell characteristics were maintained in established embryogenic cell clusters (Figure 2B C3) with a sharp increase in the number of proliferating cells. Indeed, cell clusters were maintained for four months in proliferation medium and became brownish in color (Figure 1C3). All pro-embryogenic masses (PEMs) collected during redifferentiation (Figure 2B R1, R2, R3, R4) shared the same characteristics with embryogenic cell clusters on histological and cellular levels with a higher ratio of square-shaped cells (width-to-length ratio = 0.70 ± 0.07 vs. 0.54 ± 0.06 , Table 1). Easily observable, cell cluster dark brown color quickly turned to light brown as soon as the redifferentiation step was induced (Figure 1C3,R1). A sugar-containing mucilaginous coating layer was present between cell walls (Figure 2B R1). The formation of globular embryo structures could be seen from the R4 stage in which a sugar mucilaginous coating layer was backing the newly formed embryoids (Figure 2A R4). While PEMs were characterized by a high N/C ratio, embryonic cells of globular-shaped embryos (Figure 2A E1) were characterized by a small nucleus hence a low N/C ratio (0.10 ± 0.01 vs. 0.31 ± 0.03 , Table 1) and a dense nucleus at the edge of the cell with the vacuole occupying the main cell volume (Figure 2B E1). This aspect was more highlighted in embryonic cells of the ground promeristem of torpedo-shaped embryos (Figure 2A E2) where cells were circular and of larger size ($943 \pm 55 \mu\text{m}^2$) (Figure 2B E2). Globular and torpedo-shaped embryos were white in color (Figure 1E1,E2). Globular embryos were surrounded by a protodermis (outer layer) (Figure 2A E1). Torpedo-shaped embryos were bipolar structures with easily distinguishable shoot and root poles (Figure 2A E2).

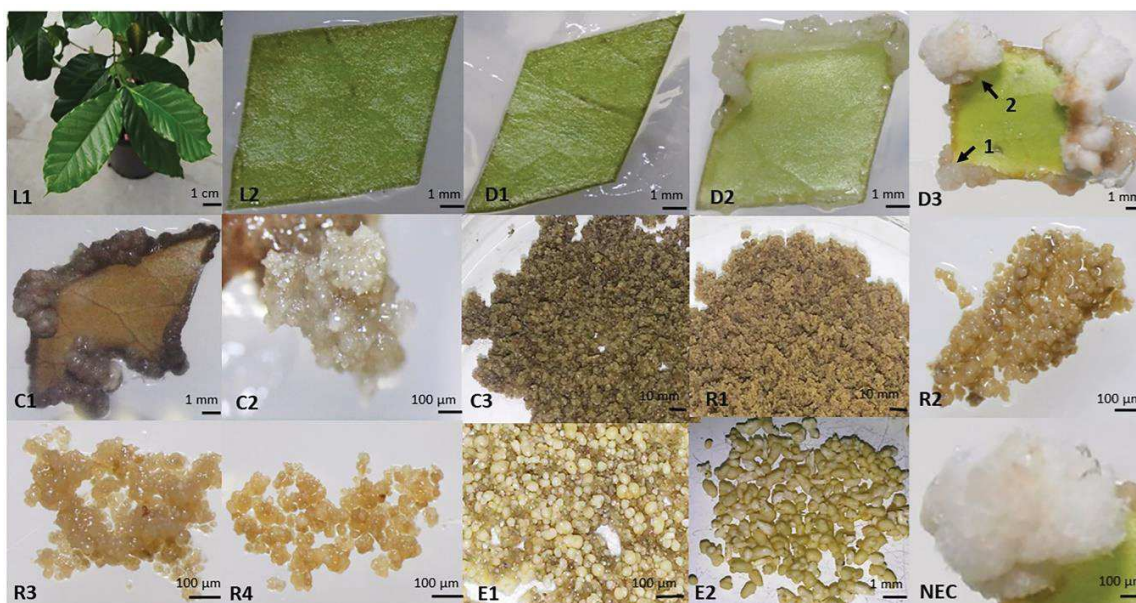


Figure 1. Collection of samples from leaf explant till embryo development covering 14 key sampled stages of the Arabica SE process: leaves from greenhouse plants (L1), explants during dedifferentiation [0 h (L2), 1 week (D1), 2 weeks (D2), 5 weeks (D3)], compact primary callus obtained 3 months after induction (C1), embryogenic callus obtained 7 months after induction (C2), established cell clusters obtained after 4 months in proliferation medium (C3), early redifferentiation from cell clusters [1 week in DIF medium without reducing cell density (R1), 24 h in DIF medium after reducing cell density (R2), 72 h (R3), 10 d (R4)], globular embryos (E1) and torpedo-shaped embryos (E2). Non-embryogenic callus (NEC) was also sampled at the same time as the embryogenic callus (C2). Images were taken using an Olympus E-5 digital camera mounted on an Olympus SZX7 stereomicroscope. The stereomicroscope was not used for (L1). Two types of calli can be seen in (D3): an organized yellowish callus (1) and a whitish spongy callus (2).

2.2. Clustering Metabolite Profiles Divides the Arabica SE Process into Five Major Phases

A total of 92 primary metabolites and 12 secondary metabolites were detected over the 14 SE sampled stages (Table S1). The heatmap generated from the Z-score values of the total primary and secondary metabolites detected showed similar patterns in some of these stages (Figure 3A). We decided to carry out a hierarchical clustering analysis to highlight stages that shared similar metabolic patterns. Strong correlations between patterns were obtained as the threshold applied (red line, Figure 3B) yielded 5 major nodes: the “Leaf” node where the L1 and L2 stages clustered together (cluster probability = 100%), the “Dedifferentiation” node where the D1, D2 and D3 stages clustered together (cluster probability = 67%), the “Callus” node where the C1, C2 and C3 stages clustered together (cluster probability = 57%), the “Redifferentiation” node where the R1, R2, R3 and R4 stages clustered together (cluster probability = 69%) and the “Embryo” node where the E1 and E2 stages clustered together (cluster probability = 87%). The obtained clusters grouped along the axis of the SE process as the obtained nodes corresponded to successive developmental phases of the SE process, i.e., the leaf phase, the leaf explant dedifferentiation phase, the callus phase, the redifferentiation phase (embryogenic cell clusters to embryoid structures), and the embryo development phase (globular to torpedo-shaped embryos).

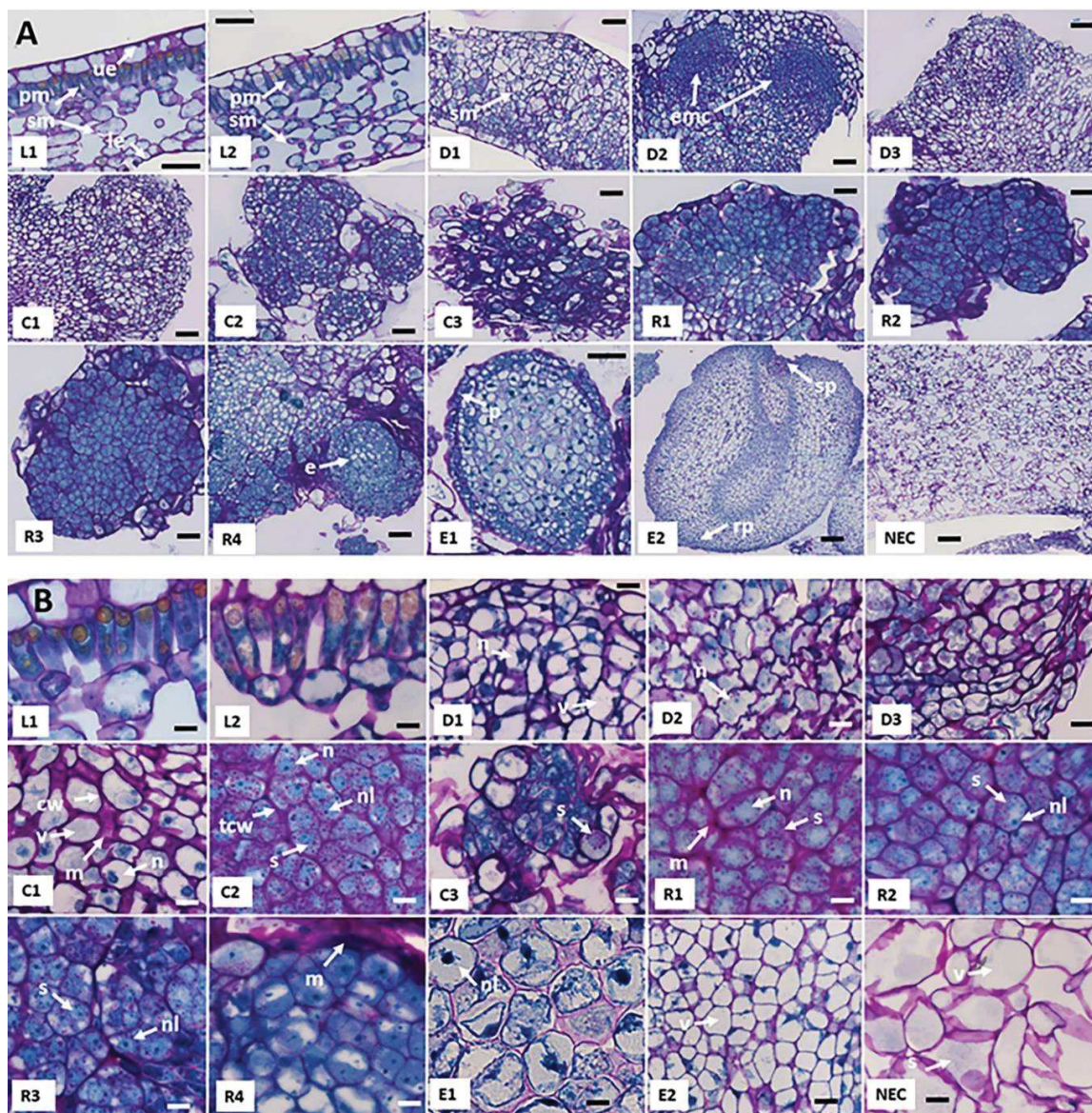


Figure 2. Characterization of 15 key sampled stages throughout the Arabidopsis SE process at a histological (A) and cellular level (B). Sampled stages correspond to: leaves from greenhouse plants (L1), explants during dedifferentiation [0 h (L2), 1 week (D1), 2 weeks (D2), 5 weeks (D3)], compact primary callus obtained 3 months after induction (C1), embryogenic callus obtained 7 months after induction (C2), established cell clusters obtained after 4 months in liquid proliferation medium (C3), pro-embryogenic masses [1 week in redifferentiation medium after auxin withdrawal (R1), 24 h in redifferentiation medium after reducing cell density (R2), 72 h (R3), 10 days (R4)], globular embryos (E1) and torpedo-shaped embryos (E2). An additional stage, the non-embryogenic callus (NEC), was also sampled at the same time as the embryogenic callus (C2) and obtained in the same culture conditions. Sections were treated with 1% periodic acid and stained with Schiff reagent (colors polysaccharides in purple) and Naphthol Blue Black (colors soluble proteins in blue). Images were taken in bright field. The scale bar is set to 50 μm in (A) and 10 μm in (B). White arrows indicate: cw: cell wall, e: embryoid structure, emc: emerging callus, le: lower epidermis, m: mucilaginous coating, n: nucleus, nl: nucleolus, p: protodermis, pm: palisade mesophyll cells, rp: root pole, s: starch, sm: spongy mesophyll cells, sp: shoot pole, tcw: thickened outer cell wall, ue: upper-epidermis, v: vacuole.

Table 1. Characteristics of the main represented cell types in each of the 15 key sampled stages of the Arabica SE process.

Cell Type	Stage	Cell Length (µm)	Cell Width (µm)	W/L Ratio	Cytoplasm Size (µm ²)	Nucleus Size (µm ²)	N/C Ratio	Cell Division	Starch	Proteins
Palisade mesophyll cells	L1	41.5 ± 1.7 ^a	13.9 ± 2.5 ^a	0.33 ± 0.04 ^a	430 ± 93 ^a	66.5 ± 25 ^a	0.17 ± 0.02 ^a	-	+	+
Palisade mesophyll cells	L2	38.2 ± 4.1 ^a	13.0 ± 2.0 ^a	0.35 ± 0.01 ^a	440 ± 99 ^a	59.1 ± 17 ^a	0.19 ± 0.03 ^a	-	+	+
Dedifferentiated cells	D1	33.9 ± 2.5 ^a	22.8 ± 4.0 ^b	0.67 ± 0.08 ^b	647 ± 71 ^b	56.8 ± 7.4 ^a	0.09 ± 0.02 ^a	+++	+	+
Dedifferentiated cells	D2	32.7 ± 1.8 ^a	22.4 ± 0.7 ^b	0.69 ± 0.02 ^b	681 ± 44 ^b	155 ± 18.3 ^b	0.24 ± 0.04 ^b	+++	+	+
Dedifferentiated cells	D3	35.8 ± 2.6 ^a	27.2 ± 5.1 ^b	0.76 ± 0.11 ^b	695 ± 91 ^b	55.4 ± 11.3 ^a	0.08 ± 0.03 ^a	+++	++	+
Primary callus cells	C1	35.9 ± 2.1 ^a	30.6 ± 2.6 ^b	0.85 ± 0.02 ^b	784 ± 56 ^b	114 ± 25.3 ^b	0.19 ± 0.03 ^a	+	++	+
Embryogenic cells	C2	20.2 ± 1.3 ^b	14.7 ± 0.5 ^a	0.73 ± 0.03 ^b	214 ± 18 ^c	70.7 ± 5.1 ^a	0.33 ± 0.02 ^b	++	+++	+++
Embryogenic cell clusters	C3	23.9 ± 1.0 ^b	12.7 ± 1.1 ^a	0.54 ± 0.06 ^a	223 ± 15 ^c	70.1 ± 10.5 ^a	0.31 ± 0.03 ^b	++++	+++	+++
Pro-embryogenic masses	R1	19.5 ± 2.7 ^b	12.1 ± 0.5 ^a	0.62 ± 0.04 ^b	144 ± 15 ^c	66.3 ± 8.4 ^a	0.45 ± 0.07 ^b	+++	+++	+++
Pro-embryogenic masses	R2	20.2 ± 1.8 ^b	13.9 ± 1.5 ^a	0.69 ± 0.54 ^b	202 ± 23 ^c	65.2 ± 10.8 ^a	0.32 ± 0.02 ^b	+++	+++	+++
Pro-embryogenic masses	R3	18.8 ± 1.3 ^b	12.5 ± 0.3 ^a	0.67 ± 0.03 ^b	165 ± 13 ^c	69.0 ± 15.7 ^a	0.41 ± 0.07 ^b	+++	+++	+++
Pro-embryogenic masses	R4	26.7 ± 1.4 ^b	16.5 ± 1.6 ^a	0.70 ± 0.07 ^b	315 ± 26 ^c	98.7 ± 6.6 ^a	0.31 ± 0.03 ^b	+++	+++	+++
Embryonic cells	E1	22.5 ± 1.0 ^b	14.5 ± 1.6 ^a	0.64 ± 0.04 ^b	236 ± 24 ^c	24.2 ± 4.2 ^c	0.10 ± 0.01 ^a	+++	+	+++
Embryonic cells	E2	42.0 ± 1.5 ^a	31.9 ± 0.7 ^b	0.76 ± 0.04 ^b	943 ± 55 ^d	88.5 ± 10.3 ^a	0.09 ± 0.01 ^a	+++	+	++

Values represent mean ± SD of 30 cells analyzed for each sampled stage. Stages correspond to: leaves from greenhouse plants (L1), explants during dedifferentiation [0 h (L2), 1 week (D1), 2 weeks (D2), 5 weeks (D3)], compact primary callus obtained 3 months after induction (C1), embryogenic callus obtained 7 months after induction (C2), established cell clusters obtained after 4 months in liquid proliferation medium (C3), pro-embryogenic masses [1 week in redifferentiation medium after auxin withdrawal (R1), 24 h in redifferentiation medium after reducing cell density (R2), 72 h (R3), 10 days (R4)], globular embryos (E1) and torpedo-shaped embryos (E2). W/L ratio corresponds to the mean of ratios between width and length for each cell. N/C ratio corresponds to the mean of ratios between nucleus and cytoplasm for each cell (nucleus-to-cytoplasm ratio). Cell division activity was estimated based on the number of cells in telophase. Starch granules were evidenced by the Schiff reagent and soluble proteins were evidenced by Naphthol Blue Black. Differences between the sampled stages were analysed with a one-way ANOVA test. Data followed by different letters in a same column are significantly different according to the Tukey post-hoc test ($P < 0.05$).

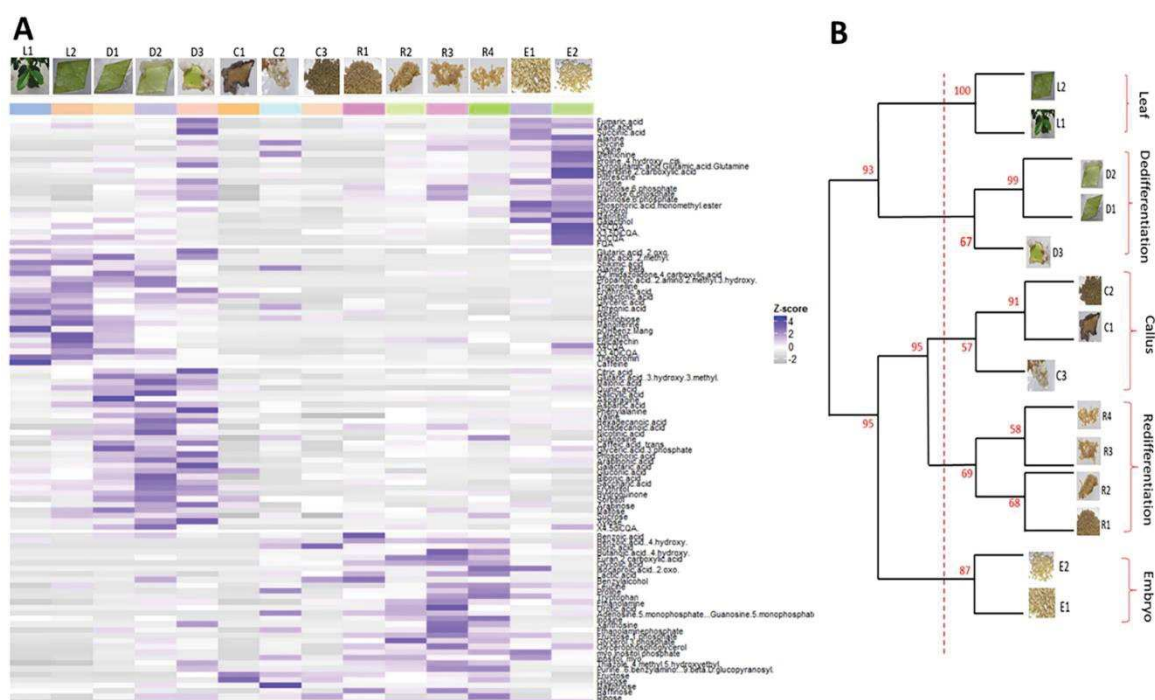


Figure 3. Profiling primary and secondary metabolites during 14 key sampled stages of the Arabica somatic embryogenesis process. (A) Heatmap generated from the Z-score values of a total of 92 primary metabolites and 12 secondary metabolites detected. Rows correspond to metabolites and columns to the sampled stages. Positive Z-score values are shown in blue and negative values in grey. (B) Hierarchical clustering of the 14 sampled stages according to the similarities in their metabolic profiles. The clustering was performed using Pearson's correlation coefficient. Cluster probabilities were calculated via a multiscale bootstrap with a total of 1000 iterations. Clustering yielded 5 major nodes: Leaf, Dedifferentiation, Callus, Redifferentiation and Embryo.

2.3. Metabolic Characterization of the Five Phases of the Arabica SE Process

For each developmental phase, primary and secondary metabolites displaying significant increased values from their respective means across the 14 stages ($Z\text{-score} > 1$) are listed in Table 2 and those displaying significant decreased values ($Z\text{-score} < -1$) are listed in Table 3. The “Leaf” phase highlighted an accumulation of two amino acids (beta-alanine, aspartic acid), metabolites implicated in the ascorbate and aldarate metabolism (galactonic and threonic acid) and also primary metabolites involved in the shikimate pathway (shikimic and quinic acid), as well as secondary metabolites production such as alkaloids (trigonelline, theobromine and caffeine) and phenolic compounds (mangiferin, catechin, epicatechin, 3-caffeoylquinic acid (3-CQA), 4-CQA, 3,4-diCQA, feruloylquinic acid (FQA)). The “Leaf” phase is also characterized by a decrease of sugars (fructose, ribose) and phosphorylated sugars (glucose-6-phosphate, fructose-6-phosphate, mannose-6-phosphate), two amino acids (leucine, lysine) and myo-inositol.

The “Dedifferentiation” phase showed an accumulation of metabolites involved in the tricarboxylic acid (TCA) cycle (aconitate, 2-oxoglutarate, succinate, fumarate, malate, citrate) and in the ascorbate and aldarate metabolism (arabinonic acid, galactaric acid, galactonic acid). It also included an accumulation of sugars (maltose, sucrose), phosphorylated sugars (glucose-6-phosphate, fructose-6-phosphate, mannose-6-phosphate), amino acids (beta-alanine, aspartic acid, asparagine, glycine, glutamine, valine), and fatty acids (hexadecanoic acid, octadecanoic acid, glyceric acid). Primary metabolites involved in secondary metabolism pathways were also accumulated such as phenylalanine, caffeic and quinic acid, nicotinic acid and xanthosine. During this phase, lower levels were obtained for some carbohydrates such as glucose, fructose and starch and one amino acid, lysine.

Table 2. Significantly over-accumulated metabolites in the five developmental phases forming the Arabica SE process.

Metabolic Pathways	Developmental Phases of Coffee Somatic Embryogenesis				
	Leaf	Dedifferentiation	Callus	Redifferentiation	Embryo
TCA cycle		Aconitic acid Citric acid Fumaric acid Glutaric acid, 2-oxo- Malic acid Succinic acid		Aconitic acid	Fumaric acid Malic acid Succinic acid
Sugar metabolism		Glucose-6-phosphate Mannose-6-phosphate Maltose Sucrose	Glucose Fructose Maltotriose Ribose	Glucose-6-phosphate Fructose-6-phosphate Mannose-6-phosphate Ribose	Glucose-6-phosphate Fructose-6-phosphate Mannose-6-phosphate Starch
Amino acid metabolism	Alanine, beta- Aspartic acid	Alanine, beta- Aspartic acid Asparagine Glycine Glutamine Valine	Glycine Methionine Proline	Leucine Lysine Proline	Alanine Glycine Leucine Lysine Methionine Proline Glutamate
Fatty acid metabolism		Hexadecanoic acid Octadecanoic acid Glyceric acid	Glycerol	Glycerol-3-phosphate	Glycerol Glycerol-3-phosphate
Polyols			Inositol, myo-	Inositol, myo-	Inositol, myo- Mannitol
N-compounds					Putrescine

Table 2. Cont.

Metabolic Pathways	Developmental Phases of Coffee Somatic Embryogenesis				
	Leaf	Dedifferentiation	Callus	Redifferentiation	Embryo
Ascorbate and aldarate metabolism	Galactonic acid Threonic acid	Arabinonic acid Galactaric acid Galactonic acid			
Shikimate pathway	Quinic acid Shikimic acid	Quinic acid Phenylalanine Caffeic acid		Tryptophan Caffeic acid Benzoic acid	Tryptophan
Alkaloid precursors		Nicotinic acid Xanthosine		Guanosine Xanthosine Inosine	
Alkaloids	Trigonelline Theobromine Caffeine				
Phenolic compounds	Mangiferin Catechin Epicatechin 3-CQA 3,4-diCQA 4-CQA FQA				3-CQA 3,5-diCQA 4-CQA 4,5-diCQA 5-CQA FQA

All primary and secondary metabolites displaying significant increased values from their respective means across the 14 stages (Z -score > 1) were listed in this table and were classified according to Kyoto Encyclopedia of Genes and Genomes (KEGG) database of metabolic pathways. CQA: caffeoylquinic acid, FQA: feruloylquinic acid, TCA: tricarboxylic acid.

Table 3. Significantly under-accumulated metabolites in the five developmental phases forming the Arabica SE process.

Metabolic Pathways	Developmental Phases of Coffee Somatic Embryogenesis				
	Leaf	Dedifferentiation	Callus	Redifferentiation	Embryo
TCA cycle			Fumaric acid Glutaric acid, 2-oxo- Malic acid		
Sugar metabolism	Fructose-6-phosphate Glucose-6-phosphate Mannose-6-phosphate Fructose Ribose	Glucose Fructose Starch	Fructose-6-phosphate Glucose-6-phosphate Mannose-6-phosphate Sucrose		Fructose
Amino acid metabolism	Leucine Lysine	Lysine	Leucine Lysine Valine	Valine	
Ascorbate and aldarate metabolism			Arabinonic acid Galactonic acid Threonic acid		
Polyols	Inositol, myo-				

All primary and secondary metabolites displaying significant decreased values from their respective means across the 14 stages (Z -score < -1) were listed in this table and were classified according to KEGG database of metabolic pathways. TCA: tricarboxylic acid.

The “Callus” phase displayed an accumulation of some sugars (maltotriose, ribose, glucose, fructose), amino acids (glycine, methionine, proline), fatty acids (glycerol), and myo-inositol. This phase induced a decrease for metabolites implicated in the TCA cycle and the ascorbate and aldarate metabolisms, as well as for phosphorylated sugars, sucrose and three amino acids (leucine, lysine, valine).

The “Redifferentiation” phase showed an accumulation of phosphorylated sugars, ribose, three amino acids (leucine, lysine, proline), glycerol-3-phosphate, myo-inositol, and aconitate. Primary metabolites involved in secondary metabolism pathways were also accumulated such as tryptophan, caffeic and benzoic acid, guanosine, inosine and xanthosine. Only valine exhibited a significant decrease for this stage.

The “Embryo” phase (globular to torpedo-shaped embryos) included an accumulation of TCA cycle metabolites, phosphorylated sugars, starch, amino acids (alanine, glycine, leucine, lysine, methionine, proline, glutamine, tryptophan), polyols (myo-inositol, mannitol), glycerol, glycerol-3-phosphate and putrescine. Phenolic compounds (3-CQA, 4-CQA, 5-CQA, 3,4-diCQA, 3,5-diCQA, FQA) were accumulated again, as in the “Leaf” phase. Only fructose was decreased.

The presence and accumulation of total chlorogenic acids during embryo redifferentiation was confirmed by multiphoton microscopy combined with emission spectral analysis of total chlorogenic acids in PEMs after 1 day (R2), 3 days (R3), 10 days (R4) and 3 weeks (E1) in redifferentiation medium (Figures 4 and S2). Emission spectral analysis allows to specifically associate emission spectra from cells with defined pure autofluorescent compounds (in this case, total chlorogenic acids) [40]. Absent in embryogenic cell clusters, total chlorogenic acids seemed to accumulate on the edge of PEMs at day 3 at the same location where the embryo structure later emerged (at day 10). Autofluorescence intensity of total chlorogenic acids was strongly associated with the number of globular-shaped embryos obtained after 3 weeks. Taken together, these findings indicate that chlorogenic acids are potent metabolic markers of the embryonic state.

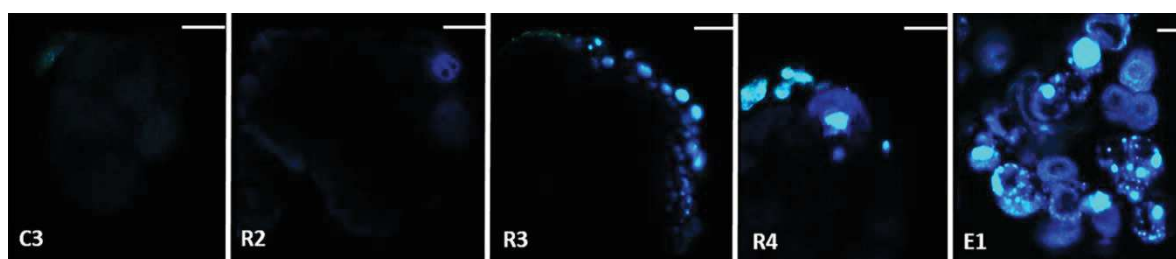


Figure 4. Chlorogenic acid localization in embryogenic cell clusters (C3) and during embryo redifferentiation after 1 day (R2), 3 days (R3), 10 days (R4) and 3 weeks (E1) in redifferentiation medium after cell density reduction. Protuberance of newly generated embryos emerge after 10 days in redifferentiation medium and globular-shaped embryos are obtained after 3 weeks. For chlorogenic acid localization, fresh samples were mounted on a glass slide in a drop of water and observed using a Zeiss LSM880 multiphoton microscope, equipped with a Chameleon Ultra II laser. Excitation wavelength was set to 720 nm and emission was observed with a band-pass filter 386–502 nm (blue). The presence and accumulation of total chlorogenic acids during embryo redifferentiation was confirmed by multiphoton microscopy combined with emission spectral analysis of total chlorogenic acids (Figure S2). The scale bar is set to 100 μ m.

2.4. Cell Hormonal Content Changes throughout the Arabica SE Process

Analysis of cell hormonal content showed variations over the five developmental phases of the SE process (Table 4). When switching from the “Leaf” phase to the “Dedifferentiation” phase, indole-3-acetic acid (IAA), 1-aminocyclopropane-1-carboxylic acid (ACC) and zeatin (*cis*- and *trans*-) (Z) levels significantly increased whereas *cis*-abscisic acid (ABA) levels significantly decreased. Moreover, indole 3-butyric acid (IBA) and isopentenyladenine (iP) levels were only detectable in the

“Dedifferentiation” phase. ABA glucose ester (ABAGE) and zeatin riboside (*cis*- and *trans*-) (ZR) levels were unaffected.

Table 4. Hormone dynamics during the five main developmental phases of the Arabica SE process.

Hormones	Developmental Phases of Coffee Somatic Embryogenesis				
	Leaf >	Dedifferentiation >	Callus >	Redifferentiation >	Embryo
ABA	583 ± 72 ^a	113 ± 16 ^b	74 ± 12 ^b	70 ± 12 ^b	600 ± 55 ^a
ABAGE	243 ± 45 ^a	121 ± 36 ^a	675 ± 41 ^b	671 ± 77 ^b	1230 ± 103 ^c
ACC	1281 ± 173 ^a	19491 ± 447 ^b	3063 ± 272 ^a	3179 ± 193 ^a	3065 ± 169 ^a
IAA	34 ± 6 ^a	182 ± 27 ^b	17 ± 7 ^a	60 ± 8 ^c	81 ± 19 ^c
IBA	N.D. ^a	4345 ± 385 ^b	N.D. ^a	N.D. ^a	N.D. ^a
iP	N.D. ^a	4557 ± 441 ^b	N.D. ^a	N.D. ^a	N.D. ^a
Z	2 ± 1 ^a	17 ± 3 ^b	2 ± 1 ^a	N.D. ^a	4 ± 2 ^b
ZR	6 ± 3 ^a	13 ± 3 ^a	22 ± 4 ^b	15 ± 5 ^a	60 ± 6 ^c

Values represent the mean of hormone levels, expressed in ng/g DW, in the sampled stages forming each of the five phases (Leaf, Dedifferentiation, Callus, Redifferentiation and Embryo) in three independent replicates ± SD. For each hormone, differences in levels between phases were analysed with a Kruskal-Wallis test. Data followed by different letters are significantly different according to Dunn post-hoc test ($P < 0.05$). Blue color corresponds to an increase in hormone levels between a phase and the previous one (column on the left) while grey color corresponds to a decrease in hormone levels. N.D. stands for not detected. ABA: *cis*-abscisic acid, ABAGE: ABA glucose ester, IAA: indole-3-acetic acid, IBA: indole 3-butyric acid, iP: isopentenyladenine, Z: zeatin (*cis*- and *trans*-), ZR: zeatin riboside (*cis*- and *trans*-).

When switching from the “Dedifferentiation” phase to the “Callus” phase, IAA, ACC and Z levels significantly decreased while ABAGE and ZR levels significantly increased. ABA was maintained at similar levels.

When switching from the “Callus” phase to the “Redifferentiation” phase, IAA levels increased again whereas ZR levels dropped. ABA, ABAGE, ACC and Z were maintained at similar levels.

When switching from the “Redifferentiation” phase to the “Embryo” phase, ABA, ABAGE, Z and ZR levels significantly increased. IAA and ACC levels were unaffected.

2.5. Most Significant Changes in Metabolic and Hormonal Profiles Occur during Phase Switches

A developmental phase switch corresponds to the transition between the last sampled stage of a developmental phase and the first sampled stage from the following phase. Since five main phases were highlighted (Leaf, Dedifferentiation, Callus, Redifferentiation, Embryo), four phase switches could be identified and summarized in Figure 5:

- (i) The leaf explant-to-dedifferentiated cell switch occurring after one week in induction medium (L2 stage to D1). During this switch, and even though no change could be observed on a morphological basis, huge changes were noticed at a biochemical level. Our results show a drastic increase in levels of metabolites involved in the TCA cycle pathway, in sugar (glucose-6-phosphate, fructose-6-phosphate, mannose-6-phosphate), amino acid and fatty acid metabolisms while shikimate, caffeine and phenolic compounds pathways were poorly present. Hormone levels also sharply increased during this phase switch mainly ACC (24-fold), IAA (6-fold), IBA (4707-fold), iP (5336-fold) and Z (6-fold), except for ABA which levels were decreased (2.2-fold).
- (ii) The dedifferentiated cells-to-established primary callus cells switch occurring three months after induction (D3 stage to C1). This switch was characterized by a sharp decrease in cell basic functions, mainly a decrease in the levels of metabolites involved in the TCA cycle pathway, sugar metabolism (phosphorylated glucose and fructose), and the ascorbate pathway which is related to the cell photosynthetic activity. Only synthesis of some amino acids was maintained (methionine, glycine). Hormone levels also sharply decreased in compact primary callus cells compared to cells in early dedifferentiation mainly IBA (1954-folds), iP (1527-fold), IAA (9-fold) and ACC (3.4-fold).

- (iii) The embryogenic cell clusters-to-PEMs switch occurring after one week in redifferentiation medium (C3 stage to R1). During this switch, a re-activation in cell primary and secondary functions was observed; levels of phosphorylated sugars (glucose-6-phosphate, fructose-6-phosphate, mannose-6-phosphate) and fatty acids (glycerol-3-phosphate) increased as well as precursors of caffeine (inosine, xanthosine) and end-products of the shikimate pathway (caffeic acid, benzoic acid, tryptophan). In parallel, IAA levels also increased in PEMs compared to embryogenic cell clusters (3.7-fold) while levels of ZR sharply decreased (1.6-fold).
- (iv) The PEMs-to-embryonic cells switch occurring after four weeks in redifferentiation medium (R4 stage to E1). During this switch, cell functions are re-established in embryonic cells with increased levels of TCA cycle metabolites, phosphorylated sugars and fatty acids, and the synthesis of a number of amino acids (alanine, glycine, leucine, lysine, methionine, proline, glutamate) and phenolic compounds (5-CQA; 3-CQA; 4-CQA; 3,5-di-CQA; 4,5-diCQA and FQA). A synthesis of putrescine and mannitol was also noticed. Our results also showed a significant increase in ABA levels (7-fold), ABAGE levels (1.9-fold), Z levels (40-fold) and ZR levels (40-fold) in embryonic cells compared to PEMs.

These results clearly show that comparing metabolic profiles between two sampled stages corresponding to a phase switch allowed obtaining a better description of these stages than a morphological or histological comparison as it made possible to distinguish between two sampled stages that are morphologically similar (e.g., leaf explant before and after one week on dedifferentiation medium) or histologically similar (e.g., dedifferentiated cells and established primary callus cells).

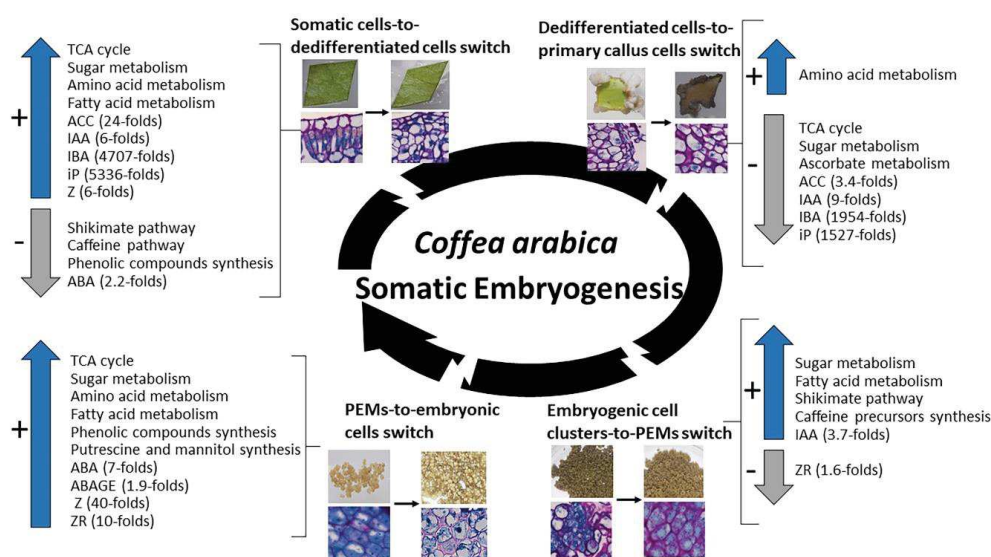


Figure 5. Metabolic pathways and hormone dynamics during the four main developmental phase switches. The four phase switches are: the leaf explant-to-dedifferentiated cell switch occurring after one week in induction medium, the dedifferentiated cells-to-established primary callus cells switch occurring three months after induction, the embryogenic cell clusters-to-pro-embryogenic masses (PEMs) switch occurring after one week in redifferentiation medium, and the PEMs-to-embryonic cells switch occurring after four weeks in redifferentiation medium. Metabolic pathways were identified based on an enrichment in significantly over-accumulated and under-accumulated metabolites, i.e., Z-score > 1 and Z-score < -1, respectively. All presented hormones differ significantly in their levels according to a two-by-two Wilcoxon mean comparison test ($P < 0.05$). Values represent mean of ratios in hormone levels of independent replicates between a sampled stage and the previous one \pm SD.

2.6. Embryogenic and Non-embryogenic Calli Differ Morphologically, Metabolically and Hormonally

Embryogenic (EC) and non-embryogenic calli (NEC) can be easily distinguished on the basis of their morphology, color and cellular characteristics (Figure 2A C2, NEC; Table 5). Somatic embryos

could not be regenerated from NEC. EC developed from compact primary calli whereas NEC developed from non-compact primary calli. EC were yellowish and friable, constituted by small cells ($214 \pm 18 \mu\text{m}^2$), isodiametric, surrounded by a thickened outer cell wall, arranged in clusters, with a dense cytoplasm (rich protein-staining) and a central nucleus with a prominent nucleolus stained in blue-black (N/C ratio = 0.33 ± 0.02) (Figures 1C2 and 2B C2). NEC were spongy and translucent, displaying cells of a much larger size ($2081 \pm 386 \mu\text{m}^2$) with the vacuole occupying a big share of the cytoplasm, and an absence of cytoplasmic organelles (Figures 1NEC and 2B NEC). The nucleocytoplasmic ratio in NEC was significantly lower than in EC (0.17 ± 0.02 vs 0.33 ± 0.02). Protein and sugar staining-surfaces were limited to cell membranes and cell walls. A characteristic attribute of EC in coffee was the abundance of starch granules whereas larger insoluble starch reserves were found in NEC.

At a biochemical level, the heatmap generated from primary and secondary metabolite profiles of both tissues (Figure 6) displayed significantly increased and decreased levels for some metabolites (Z-score > 1 and Z-score < -1, respectively) listed in Table 6. Both EC and NEC accumulated ribose, amino acids (methionine and glycine) and fatty acids (glycerol) (Table 6) and had negative Z-scores for phosphorylated sugars, metabolites related to ascorbate and aldarate pathways and valine. However, while NEC over-accumulated TCA cycle metabolites and two amino acids (β -alanine, lysine) and under-accumulated a polyol (myo-inositol) and an amino acid (proline), EC over-accumulated sugars (glucose, fructose, maltotriose), an amino acid (proline), myo-inositol, and under-accumulated TCA cycle metabolites, an amino acid (leucine) and sucrose.

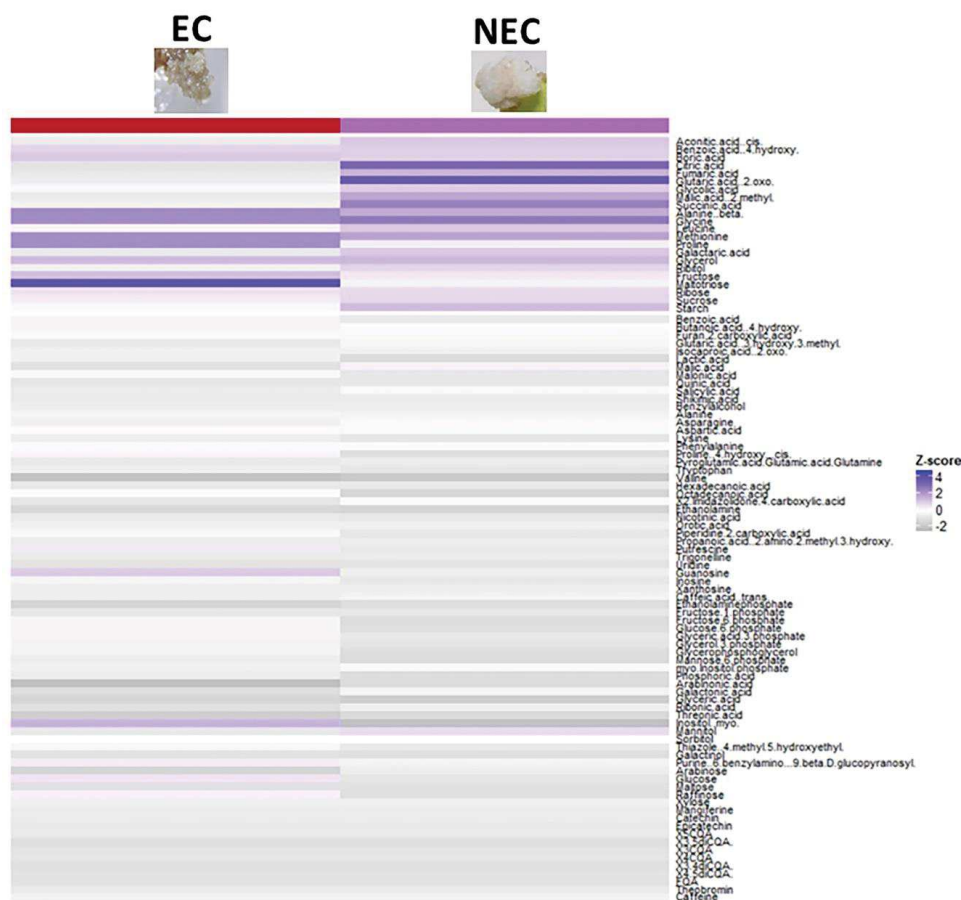


Figure 6. Profiling primary and secondary metabolites in embryogenic (EC) and non-embryogenic calli (NEC). Heatmap generated from the Z-score values of a total of 92 primary metabolites and 12 secondary metabolite detected. Rows correspond to metabolites and columns to the type of callus. Positive Z-score values are shown in blue and negative values in grey.

Table 5. Characteristics of the mainly represented cell populations in embryogenic (EC) and non-embryogenic calli (NEC).

Cell Type	Stage	Cell Length (μm)	Cell Width (μm)	W/L Ratio	Cytoplasm Size (μm ²)	Nucleus Size (μm ²)	N/C Ratio	Cell Division	Starch	Proteins
Embryogenic cells	EC	20.2 ± 1.3 ^a	14.7 ± 0.5 ^a	0.73 ± 0.03 ^a	214 ± 18.0 ^a	70.7 ± 5.1 ^a	0.33 ± 0.02 ^a	++	++	+++
Non-embryogenic cells	NEC	69.7 ± 10.7 ^b	41.3 ± 1.6 ^b	0.60 ± 0.08 ^a	2081 ± 386 ^b	364 ± 53 ^b	0.17 ± 0.02 ^b	+++	+++	+

Values represent mean ± SD of 30 cells analyzed for each sampled stage. W/L ratio corresponds to the mean of ratios between each cell width and length. N/C ratio corresponds to the mean of ratios between each cell nucleus and cell size. Cell division activity was estimated on the basis of the number of cells in telophase. Starch reserves and proteins were evidenced by the Schiff reagent and Naphthol Blue Black, respectively. Data followed by different letters are significantly different according to Student *t*-test ($P < 0.05$).

Table 6. Over-accumulated and under-accumulated metabolites in embryogenic and non-embryogenic calli.

Metabolic Pathways	Over-Accumulated Metabolites		Under-Accumulated Metabolites	
	Embryogenic Callus	Non-Embryogenic Callus	Embryogenic Callus	Non-Embryogenic Callus
TCA cycle		Citric acid Glutaric acid, 2-oxo- Malic acid Succinic acid	Fumaric acid Glutaric acid, 2-oxo- Malic acid	
Sugar metabolism	Glucose Fructose Maltotriose Ribose	Starch Ribose	Fructose-6-phosphate Glucose-6-phosphate Mannose-6-phosphate	Fructose-6-phosphate Glucose-6-phosphate Mannose-6-phosphate
Amino acid metabolism	Glycine Methionine Proline	Alanine, beta- Glycine Leucine Methionine	Leucine Valine	Proline Valine
Fatty acid metabolism	Glycerol	Glycerol		
Polyols	Inositol, myo-		Inositol, myo-	
Ascorbate and aldarate metabolism			Arabinonic acid Threonic acid Galactonic acid	Arabinonic acid Threonic acid

All primary and secondary metabolites displaying significant increased values (Z -score > 1) and decreased values (Z -score < -1) between embryogenic callus and non-embryogenic callus, were listed in this table and were classified according to KEGG database of metabolic pathways. TCA: tricarboxylic acid.

Concerning hormones, both EC and NEC had similar levels of ACC [2855–3177 ng/g DW] and IAA [19–32 ng/g DW] (Table 7). However, levels of ABA and ABAGE were significantly higher in EC than in NEC (16- and 50-fold, respectively), as well as the levels of ZR (4-fold). Z was only detected in EC (2 ± 1 ng/g DW). IBA and iP were not detected in either tissues.

Table 7. Hormone levels in embryogenic and non-embryogenic calli.

Hormones	Embryogenic Callus	Non-Embryogenic Callus
ABA	130 ± 14 ^a	8 ± 2 ^b
ABAGE	1200 ± 84 ^a	24 ± 11 ^b
ACC	3074 ± 96 ^a	3016 ± 161 ^a
IAA	25 ± 6 ^a	29 ± 3 ^a
IBA	N.D. ^a	N.D. ^a
iP	N.D. ^a	N.D. ^a
Z	2 ± 1 ^a	N.D. ^b
ZR	40 ± 5 ^a	10 ± 3 ^b

Values represent the mean ± SD of 3 independent replicates, expressed in ng/g DW. Non-parametric two-by-two Wilcoxon mean comparison test was used to calculate any significant differences in hormone levels between embryogenic and non-embryogenic calli. Data followed by different letters are significantly different ($P < 0.05$). Grey color corresponds to a decrease in hormone levels in non-embryogenic calli compared to embryogenic-calli. N.D. stands for not detected. ABA: *cis*-abscisic acid, ABAGE: ABA glucose ester, IAA: indole-3-acetic acid, IBA: indole 3-butyric acid, iP: isopentenyladenine, Z: zeatin (*cis*- and *trans*-), ZR: zeatin riboside (*cis*- and *trans*-).

3. Discussion

3.1. Arabica SE Metabolites as a Signature of Cell Fate

This study reports the first global analysis of SE metabolome and hormone dynamics, conducted on 15 key sampled stages covering the entire SE process, from leaf explant dedifferentiation to embryo development. Two pre-requisites were mandatory for this study: (i) the availability of large-scale protocols for coffee SE, offering high reliability and cell synchronicity in each sampled stage as well as a good efficiency; (ii) the availability of the latest omics technologies allowing to rapidly offer masses of information. This first global analysis applied on a model tree species for SE should serve as a reference for a wide range of plant species as the detailed sampling conducted on 15 stages provided a better field-of-view to the SE process as a whole and enabled to clarify some real black boxes [24]. The statistical approach employed in this study helped to clearly divide the SE process into five phases and four key phase switches that are strategic for the whole biological efficiency of embryo redifferentiation. Many authors recently reported the necessity of a better understanding of the SE process to clarify the current bottlenecks [22,26,41]. The fact that each of the metabolic profiles correlated with a specific cell status—leaf cells, cells in dedifferentiation, callus cells, pro-embryogenic masses, embryonic cells—is extremely relevant. This shows that metabolite fingerprints are a specific signature for each developmental phase and bring rich information on cell status, as some of these stages were undistinguishable by morphological and histological approaches. A number of papers on conifers assumed that clear correlations may exist among the different metabolome or phytohormone profiles and some specific SE stages (cell line proliferation to embryo maturation) [28,29,31]. This global analysis is a proof that metabolites could be good predictors of all cell fate transitions—from leaf explant introduction until embryo development—and could be used to control the SE process key stages. This goes beyond describing cell morphology and color, which was until now the only way to support empirical protocol optimization. Combining this study with transcriptomic approaches could give a much clearer understanding of the molecular mechanisms underlying cell reprogramming [22].

3.2. The Dedifferentiation Episode Is Characterized by Huge Changes in Cell Metabolic Pathways

SE is a process driven by exogenously supplied plant growth regulators. Although most plants require similar physical conditions (temperature, light regime) for the induction of SE, medium

composition can have a large impact on SE outcome. According to Sugimoto et al. [42], the prime characteristic of plant regeneration is cell fate reprogramming induced by wounding, stress, and hormones. Auxins and cytokinins are widely known to play essential roles in the induction of embryogenic cultures [31]. The combined addition of natural (IBA and N⁶-(2-Isopentenyl) adenine (2-iP)) and synthetic (2,4-dichlorophenoxyacetic acid (2,4-D) and 6-benzylaminopurine (6-BA)) as exogenous growth regulators in the medium sharply increased endogenous levels of IAA and Z. The synthesis of endogenous auxins is considered as a crucial early step in the switch to totipotent growth. The results reported by different authors [43,44] account for an important role of auxins in SE induction. The precursor of ethylene, ACC, also showed a sharp increase during cellular dedifferentiation in our study and has been reported to improve somatic embryo induction in *Arabidopsis thaliana* [45].

Somatic cells within the plant contain all the genetic information necessary to create a complete and functional plant (with the exception of anuclear vascular cells) [6]. During cellular dedifferentiation, the existing developmental information of somatic cells must be switched off or re-configured to make the somatic cells ready for an embryogenic program [46]. Our results showed that secondary metabolites (phenolic compounds and alkaloids) no longer accumulated during this stage, whereas their precursor levels increased (phenylalanine and caffeic acid; nicotinic acid and xanthosine, respectively). Nic-Can et al. [47] have published a demonstration on the effect of caffeine and chlorogenic acid in SE of *C. canephora*. These authors have provided evidence that those compounds inhibited the embryogenic process by affecting DNA methylation. Since leaf explants are placed in the dark, metabolite routes that are dependent on the photosynthetic activity such as the ascorbate metabolism are slowly replaced by cell respiratory pathways, mainly glycolysis and TCA cycle. Glycolysis is a key and ubiquitous metabolic pathway that allows plant cells to convert carbohydrates into the energetic coin ATP. Our results are in agreement with a number of studies that showed that the glycolytic pathway is used by cells undergoing dedifferentiation [48,49].

Callus induction is a spectacular phenomenon in which important changes in tissue structure take place. Xu et al. [50] showed that genes related to cell wall hydrolysis and lipid metabolism are quickly activated after auxin treatment. Our results revealed an accumulation of amino acids and lipid residues that could originate from the fast degradation of membranes and proteins.

3.3. Identifying Metabolic Markers of the Embryogenic Capacity

To date, many authors have distinguished EC from NEC based on morphological and histological studies [51–53]. Biochemical comparison of these tissues showed that carbohydrates, phytohormones and free amino acids differentially accumulated in EC and NEC [32,54,55]. When comparing metabolomic and hormonal profiles of EC and NEC, we were able to identify metabolic markers of the embryogenic state that can be directly used to pilot the optimization of the composition of the culture medium. An over-accumulation of maltotriose, myo-inositol, proline, ABA and Z was characteristic of EC as opposed to NEC. On the contrary, an under-accumulation of leucine and TCA cycle metabolites was found in EC NEC. These metabolites could serve as predictors of the regenerative capacity.

As observed through histological analysis, EC is characterized by the abundance of starch granules, whereas larger insoluble starch reserves were found in NEC. Metabolite profiles in our study showed that NEC accumulated starch reserves, while EC highly accumulated maltotriose, a product of starch hydrolysis. Gautier et al. [32] also observed large structures of starch reserves in NEC of Douglas-fir (*P. menziesii*) and hypothesized that NEC main fate was survival while EC was a transient state before fate transition. We highlight here for the first time that maltotriose is the preferred way to store sugars in EC for a faster usage during fate transition. Other metabolic markers of the embryogenic state, myo-inositol and proline, that were abundant in EC, were detected at low levels in NEC. High levels of inositol were also found in Norway spruce cell lines competent for SE and are involved in stress responses enabling embryo formation [56]. Guillou et al. [57] reported a better embryo-to-plant conversion in *Theobroma cacao* with medium supplemented with myo-inositol. Liang et al. [58] pointed

out that proline metabolism stimulates cell-signaling pathways through increased formation of reactive oxygen species (ROS) that promote embryo formation. Neves et al. [54] also found an accumulation of free proline in EC cells of sugarcane. Another metabolic marker of the embryogenic state was leucine which levels were severely lower in EC than in NEC. This could be due to the synthesis of leucine-rich repeat (LRR) proteins which are fused to a central nucleotide binding domain and collectively called the NB-ARC domain. The role of NB-ARC domain containing receptor kinases related to SE, e.g., SERK, was reported in several plant species [6].

Furthermore, Magnani et al. [59] reported in *Arabidopsis thaliana* that EC shut off biochemical pathways related to sugar metabolism and activate the transcriptional machinery. Our results are in agreement with this statement since no TCA cycle activity was present in EC. However, TCA cycle was highly active in NEC. This supports our above hypothesis about NEC battle for survival and makes the absence of TCA metabolites a predictor of the embryogenic state. At the hormonal level, EC had increased concentrations of ABA, ABAGE, Z and ZR compared to NEC. ABA has been previously reported to play an important role in embryo development [60]. Z plays an important role in *WUSCHEL* mediated maintenance of stem cell niche in the shoot apical meristem [61].

3.4. Metabolic and Hormonal Profiles of the Callus Phase Are Proof of Cell Totipotency

After the groundbreaking discovery that callus can be generated artificially in vitro [62] and that the balance between two plant hormones, auxin and cytokinin, determines the state of differentiation and dedifferentiation [63], the totipotency characteristic of callus cells has been widely used in both basic research and industrial applications [64]. Our results revealed a drastic decrease in endogenous hormone levels (auxins, cytokinins and the ethylene precursor ACC) after callus formation probably reflecting that the totipotency program is now established [65]. Only an increase in ABAGE levels is noted during callus phase and mainly in embryogenic calli. This suggests that cell fate is going towards an embryo expression program because ABA has been previously reported to play an important role in embryo development [60]. However, this program is still repressed since ABA is accumulated in a glucosylated form (ABAGE). Our results also showed that the totipotent state of callus is characterized by a clear shut-off of biochemical pathways related to sugar metabolism and energy production in agreement with Magnani et al. [59] in *Arabidopsis*, and a shut-off of the photosynthesis-related activity as previously reported by Ladygin et al. [66] in *Stevia rebaudiana*. The decrease reached a minimum level for phenolic compounds as previously described in ginger (*Zingiber officinale*) by Ali et al. [67]. These findings bring the first insights on the biochemical pathways involved in the expression of totipotency in callus. Consequently, exogenous hormones play an essential role to maintain the stem cell niche [31]. Only sucrose content showed a sharp decrease followed by an increase in glucose and fructose meaning that cells are getting their energy from the exogenous sucrose as recently showed by Gautier et al. [32] in Douglas-fir (*P. menziesii*).

In addition to the embryogenic callus, embryogenic cell clusters and well-established compact primary callus clustered together based on their metabolic profiles. Surprisingly, primary callus had more similarities with the embryogenic callus than with the initial stages of dedifferentiation. This means that this stage could already contain, in a repressed way, the program leading to the regeneration of a whole plant. This theory is in agreement with Ge et al. [68] who showed that differential molecular responses are already present in very early stages of EC formation. An observed higher nucleocytoplasmic ratio in primary callus and EC also supports this theory. Primary callus represents a true limiting element that needs improvement of culture conditions to reduce the time needed to obtain a characteristic embryogenic callus (7 months). The reliable and highly efficient redifferentiation potential of cell clusters proves that the exogenous auxin and cell density conditions have been well optimized in coffee SE in order to maintain a proliferating totipotent state without losing embryogenic capacity [37].

3.5. Re-Establishment of Primary and Secondary Metabolites During Redifferentiation Leading to Embryo Development

To trigger the redifferentiation process towards the development of somatic embryos, cell clusters are inoculated at a low density in a fresh medium where the exogenous auxin (2,4-D) is removed. The significant increase in endogenous IAA levels along with the accumulation of tryptophan, a precursor of IAA biosynthesis, was also highlighted by Businge et al. [28] in Norway spruce cell lines during redifferentiation. Even though the medium is highly supplemented with 6-BA during redifferentiation, Z was detected at rather low concentrations in embryos, whereas ZR levels sharply increased. The relationship between auxin and cytokinin contents is evidently decisive for embryonic development [61], as the high auxin/cytokinin ratio at maturation seems to be a prerequisite for proper formation of apical and root meristems, optimal development of root caps, hypocotyls and cotyledons [69]. Endogenous ABA which was not regulated exogenously (no ABA supply), showed significantly increased levels in globular and torpedo-shaped embryos during redifferentiation into embryos. This agrees with previous work showing a marked increase in the endogenous levels of ABA and ABAGE during maturation of larch (*Larix kaempferi*) somatic embryos [60]. In conifers, high exogenous ABA concentrations (20 μM) are needed throughout the maturation process whereas the absence, or low amounts, of ABA in the medium may lead to inhibited or aberrant development characterized by numerous malformed embryos [70].

While cell respiratory pathways (glycolysis and TCA cycle) resume their activities to produce enough energy for the formation of new early embryos, an accumulation of starch is noted in developed embryos that has been reported to contribute to the maturation process [71]. Synthesis of a number of amino acids reveals that embryos are ready to produce more complex structures, i.e., storage and polarization-related proteins [72]. Our results showed that a polyamine, putrescine, accumulated. This finding agrees with Pedroso et al. [73] who postulated that putrescine might be related to the formation and development of globular embryos in *Camellia japonica*. An accumulation of phenolic compounds is also seen during embryo development, mainly chlorogenic acids, as they are key intermediaries for lignin synthesis, in addition to their more general role in reactive oxygen species scavenging and abiotic stress tolerance [74]. Polyamines have also been involved in plant abiotic stress tolerance [75], confirming the fact that somatic embryo formation is a stress-related phenomenon. Precursors of caffeine also accumulated during redifferentiation probably meaning that caffeine would be produced in later stages. Mannitol accumulated, probably to help tolerate the osmotic stress generated by the growth medium after cell density reduction [76].

4. Materials and Methods

4.1. Tissue Culture and Sampling

An intraspecific hybrid (GPFA116) proceeding from Arabica breeding program and selected for its high cup quality, high yield and vigor, was used in this study. Somatic embryogenesis was performed in the Nestlé Research laboratories (Tours, France) based on the large-scale protocols described previously for *C. arabica* [10], and carried out in five replicates, i.e., five independent leaf introductions in February, April, June, October and December 2016 from distinct 1-year old plants obtained by SE and grown in the Nestlé Research greenhouse. Explants were then cultured in Petri dishes on T1 'dedifferentiation 1' medium, i.e., Murashige and Skoog (MS) half-strength solid medium supplemented with 2.3 μM 2,4-D, 4.9 μM IBA and 9.8 μM 2-iP for 1 month before transfer on T2 'dedifferentiation 2' medium, i.e., MS/2 solid medium supplemented with 4.6 μM 2,4-D and 17.8 μM 6-BA for 6 months until emergence of embryogenic calli. Petri dishes were placed at 25 °C in the dark. Embryogenic calli were then inoculated at a density of 10 g/L in Erlenmeyer flasks containing M 'proliferation' liquid nutritive medium, i.e., medium supplemented with 1.4 μM 2,4-D and 4.4 μM 6-BA and cultured for 4 months on shakers (120 rpm) at 25 °C in the dark, until proliferation of cell clusters. To enhance somatic embryo regeneration, cell clusters were transferred to 250-mL Erlenmeyer flasks containing 100 mL of DIF 'redifferentiation' liquid medium lacking 2,4-D at a 10 g/L density for 1 week then at a 1 g/L density for

8 weeks until development of torpedo-shaped embryos. Erlenmeyer flasks were placed on shakers (120 rpm) at 25 °C in the dark.

Fourteen sampled stages were chosen to cover the SE process from leaf explant to torpedo-shaped embryos development as shown in Figure 1: leaves from greenhouse plants (L1), explants during dedifferentiation [0 h (L2), 1 week (D1), 2 weeks (D2), 5 weeks (D3)], compact primary callus obtained 3 months after induction (C1), embryogenic callus obtained 7 months after induction (C2), established cell clusters obtained after 4 months in proliferation medium (C3), early regeneration of embryos from cell clusters [1 week in DIF medium without reducing cell density (R1), 24 h in DIF medium after reducing cell density (R2), 72 h (R3), 10 d (R4)], globular embryos (E1) and torpedo-shaped embryos (E2). Additionally, non-embryogenic callus (NEC) was also sampled at the same time as the embryogenic callus (C2). Approximately 2 g of fresh weight/sample/replicate were collected for metabolome and hormone content analysis. For these experiments, material was kept in liquid nitrogen then freeze-dried. For histological analysis, fresh samples were incubated in a fixative solution (1% glutaraldehyde, 2% paraformaldehyde and 1% caffeine in 0.1 M phosphate buffer pH 7.0) for 24 h at 4 °C.

4.2. Primary Metabolite Analysis

Primary metabolites were extracted according to Dethloff et al. [77]. Briefly, freeze-dried samples were ground in a ball mill (TissueLyser II, Qiagen, Hilden, Germany) and 30 mg of each sample were swiftly mixed with 300 µL methanol, 30 µL nonadecanoic acid methylester (CAS 1731-94-8) (2 mg/mL in chloroform) and 30 µL of ¹³C₆-sorbitol (CAS 121067-66-1) (0.2 mg/mL in methanol). 40 µL from the polar phase were dried in a ScanVac speed vacuum concentrator (Labogene, Allerød, Denmark). Dried samples were mixed with 40 µL of methoxyamine hydrochloride (CAS 593-56-6, 40 mg/mL in pyridine (CAS 110-86-1)) and incubated for 90 min at 30 °C. Then, 70 µL of N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA, CAS 25561-30-2) and 10 µL of a C₁₀, C₁₂, C₁₅, C₁₈, C₁₉, C₂₂, C₂₈, C₃₂, and C₃₆ *n*-alkane mixture were added for retention index calibration [78]. The mixture was subsequently incubated for additional 30 min at 37 °C to complete the reaction. Sample aliquots of 1 µL were injected in splitless and split 1/30 mode using an Agilent 6890 injector system set to 250 °C (Santa Clara, CA, USA). Initial oven temperature was maintained at 70 °C for 1 min, then raised to 350 °C at 9 °C/min and kept at 350 °C for 5 min. Helium carrier gas flow was set to 0.6 mL/min. All samples were run on a gas chromatography-electron impact ionization-time of flight/mass spectrometry (GC-EI-TOF/MS) instrument with an Agilent 6890N gas chromatograph and a LECO Pegasus III TOF mass spectrometer using the manufacturer's ChromaTOF software for acquisition and baseline correction (version 2.32, LECO, St. Joseph, MI, USA). Data processing and peak identity annotation was manually performed using the TagFinder visualization tool for mass spectral matching of time groups and clusters [79]. Mass spectral features were matched to the mass spectra and retention time indices of authenticated reference metabolites from the Golm Metabolome Database [80]. Criteria for manually supervised metabolite annotation were the presence of at least 3 specific mass fragments per compound and a retention index deviation <1.0% [81]. Values obtained were normalized with internal standard (¹³C₆-sorbitol) and DW. This experiment was carried out in 5 replicates. Metabolites showing a coefficient of variation >80% between replicates in at least 3 stages of the SE process were removed.

4.3. Secondary Metabolite Analysis

The freeze dried samples were ground into fine powder in a ball mill (TissueLyser II, Qiagen) and extracted by stirring (225 rpm, Rotamax 120, Heidolph, Schwabach, Germany) 15 mg of each collected sample in 600 µL of methanol (MeOH)/H₂O (80:20, *v/v*) at 4 °C, in the dark, for 3 h. After centrifugation (10 min, 8 °C, 3500 rpm), the methanolic extract was collected and filtered (0.25 µm porosity, Interchim, Montluçon, France) before analysis. Each extraction was carried out in triplicate.

Quantitative analyses were carried out on a high-performance liquid chromatography (HPLC) system (Shimadzu LC-20, Tokyo, Japan) equipped with a binary pump, a photodiode array detector (DAD) and an Eclipse XDB C18 (3.5 μm) column (100 mm \times 4.6 mm, Agilent). Extracts (10 μL) were analyzed at a flow rate of 0.6 mL/min using an elution system composed of solvents B ($\text{H}_2\text{O}/\text{MeOH}/\text{acetic acid}$, 5:90:5 *v/v/v*) and A (water/acetic acid, 98:2, *v/v*) mixed as described in Campa et al. [82]. Parallel analyses were performed in triplicate on pure standard solutions of theobromine, caffeine, mangiferin, 3,4-, 3,5- and 4,5-O-diCQA, 5-CQA and caffeic acid from Sigma-Aldrich (St Quentin Fallavier, France), glycosylated kaempferol and quercetin, rutin, (+)-catechin, (–)-epicatechin and epigallocatechin from Extrasynthese (Lyon, France) at 10, 25, 50, and 75 $\mu\text{g}/\text{mL}$. Quantification of 3-, 4- and 5-CQA, FQA and 3,4-, 3,5- and 4,5-diCQA was undertaken at 320 nm, caffeine and catechin derivatives at 280 nm, and mangiferin, kaempferol and quercetin derivatives at 360 nm. Concentration was expressed in mg/g dry weight by comparison with the standard curves established with respective standards. For 3-CQA, 4-CQA and FQA, content was calculated taking into account the 5-CQA standard curve. Each coffee leaf powder sample was extracted and analyzed twice. Values obtained were normalized with DW.

4.4. Hormone Content Analysis

Quantification of widely reported phytohormones, i.e., ABA and ABA catabolites, auxins, cytokinins, gibberellins and the ethylene precursor ACC, was conducted at the National Research Council of Canada (Saskatoon, SK, Canada) by ultra-performance liquid chromatography-electrospray tandem mass spectrometry (UPLC-ESI-MS/MS) [83] using deuterium labeled internal standards. The procedure for the hormone profiling analysis was previously described in detail [84,85]. This experiment was carried out in triplicate. Results were expressed as ng/g DW. Hormones showing a coefficient of variation >80% between replicates in at least 3 of the 14 stages of the SE process were removed.

4.5. Starch Content Assay

The starch content of 20 mg of freeze-dried powder was determined using the total starch kit GOPOD (D-glucose, κ -Gluc, Megazyme International, Wicklow, Ireland). After elimination of soluble sugars and of the soluble products of starch degradation, the residue was successively hydrolysed into glucose units with α -amylase and amyloglucosidase. The resulting D-glucose was then degraded with glucose oxidase and the resulting hydrogen peroxide quantified by spectrophotometry at 510 nm after a last enzymatic reaction. Values obtained were normalized with DW.

4.6. Histological Analysis and Cell Imaging

Fixated samples were dehydrated by successive immersions of 1 h-duration each in graded solvent series of ethanol, from 50% to 100%. Samples were then embedded in Technovit 7100 resin (Kulzer, Werheim, Germany) according to the manufacturer's instructions. The cross-sections (3.5 μm), obtained with a Leica RM2255 microtome (Leica Microsystems, Wetzlar, Germany), were oxidized in 1% periodic acid (5 min), washed with distilled water and stained in the dark for 10 min with Schiff reagent (ref. 3952016, Sigma Aldrich). After washing, sections were stained with Naphthol Blue Black [1 g in 7% (*v/v*) (ref. 1.01167, Sigma Aldrich)] for 5 min at 60 °C, treated with 7% acetic acid and finally dried for 15 min at 60 °C. A total of 9 sections per time point were analysed. Images were obtained under a Nikon Eclipse Ni-E microscope (Nikon Instruments Inc., Melville, NY, USA) with two objectives, 20 \times and 40 \times , after acquisition using NIS software (Nikon Instruments Inc.). Images were analyzed using Image J 1.47v software.

Chlorogenic acid localization in living cells was carried out according to Talamond et al. [40]. Fresh samples were mounted on a glass slide in a drop of water and observed using a Zeiss LSM880 multiphoton microscope (Zeiss, Jena, Germany), equipped with a Chameleon Ultra II laser (Coherent, Santa Clara, CA, USA) and with an 20x Plan Apo 1.0 NA objective. Excitation wavelength was set to 720 nm and emission was observed with a band-pass filter, 386-502 nm (blue). Image acquisition was performed using Zen software (Zeiss). The presence and accumulation of total chlorogenic acids

during embryo redifferentiation was confirmed by multiphoton microscopy combined with emission spectral analysis of total chlorogenic acids (Figure S2). After obtaining the spectral acquisitions, the LinearUnmixing function (Zen software) was executed to separate, pixel by pixel, the mixed signals of defined pure autofluorescent compounds (chlorogenic acids and chlorophyll), using the entire emission spectrum of each compound, plus a residual channel. In the residual channel, the intensity values represented the difference between the acquired spectral data and the fitted linear combination of the reference spectra.

4.7. Statistical Analysis

All analyses were carried out with R [86] using the default functions. First, a Principal Component Analysis (PCA) was carried out to verify homogeneity of replicates using the `prcomp` function (package `stats`) (Figure S1). For each primary or secondary metabolite, normalized values obtained at each stage were transformed to Z-scores by subtracting the mean of the values obtained at all collected stages and dividing it by the standard deviation. The Z-score transformation provides a way of standardizing data across a wide range of experiments [87]. A Z-score equal to 1 is one standard deviation above the mean. Heatmaps of Z-scores were generated using the `ComplexHeatmap` package [88]. The same data were used for the analysis of hierarchical clustering. Hierarchical clustering was performed with the `pvclust` package [89] using Pearson's correlation coefficient [90]. Cluster probabilities were calculated via a multiscale bootstrap with a total of 1000 iterations. A cluster probability is a percentage which indicates how strong the cluster is supported by data. One-way ANOVA was used to calculate any significant differences (at $P = 0.05$) in cell measurements between the 14 sampled stages of the SE process, followed by a Tukey test as post-hoc. Similarly, a Student *t*-test was used to compare cell measurements in embryogenic (EC) and non-embryogenic calli (NEC). Normality of distribution and equality of variance were verified prior to these tests. Concerning hormone content, first, the mean of values of the different sampled stages was calculated for each developmental phase then a Kruskal-Wallis test was used to calculate any significant differences (at $P = 0.05$) in mean hormone levels between phases, followed by a Dunn test as a post-hoc (`dunn.test` package). A non-parametric two-by-two Wilcoxon mean comparison test was used to assess significant differences (at $P = 0.05$) in hormone content between the four phase switches and between EC and NEC. Highlighted metabolic pathways over the four switches were identified based on an enrichment analysis using `Metaboanalyst` [91] in significantly over-accumulated and under-accumulated metabolites (Z-score >1 or <-1).

5. Conclusions

This first global analysis of SE metabolomics, applied on coffee, one of the model tree species, showed that metabolomics provides a powerful approach to investigate global metabolite patterns during the SE process. This approach helped to clearly divide the 14 sampled stages of the SE process into five important phases (Leaf, Dedifferentiation, Callus, Redifferentiation and Embryo) leading to the identification of four key developmental phase switches, which are strategic for the whole biological efficiency of embryo regeneration. This global analysis showed that metabolites provide a specific signature for each developmental phase and bring us rich information on cell fate, as some of the sampled stages were undistinguishable by conventional morphological and histological studies (e.g., the first dedifferentiation and redifferentiation events). A number of metabolic markers of the embryogenic state were identified and can be directly used as targets to pilot the optimization of the composition of the culture media. This global metabolomics analysis brings first insights to unravel the metabolic and hormonal mechanisms taking place during SE and will be combined with a transcriptomic approach we are carrying on the same sampled stages in order to give a much clearer understanding of the intimate mechanisms governing totipotency and embryogenesis. This study represents a starting point for optimizing coffee SE protocols in a rational way, putting an end to thirty

years of empirical research. These findings should be informative and useful to a wide range of plant species, offering unprecedented perspectives in plant micropropagation.

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/1422-0067/20/19/4665/s1>.

Author Contributions: H.E., B.B. and D.C. designed the research. D.B., F.G. and R.A. established tissue culture protocols. M.L., C.C., E.G. and R.A. contributed to material sampling strategy. J.K., A.E., L.L., F.M.-S. and R.A. contributed to primary metabolite analysis. C.C., C.A. and R.A. contributed to starch and secondary metabolite analysis. J.-L.V., G.C., S.L., E.D. and R.A. contributed to cell imaging and histological analysis. R.A. and B.B. performed the statistical analysis. R.A. and H.E. wrote the draft of the manuscript and all authors revised it.

Funding: R.A. benefits from a Ph.D. grant and funding from Nestlé and CIRAD.

Acknowledgments: We thank the imaging facility MRI, which is part of the UMS BioCampus Montpellier and a member of the national infrastructure France-BioImaging, supported by the French National Research Agency (ANR-10-INBS-04).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

2,4-D	2,4-Dichlorophenoxyacetic acid
2-iP	N ⁶ -(2-Isopentenyl) adenine
6-BA	6-Benzylaminopurine
ABA	<i>cis</i> -Abscisic acid
ABAGE	ABA glucose ester
ACC	1-Aminocyclopropane-1-carboxylic acid
BSTFA	N,O-Bis(trimethylsilyl)trifluoroacetamide
CQA	Caffeoylquinic acid
DAD	Photodiode array detector
DW	Dry weight
EC	Embryogenic callus
FQA	Feruloylquinic acid
GC–EI–TOF/MS	Gas chromatography–electron impact ionization–time of flight/mass spectrometry
GC-MS	Gas chromatography–mass spectrometry
HPLC	High-performance liquid chromatography
IAA	Indole-3-acetic acid
IBA	Indole 3-butyric acid
iP	Isopentenyladenine
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC-MS	Liquid chromatography–mass spectrometry
MeOH	Methanol
MS	Murashige and Skoog
N/C ratio	Nucleocytoplasmic ratio
NEC	Non-embryogenic callus
PCA	Principal Component Analysis
PEMs	Pro-embryogenic masses
SE	Somatic embryogenesis
TCA	Tricarboxylic acid
UPLC-ESI-MS/MS	Ultra-performance liquid chromatography–electrospray tandem mass spectrometry
W/L ratio	Width-to-length ratio
Z	Zeatin (<i>cis</i> - and <i>trans</i> -)
ZR	Zeatin riboside (<i>cis</i> - and <i>trans</i> -)

References

1. Verdeil, J.-L.; Alemanno, L.; Niemenak, N.; Tranbarger, T.J. Pluripotent versus totipotent plant stem cells: Dependence versus autonomy? *Trends Plant Sci.* **2007**, *12*, 245–252. [[CrossRef](#)]

2. Ikeuchi, M.; Iwase, A.; Rymen, B.; Harashima, H.; Shibata, M.; Ohnuma, M.; Breuer, C.; Morao, A.K.; de Lucas, M.; De Veylder, L.; et al. PRC2 represses dedifferentiation of mature somatic cells in *Arabidopsis*. *Nat. Plants* **2015**, *1*, 15089. [[CrossRef](#)]
3. Zeng, F.; Zhang, X.; Jin, S.; Cheng, L.; Liang, S.; Hu, L.; Guo, X.; Nie, Y.; Cao, J. Chromatin reorganization and endogenous auxin/cytokinin dynamic activity during somatic embryogenesis of cultured cotton cell. *Plant Cell Tissue Organ Cult.* **2007**, *90*, 63–70. [[CrossRef](#)]
4. Reinert, J. Morphogenese und ihre Kontrolle an Gewebekulturen aus Carotten. *Naturwissenschaften* **1958**, *45*, 344–345. [[CrossRef](#)]
5. Steward, F.C.; Mapes, M.O.; Mears, K. Growth and Organized Development of Cultured Cells. II. Organization in Cultures Grown from Freely Suspended Cells. *Am. J. Bot.* **1958**, *45*, 705–708. [[CrossRef](#)]
6. Yang, X.; Zhang, X.; Yuan, D.; Jin, F.; Zhang, Y.; Xu, J. Transcript profiling reveals complex auxin signalling pathway and transcription regulation involved in dedifferentiation and redifferentiation during somatic embryogenesis in cotton. *BMC Plant Biol.* **2012**, *12*, 110. [[CrossRef](#)]
7. Indoliya, Y.; Tiwari, P.; Chauhan, A.S.; Goel, R.; Shri, M.; Bag, S.K.; Chakrabarty, D. Decoding regulatory landscape of somatic embryogenesis reveals differential regulatory networks between *japonica* and *indica* rice subspecies. *Sci. Rep.* **2016**, *6*, 23050. [[CrossRef](#)]
8. Gaj, M.D.; Zhang, S.; Harada, J.J.; Lemaux, P.G. Leafy cotyledon genes are essential for induction of somatic embryogenesis of *Arabidopsis*. *Planta* **2005**, *222*, 977–988. [[CrossRef](#)]
9. Lelu-Walter, M.-A.; Thompson, D.; Harvengt, L.; Sanchez, L.; Toribio, M.; Pâques, L.E. Somatic embryogenesis in forestry with a focus on Europe: State-of-the-art, benefits, challenges and future direction. *Tree Genet. Genomes* **2013**, *9*, 883–899. [[CrossRef](#)]
10. Etienne, H. Somatic embryogenesis protocol: Coffee (*Coffea arabica* L. and *C. canephora* P.). In *Protocols for Somatic Embryogenesis in Woody Plants*; Jain, S.M., Gupta, P.K., Eds.; Springer: Dordrecht, The Netherlands, 2005; pp. 167–179.
11. Maximova, S.N.; Alemanno, L.; Young, A.; Ferriere, N.; Traore, A.; Guiltinan, M.J. Efficiency, genotypic variability, and cellular origin of primary and secondary somatic embryogenesis of *Theobroma cacao* L. *In Vitro Cell. Dev. Biol. Plant* **2002**, *38*, 252–259. [[CrossRef](#)]
12. Montalbán, I.A.; De Diego, N.; Moncaleán, P. Enhancing initiation and proliferation in radiata pine (*Pinus radiata* D. Don) somatic embryogenesis through seed family screening, zygotic embryo staging and media adjustments. *Acta Physiol. Plant.* **2012**, *34*, 451–460. [[CrossRef](#)]
13. Corredoira, E.; Ballester, A.; Ibarra, M.; Vieitez, A.M. Induction of somatic embryogenesis in explants of shoot cultures established from adult *Eucalyptus globulus* and *E. saligna* × *E. maidenii* trees. *Tree Physiol.* **2015**, *35*, 678–690. [[CrossRef](#)]
14. Corredoira, E.; Valladares, S.; Vieitez, A.M. Morphohistological analysis of the origin and development of somatic embryos from leaves of mature *Quercus robur*. *In Vitro Cell. Dev. Biol. Plant* **2006**, *42*, 525–533. [[CrossRef](#)]
15. Ribas, A.F.; Dechamp, E.; Champion, A.; Bertrand, B.; Combes, M.-C.; Verdeil, J.-L.; Lapeyre, F.; Lashermes, P.; Etienne, H. Agrobacterium-mediated genetic transformation of *Coffea arabica* (L.) is greatly enhanced by using established embryogenic callus cultures. *BMC Plant Biol.* **2011**, *11*, 92. [[CrossRef](#)]
16. Ikeuchi, M.; Ogawa, Y.; Iwase, A.; Sugimoto, K. Plant regeneration: Cellular origins and molecular mechanisms. *Development* **2016**, *143*, 1442–1451. [[CrossRef](#)]
17. Boutilier, K.; Angenent, G.C.; Castan, M.S.; Hui, L. Haploid Embryogenesis. Patent US20160212956A1, 28 July 2016.
18. Yang, X.; Zhang, X. Regulation of Somatic Embryogenesis in Higher Plants. *Crit. Rev. Plant Sci.* **2010**, *29*, 36–57. [[CrossRef](#)]
19. Jayasankar, S.; Bondada, B.R.; Li, Z.; Gray, D.J. Comparative Anatomy and Morphology of *Vitis vinifera* (Vitaceae) Somatic Embryos from Solid- and Liquid-Culture-Derived Proembryogenic Masses. *Am. J. Bot.* **2003**, *90*, 973–979. [[CrossRef](#)]
20. Raghavan, V. Role of 2,4-dichlorophenoxyacetic acid (2,4-D) in somatic embryogenesis on cultured zygotic embryos of *Arabidopsis*: Cell expansion, cell cycling, and morphogenesis during continuous exposure of embryos to 2,4-D. *Am. J. Bot.* **2004**, *91*, 1743–1756. [[CrossRef](#)]

21. Florez, S.L.; Erwin, R.L.; Maximova, S.N.; Guiltinan, M.J.; Curtis, W.R. Enhanced somatic embryogenesis in *Theobroma cacao* using the homologous BABY BOOM transcription factor. *BMC Plant Biol.* **2015**, *15*, 121. [[CrossRef](#)]
22. Campos, N.A.; Panis, B.; Carpentier, S.C. Somatic Embryogenesis in Coffee: The Evolution of Biotechnology and the Integration of Omics Technologies Offer Great Opportunities. *Front. Plant Sci.* **2017**, *8*, 1460. [[CrossRef](#)]
23. Bonga, J.M.; Klimaszewska, K.K.; von Aderkas, P. Recalcitrance in clonal propagation, in particular of conifers. *Plant Cell Tissue Organ Cult.* **2010**, *100*, 241–254. [[CrossRef](#)]
24. Etienne, H.; Breton, D.; Breitler, J.-C.; Bertrand, B.; Déchamp, E.; Awada, R.; Marraccini, P.; Lérant, S.; Alpizar, E.; Campa, C.; et al. Coffee Somatic Embryogenesis: How Did Research, Experience Gained and Innovations Promote the Commercial Propagation of Elite Clones from the Two Cultivated Species? *Front. Plant Sci.* **2018**, *9*, 1630. [[CrossRef](#)]
25. Aguilar-Hernández, V.; Loyola-Vargas, V.M. Advanced Proteomic Approaches to Elucidate Somatic Embryogenesis. *Front. Plant Sci.* **2018**, *9*, 1658. [[CrossRef](#)]
26. Pais, M.S. Somatic Embryogenesis Induction in Woody Species: The Future After OMICS Data Assessment. *Front. Plant Sci.* **2019**, *10*, 240. [[CrossRef](#)]
27. Fiehn, O.; Kopka, J.; Dörmann, P.; Altmann, T.; Trethewey, R.N.; Willmitzer, L. Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **2000**, *18*, 1157–1161. [[CrossRef](#)]
28. Businge, E.; Brackmann, K.; Moritz, T.; Egertsdotter, U. Metabolite profiling reveals clear metabolic changes during somatic embryo development of Norway spruce (*Picea abies*). *Tree Physiol.* **2012**, *32*, 232–244. [[CrossRef](#)]
29. Robinson, A.R.; Dauwe, R.; Ukrainetz, N.K.; Cullis, I.F.; White, R.; Mansfield, S.D. Predicting the regenerative capacity of conifer somatic embryogenic cultures by metabolomics. *Plant Biotechnol. J.* **2009**, *7*, 952–963. [[CrossRef](#)]
30. Dobrowolska, I.; Businge, E.; Abreu, I.N.; Moritz, T.; Egertsdotter, U. Metabolome and transcriptome profiling reveal new insights into somatic embryo germination in Norway spruce (*Picea abies*). *Tree Physiol.* **2017**, *37*, 1752–1766. [[CrossRef](#)]
31. Vondrakova, Z.; Dobrev, P.I.; Pesek, B.; Fischerova, L.; Vagner, M.; Motyka, V. Profiles of Endogenous Phytohormones Over the Course of Norway Spruce Somatic Embryogenesis. *Front. Plant Sci.* **2018**, *9*, 1283. [[CrossRef](#)]
32. Gautier, F.; Label, P.; Eliášová, K.; Leplé, J.-C.; Motyka, V.; Boizot, N.; Vondráková, Z.; Malbeck, J.; Trávníčková, A.; Le Metté, C.; et al. Cytological, Biochemical and Molecular Events of the Embryogenic State in Douglas-fir (*Pseudotsuga menziesii* [Mirb.]). *Front. Plant Sci.* **2019**, *10*, 118. [[CrossRef](#)]
33. Bertrand, B.; Montagnon, C.; Georget, F.; Charmetant, P.; Etienne, H. Création et diffusion de variétés de caféiers Arabica: Quelles innovations variétales? *Cah. Agric.* **2012**, *21*, 77–88.
34. Ducos, J.P.; Labbe, G.; Lambot, C.; Pétiard, V. Pilot scale process for the production of pre-germinated somatic embryos of selected robusta (*Coffea canephora*) clones. *In Vitro Cell. Dev. Biol. Plant* **2007**, *43*, 652–659. [[CrossRef](#)]
35. Etienne, H.; Bertrand, B.; Montagnon, C.; Bodadilla Landey, R.; Dechamp, E.; Jourdan, I.; Alpizar, E.; Malo, E.; Georget, F. Un exemple de transfert de technologie réussi dans le domaine de la micropropagation: La multiplication de *Coffea arabica* par embryogénèse somatique. *Cah. Agric.* **2012**, *21*, 115–124.
36. Ducos, J.P.; Alenton, R.; Reano, J.F.; Kanchanomai, C.; Deshayes, A.; Pétiard, V. Agronomic performance of *Coffea canephora* P. trees derived from large-scale somatic embryo production in liquid medium. *Euphytica* **2003**, *131*, 215–223. [[CrossRef](#)]
37. Landey, R.B.; Cenci, A.; Georget, F.; Bertrand, B.; Camayo, G.; Dechamp, E.; Herrera, J.C.; Santoni, S.; Lashermes, P.; Simpson, J.; et al. High Genetic and Epigenetic Stability in *Coffea arabica* Plants Derived from Embryogenic Suspensions and Secondary Embryogenesis as Revealed by AFLP, MSAP and the Phenotypic Variation Rate. *PLoS ONE* **2013**, *8*, e56372.
38. Georget, F.; Courtel, P.; Garcia, E.M.; Hidalgo, M.; Alpizar, E.; Breitler, J.-C.; Bertrand, B.; Etienne, H. Somatic embryogenesis-derived coffee plantlets can be efficiently propagated by horticultural rooted mini-cuttings: A boost for somatic embryogenesis. *Sci. Hortic.* **2017**, *216*, 177–185. [[CrossRef](#)]

39. Verdeil, J.-L.; Hocher, V.; Huet, C.; Grosdemange, F.; Escoute, J.; Ferrière, N.; Nicole, M. Ultrastructural changes in coconut calli associated with the acquisition of embryogenic competence. *Ann. Bot.* **2001**, *88*, 9–18. [[CrossRef](#)]
40. Talamond, P.; Verdeil, J.-L.; Conéjéro, G. Secondary metabolite localization by autofluorescence in living plant cells. *Molecules* **2015**, *20*, 5024–5037. [[CrossRef](#)]
41. Díaz-Sala, C. Molecular Dissection of the Regenerative Capacity of Forest Tree Species: Special Focus on Conifers. *Front. Plant Sci.* **2019**, *9*, 1943. [[CrossRef](#)]
42. Sugimoto, K.; Xu, L.; Paszkowski, U.; Hayashi, M. Multifaceted Cellular Reprogramming at the Crossroads Between Plant Development and Biotic Interactions. *Plant Cell Physiol.* **2018**, *59*, 651–655. [[CrossRef](#)]
43. Tao, L.; Zhao, Y.; Wu, Y.; Wang, Q.; Yuan, H.; Zhao, L.; Guo, W.; You, X. Transcriptome profiling and digital gene expression by deep sequencing in early somatic embryogenesis of endangered medicinal *Eleutherococcus senticosus* Maxim. *Gene* **2016**, *578*, 17–24. [[CrossRef](#)] [[PubMed](#)]
44. Trontin, J.-F.; Klimaszewska, K.; Morel, A.; Hargreaves, C.; Lelu-Walter, M.-A. Molecular Aspects of Conifer Zygotic and Somatic Embryo Development: A Review of Genome-Wide Approaches and Recent Insights. *Methods Mol. Biol.* **2016**, *1359*, 167–207. [[PubMed](#)]
45. Nowak, K.; Wójcikowska, B.; Gaj, M.D. ERF022 impacts the induction of somatic embryogenesis in *Arabidopsis* through the ethylene-related pathway. *Planta* **2015**, *241*, 967–985. [[CrossRef](#)] [[PubMed](#)]
46. Zavattieri, M.A.; Frederico, A.M.; Lima, M.; Sabino, R.; Arnholdt-Schmitt, B. Induction of somatic embryogenesis as an example of stress-related plant reactions. *Electro. J. Biotech.* **2010**, *13*, 12–13. [[CrossRef](#)]
47. Nic-Can, G.I.; Galaz-Ávalos, R.M.; De-la-Peña, C.; Alcazar-Magaña, A.; Wrobel, K.; Loyola-Vargas, V.M. Somatic Embryogenesis: Identified Factors that Lead to Embryogenic Repression. A Case of Species of the Same Genus. *PLoS ONE* **2015**, *10*, e0126414. [[CrossRef](#)] [[PubMed](#)]
48. Gomez-Garay, A.; Lopez, J.A.; Camafeita, E.; Bueno, M.A.; Pintos, B. Proteomic perspective of *Quercus suber* somatic embryogenesis. *J. Proteom.* **2013**, *93*, 314–325. [[CrossRef](#)] [[PubMed](#)]
49. De Carvalho Silva, R.; Carmo, L.S.T.; Luis, Z.G.; Silva, L.P.; Scherwinski-Pereira, J.E.; Mehta, A. Proteomic identification of differentially expressed proteins during the acquisition of somatic embryogenesis in oil palm (*Elaeis guineensis* Jacq.). *J. Proteom.* **2014**, *104*, 112–127. [[CrossRef](#)]
50. Xu, C.; Cao, H.; Xu, E.; Zhang, S.; Hu, Y. Genome-Wide Identification of *Arabidopsis* LBD29 Target Genes Reveals the Molecular Events behind Auxin-Induced Cell Reprogramming during Callus Formation. *Plant Cell Physiol.* **2018**, *59*, 744–755. [[CrossRef](#)]
51. Gatica-Arias, A.M.; Arrieta-Espinoza, G.; Espinoza Esquivel, A.M. Plant regeneration via indirect somatic embryogenesis and optimisation of genetic transformation in *Coffea arabica* L. cvs. Caturra and Catuaí. *Electron. J. Biotechnol.* **2008**, *11*, 1–12. [[CrossRef](#)]
52. Pádua, M.S.; Paiva, L.V.; Silva, L.C.D.; Livramento, K.G.D.; Alves, E.; Castro, A.H.F. Morphological characteristics and cell viability of coffee plants calli. *Ciênc. Rural* **2014**, *44*, 660–665. [[CrossRef](#)]
53. Silva, A.T.; Barduche, D.; do Livramento, K.G.; Ligterink, W.; Paiva, L.V. Characterization of a Putative Serk-Like Ortholog in Embryogenic Cell Suspension Cultures of *Coffea arabica* L. *Plant Mol. Biol. Rep.* **2014**, *32*, 176–184. [[CrossRef](#)]
54. Neves, N.; Segura-Nieto, M.; Blanco, M.A.; Sánchez, M.; González, A.; González, J.L.; Castillo, R. Biochemical characterization of embryogenic and non-embryogenic calluses of sugarcane. *In Vitro Cell. Dev. Biol. Plant* **2003**, *39*, 343–345. [[CrossRef](#)]
55. Jeyaseelan, M.; Rao, M.V. Biochemical studies of embryogenic and non-embryogenic callus of *Cardiospermum halicacabum* L. *Indian J. Exp. Biol.* **2005**, *43*, 555–560.
56. Vestman, D.; Larsson, E.; Uddenberg, D.; Cairney, J.; Clapham, D.; Sundberg, E.; von Arnold, S. Important processes during differentiation and early development of somatic embryos of Norway spruce as revealed by changes in global gene expression. *BMC Proc.* **2011**, *5*, P78. [[CrossRef](#)]
57. Guillou, C.; Fillodeau, A.; Brulard, E.; Breton, D.; De Faria Maraschin, S.; Verdier, D.; Simon, M.; Ducos, J.-P. Indirect somatic embryogenesis of *Theobroma cacao* L. in liquid medium and improvement of embryo-to-plantlet conversion rate. *In Vitro Cell. Dev. Biol. Plant* **2018**, *54*, 377–391. [[CrossRef](#)] [[PubMed](#)]
58. Liang, X.; Zhang, L.; Natarajan, S.K.; Becker, D.F. Proline mechanisms of stress survival. *Antioxid. Redox Signal.* **2013**, *19*, 998–1011. [[CrossRef](#)] [[PubMed](#)]
59. Magnani, E.; Jiménez-Gómez, J.M.; Soubigou-Taconnat, L.; Lepiniec, L.; Fiume, E. Profiling the onset of somatic embryogenesis in *Arabidopsis*. *BMC Genom.* **2017**, *18*, 998. [[CrossRef](#)]

60. Von Aderkas, P.; Lelu, M.-A.; Label, P. Plant growth regulator levels during maturation of larch somatic embryos. *Plant Physiol. Biochem.* **2001**, *39*, 495–502. [[CrossRef](#)]
61. Su, Y.-H.; Liu, Y.-B.; Zhang, X.-S. Auxin–Cytokinin Interaction Regulates Meristem Development. *Mol. Plant* **2011**, *4*, 616–625. [[CrossRef](#)]
62. White, P.R. Potentially Unlimited Growth of Excised Plant Callus in an Artificial Nutrient. *Am. J. Bot.* **1939**, *26*, 59–64. [[CrossRef](#)]
63. Skoog, F.; Miller, C.O. Chemical regulation of growth and organ formation in plant tissues cultured in vitro. *Symp. Soc. Exp. Biol.* **1957**, *11*, 118–130. [[PubMed](#)]
64. Bourgaud, F.; Gravot, A.; Milesi, S.; Gontier, E. Production of plant secondary metabolites: A historical perspective. *Plant Sci.* **2001**, *161*, 839–851. [[CrossRef](#)]
65. Guan, Y.; Li, S.-G.; Fan, X.-F.; Su, Z.-H. Application of Somatic Embryogenesis in Woody Plants. *Front. Plant Sci.* **2016**, *7*, 938. [[CrossRef](#)] [[PubMed](#)]
66. Ladygin, V.G.; Bondarev, N.I.; Semenova, G.A.; Smolov, A.A.; Reshetnyak, O.V.; Nosov, A.M. Chloroplast ultrastructure, photosynthetic apparatus activities and production of steviol glycosides in *Stevia rebaudiana* in vivo and in vitro. *Biol. Plant.* **2008**, *52*, 9–16. [[CrossRef](#)]
67. Ali, A.M.A.; El-Nour, M.E.M.; Yagi, S.M. Total phenolic and flavonoid contents and antioxidant activity of ginger (*Zingiber officinale* Rosc.) rhizome, callus and callus treated with some elicitors. *J. Genet. Eng. Biotechnol.* **2018**, *16*, 677–682. [[CrossRef](#)] [[PubMed](#)]
68. Ge, F.; Hu, H.; Huang, X.; Zhang, Y.; Wang, Y.; Li, Z.; Zou, C.; Peng, H.; Li, L.; Gao, S.; et al. Metabolomic and Proteomic Analysis of Maize Embryonic Callus induced from immature embryo. *Sci. Rep.* **2017**, *7*, 1004. [[CrossRef](#)]
69. Svobodová, H.; Albrechtová, J.; Kumstýřová, L.; Lipavská, H.; Vágner, M.; Vondráková, Z. Somatic embryogenesis in Norway spruce: Anatomical study of embryo development and influence of polyethylene glycol on maturation process. *Plant Physiol. Bioch.* **1999**, *37*, 209–221. [[CrossRef](#)]
70. Fischerova, L.; Fischer, L.; Vondrakova, Z.; Vagner, M. Expression of the gene encoding transcription factor PaVP1 differs in *Picea abies* embryogenic lines depending on their ability to develop somatic embryos. *Plant Cell Rep.* **2008**, *27*, 435–441. [[CrossRef](#)]
71. Iraqi, D.; Tremblay, F.M. Analysis of carbohydrate metabolism enzymes and cellular contents of sugars and proteins during spruce somatic embryogenesis suggests a regulatory role of exogenous sucrose in embryo development. *J. Exp. Bot.* **2001**, *52*, 2301–2311. [[CrossRef](#)]
72. Dodeman, V.L.; Ducreux, G.; Kreis, M. Zygotic embryogenesis versus somatic embryogenesis. *J. Exp. Bot.* **1997**, *48*, 1493–1509.
73. Pedroso, M.C.; Primikiris, N.; Roubelakis-Angelakis, K.A.; Pais, M.S. Free and conjugated polyamines in embryogenic and non-embryogenic leaf regions of camellia leaves before and during direct somatic embryogenesis. *Physiol. Plant.* **1997**, *101*, 213–219. [[CrossRef](#)]
74. Grace, S.C.; Logan, B.A. Energy dissipation and radical scavenging by the plant phenylpropanoid pathway. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2000**, *355*, 1499–1510. [[CrossRef](#)]
75. Gill, S.S.; Tuteja, N. Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant Physiol. Biochem.* **2010**, *48*, 909–930. [[CrossRef](#)] [[PubMed](#)]
76. Ikeda-Iwai, M.; Umehara, M.; Satoh, S.; Kamada, H. Stress-induced somatic embryogenesis in vegetative tissues of *Arabidopsis thaliana*. *Plant J.* **2003**, *34*, 107–114. [[CrossRef](#)] [[PubMed](#)]
77. Dethloff, F.; Erban, A.; Orf, I.; Alpers, J.; Fehrle, I.; Beine-Golovchuk, O.; Schmidt, S.; Schwachtje, J.; Kopka, J. Profiling methods to identify cold-regulated primary metabolites using gas chromatography coupled to mass spectrometry. *Methods Mol. Biol.* **2014**, *1166*, 171–197.
78. Erban, A.; Schauer, N.; Fernie, A.R.; Kopka, J. Nonsupervised construction and application of mass spectral and retention time index libraries from time-of-flight gas chromatography-mass spectrometry metabolite profiles. *Methods Mol. Biol.* **2007**, *358*, 19–38. [[PubMed](#)]
79. Luedemann, A.; Strassburg, K.; Erban, A.; Kopka, J. TagFinder for the quantitative analysis of gas chromatography—Mass spectrometry (GC-MS)-based metabolite profiling experiments. *Bioinformatics* **2008**, *24*, 732–737. [[CrossRef](#)]
80. Kopka, J.; Schauer, N.; Krueger, S.; Birkemeyer, C.; Usadel, B.; Bergmüller, E.; Dörmann, P.; Weckwerth, W.; Gibon, Y.; Stitt, M.; et al. GMD@CSB.DB: The Golm Metabolome Database. *Bioinformatics* **2005**, *21*, 1635–1638. [[CrossRef](#)]

81. Strehmel, N.; Hummel, J.; Erban, A.; Strassburg, K.; Kopka, J. Retention index thresholds for compound matching in GC–MS metabolite profiling. *J. Chromatogr. B* **2008**, *871*, 182–190. [[CrossRef](#)]
82. Campa, C.; Urban, L.; Mondolot, L.; Fabre, D.; Roques, S.; Lizzi, Y.; Aarrouf, J.; Doulebeau, S.; Breitler, J.-C.; Letrez, C.; et al. Juvenile Coffee Leaves Acclimated to Low Light Are Unable to Cope with a Moderate Light Increase. *Front. Plant Sci.* **2017**, *8*, 1126. [[CrossRef](#)]
83. N.R.C. Canada. Plant Hormone Profiling. Available online: <https://nrc.canada.ca/en/research-development/products-services/technical-advisory-services/plant-hormone-profiling> (accessed on 20 June 2019).
84. Lulsdorf, M.M.; Yuan, H.Y.; Slater, S.M.H.; Vandenberg, A.; Han, X.; Zaharia, L.I.; Abrams, S.R. Endogenous hormone profiles during early seed development of *C. arietinum* and *C. anatolicum*. *Plant Growth Regul.* **2013**, *71*, 191–198. [[CrossRef](#)]
85. Chauvaux, N.; Van Dongen, W.; Esmans, E.L.; Van Onckelen, H.A. Quantitative analysis of 1-aminocyclopropane-1-carboxylic acid by liquid chromatography coupled to electrospray tandem mass spectrometry. *J. Chromatogr. A* **1997**, *775*, 143–150. [[CrossRef](#)]
86. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
87. Cheadle, C.; Vawter, M.P.; Freed, W.J.; Becker, K.G. Analysis of Microarray Data Using Z Score Transformation. *J. Mol. Diagn.* **2003**, *5*, 73–81. [[CrossRef](#)]
88. Gu, Z.; Eils, R.; Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **2016**, *32*, 2847–2849. [[CrossRef](#)] [[PubMed](#)]
89. CRAN—Package Pvclust. Available online: <https://cran.r-project.org/web/packages/pvclust/index.html> (accessed on 20 June 2019).
90. Suzuki, R.; Shimodaira, H. Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **2006**, *22*, 1540–1542. [[CrossRef](#)] [[PubMed](#)]
91. MetaboAnalyst. Available online: <https://www.metaboanalyst.ca/MetaboAnalyst/faces/upload/EnrichUploadView.xhtml> (accessed on 19 August 2019).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Chapter IV

Global transcriptome profiling reveals differential regulatory, metabolic and hormonal networks during somatic embryogenesis in *Coffea arabica*

Paper to be submitted in *BMC Genomics*

Background

In this chapter, we will present a complementary work to the previous one on metabolomics, this one aiming at describing the transcriptomic events occurring during coffee somatic embryogenesis. Transcriptomics is probably the easiest way to study changes on a genome-wide scale. In order to guarantee total complementarity in and to extract as much information as possible that allows a rational optimization of SE, this study was carried out on the same experiments as the previous one, i.e. on the basis of the same sampling.

Similarly to the metabolomics approach, we have the following objectives for this study (we will develop adapted statistical pipelines to reach them): identify i) the developmental phases of coffee SE, ii) the key developmental switches, iii) transcriptomic fingerprints of each phase, if possible of each developmental stage, iv) a list of key genes (candidate ‘markers’ to pilot SE optimization) showing differential expression during the key developmental switches (hence associated with cell fate) or associated with the acquisition of embryogenic capacity, v) establish some gene co-expression networks schematizing the gene and metabolic regulations characterizing each phase of development.

This study was allowed thanks to a granted access to one of the most well-annotated Arabica reference genome (ACGC, unpublished results) combined with a sequencing of the highest quality (Illumina 2500, 125 bp x 2 (paired-end)) realized in the Nestlé Research facilities in Lausanne (Patrick Descombes team). RNA-seq generated a total of 90 million high quality reads per sample, approximately 74% of which were uniquely mapped to the Arabica genome.

We expect that this transcriptomic approach, combined with the metabolic one, will represent the first most-detailed blueprint of the mechanisms regulating SE.

1 To be submitted to BMC Genomics

2

3 **Global transcriptome profiling reveals differential regulatory,** 4 **metabolic and hormonal networks during somatic embryogenesis in** 5 ***Coffea arabica***

6

7 Rayan Awada^{1,2,3*}, Maud Lepelley¹, Aline Charpagne^{4,5}, Victoria Berry¹, Claudine Campa^{3,6}, Jean-
8 Christophe Breitler^{2,3}, Sophie Lérans^{2,3}, David Breton¹, Dominique Crouzillat¹, Benoît Bertrand^{2,3}, and
9 Hervé Etienne^{2,3}

10 ¹ Nestlé Research—Plant Science Research Unit, 101 avenue Gustave Eiffel, F-37097 Tours Cedex 2,
11 France

12 ² CIRAD, UMR IPME, F-34398 Montpellier, France

13 ³ UMR IPME, University of Montpellier, CIRAD, IRD, F-34398 Montpellier, France

14 ⁴ Nestlé Research, Vers-chez-les-Blanc, 1000 Lausanne 26, Switzerland

15 ⁵ Sophia Genetics, 9 chemin des mines, 1202 Genève, Switzerland

16 ⁶ IRD, UMR IPME, F-34398 Montpellier, France

17

18 * Correspondence: rayan.awada@cirad.fr

19

20 **Abstract**

21

22 **Background**

23 Somatic embryogenesis (SE) is one of the most promising processes for large-scale dissemination of elite varieties.
24 However, for many plant species, optimizing SE protocols still relies on a trial-and-error approach. We report here
25 the first global scale transcriptome profiling performed on all fundamental industrial stages of SE in coffee to unravel
26 mechanisms regulating cell fate and totipotency.

27

28 **Results**

29 RNA-seq of 48 samples (12 developmental stages x 4 biological replicates) generated a total of 90 million high quality
30 reads per sample, approximately 74% of which were uniquely mapped to Arabica genome. Firstly, a statistical
31 approach on transcript data clearly grouped the developmental stages of SE into seven important phases (Leaf,
32 Dedifferentiation, Primary callus, Embryogenic callus, Embryogenic cell clusters, Redifferentiation and Embryo)
33 leading to the identification of six key developmental switches, which are strategic for the overall biological efficiency
34 of embryo regeneration. Differential gene expression and functional analysis showed that genes encoding transcription
35 factors, stress-related genes, metabolism-related genes and hormone signaling-related genes were significantly
36 enriched. Secondly, environmental drivers classically used to master SE, i.e. light, growth regulators and cell density,
37 were clearly perceived at the molecular level at different developmental stages. Thirdly, expression profiles of auxin-
38 related genes, transcription factor-related genes and secondary metabolism-related genes were analyzed during SE.
39 Gene co-expression networks were also inferred. Auxin-related genes were up-regulated during dedifferentiation and
40 redifferentiation while transcription factor-related genes were switched-on from the embryogenic callus and forwards.
41 Secondary metabolism-related genes were switched-off during dedifferentiation and switched back on at the onset of
42 redifferentiation. Secondary metabolites and endogenous IAA content were highly related to their respective gene
43 expression. Lastly, comparing Arabica embryogenic and non-embryogenic cell transcriptomes enabled the
44 identification of biological processes involved in the acquisition of embryogenic capacity.

45

46 **Conclusion**

47 The present analysis showed that transcript fingerprints are discriminant signatures of cell fate and are under direct
48 influence of environmental drivers. A total of 23 molecular markers were successfully identified overall the 12
49 developmental phases and should now be tested in many plant species for optimizing SE protocols in a rational way.

50

51 **Keywords:** cell fate, coffee, molecular markers, molecular networks, somatic embryogenesis, totipotency,
52 transcriptomics.

53

54 **Background**

55

56 Somatic embryogenesis (SE) is a developmental process where a plant somatic cell can dedifferentiate into
57 a totipotent embryogenic stem cell that has the ability to redifferentiate and give rise to an embryo under
58 appropriate culture conditions [1,2,3]. This regenerated embryo can further develop into a whole plant
59 having the same characteristics as the mother plant. Since its first description in carrot by Reinert [4] and
60 Steward et al. [5], this process has been reported in a wide range of plant species, both annual [6,7,8] and
61 perennial [9,10,11,12,13]. SE has demonstrated significant benefits when applied to forest tree species,
62 bringing important advantages including clonal mass propagation, cryopreservation of valuable germplasm
63 and genetic transformation [14,15]. SE is particularly important for plants that have a long life cycle (woody
64 species) and are difficult to propagate by conventional horticultural methods such as cuttings [16].

65 The ability of a somatic cell to undergo embryogenesis *in vitro* is both an inherent and an acquired
66 characteristic that requires the right combination between an explant type and the culture environment [17].
67 The most efficient treatments used to induce SE are diverse and range from application of exogenous growth
68 regulators to abiotic stress. Under the appropriate conditions, the explant produces differentiated embryos,
69 either directly from the explant or indirectly from callus [18]. The direct SE is often described as a low
70 yield method and the indirect SE as a high yield method [10]. The morphological and cellular changes that
71 occur during *in vitro* embryogenesis have been well-described [1,19,20]. Briefly, nine developmental stages
72 have been characterized in the indirect SE of dicots i.e. explant, primary then embryogenic callus,
73 embryogenic cell clusters, pro-embryogenic masses, globular-shaped embryos, heart-shaped embryos,
74 torpedo-shaped embryos and cotyledonary embryos, before development into a whole plant [1,19,20].
75 Unlike the detailed knowledge about morphological and histological events, little is known about the
76 underlying molecular mechanisms involved in the successful transition between the different
77 developmental stages occurring during SE and in the expression of totipotency.

78 Research on SE remains mainly empirical, characterized by a low-throughput trial-and-error approach. A
79 set of drawbacks have been reported especially a strong genotypic effect, a difficulty in obtaining
80 embryogenic callus, a weak quality of regenerated embryos, and more generally a lack of efficiency of
81 certain steps [9,21,22] leading to hitherto prohibitive production costs and an overall slow technical
82 progress. Although SE has already been widely described in a number of woody species [10,11,12,13]
83 propagating adult woody plants remains a laborious, intensive and tricky operation. Many authors reported
84 that the lack of knowledge on the mechanisms underlying the reprogramming of somatic cells represents
85 the greatest limitation to the improvement of the SE processes [23,24].

86 Many authors consider that applying cutting-edge -omics technologies (genomics, transcriptomics,
87 proteomics, epigenomics, metabolomics and phenomics) on SE can have a tremendous impact on the
88 knowledge of the underlying molecular mechanisms [24,25,26]. Transcriptomic data can provide a wealth
89 of information in the description and elucidation of physiological responses to environmental conditions in
90 plants [27]. However, to date, very few authors have carried out an RNA-seq approach applied to SE.
91 Authors have mainly studied the early events of SE induction in annual plants such as Arabidopsis [28],
92 cotton [29], rice [30], maize [31]. In woody plants, Gautier et al. [32] carried out a comparison between
93 embryogenic callus (EC) and non-embryogenic callus (NEC) in Douglas-fir (*P. menziesii*), and more
94 recently, Chen et al. [33] compared EC, NEC and globular embryos (GE) in *Dimocarpus longan*.

95 Coffee is one of the world's favorite beverages. It has a great economic impact in many producing countries,
96 especially in South America and Vietnam [34]. Today, SE applied to coffee is one of the most advanced
97 technologies in plant mass vegetative propagation [24]. Thirty years of research on coffee SE has led to a
98 successful large-scale dissemination of *Coffea arabica* F1 hybrids and *C. canephora* cv. Robusta clones
99 [22]. Even though good biological efficiencies characterize the processes established in the two cultivated
100 species, with low genotypic effects [35,36] and controlled somaclonal variation [37,38], production cost
101 associated to SE remain high and still cannot compete with production costs of seedlings or traditional
102 cuttings [39].

103 Indeed, as for other species, coffee SE research remains mainly empirical, also characterized by an overall
104 slow technical progress. For example, ten years of laborious empirical research were needed to develop
105 culture conditions for the mass redifferentiation of embryogenic clusters into somatic embryos in liquid
106 nutrient media [24]. Some SE developmental switches are considered as real black boxes due to the lack of
107 knowledge about the involved cellular and molecular events. While a scale-up is needed to meet increasing
108 market demand estimated at 50–100 million coffee vitroplants per year current overall production cannot
109 meet this huge demand [24]. Nevertheless, a rational optimization based on a detailed knowledge of
110 associated molecular mechanisms seems possible.

111 In this paper, we describe how we took advantage of the latest -omics technologies (new Arabica genome
112 sequence, transcriptomics) combined with a reliable, synchronized and efficient large-scale Arabica SE
113 propagation process to draw global transcript profiles of 12 key developmental stages. A robust statistical
114 method based on transcript modulations was used to identify the main developmental phase switches and
115 the main biological processes involved in these phase switches. Differentially expressed genes (DEGs)
116 involved in three of the identified biological processes were further studied, i.e. genes encoding
117 transcription factors, genes related to phytohormone biosynthesis and response, and genes encoding
118 secondary metabolites. Co-expression networks between these gene families were also revealed. Lastly,
119 comparing Arabica embryogenic and non-embryogenic calli enabled the identification of biological
120 processes implicated in the acquisition of the embryogenic capacity.

121

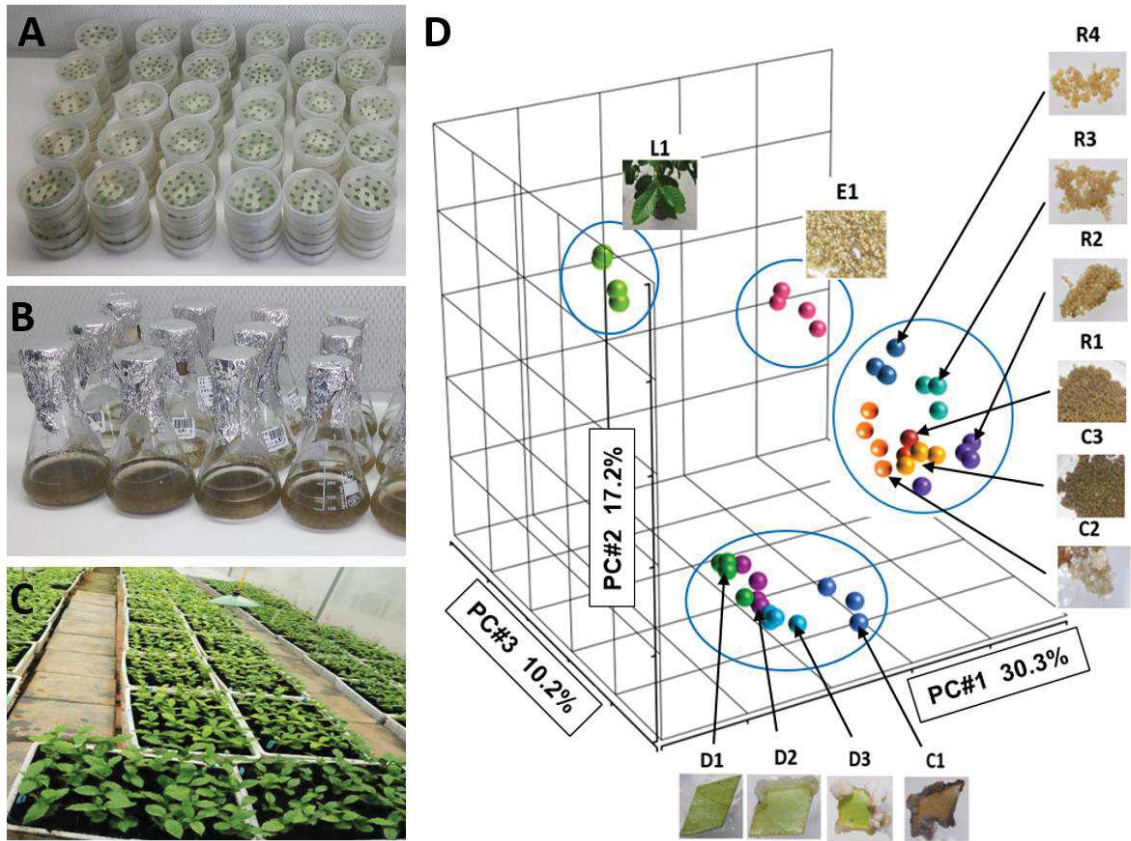
122 **Results**

123

124 **A high sampling quality allowed a high read reliability between replicates**

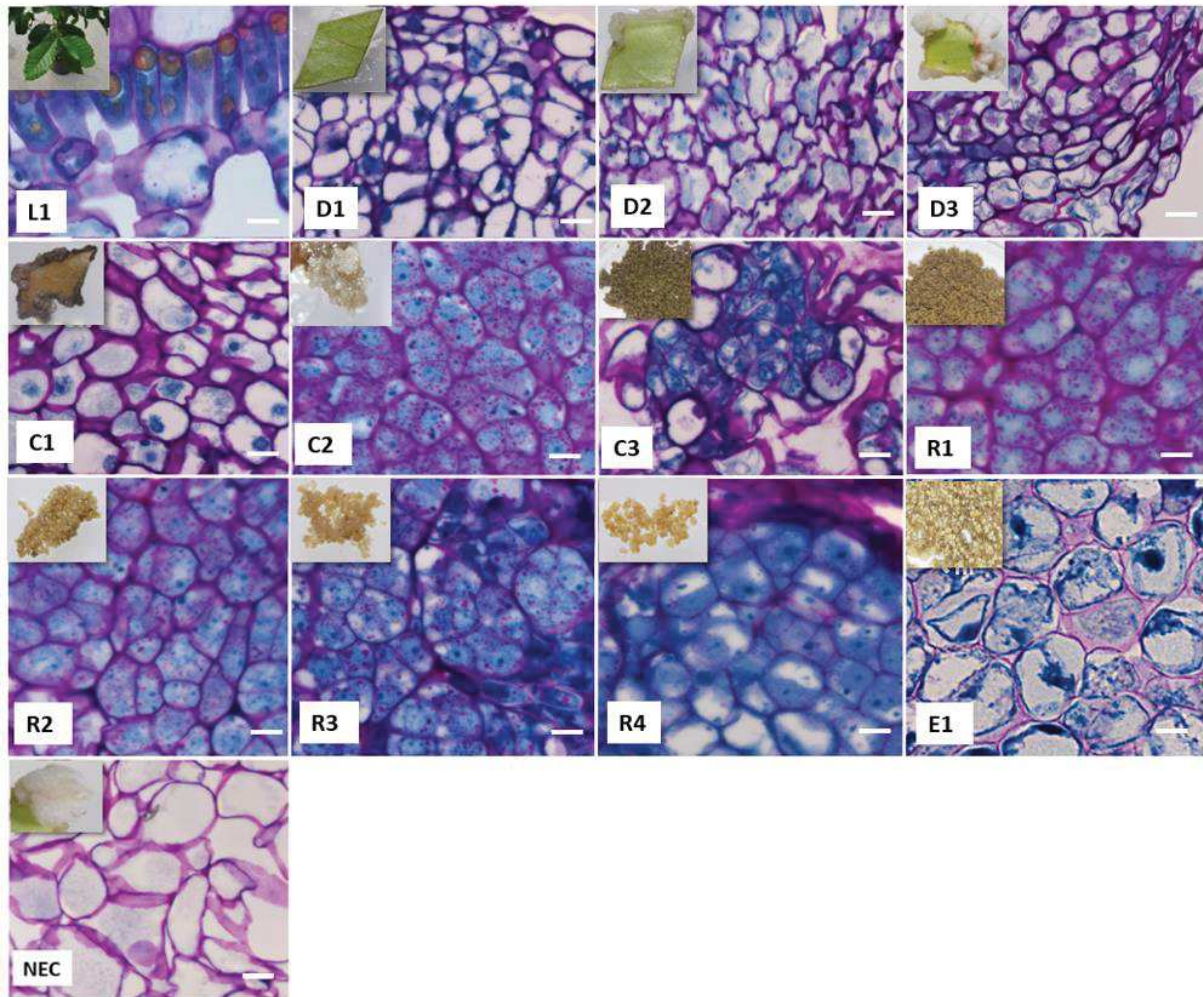
125 The availability of efficient large-scale propagation protocols for coffee SE, currently used at the
126 commercial level, allowed the introduction of more than 1,000 leaf explants on dedifferentiation medium
127 for each of the 4 replicates (Fig. 1A) as well as a total of 20 independent established embryogenic cell lines
128 (Fig. 1B). The quality of sampling was validated by the observation that all obtained cell lines were high-
129 yielding and time-synchronized during embryo regeneration and that torpedo-shaped embryos harvested
130 from all 4 biological replicates were successfully converted into plantlets (Fig. 1C). Twelve developmental
131 stages were selected for RNA-seq analysis and summarized in Fig. 2. A three-dimensional Principal
132 Component Analysis (PCA) was carried out to examine the reliability between the four biological replicates
133 for each developmental stage based on the normalized expression values of all 41,569 genes (Fig. 1D). The
134 three axes of the PCA accounted for 57.7% of total inertia. The 12 developmental stages were successively
135 aligned along the three axes in a counter clockwise form. The PCA yielded 4 main groups that could be
136 easily discriminated: the leaf group (L1), the leaf dedifferentiation group (D1, D2, D3, C1), a group
137 containing the embryogenic callus (C2), its proliferation (C3) and its early redifferentiation (R1, R2, R3,
138 R4), and lastly the embryo group (E1). The PCA also showed the robustness of the biological replicates as
139 the four replicates of each developmental stage globally clustered together. Replicates of the embryogenic

140 callus group were the most distant as this material tends to quickly redifferentiate in pro-embryos when
141 cultivated at low cellular density in liquid medium.



142
143 **Figure 1.** The availability of large-scale protocols for coffee somatic embryogenesis (SE) offered
144 high reliability and development synchronization at each developmental stage as well as a good
145 biological efficiency. (A) More than 1,000 leaf explants per replicate were introduced on
146 dedifferentiation medium for this study. (B) Fifty grams of cell clusters were needed to sample the
147 embryogenic cell cluster stage (C3) and the whole redifferentiation episode. (C) Embryos were
148 successfully converted into plantlets hence validating the quality of all previous sampled stages. (D)
149 RNA-seq was carried on 12 developmental stages and a three-dimensional PCA was drawn based on
150 normalized expression values of all 41,569 genes. This analysis showed the robustness of the
151 biological replicates, i.e. the four replicates of each stage globally clustered together, and
152 discriminated the different Arabica SE stages.

153

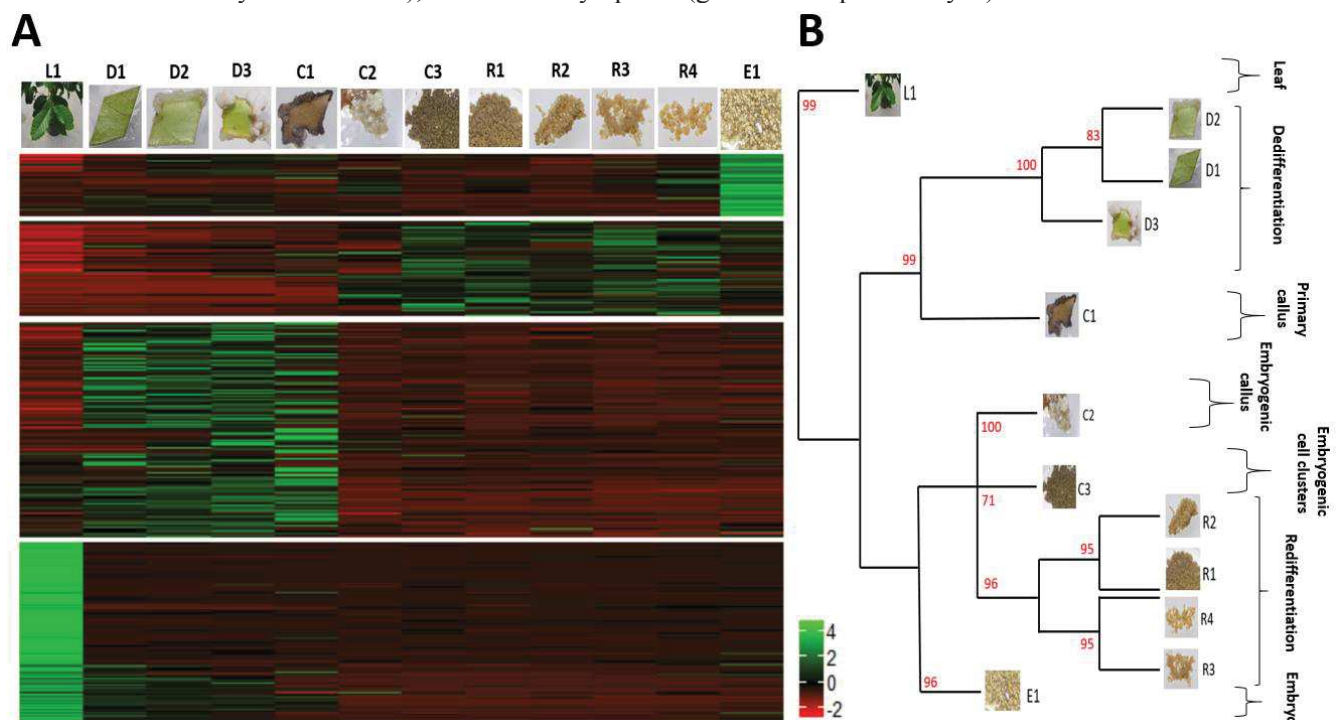


154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171

Figure 2. Characterization of the 12 sampled key developmental stages throughout the *Arabidopsis* somatic embryogenesis (SE) process at both a morphological and cellular level. Developmental stages correspond to: leaves from greenhouse plants i.e. explant (L1), explants during dedifferentiation [1 week (D1), 2 weeks (D2), 5 weeks (D3)], compact primary callus obtained 3 months after induction (C1), embryogenic callus obtained 7 months after induction (C2), established cell clusters obtained after 4 months in liquid proliferation medium (C3), pro-embryogenic masses [1 week in redifferentiation medium after auxin withdrawal (R1), 24 h in redifferentiation medium after reducing cell density (R2), 72 h (R3), 10 days (R4)] and globular embryos (E1) obtained after 3 weeks of culture. An additional stage, the non-embryogenic callus (NEC), was also sampled at the same time as the embryogenic callus (C2) and obtained in the same culture conditions. Sections were treated with 1% periodic acid and stained with Schiff reagent (colors polysaccharides in purple) and Naphthol Blue Black (colors soluble proteins in blue). Histological images were taken in bright field. The scale bar is set to 10 μm . Phenotyping images were taken using an Olympus E-5 digital camera mounted on an Olympus SZX7 stereomicroscope. The stereomicroscope was not used for (L1). Morphological and cellular characterization of the different developmental stages was previously described in Awada et al. [56].

172 **Clustering DEG profiles divides the Arabica SE process into seven developmental phases**

173 A high resolution analysis of differentially expressed genes (DEGs) during SE was achieved using RNA-
 174 seq technology. A total of 10,384 DEGs were obtained over the 12 SE developmental stages. The heatmap
 175 generated from the normalized counts of the total DEGs over the 12 SE developmental stages showed
 176 similar transcript profiles in some of these stages (Fig. 3A). We carried out a hierarchical clustering analysis
 177 to highlight stages that shared similar transcript profiles (Fig. 3B). Strong correlations between profiles
 178 were obtained resulting in seven major nodes: the “Leaf” node including the L1 stage transcript profile was
 179 distinct from other profiles (cluster probability = 99%); the “Dedifferentiation” node where the D1, D2 and
 180 D3 stages clustered together (cluster probability = 100%); the “Primary Callus”, “Embryogenic callus” and
 181 “Embryogenic cell clusters” nodes where the C1, C2 and C3 stages clustered distinctively even though they
 182 are very similar at the morphological level (cluster probability = 99%, 100%, 71% respectively); the
 183 “Redifferentiation” node where the R1, R2, R3 and R4 stages clustered together (cluster probability = 96%);
 184 and the “Embryo” node where the E1 stage transcript profile showed to be different than the
 185 redifferentiation stages (cluster probability = 96%). The obtained clusters revealed a developmental
 186 pattern as the obtained nodes corresponded to successive developmental phases of the SE process of the
 187 SE process, i.e., the leaf phase, the leaf explant dedifferentiation phase, the primary callus phase, the
 188 embryogenic callus phase, the embryogenic cell clusters phase, the redifferentiation phase (embryogenic
 189 cell clusters to embryoid structures), and the embryo phase (globular-shaped embryos).

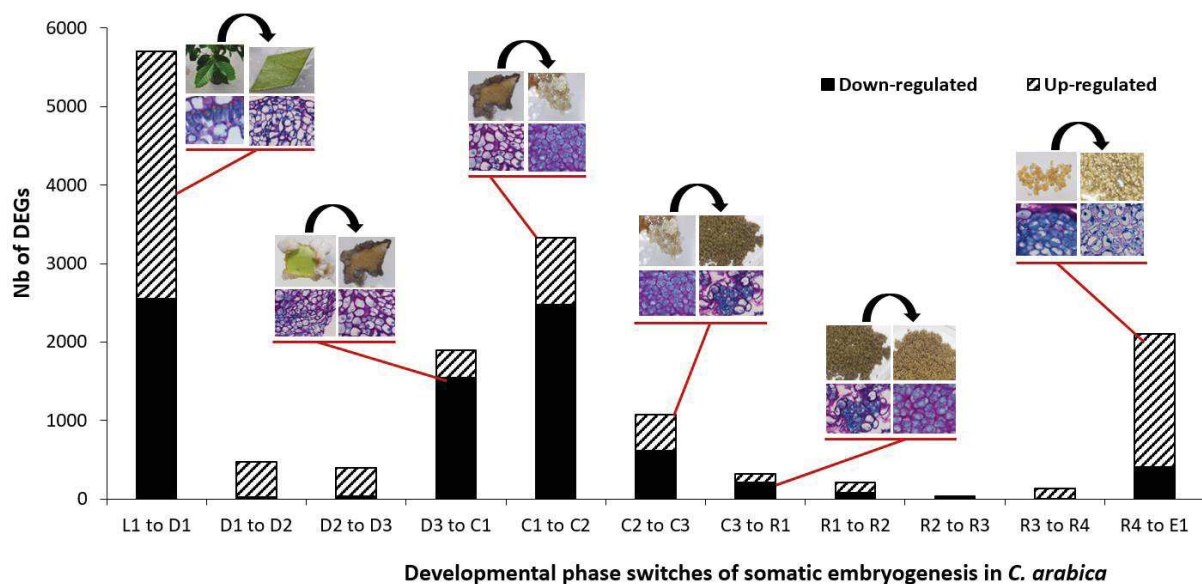


190 **Figure 3.** Profiling of differentially expressed genes during the 12 key developmental stages of the
 191 Arabica coffee somatic embryogenesis (SE) process. (A) Heatmap generated from the normalized
 192 counts of 10,384 DEGs over the 12 developmental stages. Rows correspond to DEGs and columns
 193 to the developmental stages. Normalized counts of each gene were transformed in order to follow a
 194 standard normal distribution. Positive expression values are shown in green and negative values in
 195 red. (B) Hierarchical clustering of the 12 SE developmental stages according to the similarities in
 196 their transcript profiles. The clustering was performed using Pearson’s correlation coefficient.
 197 Cluster probabilities were calculated via a multiscale bootstrap with a total of 1,000 iterations.
 198

199 Clustering yielded 7 major nodes: Leaf, Dedifferentiation, Primary callus, Embryogenic callus,
 200 Embryogenic cell clusters, Redifferentiation and Embryo.

201 **Transcriptional characterisation of six major developmental phase switches during**
 202 **Arabica SE**

203 The number of genes up- or down-regulated at different developmental stages is shown in Fig. 4. The
 204 number of DEGs were particularly higher during the transition from a developmental phase to another,
 205 hence characterizing a developmental phase switch. Our transcript profiling revealed seven developmental
 206 phases (Fig. 3B) and six developmental phase switches (Fig. 4). In order, the switches are (Fig. 4) : Leaf-
 207 to-Dedifferentiation (L1 to D1), Dedifferentiation-to-Primary callus (D3 to C1), Primary callus-to-
 208 Embryogenic callus (C1 to C2), Embryogenic callus-to-Embryogenic cell clusters (C2 to C3), Embryogenic
 209 cell clusters-to-Redifferentiation (C3 to R1) and Redifferentiation to Embryo (R4 to E1). During the first
 210 phase switch (L1 to D1), the highest number of DEGS was found (5,701), 55% of these genes were up-
 211 regulated and 45% were down-regulated. The four next switches, i.e. D3 to C1, C1 to C2, C2 to C3 and C3
 212 to R1, had 1,894; 3,326; 1,074 and 323 DEGs respectively, and at least 75% of the genes were down-
 213 regulated for each of these switches. The last developmental phase switch (R4 to E1) had 2,098 DEGs and
 214 80% of them were up-regulated. It is interesting to note that the passages from D1 to D3 that characterize
 215 early stages of dedifferentiation do not appear to be remarkable switches. This is also the case for R1 to
 216 R4 that characterize redifferentiation and where remarkable switches did not occur.



217 **Figure 4. Distribution of differentially expressed genes among key developmental phase**
 218 **switches of Arabica somatic embryogenesis (SE).** Up-regulated (striped) and down-regulated
 219 (black) gene numbers are given for 12 developmental stages of the Arabica SE process. Red lines
 220 correspond to the identified developmental phase switches: Leaf-to-Dedifferentiation (L1 to D1),
 221 Dedifferentiation-to-Primary callus (D3 to C1), Primary callus-to-Embryogenic callus (C1 to C2),
 222 Embryogenic callus-to-Embryogenic cell clusters (C2 to C3), Embryogenic cell clusters-to-
 223 Redifferentiation (C3 to R1) and Redifferentiation to Embryo (R4 to E1).
 224

225 For the six key developmental phase switch, DEGs were compared against The Arabidopsis Information
 226 Resource database (TAIR, www.arabidopsis.org) using BLASTP with an e-value cut-off of 1×10^{-4} .

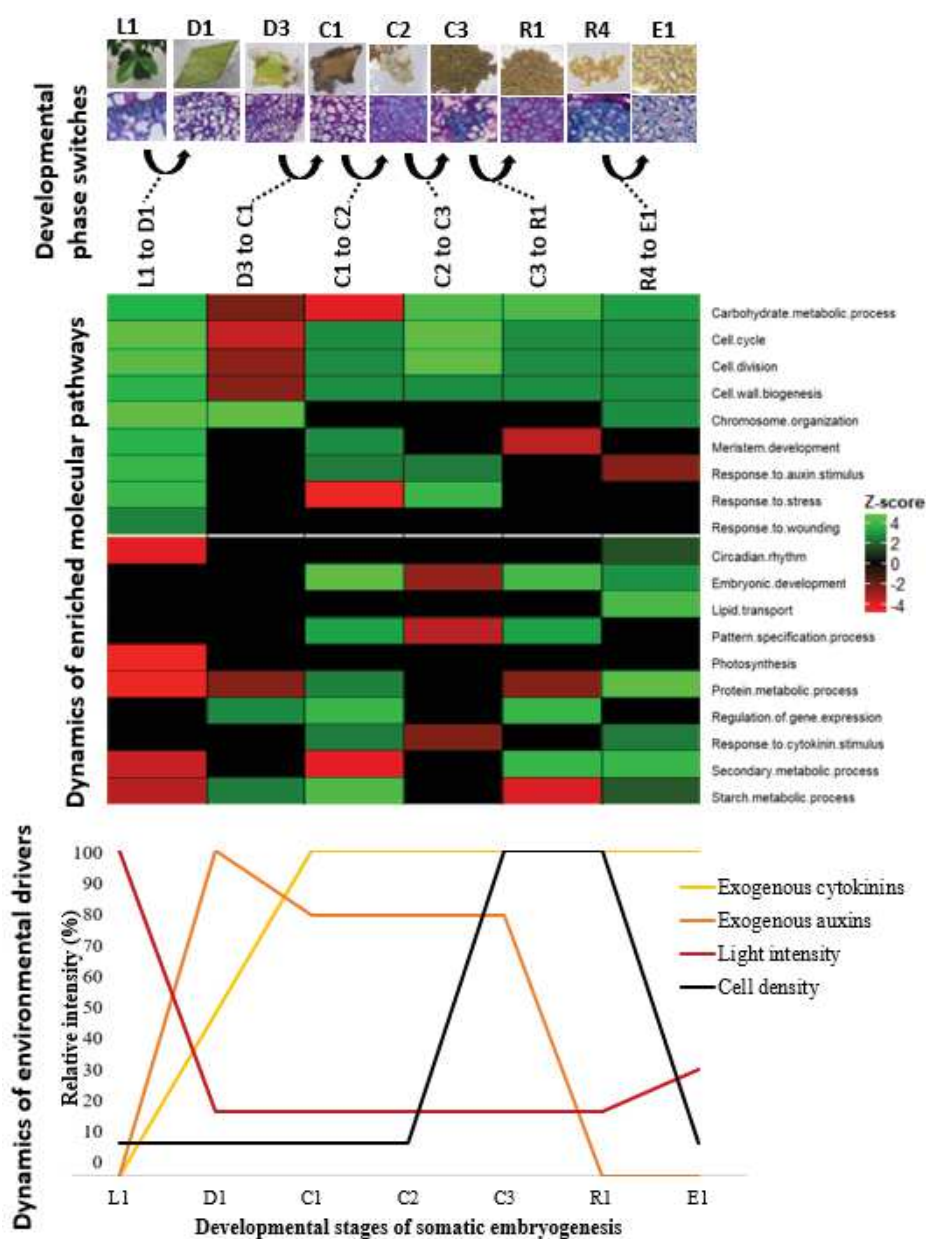
227 Approximately 50% of the Arabica DEGs had an Arabidopsis orthologue since the Arabica species is an
228 allotetraploid species containing two diploid genomes that originated from two different diploid wild
229 ancestor species ($2n = 22$), *C. canephora* and *C. eugenioides* [41]. PAGE tool yielded significant Gene
230 Ontology (GO) terms that could be classified into 19 biological processes (Fig. 5). During the L1 to D1
231 switch corresponding to the transition from leaf cells to dedifferentiated cells, the PAGE tool showed an
232 up-regulation in genes encoding the carbohydrate metabolic process, cell cycle and division, cell wall
233 biogenesis, chromosome organization, meristem development, response to auxin stimulus, stress and
234 wounding. A down-regulation of genes encoding the circadian rhythm, photosynthesis, protein, starch and
235 secondary metabolic processes was also noted. The second phase switch (D3 to C1), corresponding to the
236 formation of primary callus cells, was characterized by a sharp up-regulation of genes encoding
237 chromosome organization, regulation of gene expression and starch metabolism; and a down-regulation of
238 genes encoding carbohydrate metabolic process, cell cycle, division, cell wall biogenesis and protein
239 metabolism. The apparition of embryogenic cells during the C1 to C2 switch was characterized by a sharp
240 up-regulation of genes encoding embryonic development in addition to cell cycle and division, cell wall
241 biogenesis, meristem development, regulation of gene expression, response to auxin and cytokinin, pattern
242 specification, protein and starch metabolism; and a down-regulation of genes encoding the carbohydrate
243 metabolism, response to stress and the secondary metabolic process. The fourth developmental phase switch
244 (C2 to C3) corresponding to the transition of embryogenic cells on solid medium to embryogenic cell
245 clusters in liquid medium was characterized by a sharp up-regulation of genes encoding carbohydrate
246 metabolic process, cell cycle and division, cell wall biogenesis, response to stress and response to auxin
247 stimulus; and a down-regulation of genes encoding embryonic development, pattern specification and
248 response to cytokinin. The formation of proembryogenic masses during the fifth phase switch (C3 to R1)
249 was characterized by a sharp up-regulation of genes encoding the carbohydrate metabolism, secondary
250 metabolism, embryonic development, regulation of gene expression, pattern specification in addition to
251 enhanced cell cycle and division processes; and a down-regulation of genes encoding the meristem
252 development, protein and starch metabolic processes. The last phase switch (R4 to E1) corresponding to
253 the formation of globular embryos was characterized by an up-regulation of genes encoding the response
254 to cytokinin, protein and starch metabolism, secondary metabolism, lipid transport, embryonic
255 development, circadian rhythm in addition to enhanced carbohydrate metabolism, chromosome
256 organization, cell wall biogenesis, cell cycle and division. Genes encoding the response to auxin were
257 down-regulated.

258 **Effects of environmental drivers on the regulation of biological processes during SE**

259 SE is a process involving a number of exogenous artificial factors that are perceived as environmental cues
260 by plant cell or tissue culture. These environmental drivers (light, exogenous growth regulators mainly
261 auxins and cytokins, cell density) were widely reported as factors influencing the success of the SE process
262 [6,9]. In this study, variations in these drivers make it possible to study their influence on gene expression
263 and thereby verify whether they are important drivers. In Fig. 5 we showed that environmental drivers are
264 tightly related to the regulation of genes involved in crucial biological processes. For example, at the
265 beginning of the SE process, light exposure was suppressed through a transition from a 12h/12h photoperiod
266 in the greenhouse (L1) to obscurity conditions in the culture chamber (D1) before being partially restored
267 during globular embryo formation under light conditions (E1). Light intensity variations were directly
268 linked to the variation of expression of genes involved in photosynthesis and circadian rhythm. Exogenous
269 auxins were present at high concentrations in the dedifferentiation medium (D1, C1, C2) as well as in the
270 embryogenic cell cluster proliferation medium (C3) to maintain the embryogenic capacity and were
271 removed from the redifferentiation medium (R1, E1). We observed that genes involved in the response to
272 the auxin stimulus followed the same pattern as they were up-regulated during dedifferentiation (D1),
273 embryogenic cells (C2) and cell clusters (C3) and down-regulated in globular embryos (E1). Similarly,

274 genes involved in the response to cytokinin stimulus are up-regulated in embryogenic cells (C2) and in
 275 embryonic cells (E1) when exogenous cytokinin was present at high concentrations in their respective
 276 media. High cell density also plays a crucial role mainly during the proliferation of embryogenic cell
 277 clusters as it inhibits the embryonic pathway and enhances biomass proliferation. This was clearly translated
 278 in a down-regulation of genes involved in embryonic development and pattern specification processes in
 279 embryogenic cell clusters while an up-regulation of the same genes was noted when cell density was sharply
 280 decreased during redifferentiation. Genes involved in the response to wounding were also up-regulated
 281 during the transition from entire leaves (L1) to dissected leaf squares (D1) needed to build the explant and
 282 induce the dedifferentiation mechanisms.

283

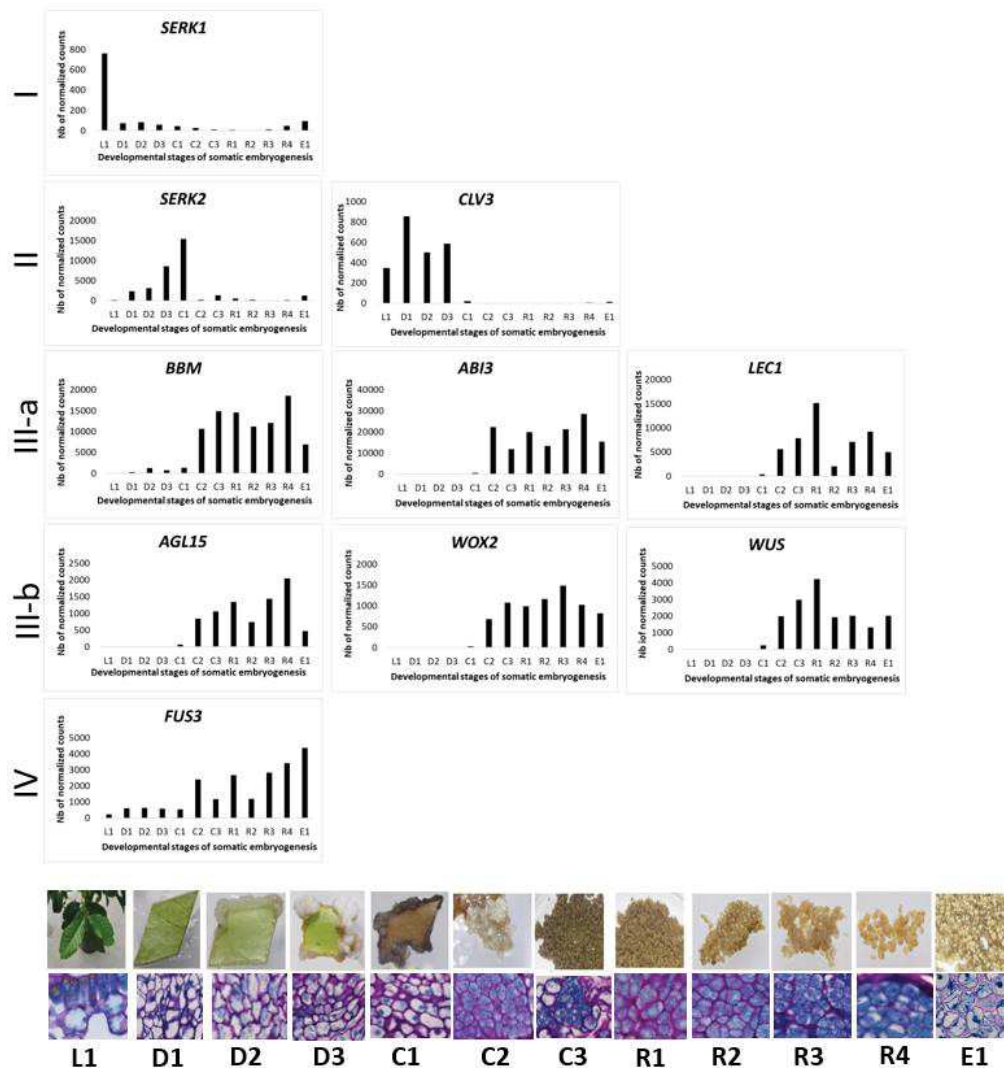


284

285 **Figure 5. Functional categorization of the Arabica differentially expressed genes (DEGs)**
286 **during the six key developmental phase switches based on their Arabidopsis orthologues.**
287 DEGs were first compared against The Arabidopsis Information Resource database
288 (TAIR, www.arabidopsis.org) using BLASTP with an e-value cut-off of 1×10^{-4} . Functional
289 categorization of Arabidopsis orthologues in all 6 developmental phase switches was performed
290 through Parametric Analysis of Gene set Enrichment (PAGE) tool in agriGO v2.0 [42]. PAGE tool
291 yielded significant Gene Ontology (GO) terms that could be classified into 19 biological processes.
292 Rows correspond to biological processes and columns to the developmental phase switches.
293 Positive Z-score values are shown in green and negative values in red. Schematic representation of
294 the dynamics of environmental drivers during the Arabica SE process are also shown under the
295 heatmap. The intensity levels of environmental drivers are expressed through their relative
296 intensity.
297

298 **Analysis of genes encoding transcription factors during Arabica SE**

299 Many authors reported the implication of some transcription factors (TFs) in the induction of somatic
300 embryos in different species, among these TFs are ABA INSENSITIVE 3 (ABI3) [43], AGAMOUS LIKE
301 15 (AGL15) [44,45], BABY BOOM (BBM) [46], LEAFY COTYLEDON (LEC1/LEC2) [47,48],
302 WUSCHEL-RELATED HOMEBOX (WUS/WOX2) [49,50], SOMATIC EMBRYOGENESIS
303 RECEPTOR KINASE (SERK) [51,52], CLAVATA 3 (CLV3) [53] and FUSCA 3 (FUS3) [54]. However,
304 the time points at which the genes encoding these TFs are highly expressed remain unknown. We showed
305 in Fig. 6 the detailed kinetics course of these genes during the whole Arabica SE process. Four types of
306 expression profiles were obtained: i) *SERK1* was highly active in the differentiated cells belonging to leaf
307 tissues and sharply decreased during the initial leaf-to-dedifferentiation phase switch (L1 to D1); ii) *CLV3*
308 was up-regulated at the first dedifferentiation stage (D1) while *SERK2* was switched on at this same stage
309 and was gradually up-regulated to reach a maximum in the primary callus cells. *SERK2* had a high
310 expression level while *CLV3* had a low expression profile. Both genes were switched off during the primary
311 callus-to-embryogenic callus switch, hence not expressed in embryogenic cells, embryogenic cell clusters
312 and embryos; iii) *BBM*, *ABI3*, *LEC1*, *AGL15*, *WOX2* and *WUS* which were initially switched-off in
313 differentiated cells of leaf explant were induced during the primary callus-to-embryogenic callus switch
314 and remain expressed in embryogenic cells, cell clusters and pro-embryos. These six genes could be further
315 classified in two groups: highly expressed genes (*BBM*, *ABI3*, *LEC1*) and lowly expressed genes (*AGL15*,
316 *WOX2*, *WUS*); iv) *FUS3* was expressed during the whole SE, first at a low level during the 5 first
317 developmental stages, then at a high level in embryogenic tissues and during the embryo redifferentiation
318 phase, with the highest expression in globular-shaped embryos (E1).



319

320 **Figure 6. Expression profiles of genes encoding transcription factors during the key**
 321 **developmental stages of somatic embryogenesis in *C. arabica*. *SERK1*, *SERK2*, *CLV3*, *BBM*,**
 322 ***ABI3*, *LEC1*, *AGL15*, *WOX2*, *WUS* and *FUS3* were studied. Genes were classified into 4 types of**
 323 **expression profiles (I, II, III, IV). Class III was subdivided into 2 subclasses (a & b) corresponding**
 324 **to high and low gene expression respectively. The x-axis corresponds to the different**
 325 **developmental stages and the y-axis corresponds to the number of normalized counts for each gene**
 326 **after RNA-seq read mapping on the Arabica reference genome [ref.] followed by DESeq2 [54]**
 327 **normalization.**

328

329 Analysis of genes encoding hormone signaling pathways during somatic embryogenesis

330 The role of hormones in the induction and maturation of somatic embryos has been widely reported, in
 331 particular that of auxin, cytokinin, abscisic acid and ethylene [55]. As we have previously studied the
 332 dynamics of these endogenous hormones during the different key steps of SE [56], we took the opportunity
 333 to analyze in the present work the kinetics of some hormone-related DEGs during the same SE key
 334 developmental steps (Fig. 7). *Aux/IAA* repressor genes (*IAA29*, *IAA30*, *IAA31*) have been reported as

335 contributing to SE induction [57]. Here, we showed that the most highly expressed gene in *C. arabica* was
336 *IAA29*. This gene was induced at the beginning of the dedifferentiation phase, after exogenous auxin supply
337 in the medium, and reached its highest expression in embryogenic calli. Logically, this gene was down-
338 regulated when exogenous auxin was removed from the medium but was still expressed at a low-level
339 during the redifferentiation phase leading to embryo formation.

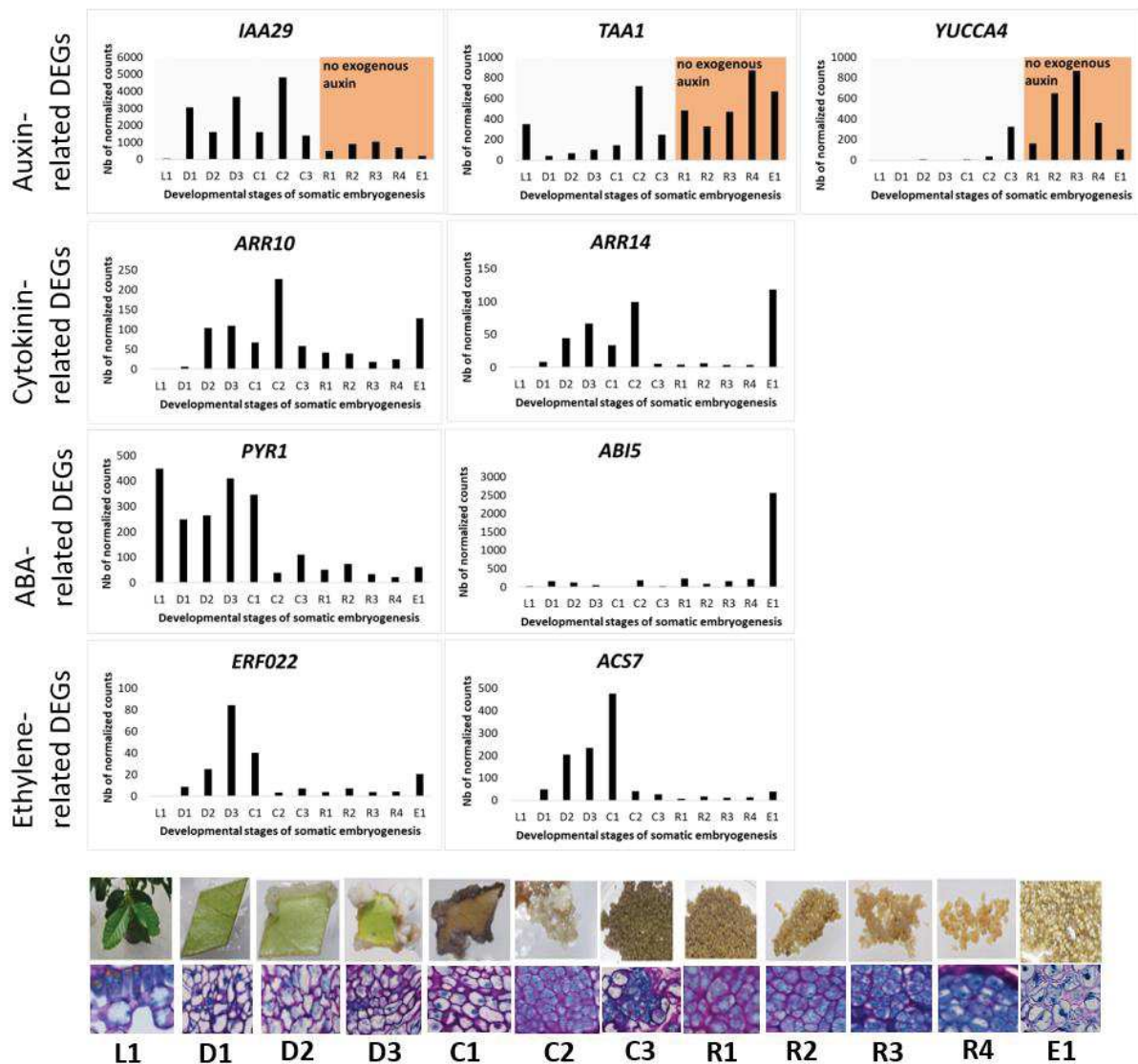
340 Both *Tryptophan Aminotransferase of Arabidopsis (TAA)* and *YUCCA* family genes are needed in the
341 tryptophan-dependent indole-3-acetic acid (IAA) biosynthesis pathway [58]. TAA plays a role in the
342 conversion of tryptophan to indole-3-pyruvic acid (IPA) and YUCCA plays a role in the conversion of IPA
343 to IAA [58]. For Arabica cells, *TAA1* and *YUCCA4* were the most highly expressed in their respective gene
344 families. Our results showed that *TAA1* is highly up-regulated in embryogenic callus cells and during
345 redifferentiation just after exogenous auxin removal (from R1 stage and forwards) followed by *YUCCA4*
346 which is highly up-regulated in R2 and R3 stages indicating auxin biosynthesis in developing embryos.
347 *YUCCA4* is also active in embryogenic cell clusters probably due to an insufficient amount of exogenous
348 auxin in the medium. Transcript profiling revealed that auxin-related genes showed diversity in their
349 expression profiles and were further detailed in Fig. 8.

350
351 Cytokinin-response regulators like the *Arabidopsis Response Regulators (ARRs)* [59], were also
352 investigated. Two type-B *ARRs* were differentially expressed during Arabica SE: *ARR10* and *ARR14*. Both
353 genes had a similar expression profile. They were induced during dedifferentiation and reached a maximum
354 of expression in embryogenic calli before being sharply down-regulated during the redifferentiation phase
355 until embryo formation where they were both again sharply up-regulated. *WUSCHEL*, whose expression
356 was sharply up-regulated in embryogenic callus cells (Fig. 6), was reported to be a direct target of *ARR10*
357 in *Arabidopsis* [60].

358
359 *PYR1* gene encodes for a protein that plays a major role in the abscisic acid (ABA) receptor
360 PYR/PYL/RCAR [61]. Our results showed that *PYR1* was highly expressed in leaf, during the
361 dedifferentiation phase and in primary callus; and was sharply down-regulated in embryogenic callus and
362 during the further redifferentiation stages. Downstream genes encoding TFs of the ABA response-pathway
363 were induced in embryogenic callus (*ABI3*, Fig. 6) or later during embryo formation (*ABI5*, Fig. 7).

364
365 *Ethylene Responsive Factor 022 (ERF022)* was reported to promote the formation of somatic embryos in
366 *Arabidopsis* through the ethylene-related pathway and to be co-expressed with *1-aminocyclopropane-1-*
367 *Carboxylate Synthase 7 (ACS7)* involved in ethylene biosynthesis [62]. Our results showed that *ERF022*
368 was highly activated during the dedifferentiation phase and was also up-regulated during the embryo
369 formation stage (E1) and down-regulated the expression of *ACS7* when expressed in *C. arabica*. Ethylene
370 has been reported to inhibit the formation of somatic embryos when present in culture medium [63]. *ACS7*
371 reached its highest expression in primary callus and was sharply down-regulated in embryogenic callus
372 while *BBM*, a gene also encoding an ethylene-responsive TF, was induced in embryogenic callus (Fig. 6).

373



374

375 **Figure 7. Expression profiles of hormone-related genes during the key developmental stages**
 376 **of somatic embryogenesis (SE) in *C. arabica*.** Auxin-related genes (*IAA29*, *TAA1* and *YUCCA4*),
 377 cytokinin-related genes (*ARR10*, *ARR14*), ABA-related genes (*PYR1*, *ABI5*) and ethylene-related
 378 genes (*ERF022*, *ACS7*) were investigated. The x-axis corresponds to the different SE
 379 developmental stages and the y-axis corresponds to the number of normalized counts for each gene
 380 after RNA-seq read mapping on the Arabica reference genome followed by DESeq2 [54]
 381 normalization. The colored area corresponds to total exogenous auxin removal from the medium.

382 **Auxin, transcription factors and secondary metabolism pathways are highly modulated**
 383 **during somatic embryogenesis**

384 Among the different biological processes taking place in SE and identified in Fig. 5, we chose to focus on
 385 three of them: auxin, transcription factors and secondary metabolism pathways since their variation appear
 386 fundamental in the SE process. All auxin-related DEGs, SE transcription factor-related DEGs and
 387 secondary metabolism-related DEGs were selected to perform a cluster analysis. For didactic purposes, we

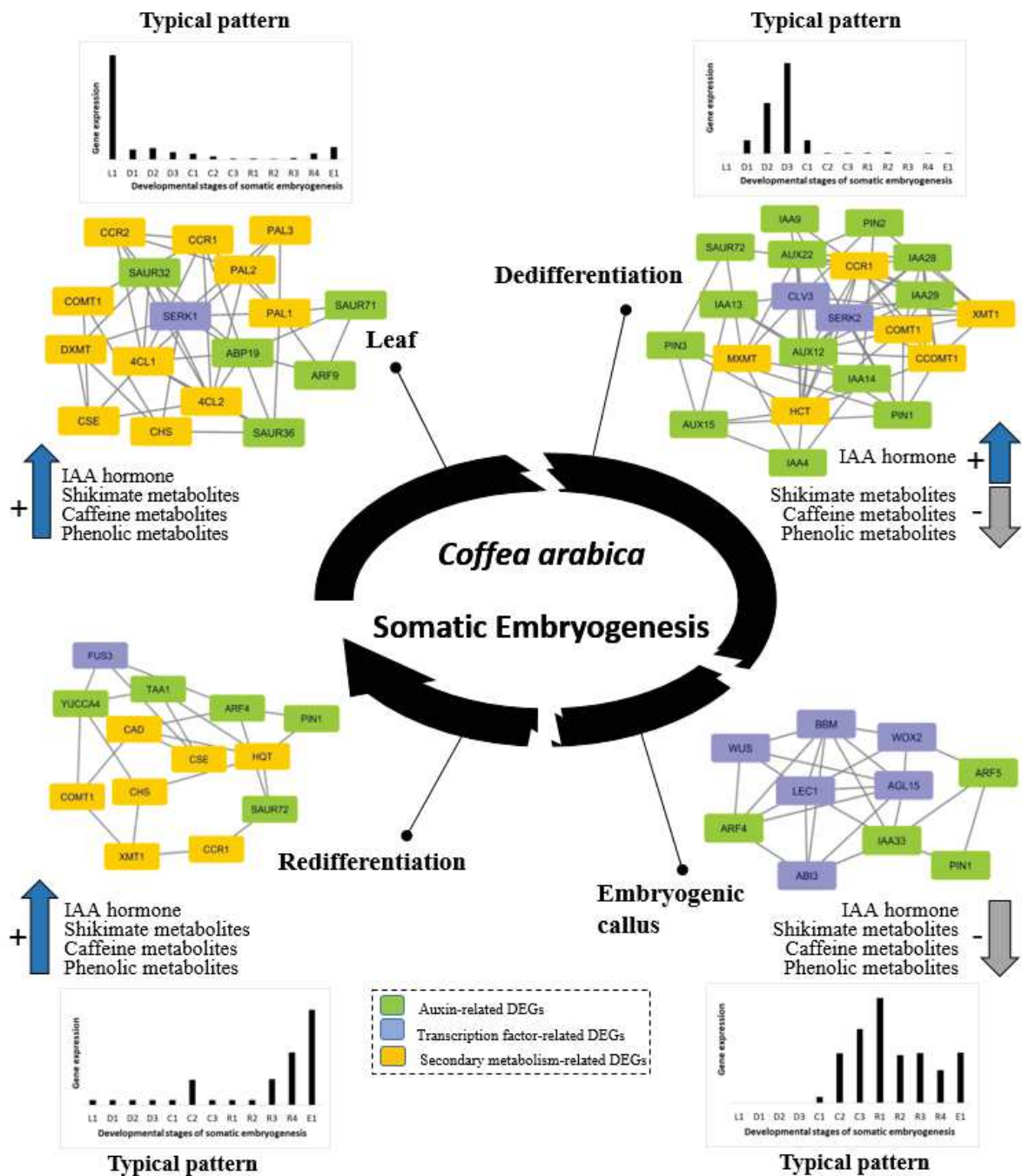
388 chose to generate four clusters; each cluster corresponding to a type of expression profile. The four clusters
389 obtained could be divided into four typical patterns (Fig. 8): i) Genes highly active during the leaf phase
390 and sharply down-regulated during dedifferentiation; ii) Genes highly up-regulated during dedifferentiation
391 and sharply down-regulated during the embryogenic callus phase, iii) Genes induced during the
392 embryogenic callus phase, iv) Genes up-regulated during redifferentiation and embryo formation. For each
393 cluster, a gene co-expression network was inferred. The four resulting networks (Leaf, Dedifferentiation,
394 Embryogenic callus and Redifferentiation) were summarized in Table 1. As an indication, in addition to
395 transcriptomic data, schematic kinetics of IAA and secondary metabolite endogenous contents were also
396 presented (Supplementary data).

397 Eleven genes related to secondary metabolism were expressed in leaf cells and defined the “Leaf” network.
398 On a metabolic level, shikimate, caffeine and phenolic metabolites were accumulated indicating a tight
399 relation between gene expression and metabolite accumulation. Five genes involved in endogenous auxin
400 response were also expressed in leaf cells while endogenous Indole-3-acetic acid (IAA) was also present.
401 *SERK1* was the only gene encoding a SE-related TF and expressed in leaf cells. *SERK1* appeared at the
402 center of the gene co-expression network as it shared the most edges with other genes (Fig. 8).

403 The “Dedifferentiation” network was mainly composed of auxin-related genes (13 genes) which were up-
404 regulated during the dedifferentiation phase while six genes related to secondary metabolism were also
405 expressed during this phase. Similarly, endogenous IAA was over-accumulated compared to the leaf phase
406 while shikimate, caffeine and phenolic metabolites were under-accumulated (Fig. 8). Precursors of these
407 metabolites were over-accumulated indicating a probable inhibition in their respective biosynthesis
408 pathways. *SERK2* and *CLV3*, two genes encoding SE-related TFs, were up-regulated during the
409 dedifferentiation phase and appeared at the center of the gene co-expression network as they shared the
410 most edges with other genes.

411 The “Embryogenic callus” network was a reduced network, mostly composed of genes encoding TFs and
412 four auxin-related genes. Edges were equally shared between genes encoding TFs. No secondary
413 metabolism-related DEGs were up-regulated in this network. During this phase, endogenous IAA as well
414 as shikimate, caffeine and phenolic metabolites were under-accumulated (Fig. 8) confirming the tight
415 relation between genes and metabolites.

416 The “Redifferentiation” network was equally shared between secondary metabolism-related genes and
417 auxin-related genes (Fig. 8). *TAA1* and *YUCCA4* were also up-regulated indicating an activation of
418 endogenous auxin biosynthesis. This was confirmed by the levels of endogenous IAA which were over-
419 accumulated compared to the levels of IAA in embryogenic callus. Similarly, secondary metabolites were
420 over-accumulated again compared to their levels in embryogenic callus and tended to reach the same levels
421 as in leaf cells. *FUS3* was up-regulated during the redifferentiation phase and was directly linked to *TAA1*
422 and *YUCCA4* indicating a role in the activation of endogenous auxin biosynthesis leading to embryo
423 formation.



424
 425
 426
 427
 428
 429
 430
 431
 432
 433

Figure 8. Co-expression analysis of auxin-related DEGs (green), SE transcription factor-related DEGs (violet) and secondary metabolism-related DEGs (orange) during the four main developmental phases of somatic embryogenesis. Four clusters of genes were generated. An average expression profile of the genes present in each cluster was also generated (bar diagrams). For each cluster, a gene co-expression network was inferred using the ARACNE algorithm [63]. All networks were visualized in Cytoscape [64]. Co-expressed genes shared edges. Schematic kinetics of IAA and secondary metabolite endogenous content were also presented (Supplementary data).

434 **Table 1. Generated gene co-expression networks and genes belonging to each network.**
 435 Auxin-related genes, transcription factor-related genes and secondary metabolism-related genes
 436 were clustered into four types of expression profiles and co-expression networks were inferred.
 437 Four co-expression networks were obtained: Leaf, Dedifferentiation, Embryogenic callus,
 438 Redifferentiation.

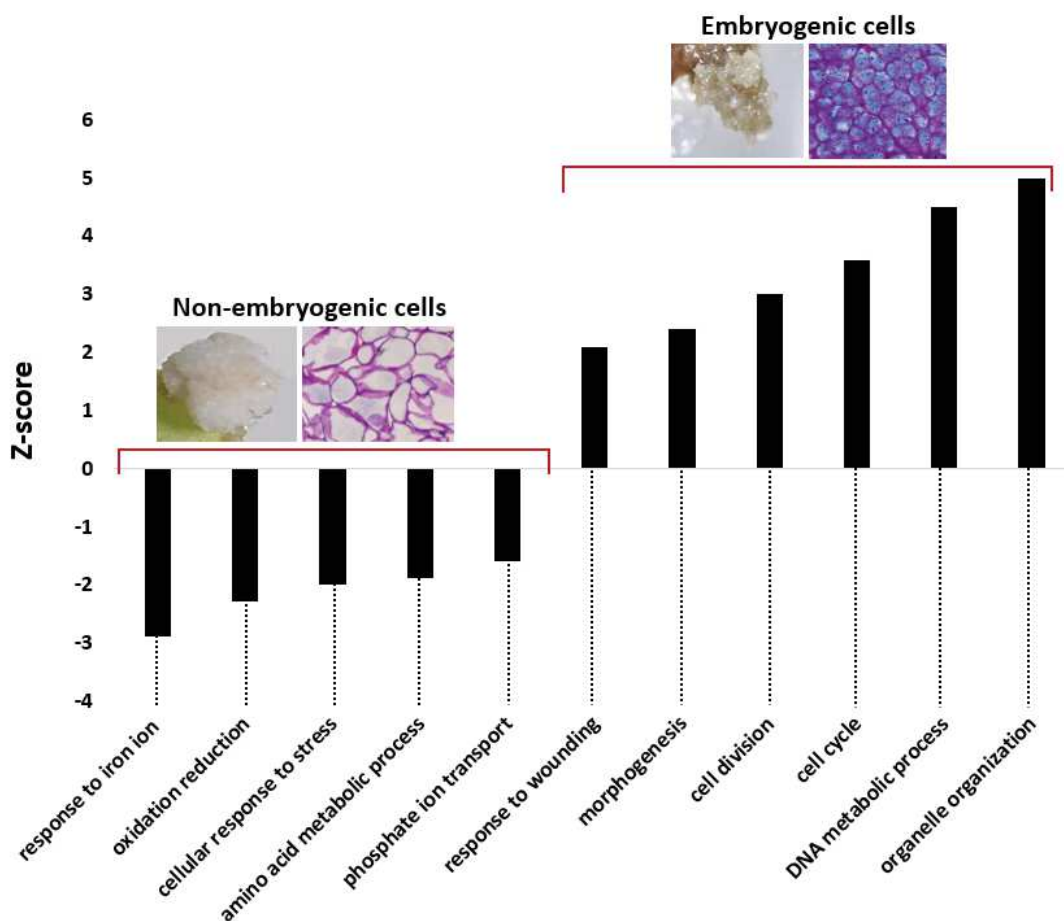
Co-expression network	Genes belonging to the co-expression network		
	Transcription factor-related genes	Auxin-related genes	Secondary metabolism-related genes
Leaf	<i>SERK1</i>	<i>Auxin Response Factor 9 (ARF9)</i> <i>Small Auxin Up RNA 32 (SAUR32)</i> <i>SAUR36</i> <i>SAUR71</i> <i>Auxin Binding Protein 19 (ABP19)</i>	<i>Phenylalanine Ammonia Lyase 1 (PAL1)</i> <i>PAL2</i> <i>PAL3</i> <i>4-cinnamoyl CoA ligase 1 (4CL1)</i> <i>4CL2</i> <i>Chalcone Synthase (CHS)</i> <i>Caffeine Synthase (DXMT)</i> <i>Caffeoyl Shikimate Esterase (CSE)</i> <i>Cinnamoyl-CoA Reductase 1 (CCR1)</i> <i>CCR2</i> <i>Caffeic acid O-MethylTransferase (COMT1)</i>
Dedifferentiation	<i>SERK2</i> <i>CLV3</i>	<i>AUX12</i> <i>AUX15</i> <i>AUX22</i> <i>IAA4</i> <i>IAA9</i> <i>IAA13</i> <i>IAA14</i> <i>IAA28</i> <i>IAA29</i> <i>PIN-FORMED1 (PIN1)</i> <i>PIN2</i> <i>PIN3</i> <i>SAUR72</i>	<i>Xanthosine Methyltransferase 1 (XMT1)</i> <i>MXMT</i> <i>CCR1</i> <i>COMT1</i> <i>Caffeoyl-CoA-O-MethylTransferase 1 (CCOMT1)</i> <i>Hydroxycinnamoyl-CoA shikimate/quinic acid hydroxycinnamoyl Transferase (HCT)</i>
Embryogenic callus	<i>ABI3</i> <i>BBM</i> <i>WUS</i> <i>WOX2</i> <i>AGL15</i> <i>LEC1</i>	<i>ARF4</i> <i>ARF5</i> <i>IAA33</i> <i>TAA1</i>	
Redifferentiation	<i>FUS3</i>	<i>ARF4</i> <i>TAA1</i> <i>YUCCA4</i> <i>SAUR72</i> <i>PIN1</i>	<i>CHS</i> <i>XMT1</i> <i>CSE</i> <i>CCR1</i> <i>COMT1</i> <i>Hydroxycinnamoyl-CoA Quinate hydroxycinnamoyl Transferase (HQT)</i> <i>Cinnamyl-Alcohol Dehydrogenase (CAD)</i>

439

440 **Embryogenic and non-embryogenic calli strongly differ at a transcriptomic level**

441 Why some of the thousands of cells evolve into an embryogenic callus while others proliferate
442 undifferentiated is a fundamental question. Our study allows us to address this issue. We showed that
443 embryogenic (EC) and non-embryogenic calli (NEC) can be easily distinguished on the basis of their
444 morphology, color and cellular characteristics (Fig. 2 C2, NEC). Somatic embryos could not be regenerated
445 from NEC. EC are yellowish and friable, constituted by small isodiametric cells, arranged in clusters, with
446 a dense cytoplasm and a central nucleus with a prominent nucleolus. NEC are spongy and translucent,
447 displaying cells of a much larger size with the vacuole occupying a big share of the cytoplasm, an absence
448 of cytoplasmic organelles and a low nucleocytoplasmic ratio.

449 At a transcriptomic level, a total of 499 DEGs were obtained between embryogenic cells and non-
450 embryogenic cells. Surprisingly, most of these genes (414 DEGs) were down-regulated in embryogenic
451 cells compared to non-embryogenic cells while only 85 were up-regulated. PAGE tool yielded significant
452 GO terms that could be classified into 11 biological processes (Fig. 9). Compared to non-embryogenic cells,
453 PAGE tool showed in embryogenic cells a down-regulation of genes encoding processes including mainly
454 response to iron ion, oxidation reduction, cellular response to stress, amino acid metabolic process and
455 phosphate ion transport while it showed an up-regulation of genes involved in organelle organization, DNA
456 metabolic process, cell cycle and division, morphogenesis and response to wounding.



457

458 **Figure 9. Functional categorization of the Arabica differentially expressed genes (DEGs) in**
459 **embryogenic calli compared to non-embryogenic calli, based on their Arabidopsis**
460 **orthologues.** An additional stage, the non-embryogenic callus (NEC), was sampled at the same
461 time as the embryogenic callus (C2). DEGs were compared against The Arabidopsis Information
462 Resource database (TAIR, www.arabidopsis.org) using BLASTP with an e-value cut-off of $1 \times$
463 10^{-4} . Functional categorization of Arabidopsis orthologues was performed through Parametric
464 Analysis of Gene set Enrichment (PAGE) tool in agriGO v2.0 [42]. PAGE tool yielded significant
465 Gene Ontology (GO) terms that could be classified into 11 biological processes. The x-axis
466 corresponds to the different biological processes obtained and the y-axis corresponds to the Z-score
467 values yielded by the PAGE tool.
468

469 Discussion

470
471 *Somatic embryogenesis specific gene expression profiles as a signature of cell fate*

472 This study reports the first global analysis of SE gene expression analysis, conducted on 12 key
473 developmental stages covering the whole SE process from leaf explant dedifferentiation to embryo
474 formation. Two pre-requisites were key for this study: (i) the availability of large-scale protocols, offering
475 high reliability and cell synchronicity in each developmental stage as well as a good biological efficiency,
476 (ii) the availability of recent -omics technologies allowing to rapidly offer masses of information. This first
477 global analysis applied on SE should serve as a reference for a wide range of plant species as the intensive
478 sampling conducted on the successive key developmental stages, which are conserved among species,
479 provided an overview of the SE process and enabled to clarify some real black boxes [24]. The statistical
480 approach employed in this study to analyze transcriptomic data allowed to cluster the whole SE process
481 into seven main developmental phases and six key developmental phase switches which are the basis of
482 SE. Many authors recently reported the necessity of a better understanding of the SE process to clarify the
483 current bottlenecks [9,22,24,26]. The fact that each of the transcript profiles corresponded with a specific
484 cell development stage (leaf cells, cells in dedifferentiation, primary callus cells, embryogenic cells,
485 embryogenic cell clusters, pro-embryogenic masses, embryonic cells) demonstrates the great resolving
486 power of the transcriptomic approach. Each developmental stage has a highly specific signature based on
487 transcript fingerprints are a highly specific signature for each developmental stage and the transcriptomic
488 approach bring rich information on cell physiology, as many of these stages are undistinguishable by
489 morphological and histological approaches (e.g. early stages of dedifferentiation and redifferentiation). This
490 statistical approach in transcriptomics (Fig. 3) is even more robust than the one previously applied in
491 metabolomics since all callus stages (primary callus, embryogenic callus, embryogenic cell clusters) had
492 the same metabolic profile and could not be distinguished separately [56]. A number of studies in different
493 species assumed that clear correlations may exist among the different transcriptome profiles and some
494 specific SE stages [28,29,31]. This global analysis is a proof of concept that transcripts are good markers
495 of all cell fate transitions and, in the near future, could be used to understand and better pilot the
496 optimization of the SE culture conditions by using it as a milestone of successful developmental stages.
497 This goes beyond morphological and histological descriptions, which were until now the most common
498 way to support empirical protocol optimization. We believe that this global scale transcriptome study,
499 combined with a metabolic approach, will give a much clearer understanding of the molecular mechanisms
500 underlying cell reprogramming [24].
501

502 *Genes encoding regulatory, metabolic, hormonal and stress-related pathways are the most differentially*
503 *expressed during coffee somatic embryogenesis*
504

505 A number of genes were strongly up or down-regulated during the six identified developmental phase
506 switches. The functional analysis performed on these DEGs showed that the obtained GO terms could be
507 classified in 19 biological processes, themselves reclassified into six families: hormonal pathways (mainly
508 auxin and cytokinin), metabolic pathways (carbohydrate, starch, protein and secondary metabolism),
509 regulatory pathways (regulation of gene expression, pattern specification, embryo development), stress-
510 related pathways (response to stress and wounding), mitosis-related pathways (cell cycle and division, cell
511 wall and chromatin organization), and photosynthesis-related pathways (photosynthesis and circadian
512 cycle). The transition of a leaf cell to an embryogenic cell is a long process of cell division and organization
513 occurring in the dark, hence up-regulated genes related to mitosis and down-regulated genes related to
514 photosynthesis were expected [65]. SE is a process driven by exogenously supplied plant growth regulators
515 [66]. Although most plants require similar physical conditions (temperature, light regime) for the induction
516 of SE, only a specific medium composition can trigger and then support the process. According to Sugimoto
517 et al. [67], the prime characteristic of plant regeneration is cell fate reprogramming induced by wounding,
518 stress, and hormones which is in agreement with our results. Auxins and cytokinins are widely known to
519 play essential roles in the induction of embryogenic cultures [22,55,56,68]. Additionally, many authors
520 reported the implication of some transcription factors (TFs) in the induction of somatic embryos in different
521 species, more particularly ABI3 [43], AGL15 [44,45], BBM [46], LEC1/LEC2 [47,48], WUS/WOX2
522 [49,50], SERK [51,52], CLV3 [53] and FUS3 [54]. However, the time points at which the genes encoding
523 these TFs are highly active were still unknown. In this study, we showed for the first time the kinetics of
524 these genes during the successive SE steps. It has been proposed that TFs together with auxin and cytokinin
525 play a crucial role in the maintenance of stem cell niche in the shoot apical meristem in totipotent cells
526 in Arabidopsis [69] and in the cell pattern specification during the transition from totipotent-to-
527 embryonic cell at the onset of redifferentiation [70]. Our results confirmed that an exogenous supply in
528 auxin and cytokinin allowed an enrichment in transcripts of genes related to meristem development and
529 TFs-related genes in coffee embryogenic cells while an enrichment in transcripts of genes related to
530 embryonic cell pattern specification and embryonic formation was evidenced at the onset of
531 redifferentiation. For metabolic pathways, we have previously showed that carbohydrates, starch, amino
532 acids and secondary metabolites are differentially accumulated during the different SE steps [56]. Here,
533 we confirmed that DEGs involved in metabolic pathways are tightly related to the accumulation of their
534 respective metabolites.

535
536 *Genes regulating cell fate are highly modulated by environmental drivers during somatic embryogenesis*
537

538 This study is a solid proof that environmental drivers are the main regulators of cell fate as they are
539 tightly related to the regulation of genes involved in crucial biological processes. This study allowed to
540 measure the direct effects of environmental drivers classically used to master SE, particularly light, growth
541 regulators and cell density [6,9] as these drivers were clearly perceived at a molecular level. Environmental
542 drivers, conventionally named ‘culture conditions’, are classically optimized in an empirical way in order
543 to obtain the adequate nutritive medium for a particular genotype for SE induction [22,24]. Since these
544 drivers can be perceived at a molecular level, the genes modulated by the drivers can be considered of huge
545 interest since they can be used to pilot SE optimization in a rational way. It is the case for the genes involved
546 in photosynthesis and circadian rhythm which are tightly linked to light intensity, genes involved in the
547 response to auxin and cytokinin stimulus, and genes involved in embryo pattern specification which are
548 tightly linked to cell density. Many authors have focused on studying the complex gene networks involved
549 in the response to growth regulators, mainly auxin and cytokinin, in order to understand the expression of
550 cell totipotency in Arabidopsis and in cotton [6,70]. The identified genes in these model plants were also
551 found in coffee and had similar patterns in the studied developmental stages, indicating conserved pathways
552 of cell totipotency between species. We believe that coffee SE could be used as a model system to better
553 understand fundamental mechanisms for the response to growth regulators in woody plants.

554 *Genes encoding hormones and transcription factors play a major role in the expression of totipotency*

555 The role of hormones in the induction and maturation of somatic embryos has been widely reported,
556 particularly auxin, cytokinin, abscisic acid and ethylene [3,22,55]. In a previous work, we studied the
557 dynamics of these endogenous hormones during the different key steps of SE confirming their major role
558 in the acquisition and expression of totipotency [56]. SE is a process driven by exogenously supplied plant
559 growth regulators. On a transcriptomic level, the up-regulation of a total of 13 *AUX/IAA* genes which are
560 repressor genes [70] as well as efflux carrier genes (*PIN*) showed that the deprogramming process leading
561 to the formation of undifferentiated cells in *C. arabica* was highly dependent on exogenously supplied
562 auxin. Endogenous IAA concentrations are known to be tightly linked to the *YUCCA* gene expression in
563 Arabidopsis [71]. In coffee, once auxin was removed from the medium to allow cell redifferentiation, all
564 *AUX/IAA* genes were switched-off and *YUCCA* genes were switched-on enabling synthesis of endogenous
565 IAA. Our results confirmed the concomitant increase in *YUCCA* gene expression and endogenous IAA
566 during redifferentiation.

567 *ARF5* appeared to be only up-regulated in embryogenic cells and could now be considered as a molecular
568 marker of this developmental stage in coffee. Wójcikowska et al. [72] showed in Arabidopsis that *ARF5*
569 regulates the expression of numerous genes that are involved in somatic embryo formation
570 including *LEC2 (LEAFY COTYLEDON2)*, which is an activator of the *YUC1, YUC4, YUC10 (YUCCA)*
571 genes involved in auxin biosynthesis during SE. *ARF5* was also reported to be involved in the cytokinin
572 response pathway in Arabidopsis, tightly linked to *ARR10* during the patterning and cell organization of
573 meristem cells [72]. We also confirmed an up-regulation of *ARR10* in coffee embryogenic cells which
574 occurred at the same time as the up-regulation of *ARF5* gene. This shows that SE pathways are highly
575 conserved between species and the up-regulation of *ARR10* could also be considered as a molecular marker
576 of the embryogenic potential (Table 2).

577 While the abscisic acid (ABA) was not regulated exogenously (no ABA supply) in the coffee SE process,
578 an activation of *ABI3* gene in embryogenic cells was noted and followed by the activation of *ABI5* in
579 embryonic cells. *ABI3* was reported to play a major role in the regulation of SE induction in many species
580 [22,55], while *ABI5* was reported to inhibit seed germination and promote embryo maturation in conifers
581 [10]. This shows once again that SE pathways are highly conserved between species and the activation of
582 *ABI3* can be considered as a molecular marker of embryogenic cells while the activation of *ABI5* is a marker
583 of the embryo maturation process (Table 2).

584 The formation and proliferation of embryogenic cells represent the most crucial stage for the success of the
585 SE process in all plant species because the redifferentiation episode directly depends on their abundance.
586 Embryogenic cell formation represents a true bottleneck that needs improvement of culture conditions in
587 order to reduce the long time needed (7 months). We established, for the first time, the kinetics of expression
588 of genes encoding for widely reported TFs implicated in SE. These genes (summarized in Table 2) represent
589 candidates to be used as molecular markers to pilot SE optimization [24].

590 *Identifying molecular markers of the embryogenic capacity*

591
592 Many authors have distinguished embryogenic cells from non-embryogenic cells based on morphological
593 and histological studies [1,19]. When comparing transcript profiles, a huge majority (414) of the obtained
594 DEGs were down-regulated in embryogenic cells compared to non-embryogenic ones while only 85 were
595 up-regulated. This is in agreement with Yang et al. [6] who showed in cotton that the existing developmental
596 information of somatic cells must be switched-off, most probably by an epigenetic regulation, in order to
597 express the embryogenic capacity. The functional analysis performed on these DEGs allowed the

598 identification of genes implicated in biological processes that can be used as molecular markers of the
599 embryogenic state. These molecular markers could serve as predictors of the regenerative capacity meaning
600 that they can be used to quickly select or eliminate cell lines based on the presence or absence of these
601 markers. Since coffee embryogenic callus is a compact quick proliferating structure, its constitutive
602 embryogenic cells show a functional mitotic activity, an up-regulation of genes encoding organelle and
603 DNA organization, morphogenesis, cell cycle and division. In addition, an up-regulation of genes related
604 to wounding (*WIND* genes) [76] in embryogenic cells suggests that embryogenic cells resulted from a
605 controlled stress-related pathway while non-embryogenic cells resulted from an uncontrolled stress-related
606 pathway implicating a sharp up-regulation of genes involved in stress and oxidation processes [77].
607 Histological studies confirmed that non-embryogenic callus is a spongy and oxydated callus with the
608 presence of numerous vacuolated and degenerating cells. Non-embryogenic cells were characterized by an
609 up-regulation of genes encoding to respond to metal ion, oxidation reduction and phosphate ion transport,
610 which is in agreement with a number of studies on conifers [9,21,32] suggesting that non-embryogenic cell
611 main fate was survival while embryogenic cells were mainly a transient state before cell fate transition.
612 This means that the markers of the embryogenic state are conserved between species in woody plants. We
613 suggest that reducing oxidative stress, by improving gaseous O₂/CO₂ exchange and reducing ethylene, can
614 increase embryogenic capacity.

615 In our study, transcripts of genes encoding amino acids were also more abundant in non-embryogenic cells,
616 probably due to the fact that genes in embryogenic cells were involved in the synthesis of more complex
617 structures (proteins, DNA) as reported in other woody species [29,32,56].

618 *Secondary metabolism-related genes are switched-off during dedifferentiation and switched back on at*
619 *the onset of redifferentiation*

620 Somatic cells within the plant contain all the genetic information necessary to create a complete and
621 functional plant [6]. During cellular dedifferentiation, the existing developmental information of somatic
622 cells must be switched off or re-configured to make the somatic cells ready for an embryogenic program
623 [75]. Our results clearly showed that genes encoding phenolic compounds and alkaloids were sharply down-
624 regulated during this stage and completely switched off in embryogenic cells. Nic-Can et al. [78] have
625 provided evidence that those compounds inhibited the embryogenic process by affecting DNA methylation
626 in *C. canephora*. Magnani et al. [79] also reported a shut-off in biochemical pathways in Arabidopsis in
627 order to activate the transcriptional machinery. Conversely, our results showed an up-regulation of the
628 secondary metabolism-related genes at the early days of redifferentiation followed by an accumulation of
629 their respective phenolic compounds, mainly chlorogenic acids, as they are key intermediaries for cell wall
630 biosynthesis, in addition to their more general role in abiotic stress tolerance since the embryo formation
631 phenomenon was widely reported as a stress-related phenomenon [75,80]. Therefore, genes encoding
632 chlorogenic acids (*HQT*, *COMT1*) [81] can be considered as molecular markers of the redifferentiation
633 pattern of coffee SE. Transcripts of genes encoding precursors of caffeine (*XMT1*) [82] also accumulated
634 during redifferentiation probably meaning that caffeine would be produced in later embryo developmental
635 stages.

636
637 *Molecular markers of cell fate to pilot the optimization of somatic embryogenesis*
638

639 We identified a set of molecular markers during the transitions of cell fate occurring during coffee SE that
640 are summarized in Table 2. A molecular marker is firstly defined by a clear expression pattern, i.e. a gene
641 switch-on (or sharp up-regulation) followed by a switch-off (or sharp down-regulation) at one or several
642 developmental stages. Secondly, a molecular marker is chosen based on its gene expression level, i.e. a
643 gene with high expression value is preferable to a gene with low expression value. Finally, a molecular

644 marker can be validated through comparison between non-optimal conditions or developmental stages. In
 645 order to be able to implement a detailed sampling of all developmental stages of the coffee SE process, this
 646 study was limited to only one genotype. Further RT-qPCR analyses realized on a set of genotypes more or
 647 less recalcitrant to the induction of somatic embryos, will be crucial to validate the candidate molecular
 648 markers. Once a marker is validated, it can be used to efficiently pilot the SE process optimization in a
 649 reliable, cost-effective and faster way by testing a number of contrasted culture conditions in order to select
 650 the optimal, particularly in the case of recalcitrant genotypes or species.

651
 652 **Table 2.** Transcriptomic markers characterizing the different developmental phases of coffee somatic
 653 embryogenesis.

Transcriptomic markers	Developmental phases						
	Leaf cells	Cells in dedifferentiation	Primary callus cells	Embryogenic cells	Embryogenic cell clusters	Cells in redifferentiation	Embryonic cells
Transcription factor-related genes	<i>SERK1</i>	<i>SERK2</i> <i>CLV3</i>		<i>BBM</i> <i>ABI3</i> <i>LEC1</i> <i>AGL15</i> <i>WOX2</i> <i>WUS</i>		<i>FUS3</i>	<i>FUS3</i>
Auxin-related genes		<i>AUX/IAA</i>		<i>ARF5</i>	<i>YUCCA4</i>	<i>YUCCA4</i>	
Cytokinin related genes				<i>ARR10</i>			<i>ARR10</i>
ABA-related genes							<i>ABI5</i>
Ethylene-related genes			<i>ACS7</i>				
Secondary metabolism-related genes	<i>PAL</i> <i>CHS</i> <i>4CL</i> <i>DXMT</i>					<i>CHS</i> <i>4CL</i> <i>COMT1</i> <i>HQT</i>	<i>CHS</i> <i>4CL</i> <i>COMT1</i> <i>HQT</i> <i>XMT</i>

654
 655
 656
 657
 658
 659
 660
 661
 662
 663

664 **Conclusion**

665

666 This first global scale transcriptome analysis of SE, applied on coffee, showed that transcriptomics using
667 the RNAseq technology provides a powerful approach to investigate global transcript patterns. This
668 approach helped to clearly identify seven important developmental phases with very contrasted and specific
669 patterns, leading to the characterization of six key developmental phase switches, which are strategic for
670 the whole biological efficiency of somatic embryo regeneration. Through this global transcriptome profiling
671 approach of SE, we proved that transcriptomics can assign a specific signature to each developmental stage
672 and provides valuable information about cell fate. It also allowed to measure at the molecular level and for
673 each developmental stage the direct effects of environmental drivers, particularly light and growth
674 regulators, used to control the regeneration process. Together with the metabolomics approach, this study
675 gives a much clearer understanding of the intimate mechanisms governing totipotency and SE. It represents
676 a starting point for optimizing coffee SE protocols in a rational way, putting an end to thirty years of
677 empirical research. The 23 identified transcriptomic markers which are specific of the different
678 developmental phases, including the strategic ‘embryogenic state’, should be directly used as targets to pilot
679 this optimization, particularly in recalcitrant genotypes or species. We consider that these findings are
680 informative and useful to a wide range of plant species, offering unprecedented perspectives in plant
681 micropropagation through SE.

682

683 **Material and Methods**

684

685 *Tissue culture and sampling*

686

687 An intraspecific hybrid (GPFA116) proceeding from Arabica breeding program was used in this study. SE
688 was performed in the Nestlé Research laboratories (Tours, France) based on the large-scale protocols
689 described previously for *C. arabica* [84], and carried out in 4 replicates i.e., four independent leaf
690 introductions in April, June, October and December 2016 from distinct 1-year old plants and grown in the
691 Nestlé Research greenhouse. Explants were then cultured in Petri dishes on T1 ‘dedifferentiation 1’
692 medium, i.e. Murashige and Skoog (MS) half-strength solid medium supplemented with 0.5 mg/mL 2,4-D
693 (2,4-dichlorophenoxyacetic acid), 1 mg/mL IBA (indole 3-butyric acid) and 2 mg/mL 2iP (N⁶-(2-
694 Isopentenyl) adenine) for 1 month before transfer on T2 ‘dedifferentiation 2’ medium, i.e. MS/2 solid
695 medium supplemented with 1 mg/mL 2,4-D and 4 mg/mL BA (6-benzylaminopurine) for 6 months until
696 embryogenic callus formation. Petri dishes were placed at 25 °C in the dark. Embryogenic calli were then
697 inoculated at a rate of 10 g/L in 250-ml Erlenmeyer flasks containing M ‘proliferation’ liquid nutritive
698 medium, i.e. medium supplemented with 0.3 mg/L 2,4-D and 1 mg/L BA and cultured for 4 months on
699 shakers (120 rpm) at 25 °C in the dark, until proliferation of cell clusters. To stimulate regeneration of early
700 somatic embryos, cell clusters were transferred to 250-mL Erlenmeyer flasks containing DIF
701 ‘redifferentiation’ liquid medium lacking the auxin 2,4-D at a rate of 10 g/L medium for 1 week then at a
702 rate of 1 g/L for 4 weeks until formation of globular-shaped embryos. Erlenmeyer flasks were placed on
703 shakers (120 rpm) at 25 °C under indirect light (120-150 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$) during redifferentiation.

704 Twelve sampling points were chosen to cover the SE process from leaf dedifferentiation till torpedo-shaped
705 embryos development as shown in Fig. 2: leaves from greenhouse plants (L1), explants during
706 dedifferentiation [1 week (D1), 2 weeks (D2), 5 weeks (D3)], compact primary callus obtained 3 months
707 after induction (C1), embryogenic callus obtained 7 months after induction (C2), established cell clusters
708 obtained after 4 months in liquid proliferation medium (C3), pro-embryogenic masses [1 week in
709 redifferentiation medium after auxin withdrawal (R1), 24 h in redifferentiation medium after reducing cell
710 density (R2), 72 h (R3), 10 days (R4)] and globular embryos obtained after 3 weeks of culture (E1). An

711 additional stage was also sampled at the same time and in the same culture conditions as the embryogenic
712 callus (C2), the non-embryogenic callus (NEC). Approximately 1 g of fresh weight/sample/replicate was
713 collected for transcriptome analysis and instantly kept in liquid nitrogen before storage at -80 °C until
714 further analysis. A detailed morphological and cellular characterization of the different developmental
715 stages was given in Awada et al. [56].

716 ***RNA extraction***

717 Frozen tissues were ground to a fine powder in liquid nitrogen and total RNA were extracted using
718 RNeasy plant Mini Kit (QIAGEN, Germantown, MD - USA) and treated with RNase free DNaseI
719 (QIAGEN) according to manufacturer's instructions. The quality and quantity of total RNA were
720 analyzed by using Agilent 2100 Bioanalyzer RNA chip (Agilent Technologies Inc., Santa Clara, CA -
721 USA). The RNA samples with a RNA integrity number (RIN) higher than 7.0 were selected and used
722 for the subsequent analysis.

723

724 ***Illumina sequencing***

725

726 RNA sequencing was carried out by Nestlé Research (Lausanne, Switzerland). The cDNA libraries were
727 generated using the TruSeq Stranded mRNA Kit (Illumina) followed by PCR amplification for
728 sequencing on Illumina HiSeq 2500. Paired-end cDNA libraries were generated from all samples and
729 sequencing was performed to generate the ~125bp paired-end reads. Quality control and assessment of
730 raw Illumina reads in FASTQ format were done by FastQC software (v0.11.5) to obtain per base quality,
731 Guanine-Cytosine (GC) content, and sequence length distribution. Clean reads were obtained by
732 removing the low-quality reads, adapters, and poly-N-containing reads from raw data. Approximately
733 95% of high quality reads were obtained in each library from generated data. An average of 90 million
734 paired-end reads was obtained for each library.

735

736 ***Read mapping and differential gene expression analysis***

737

738 The pre-processed reads were aligned to the *C. arabica* genome sequence (ACGC confidential data
739 (2014) – unpublished) using the STAR (v2.5.3a) software. The quantification of uniquely mapped reads
740 to each gene locus was performed with a maximum of 10 mismatches per paired-end alignment using
741 Partek E/M algorithm initially described by Xing et al. [83] and a Principal Component Analysis (PCA)
742 was carried out to verify homogeneity of replicates. On average, 74% of sequenced reads per sample were
743 uniquely mapped to the reference genome. DESeq2 [54] was used to standardize reads across libraries
744 and for differential expression analysis. Differential expression was considered at a threshold value
745 of $P \leq 0.001$ and the absolute value of $\log_2\text{Ratio} \geq 3$ (≥ 2 for EC-NEC comparison) based on the
746 FDR < 0.05 . The heatmap of all differentially expressed genes was generated in R [84] using the
747 ComplexHeatmap package [85]. The same data were used for the analysis of hierarchical clustering.
748 Hierarchical clustering was performed with the pvclust package [86] using Pearson's correlation coefficient
749 [87]. Cluster probabilities were calculated via a multiscale bootstrap with a total of 1,000 iterations. A
750 cluster probability is a percentage which indicates how strong the cluster is supported by data.

751 ***Functional gene expression analysis***

752

753 All differentially expressed genes (DEGs) were compared against The Arabidopsis Information Resource
754 database (TAIR, www.arabidopsis.org) using BLASTP with an e-value cut-off of 1×10^{-4} . The resulting
755 annotation was used to perform gene ontology (GO) analysis using the Parametric Analysis of Gene set
756 Enrichment (PAGE) tool in agriGO v2.0 with default functions [42]. Significant GO terms were found

757 using the default FDR <0.05 cutoff value. Obtained Z-scores were plotted in a heatmap generated in R [84]
758 using the ComplexHeatmap package [85] or in a bar diagram in Excel.

759

760 ***Gene co-expression analysis and network construction***

761

762 A cluster analysis was performed on auxin-related DEGs, SE transcription factor-related DEGs and
763 secondary metabolism-related DEGs. Hierarchical clustering was accomplished by pvclust [87] with k-
764 means method using Pearson's correlation distance between DEG expression profiles. Four clusters
765 representing the four major types of expression profiles were generated. For each cluster, the ARACNE
766 algorithm [63] was used to infer the gene co-expression network between genes of the same cluster. The
767 ARACNE procedure starts by assigning to each pair of nodes (pair of genes) a weight equal to their mutual
768 information then all edges (link between a pair of nodes) are drawn followed by a removal of the weakest
769 edges based on the assigned weight. All networks were visualized in Cytoscape [64].

770

771

772

773 **Availability of data and materials**

774 All data generated or analysed during this study are included in this published article and its supplementary
775 information files.

776

777 **References**

778

779 1. Verdeil J-L, Alemanno L, Niemenak N, Tranbarger TJ. Pluripotent versus totipotent plant stem cells:
780 dependence versus autonomy? Trends in Plant Science. 2007 Jun 1;12(6):245–52.

781

782 2. Ikeuchi M, Iwase A, Rymen B, Harashima H, Shibata M, Ohnuma M, et al. PRC2 represses
783 dedifferentiation of mature somatic cells in *Arabidopsis*. Nature Plants. 2015 Jul;1(7):15089.

784

785 3. Zeng F, Zhang X, Jin S, Cheng L, Liang S, Hu L, et al. Chromatin reorganization and endogenous
786 auxin/cytokinin dynamic activity during somatic embryogenesis of cultured cotton cell. Plant Cell Tiss
787 Organ Cult. 2007 Jul 1;90(1):63–70.

788

789 4. Reinert J. Morphogenese und ihre Kontrolle an Gewebekulturen aus Carotten. Naturwissenschaften.
790 1958 Jan 1;45(14):344–5.

791

792 5. Steward FC, Mapes MO, Mears K. Growth and Organized Development of Cultured Cells. II.
793 Organization in Cultures Grown from Freely Suspended Cells. American Journal of Botany.
794 1958;45(10):705–8.

795

796 6. Yang X, Zhang X, Yuan D, Jin F, Zhang Y, Xu J. Transcript profiling reveals complex auxin signalling
797 pathway and transcription regulation involved in dedifferentiation and redifferentiation during somatic
798 embryogenesis in cotton. BMC Plant Biol. 2012 Jul 20;12:110.

799

- 800 7. Novak FJ, Daskalov S, Brunner H, Nesticky M, Afza R, Dolezelova M, et al. Somatic Embryogenesis in
801 Maize and Comparison of Genetic Variability Induced by Gamma Radiation and Tissue Culture
802 Techniques. *Plant Breeding*. 1988;101(1):66–79.
803
- 804 8. Gaj MD, Zhang S, Harada JJ, Lemaux PG. Leafy cotyledon genes are essential for induction of somatic
805 embryogenesis of Arabidopsis. *Planta*. 2005 Dec;222(6):977–88.
806
- 807 9. Lelu-Walter M-A, Thompson D, Harvengt L, Sanchez L, Toribio M, Pâques LE. Somatic embryogenesis
808 in forestry with a focus on Europe: state-of-the-art, benefits, challenges and future direction. *Tree Genetics
809 & Genomes*. 2013 Aug 1;9(4):883–99.
810
- 811 10. Etienne H. Somatic embryogenesis protocol: Coffee (*Coffea arabica* L. and *C. canephora* P.). In
812 *Protocols for Somatic Embryogenesis in Woody Plants*; Jain, S.M., Gupta, P.K., Eds.; Springer: Dordrecht,
813 The Netherlands, 2005; pp. 167–179.
814
- 815 11. Maximova SN, Alemanno L, Young A, Ferriere N, Traore A, Gultinan MJ. Efficiency, genotypic
816 variability, and cellular origin of primary and secondary somatic embryogenesis of *Theobroma cacao* L. In
817 *In Vitro CellDevBiol-Plant*. 2002 May 1;38(3):252–9.
818
- 819 12. Montalbán IA, De Diego N, Moncaleán P. Enhancing initiation and proliferation in radiata pine (*Pinus
820 radiata* D. Don) somatic embryogenesis through seed family screening, zygotic embryo staging and media
821 adjustments. *Acta Physiol Plant*. 2012 Mar 1;34(2):451–60.
822
- 823 13. Corredoira E, Ballester A, Ibarra M, Vieitez AM. Induction of somatic embryogenesis in explants of
824 shoot cultures established from adult *Eucalyptus globulus* and *E. saligna* × *E. maidenii* trees. *Tree Physiol*.
825 2015 Jun 1;35(6):678–90.
826
- 827 14. Corredoira E, Valladares S, Vieitez AM. Morphohistological analysis of the origin and development of
828 somatic embryos from leaves of mature *Quercus robur*. In *In Vitro CellDevBiol-Plant*. 2006 Nov 1;42(6):525–
829 33.
830
- 831 15. Ribas AF, Dechamp E, Champion A, Bertrand B, Combes M-C, Verdeil J-L, et al. Agrobacterium-
832 mediated genetic transformation of *Coffea arabica* (L.) is greatly enhanced by using established
833 embryogenic callus cultures. *BMC Plant Biol*. 2011 May 19;11:92.
834
- 835 16. Ikeuchi M, Ogawa Y, Iwase A, Sugimoto K. Plant regeneration: cellular origins and molecular
836 mechanisms. *Development*. 2016 01;143(9):1442–51.
837
- 838 17. Boutilier K, Angenent GC, Castan MS, Hui L. Haploid embryogenesis [Internet]. US20160212956A1,
839 2016 [cited 2019 Mar 5]. Available from: <https://patents.google.com/patent/US20160212956/en>
840
- 841 18. Yang X, Zhang X. Regulation of Somatic Embryogenesis in Higher Plants. *Critical Reviews in Plant
842 Sciences*. 2010 Jan 25;29(1):36–57.
843

- 844 19. Jayasankar S, Bondada BR, Li Z, Gray DJ. Comparative Anatomy and Morphology of *Vitis vinifera*
845 (Vitaceae) Somatic Embryos from Solid- and Liquid-Culture-Derived Proembryogenic Masses. *American*
846 *Journal of Botany*. 2003;90(7):973–9.
847
- 848 20. Raghavan V. Role of 2,4-dichlorophenoxyacetic acid (2,4-D) in somatic embryogenesis on cultured
849 zygotic embryos of *Arabidopsis*: cell expansion, cell cycling, and morphogenesis during continuous
850 exposure of embryos to 2,4-D. *Am J Bot*. 2004 Nov;91(11):1743–56.
851
- 852 21. Florez SL, Erwin RL, Maximova SN, Guiltinan MJ, Curtis WR. Enhanced somatic embryogenesis in
853 *Theobroma cacao* using the homologous BABY BOOM transcription factor. *BMC Plant Biol*. 2015 May
854 16;15:121.
855
- 856 22. Campos NA, Panis B, Carpentier SC. Somatic Embryogenesis in Coffee: The Evolution of
857 Biotechnology and the Integration of Omics Technologies Offer Great Opportunities. *Front Plant Sci*.
858 2017;8:1460.
859
- 860 23. Bonga JM, Klimaszewska KK, von Aderkas P. Recalcitrance in clonal propagation, in particular of
861 conifers. *Plant Cell Tiss Organ Cult*. 2010 Mar 1;100(3):241–54.
862
- 863 24. Etienne H, Breton D, Breitler J-C, Bertrand B, Déchamp E, Awada R, et al. Coffee Somatic
864 Embryogenesis: How Did Research, Experience Gained and Innovations Promote the Commercial
865 Propagation of Elite Clones From the Two Cultivated Species? *Front Plant Sci*. 2018;9:1630.
866
- 867 25. Aguilar-Hernández V, Loyola-Vargas VM. Advanced Proteomic Approaches to Elucidate Somatic
868 Embryogenesis. *Front Plant Sci* [Internet]. 2018 [cited 2019 Jun 19];9. Available from:
869 <https://www.frontiersin.org/articles/10.3389/fpls.2018.01658/full>
870
- 871 26. Pais MS. Somatic Embryogenesis Induction in Woody Species: The Future After OMICs Data
872 Assessment. *Front Plant Sci*. 2019;10:240.
873
- 874 27. Nejat N, Ramalingam A, Mantri N. Advances in Transcriptomics of Plants. *Adv Biochem Eng*
875 *Biotechnol*. 2018;164:161–85.
876
- 877 28. Hofmann F, Schon MA, Nodine MD. The embryonic transcriptome of *Arabidopsis thaliana*. *bioRxiv*.
878 2018 Nov 27;479584.
879
- 880 29. Cao A, Zheng Y, Yu Y, Wang X, Shao D, Sun J, et al. Comparative Transcriptome Analysis of SE
881 initial dedifferentiation in cotton of different SE capability. *Sci Rep*. 2017 Aug 17;7(1):1–13.
882
- 883 30. Indoliya Y, Tiwari P, Chauhan AS, Goel R, Shri M, Bag SK, et al. Decoding regulatory landscape of
884 somatic embryogenesis reveals differential regulatory networks between *japonica* and *indica* rice
885 subspecies. *Scientific Reports*. 2016 Mar 14;6:23050.
886

- 887 31. Salvo SAGD, Hirsch CN, Buell CR, Kaeppler SM, Kaeppler HF. Whole Transcriptome Profiling of
888 Maize during Early Somatic Embryogenesis Reveals Altered Expression of Stress Factors and
889 Embryogenesis-Related Genes. PLOS ONE. 2014 Oct 30;9(10):e111407.
890
- 891 32. Gautier F, Label P, Eliášová K, Leplé J-C, Motyka V, Boizot N, et al. Cytological, Biochemical and
892 Molecular Events of the Embryogenic State in Douglas-fir (*Pseudotsuga menziesii* [Mirb.]). Front Plant
893 Sci. 2019;10:118.
894
- 895 33. Chen Y, Xu X, Liu Z, Zhang Z, XuHan X, Lin Y, et al. Global scale transcriptome analysis reveals
896 differentially expressed genes involve in early somatic embryogenesis in *Dimocarpus longan* Lour. BMC
897 Genomics. 2020 Jan 2;21(1):4.
898
- 899 34. Bertrand B, Montagnon C, Georget F, Charmetant P, Etienne H. Création et diffusion de variétés de
900 caféiers Arabica : quelles innovations variétales ? Cahiers Agricultures. 2012;21:77–88.
901
- 902 35. Ducos JP, Labbe G, Lambot C, Pétiard V. Pilot scale process for the production of pre-germinated
903 somatic embryos of selected robusta (*Coffea canephora*) clones. In Vitro Cell Dev Biol-Plant. 2007 Dec
904 1;43(6):652–9.
905
- 906 36. Etienne H, Bertrand B, Montagnon C, Bodadilla Landey R, Dechamp E, Jourdan I, et al. Un exemple
907 de transfert de technologie réussi dans le domaine de la micropropagation : la multiplication de *Coffea*
908 *arabica* par embryogenèse somatique. Cahiers Agricultures. 2012;21:115–24.
909
- 910 37. Ducos JP, Alenton R, Reano JF, Kanchanomai C, Deshayes A, Pétiard V. Agronomic performance of
911 *Coffea canephora* P. trees derived from large-scale somatic embryo production in liquid medium.
912 Euphytica. 2003 May 1;131(2):215–23.
913
- 914 38. Landey RB, Cenci A, Georget F, Bertrand B, Camayo G, Dechamp E, et al. High Genetic and Epigenetic
915 Stability in *Coffea arabica* Plants Derived from Embryogenic Suspensions and Secondary Embryogenesis
916 as Revealed by AFLP, MSAP and the Phenotypic Variation Rate. PLOS ONE. 2013 Feb 13;8(2):e56372.
917
- 918 39. Georget F, Courtel P, Garcia EM, Hidalgo M, Alpizar E, Breitler J-C, et al. Somatic embryogenesis-
919 derived coffee plantlets can be efficiently propagated by horticultural rooted mini-cuttings: A boost for
920 somatic embryogenesis. Scientia Horticulturae. 2017;216:177–85.
921
- 922 40. TAIR - Home Page [Internet]. [cited 2020 Jan 31]. Available from: <https://www.arabidopsis.org/>
923
- 924 41. Lashermes P, Combes MC, Robert J, Trouslot P, D'Hont A, Anthony F, et al. Molecular characterisation
925 and origin of the *Coffea arabica* L. genome. Mol Gen Genet. 1999 Mar;261(2):259–66.
926
- 927 42. Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, et al. agriGO v2.0: a GO analysis toolkit for the agricultural
928 community, 2017 update. Nucleic Acids Res. 2017 03;45(W1):W122–9.
929

- 930 43. Shiota H, Satoh R, Watabe K, Harada H, Kamada H. C-ABI3, the carrot homologue of the Arabidopsis
931 ABI3, is expressed during both zygotic and somatic embryogenesis and functions in the regulation of
932 embryo-specific ABA-inducible genes. *Plant Cell Physiol.* 1998 Nov;39(11):1184–93.
933
- 934 44. Thakare D, Tang W, Hill K, Perry SE. The MADS-domain transcriptional regulator AGAMOUS-
935 LIKE15 promotes somatic embryo development in Arabidopsis and soybean. *Plant Physiol.* 2008
936 Apr;146(4):1663–72.
937
- 938 45. Zhai L, Xu L, Wang Y, Zhu X, Feng H, Li C, et al. Transcriptional identification and characterization
939 of differentially expressed genes associated with embryogenesis in radish (*Raphanus sativus* L.). *Sci Rep.*
940 2016 Feb 23;6(1):1–13.
941
- 942 46. Boutilier K, Offringa R, Sharma VK, Kieft H, Ouellet T, Zhang L, et al. Ectopic expression of BABY
943 BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell.* 2002 Aug;14(8):1737–49.
944
- 945 47. Lotan T, Ohto M, Yee KM, West MA, Lo R, Kwong RW, et al. Arabidopsis LEAFY COTYLEDON1
946 is sufficient to induce embryo development in vegetative cells. *Cell.* 1998 Jun 26;93(7):1195–205.
947
- 948 48. Braybrook SA, Harada JJ. LECs go crazy in embryo development. *Trends Plant Sci.* 2008
949 Dec;13(12):624–30.
950
- 951 49. Méndez-Hernández HA, Ledezma-Rodríguez M, Avilez-Montalvo RN, Juárez-Gómez YL, Skeete A,
952 Avilez-Montalvo J, et al. Signaling Overview of Plant Somatic Embryogenesis. *Front Plant Sci.*
953 2019;10:77.
954
- 955 50. Trontin J-F, Klimaszewska K, Morel A, Hargreaves C, Lelu-Walter M-A. Molecular Aspects of Conifer
956 Zygotic and Somatic Embryo Development: A Review of Genome-Wide Approaches and Recent Insights.
957 *Methods Mol Biol.* 2016;1359:167–207.
958
- 959 51. Silva AT, Barduche D, do Livramento KG, Ligterink W, Paiva LV. Characterization of a Putative Serk-
960 Like Ortholog in Embryogenic Cell Suspension Cultures of *Coffea arabica* L. *Plant Mol Biol Rep.* 2014
961 Feb 1;32(1):176–84.
962
- 963 52. Pérez-Pascual D, Jiménez-Guillen D, Villanueva-Alonzo H, Souza-Perera R, Godoy-Hernández G,
964 Zúñiga-Aguilar JJ. Ectopic expression of the *Coffea canephora* SERK1 homolog-induced differential
965 transcription of genes involved in auxin metabolism and in the developmental control of embryogenesis.
966 *Physiol Plant.* 2018 Aug;163(4):530–51.
967
- 968 53. Yadav RK, Perales M, Gruel J, Girke T, Jönsson H, Reddy GV. WUSCHEL protein movement mediates
969 stem cell homeostasis in the Arabidopsis shoot apex. *Genes Dev.* 2011 Oct 1;25(19):2025–30.
970
- 971 54. Luerksen H, Kirik V, Herrmann P, Miséra S. FUSCA3 encodes a protein with a conserved VP1/AB13-
972 like B3 domain which is of functional importance for the regulation of seed maturation in Arabidopsis
973 thaliana. *Plant J.* 1998 Sep;15(6):755–64.
974

- 975 55. Vondrakova Z, Dobrev PI, Pesek B, Fischerova L, Vagner M, Motyka V. Profiles of Endogenous
976 Phytohormones Over the Course of Norway Spruce Somatic Embryogenesis. *Front Plant Sci.* 2018;9:1283.
977
- 978 56. Awada R, Campa C, Gibault E, Déchamp E, Georget F, Lepelley M, et al. Unravelling the Metabolic
979 and Hormonal Machinery During Key Steps of Somatic Embryogenesis: A Case Study in Coffee.
980 *International Journal of Molecular Sciences.* 2019 Jan;20(19):4665.
981
- 982 57. Gliwicka M, Nowak K, Balazadeh S, Mueller-Roeber B, Gaj MD. Extensive Modulation of the
983 Transcription Factor Transcriptome during Somatic Embryogenesis in *Arabidopsis thaliana*. *PLOS ONE.*
984 2013 Jul 17;8(7):e69261.
985
- 986 58. Zhao Y. Auxin Biosynthesis: A Simple Two-Step Pathway Converts Tryptophan to Indole-3-Acetic
987 Acid in Plants. *Molecular Plant.* 2012 Mar 1;5(2):334–8.
988
- 989 59. Brenner WG, Ramireddy E, Heyl A, Schmülling T. Gene regulation by cytokinin in *Arabidopsis*. *Front*
990 *Plant Sci.* 2012;3:8.
991
- 992 60. Zubo YO, Blakley IC, Yamburenko MV, Worthen JM, Street IH, Franco-Zorrilla JM, et al. Cytokinin
993 induces genome-wide binding of the type-B response regulator ARR10 to regulate growth and development
994 in *Arabidopsis*. *Proc Natl Acad Sci U S A.* 2017 Jul 18;114(29):E5995–6004.
995
- 996 61. Santiago J, Dupeux F, Round A, Antoni R, Park S-Y, Jamin M, et al. The abscisic acid receptor PYR1
997 in complex with abscisic acid. *Nature.* 2009 Dec 3;462(7273):665–8.
998
- 999 62. Nowak K, Wójcikowska B, Gaj MD. ERF022 impacts the induction of somatic embryogenesis in
1000 *Arabidopsis* through the ethylene-related pathway. *Planta.* 2015;241(4):967–85.
1001
- 1002 63. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Favera RD, et al. ARACNE: An
1003 Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC*
1004 *Bioinformatics.* 2006 Mar 20;7(Suppl 1):S7.
1005
- 1006 64. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software
1007 environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003
1008 Nov;13(11):2498–504.
1009
- 1010 65. Jiménez VM. Regulation of in vitro somatic embryogenesis with emphasis on to the role of endogenous
1011 hormones. *Revista Brasileira de Fisiologia Vegetal.* 2001;13(2):196–223.
1012
- 1013 66. Rose RJ, Nolan KE. Genetic Regulation of Somatic Embryogenesis with Particular Reference to
1014 *Arabidopsis thaliana* and *Medicago truncatula*. *In Vitro Cellular & Developmental Biology Plant.*
1015 2006;42(6):473–81.
1016
- 1017 67. Sugimoto K, Xu L, Paszkowski U, Hayashi M. Multifaceted Cellular Reprogramming at the Crossroads
1018 Between Plant Development and Biotic Interactions. *Plant Cell Physiol.* 2018 01;59(4):651–5.
1019

- 1020 68. Zimmerman JL. Somatic Embryogenesis: A Model for Early Development in Higher Plants. *The Plant*
1021 *Cell*. 1993;5(10):1411–23.
1022
- 1023 69. Su Y-H, Liu Y-B, Zhang X-S. Auxin–Cytokinin Interaction Regulates Meristem Development.
1024 *Molecular Plant*. 2011 Jul 1;4(4):616–25.
1025
- 1026 70. Wickramasuriya AM, Dunwell JM. Global scale transcriptome analysis of *Arabidopsis* embryogenesis
1027 *in vitro*. *BMC Genomics*. 2015 Apr 16;16:301.
1028
- 1029 71. Luo J, Zhou J-J, Zhang J-Z. Aux/IAA Gene Family in Plants: Molecular Structure, Regulation, and
1030 Function. *International Journal of Molecular Sciences*. 2018 Jan;19(1):259.
1031
- 1032 72. Wójcikowska B, Gaj MD. Expression profiling of AUXIN RESPONSE FACTOR genes during somatic
1033 embryogenesis induction in *Arabidopsis*. *Plant Cell Rep*. 2017;36(6):843–58.
1034
- 1035 73. von Aderkas P, Lelu M-A, Label P. Plant growth regulator levels during maturation of larch somatic
1036 embryos. *Plant Physiology and Biochemistry*. 2001 Jun 1;39(6):495–502.
1037
- 1038 74. Fischerova L, Fischer L, Vondrakova Z, Vagner M. Expression of the gene encoding transcription factor
1039 PaVP1 differs in *Picea abies* embryogenic lines depending on their ability to develop somatic embryos.
1040 *Plant Cell Rep*. 2008 Mar;27(3):435–41.
1041
- 1042 75. Zavattieri MA, Frederico AM, Lima M, Sabino R, Arnholdt-Schmitt B. Induction of somatic
1043 embryogenesis as an example of stress-related plant reactions. *Electro Journal of Biotech*. 2010 Jan
1044 15;13(1):12–13.
1045
- 1046 76. Iwase A, Mita K, Nonaka S, Ikeuchi M, Koizuka C, Ohnuma M, et al. WIND1-based acquisition of
1047 regeneration competency in *Arabidopsis* and rapeseed. *J Plant Res*. 2015 May;128(3):389–97.
1048
- 1049 77. Zhou T, Yang X, Guo K, Deng J, Xu J, Gao W, et al. ROS Homeostasis Regulates Somatic
1050 Embryogenesis via the Regulation of Auxin Signaling in Cotton. *Mol Cell Proteomics*. 2016
1051 Jun;15(6):2108–24.
1052
- 1053 78. Nic-Can GI, Galaz-Ávalos RM, De-la-Peña C, Alcazar-Magaña A, Wrobel K, Loyola-Vargas VM.
1054 Somatic Embryogenesis: Identified Factors that Lead to Embryogenic Repression. A Case of Species of the
1055 Same Genus. *PLoS ONE*. 2015;10(6):e0126414.
1056
- 1057 79. Magnani E, Jiménez-Gómez JM, Soubigou-Taconnat L, Lepiniec L, Fiume E. Profiling the onset of
1058 somatic embryogenesis in *Arabidopsis*. *BMC Genomics*. 2017 29;18(1):998.
1059
- 1060 80. Grace SC, Logan BA. Energy dissipation and radical scavenging by the plant phenylpropanoid pathway.
1061 *Philos Trans R Soc Lond B Biol Sci*. 2000 Oct 29;355(1402):1499–510.
1062

- 1063 81. Lepelley M, Cheminade G, Tremillon N, Simkin A, Caillet V, McCarthy J. Chlorogenic acid synthesis
1064 in coffee: An analysis of CGA content and real-time RT-PCR expression of HCT, HQT, C3H1, and
1065 CCoAOMT1 genes during grain development in *C. canephora*. *Plant Science*. 2007 May 1;172(5):978–96.
1066
- 1067 82. Perrois C, Strickler SR, Mathieu G, Lepelley M, Bedon L, Michaux S, Husson J, Mueller L & Privat I.
1068 Differential regulation of caffeine metabolism in *Coffea arabica* (Arabica) and *Coffea canephora* (Robusta).
1069 *Planta* 2015; 241 :179–191.
- 1070 83. Xing Y, Yu T, Wu YN, Roy M, Kim J, Lee C. An expectation-maximization algorithm for probabilistic
1071 reconstructions of full-length isoforms from splice graphs. *Nucleic Acids Res*. 2006 May 1;34(10):3150–
1072 60.
- 1073
- 1074 84. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical
1075 Computing: Vienna, Austria, 2019.
- 1076
- 1077 85. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional
1078 genomic data. *Bioinformatics*. 2016 15;32(18):2847–9.
- 1079
- 1080 86. CRAN - Package pvclust [Internet]. [cited 2019 Jun 20]. Available from: [https://cran.r-](https://cran.r-project.org/web/packages/pvclust/index.html)
1081 [project.org/web/packages/pvclust/index.html](https://cran.r-project.org/web/packages/pvclust/index.html)
- 1082
- 1083 87. Suzuki R, Shimodaira H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering.
1084 *Bioinformatics*. 2006 Jun 15;22(12):1540–2.
- 1085

1086 **Acknowledgments**

1087 We are particularly grateful to Patrick Descombes, senior expert in Genomics at Nestlé Research, for giving
1088 access to the next generation sequencing facilities and continuous support for this work. We are also grateful
1089 to Fabrizio Arigoni for supporting this collaboration. We also would like to thank Estelle Gibault who
1090 contributed to tissue culture and sampling. Finally, we would like to thank Romain Guyot from IRD for the
1091 identification of orthologous genes in *Arabidopsis*, Olivier Darracq for contributing to PCA analyzes and
1092 the ACGC consortium for giving access to the unpublished Arabica genome.

1093

1094 **Funding**

1095 R.A. benefits from a Ph.D. grant and funding from Nestlé and CIRAD.

1096

1097 **Author information**

1098 **Affiliations**

1099 *Nestlé Research—Plant Science Research Unit, 101 avenue Gustave Eiffel, F-37097 Tours Cedex 2, France*
1100 Rayan Awada, Maud Lepelley, Victoria Berry, David Breton & Dominique Cruzillat

1101

1102 *CIRAD, UMR IPME, F-34398 Montpellier, France*

1103 Rayan Awada, Jean-Christophe Breitler, Sophie Lérant, Benoît Bertrand & Hervé Etienne

1104

1105 *UMR IPME, University of Montpellier, CIRAD, IRD, F-34398 Montpellier, France*

1106 Rayan Awada, Claudine Campa, Jean-Christophe Breitler, Sophie Lérant, Benoît Bertrand & Hervé

1107 Etienne

1108
1109 *Nestlé Research, Vers-chez-les-Blanc, 1000 Lausanne 26, Switzerland*
1110 Aline Charpagne

1111
1112 *Sophia Genetics, 9 chemin des mines, 1202 Genève, Switzerland*
1113 Aline Charpagne

1114
1115 *IRD, UMR IPME, F-34398 Montpellier, France*
1116 Claudine Campa

1117
1118 **Contributions**

1119 H.E., B.B. and D.C. designed the research. M.L. set up the experimental design for the transcriptomic study.
1120 D.B. and R.A. established tissue culture protocols. V.B., J-C.B., S.L. and R.A. performed RNA extractions.
1121 A.C. performed RNA-seq experiment. M.L. coordinated the genomics activities and performed
1122 bioinformatics analysis including read mapping, transcriptomes quantification and related exploratory
1123 analysis. B.B and R.A. performed differential gene expression analysis. R.A. performed functional, cluster
1124 and network analysis. C.C. helped extracting secondary metabolism-related genes. R.A. and H.E. wrote the
1125 first draft of the manuscript and all authors revised it.

1126
1127 **Corresponding author**

1128 Correspondence to Rayan Awada.

1129
1130 **Ethics declarations**

1131 **Ethics approval and consent to participate**
1132 Not applicable.

1133
1134 **Consent for publication**
1135 Not applicable.

1136
1137 **Competing interests**
1138 The authors declare that they have no competing interests.

1139
1140
1141

Chapter V

General discussion, conclusion and perspectives

General discussion

At the beginning of this PhD thesis, two ambitious objectives were set: (i) to identify transcriptomic and metabolic candidates characterizing the different developmental stages of SE, (ii) to achieve a proof of concept of a system able to transfer the low-throughput research on SE to a high-throughput level. At the end of this PhD work, we can say that we have reached these objectives. Indeed, we had many concerns at the beginning of this project mainly about the time-consuming sampling (12 months to sample all developmental stages), the risk of heterogeneity and weak reliability of biological replicates since leaf introductions in *in vitro* culture were made at different months, the quality of the extracted RNA (difficulty to obtain non-degraded RNA in trees like coffee which are rich in phenolic compounds, and more particularly when using leaf, callus and embryo tissues), and the feasibility of SE in a miniaturized system. A great deal of attention and care has been taken to address these concerns. Nevertheless, we were quickly reassured by the preliminary results of the transcriptomic pipeline confirming the great quality of extracted RNA which were later mapped on the most well-annotated Arabica reference genome, the reliability between biological replicates and the discrimination between developmental stages. In addition, embryo regeneration in 24-well plates was clearly encouraging and leading to the success of the first part of the thesis. Concerning transcriptomics and metabolomics data analysis, we decided to apply a robust statistical strategy covering the whole SE process in order to identify the main developmental switches, where the most important cell reconfigurations were occurring, without any *a priori* (i.e. avoid using very conventional strategies, or target genes already listed in the bibliography). This strategy has helped us avoid drowning in data analysis. A list of molecular markers has been the fruit of this strategy and is ready to be validated on a pilot scale. Concerning the screening of active compounds on SE, a number of collaborations with private and public partners (INRA Versailles, ECOM laboratory in Sebacco, Nicaragua) are starting to emerge with the objective to scale the innovative miniaturized screening system developed in this thesis to high-throughput platforms. Thus, this PhD work is different from current ongoing researches on SE as it did not only suggest a number of fundamental perspectives but also has led to the elucidation of perspectives which can be directly applied. These applied perspectives are urgently needed by the coffee industry since no evident technical progress on SE has been made in recent years, as also observed with numerous species. This PhD work also resulted in a rich global database which can be analyzed again with different approaches or focuses, and in a rich cryopreserved sample library (only a part of the samples has been used for our studies), ready to be tested with other -omics technologies, giving unprecedented perspectives on SE.

In the following sections, we will explain (i) what were the strengths and limitations of this thesis, (ii) how did the omics tools contribute in characterizing cell fate, (iii) what were the gene and metabolite candidates identified and how it will be possible to validate them, (iv) what are the steps remaining to reach a high-throughput approach, (v) how to define cell totipotency, (vi) what is the link between SE use and coffee breeding and (vii) what are the applied and fundamental perspectives of this work.

1. The key strengths of this thesis

Five key strengths were the highlight of this thesis:

- (i) **The use of efficient and reliable somatic embryogenesis protocols.** This study could not be carried out earlier as only the recent establishment of large-scale protocols for coffee SE on a commercial level ensured both a high level of cell synchronicity and high biological efficiency at each developmental stage. These qualities were not found until now in the coffee SE protocols published (Nic-Can et al., 2013; Ayil-Gutiérrez et al., 2013) mostly employing direct SE. The availability of these large-scale protocols for indirect SE offered a high reliability between replicates as shown in the PCA of total transcripts between biological replicates at each developmental stage. Each developmental stage could be discriminated from the other stages. The same protocol is now employed on all Arabica hybrid varieties commercially used or under selection with an acceptable efficiency as batches with a high number of embryos are obtained at the end of redifferentiation even though the initial quantity of embryogenic callus obtained after 7 months of dedifferentiation is still low for some genotypes (Etienne et al., 2018). The use of bioreactors during embryo conversion to plantlets helped to sharply decrease production costs through decreasing the need of manpower (Ducos et al., 2007; Etienne et al., 2012). Somaclonal variations are now well-understood and mastered (Bobadilla et al., 2015) as high levels of exogenous auxins previously employed during the proliferation of embryogenic cell clusters, appeared to be the main factor of these variations.
- (ii) **Applying cutting-edge –omics technologies to study somatic embryogenesis.** A number of authors (Campos et al., 2017; Mendez-Hernandez et al., 2019; Pais, 2019) are encouraging researchers to consider the recent boom of cutting-edge -omics technologies in order to elucidate the underlying molecular mechanisms during SE in numerous species. Omics tools can have a tremendous impact by providing a wealth of information in the description and elucidation of physiological responses to environmental conditions. We chose to apply the metabolomics and transcriptomic approaches during this study coupled

with more conventional approaches such as histology and cell imaging in order to correspond molecular profiles with their respective cell phenotypes.

- (iii) **Building a transcriptomic and metabolomic events movie during somatic embryogenesis.** We chose to perform a detailed sampling covering the whole SE process from leaf explant dedifferentiation until the development of torpedo-shaped embryos. A total of 14 successive developmental stages and 5 biological replicates were sampled to carry out metabolomics while 12 developmental stages and 4 biological replicates were sampled for transcriptomics. This is the first study with this amount of sampling points in SE. Until now, research has been conducted on either the early days of dedifferentiation with 6 sampling points (Indoliya et al., 2016) or the early days of embryo development with 4 sampling points (Businge et al., 2012). During the redifferentiation episode, our sampling points were very close to each other in order to visualize the early genes responsible for the early dedifferentiation and redifferentiation events.
- (iv) **The possible access to a high quality Arabica genome sequence.** The availability of a *C. arabica* reference genome recently annotated by the Arabica Coffee Genome Consortium (ACGC, unpublished data) which was sequenced using the long-read Pacific Bioscience technology (Mueller et al., 2015) providing high coverage. This reference genome enables a better read mapping than the diploid *C. canephora* ($2n = 2x = 22$) reference genome used until now (Denoeud et al., 2014) not adapted to the allotetraploid Arabica species ($2n = 4x = 44$). In further studies, for each differentially expressed gene, it will be possible to compare the regulation of gene expression at the subgenome level (*C. canephora* versus *C. eugenoides*). In addition, our read sequencing was carried out in 125 bp x 2 (paired-end) enabling to improve mapping quality. The RNA-seq generated a total of 90 million high quality reads per sample, approximately 74% of which were uniquely mapped to Arabica genome.
- (v) **The access to metabolomics tools and databases.** In the frame of a new partnership, our metabolomics study was performed in the laboratory of Joachim Kopka in the Max Planck Institute of Molecular Plant Physiology, Golm (Germany) which owns the most updated database for compounds identified by GC-MS, the Golm Metabolome Database (Kopka et al., 2005, Dethloff et al., 2014).

2. The key limitations of this thesis

Three key limitations can be addressed to our study:

- (i) **The study was performed on one genotype.** In order to be technically and financially able to implement a detailed sampling of all developmental stages of coffee SE, this study was limited to one Arabica genotype, GPFA116 selected by the Nestlé Research center located in Tours. This genotype was a result of the Nestlé Research breeding program and its large-scale propagation would enable a great agronomic impact for the company. This genotype is not considered as a recalcitrant genotype in the laboratory and can yield a sufficient quantity of embryogenic callus after leaf explant dedifferentiation as well as sufficient embryos after redifferentiation allowing a detailed fundamental study to understand underlying mechanisms of SE. A responsive genotype of this type needed to be used to be able to extract significant information from thousands of data. This significant information is translated into molecular markers and can now be tested on numerous genotypes, more or less recalcitrant in order to validate their significance and genericity.
- (ii) **No comparison of different culture conditions effects on SE was performed.** Since the coffee SE process is a significantly long process (12 months between leaf explant dedifferentiation and development of torpedo-shaped embryos and another 8-month period in nursery must be added in order to obtain fully developed plants ready for field planting), only chronologically sequenced and successful stages (i.e. corresponding to expected developmental stages all along the optimal SE process) were chosen to be sampled in order to ensure the detailed sampling and continuity between developmental stages. It was very difficult to conduct at the same time experiments for a same development stage (or time period during coffee SE) aimed at comparing the efficiency of different treatments/environments. Only two contrasted materials were sampled at a same time, the embryogenic vs. the non-embryogenic callus. As the availability of embryogenic callus represents a limiting factor in the coffee SE process because it is obtained after an extremely long culture period (7 months) and also because its availability in quantity and quality is crucial for the success of the further embryo redifferentiation stages, important research efforts were focused on this developmental stage. Many authors have oriented their research only on studying those two contrasted biological materials in various species, as a result of which the bibliography characterizing the embryogenic tissues in comparison to its non-embryogenic counterpart is abundant (Corredoira et al., 2015;

Gautier et al., 2019; Pais, 2019). Ideally, it would have been highly informative to perform –omic analyzes on material resulting from contrasted culture conditions, i.e. modification of the nutritive culture medium, environmental variations (light, temperature, gas exchanges, container type, cell density...) and sampled at a same developmental stage: embryogenic vs. non-embryogenic callus, proliferation of cell clusters with more or less auxin, redifferentiation into embryos with high or low cell density.

- (iii) **The necessary choices when analyzing massive data and the inevitable loss of information.** A number of choices had to be made in order to extract the significant information from these masses of data generated by the omics technologies, mainly transcriptomics. The choice of precise objectives and clear strategy as well as an adapted statistical analysis pipeline prevented drowning during the data analysis step. The robust statistical approach employed both in metabolomics and transcriptomics was very helpful to highlight the main developmental phase switches where the most significant changes were occurring, especially because we wanted to analyze this mass of data without taking any *a priori*. Most often the development phase duration is arbitrarily estimated by *in vitro* culture specialists on the basis of the environmental conditions they apply without taking into account the physiological reality of the plant material. Moreover, we did not want to specialize in the fine study of the regulation of a particular gene, based on previous works. Indeed, even when applying statistical approaches or filters on differentially expressed genes (DEGs), some choices were also needed to be made in order to reduce the mass of data without losing much information. For example, a $\log_2\text{Ratio} > |2|$ was chosen in order to compare DEGs between embryogenic vs. non-embryogenic cells which resulted in a reasonable number of DEGs obtained (499) while a $\log_2\text{Ratio} > |3|$ left us with 0 DEGs.

3. How did the “omics” tools contribute to the characterization of cell fate?

A number of arguments could be made in favor of the application of metabolomics and transcriptomics to characterize the cell fate during somatic embryogenesis:

- (i) **Combining -omics tools with statistical approaches made possible the implementation of a global approach, covering the whole coffee SE process.** For metabolomics, this global approach was applied on 12 developmental stages and 5 independent biological replicates and resulted in the identification and pattern analysis of 104 primary and secondary metabolites. For transcriptomics, this global approach was applied on 14 developmental stages and 4 independent biological replicates and resulted in analyzing

more than 45,000 genes for each developmental stage, with more than 10,000 genes being differentially expressed in total. This global analysis on successive developmental stages is a pioneer work applied on SE, all plant species included, and is different and complementary to the ones done in previous works where authors focused on studying contrasted situations (e.g. embryogenic vs. non-embryogenic cells) (Cheng et al., 2016; Gautier et al., 2019; Chen et al., 2020), or deciphering the overexpression of some particular master regulator genes (e.g. *BBM*, Boutilier et al., 2002; Florez et al., 2015).

- (ii) **The pattern regulating the cell fate seem conserved between plant species.** Our first results from omics analyzes and further comparisons with already published data allowed us to quickly notice that the pathways regulating cell fate are conserved between plant species, as similar genes and metabolites were identified in a same developmental stage in different species. Furthermore, the dynamics of these genes and metabolites appeared also similar. For example, the same transcription factor-related genes were found to be up-regulated in embryogenic cells in a number of species (Yang & Zhang et al., 2010; Florez et al., 2015; Trontin et al., 2016; Chen et al., 2020) and hormone contents were similar during the early redifferentiation and embryo development between coffee and conifers (Vondrakova et al., 2011; Corredoira et al., 2015). This means that our results can serve as a reference for the whole community working on SE.
- (iii) **Evidence of the perception of environmental cue variations at the molecular level.** Omics technologies also allowed us to prove that the environmental drivers of SE (light exposure, growth regulators, cell density) play a major role in regulating cell fate as the variations in these drivers applied to monitor SE were directly visualized at a molecular level. For example, a reduction of light exposure during early dedifferentiation down-regulated genes related to photosynthesis and circadian cycle. This is also valid for the genes related to the response to auxin and cytokinin as well as embryo pattern specification genes which were negatively linked to cell density. This means that the exogenous drivers applied to date in an empirical way allowed an acceptable efficiency of coffee SE protocols but from now on, they can be optimized in a rational way, for example testing growth regulator concentrations can now be monitored on a transcriptional level by RT-qPCR. This monitoring can also be done on a metabolic level since we showed that endogenous hormones followed the variations of the exogenous ones. This rational approach should clearly accelerate the optimization of SE protocols which is suffering from an overall slow technical progress due to the sole use of a trial-and-error approach.

- (iv) **Characterization of the key SE developmental switches.** Many authors recently reported the necessity of a better understanding of the SE molecular mechanisms to clarify the current bottlenecks classically reported in this regeneration process (Campos et al., 2017; Etienne et al., 2018; Pais, 2019) Combining omics technologies with a robust statistical approach could help to understand the current bottlenecks of the SE process. Global transcript patterns helped to clearly identify seven important developmental phases (leaf cells, cells in dedifferentiation, primary callus cells, embryogenic cells, embryogenic cell clusters, pro-embryogenic masses, embryonic cells) with very specific patterns, leading to the characterization of six key developmental phase switches, which are strategic for the whole biological efficiency of somatic embryo regeneration. Using metabolomics data, the same developmental phases as the ones identified with transcriptomic data were found with the exception of the callus phase where primary callus, embryogenic callus and embryogenic cell clusters belonged to the same developmental phase while they were clearly discriminated with transcriptomic data. Since globally the same developmental phases were identified in both metabolomic and transcriptomic approaches, the developmental phase switches unambiguously represent the key transitions between two cell fates and where the biological efficiency of the whole SE process is determined. In this PhD work, we showed that each developmental phase switch is characterized by huge reconfigurations in gene expression, hormone contents and metabolite accumulation.
- (v) **A specific transcriptomic and metabolomic signature has been assigned at each developmental stage.** The fact that each of the metabolic and transcript profiles corresponded with a specific cell developmental stage demonstrates the great resolving power of -omics approaches. Each developmental stage has a highly specific signature based on metabolic and transcript fingerprints which bring rich information on cell physiology, as many of these stages are undistinguishable by morphological and histological approaches (e.g. during early stages of dedifferentiation and redifferentiation). This global analysis is a proof of concept that transcripts and metabolites are good markers of all cell fate transitions and, in the near future, could be used to understand and better pilot the optimization of the SE culture conditions by using our study as a milestone of successful developmental stages. This goes beyond morphological and histological descriptions, which were until now the most common way to support empirical protocol optimization. We bring here a proof of the existence of clear correlations between the contrasted molecular profiles and specific SE stages confirming a number of studies in

different species having suggested this hypothesis (Robinson et al., 2009; Vondrakova et al., 2011; Businge et al., 2012; Yang et al., 2012; Indoliya et al., 2016).

- (vi) **Identification of molecular markers.** This study allowed the identification of a number of candidate genes and metabolites (see section 4) that should be used as markers of specific developmental phases, paving the way for a high possibility to rationally pilot the optimization of SE protocols and putting an end to empirical research applied since the pioneer works of Stuart et al. (1958) and Reinert (1958) on carrot, and those of Dublin on coffee in the eighties (Dublin, 1984).

4. Identification of transcriptomic and metabolic markers during the key developmental phases of somatic embryogenesis

We identified a set of molecular markers (metabolic and transcriptomic) during the transitions of cell fate occurring during coffee SE that are summarized in Table 2. A molecular marker is firstly defined by a clear pattern. Transcriptomic candidates were selected on the basis of a pattern revealing a switch-on (or sharp up-regulation) followed by a switch-off (or sharp down-regulation) at one or several developmental stages. Transcriptomic candidates were also chosen based on their gene expression level, i.e. a gene with high expression value is preferable to a gene with low expression value. Similarly, metabolic candidates were selected on the basis of a clear pattern marked by an over-accumulation of a metabolite in a developmental phase followed by an under-accumulation in the following developmental phase. Some molecular markers were also identified through a comparison between embryogenic and non-embryogenic cells. As our study was limited to only one genotype, further RT-qPCR analyses realized on samples proceeding from i) a set of genotypes more or less recalcitrant to the induction of somatic embryo or ii) compared culture conditions with contrasted biological efficiencies, will be crucial to validate the candidate genes. Indeed, these markers should also be validated on Robusta clones since the Robusta cultivated species corresponds to 40% of the coffee business. The CIRAD/ECOM partnership enabled the *in vitro* culture of a panel of Arabica genotypes (more or less recalcitrant to the current SE protocols) in the Atlantic SA/ECOM commercial propagation laboratory based in Sebaco (Nicaragua). The biological performance at each stage of the SE process is well known for a wide range of more or less recalcitrant Arabica hybrids. It would be easy to verify the correlation between the expression level of marker genes and the level of biological performances. RT-qPCR tests can be easily made in any facility able to perform RNA extractions and

cDNA synthesis. Concerning candidate metabolites, some of them can be directly added to the medium in order to evaluate their effect at different concentrations. This is supported by the fact that some of candidate metabolites are already part of the media composition in current protocols and testing them at different concentrations can be key to validating them. Preliminary validations of the positive effects of proline, maltotriose and myo-inositol are currently being tested in the Nicaraguan *in vitro* laboratory. Once transcriptomic markers are validated, they should be used to efficiently pilot the SE process optimization in a rational, reliable, cost-effective and faster way by testing a number of contrasted culture conditions in order to select the optimal one, particularly in the case of recalcitrant genotypes or species (**Figure 18**). Similarly, once metabolic markers are validated, they can be directly used in order to increase the amount of embryogenic cells or the number of regenerated embryos. For the metabolic markers that do not enter in the media composition, compounds able to modulate the marker activity and possible to be added to media composition should be tested (e.g. adding glucose to increase glucose-6-phosphate activity (identified marker)). Indeed, this approach is less direct and could be more time-consuming.

Table 2. Up-regulated transcriptomic markers and over-accumulated metabolic markers characterizing the different developmental phases of coffee somatic embryogenesis.

Molecular markers	Developmental phases						
	Leaf cells	Cells in dedifferentiation	Primary callus cells	Embryogenic cells	Embryogenic cell clusters	Cells in redifferentiation	Embryonic cells
Transcriptomic markers							
Transcription factor-related genes	<i>SERK1</i>	<i>SERK2</i> <i>CLV3</i>		<i>BBM</i> <i>ABI3</i> <i>LEC1</i> <i>AGL15</i> <i>WOX2</i> <i>WUS</i>		<i>FUS3</i>	<i>FUS3</i>
Auxin-related genes		<i>AUX/IAA</i>		<i>ARF5</i>	<i>YUCCA4</i>	<i>YUCCA4</i>	
Cytokinin related genes				<i>ARR10</i>			<i>ARR10</i>
ABA-related genes							<i>ABI5</i>
Ethylene-related genes			<i>ACS7</i>				
Secondary metabolism-related genes	<i>PAL</i> <i>CHS</i> <i>4CL</i> <i>DXMT</i>					<i>CHS</i> <i>4CL</i> <i>COMT1</i> <i>HQT</i>	<i>CHS</i> <i>4CL</i> <i>COMT1</i> <i>HQT</i> <i>XMT</i>
Metabolic markers							
Primary metabolites		Glucose-6-phosphate TCA cycle		Maltotriose Myo-inositol Proline Starch Leucine (u) TCA cycle (u)	Starch	Glucose-6-phosphate Myo-inositol Tryptophan Proline	Glucose-6-phosphate TCA cycle Myo-inositol Proline Putresceine
Secondary metabolites	3-CQA 4-CQA Caffeine					5-CQA	5-CQA
Hormones		ACC IAA Zeatin	ACC	ABA Zeatin		IAA	ABA Zeatin

(u) = under-accumulated metabolic markers

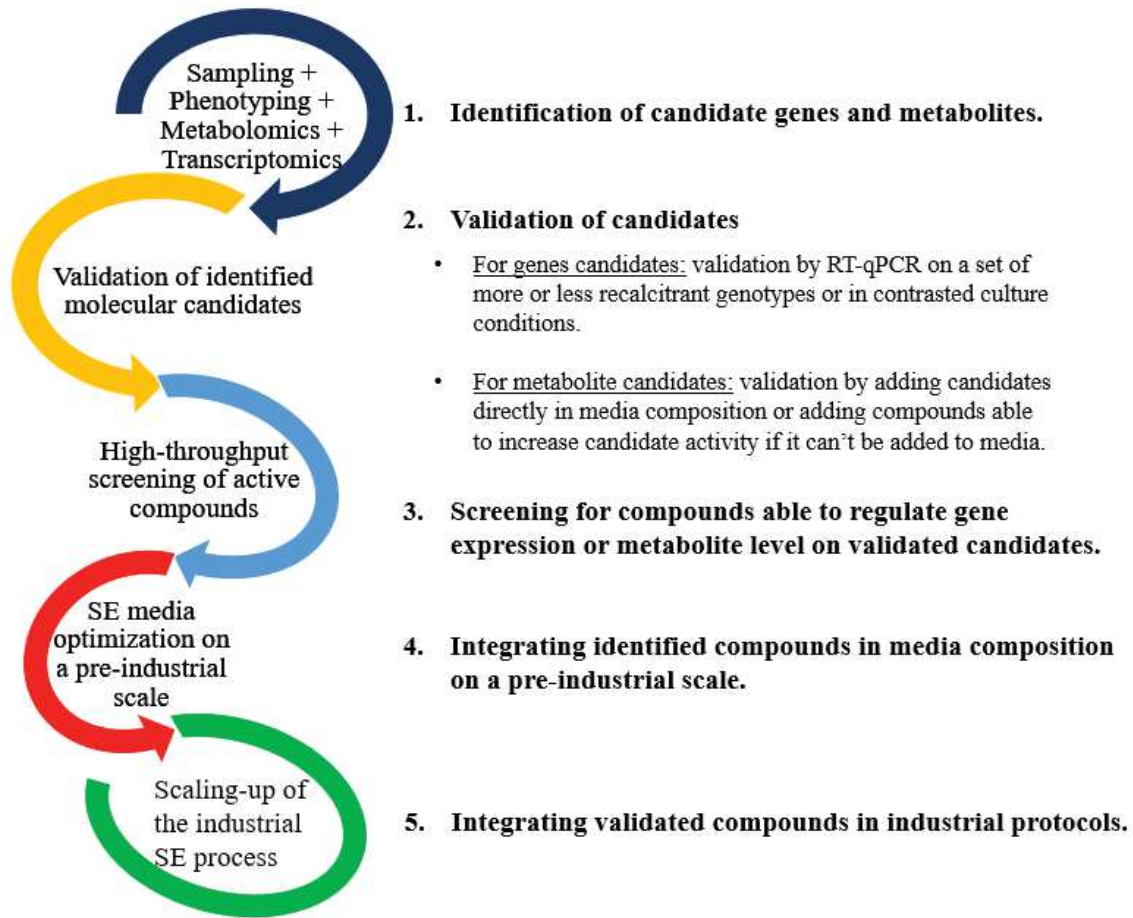


Figure 18. Schematic representation of the research pipeline leading to the scaling-up of the Arabica somatic embryogenesis (SE) process.

5. Is the high-throughput optimization of somatic embryogenesis still a dream for biotechnologists?

We proved in this study that a major step of SE processes, i.e. embryo regeneration from embryogenic cell clusters, can be successfully miniaturized and automated allowing easy, quick, cheap and reliable testing of numerous compounds at varied concentrations. Switching from conventional 250-ml Erlenmeyer flask culture containers filled with 100 ml of culture medium to 24-well plates with 1 ml of available nutritive medium allowed a miniaturization factor of 100. With this strong downscaling, consideration could be given to testing the effectiveness of a much larger number of molecules at different concentrations, while reducing the space required (2.5 cm² required per well compared to 55.4 cm² required per flask and the possibility of stacking multi-well plates). In addition, automation led to a considerable gain in time and reduced manpower, where only 1 working full-time employee

(FTE) was needed to supervise the pipetting platform. For instance, the pilot screening carried out using HDACi treatments required 36 plates, 864 wells, i.e. 864 independent culture environments. The same experiment in flasks was simply unmanageable, as 1 FTE can only inoculate 100 flasks a day, not to mention the space required, whereas inoculation of several hundred wells can be done within hours. Many authors have also shown the importance of miniaturization and automation in animal or plant cell screening systems (Mayr & Fuerst, 2008; Hughes et al., 2011). For instance, Rodriguez-Furlan *et al.* (2016) also showed that the 24-well microplate format used to screen for compounds modulating *A. thaliana* seedling growth required a minimum of sample handling and allowed automatic acquisition of quantitative data and rapid monitoring of morphological traits (leaf and root growth). This container format, along with automation, allowed the screening of 10,000 structurally diverse compounds in a limited time period.

Our focus was turned to the SE differentiation phase, where research tends to maximize embryo yields which is the most “spectacular” step, but other key phases of the coffee SE process could also be easily optimized as a result of the “automated screening of active compound” method, such as the explant-to-callus phase, or the proliferation phase.

Some perspectives of this work can be directly applied and are further discussed in the “Perspectives” section.

6. How can cell totipotency be defined?

Researchers agree on the fact that SE is the best illustration of the totipotency phenomenon. The classical hypothesis for SE initiation affirms that molecular reprogramming of somatic cells is essential. In this process, already differentiated cells regain the totipotency capacity (Fehér et al., 2003; Yang & Zhang, 2010) under high auxin supply (undifferentiated cells are classically maintained under high auxin levels) and under the right stimuli can differentiate into new embryos. Although this is the most accepted theory about totipotency in SE, since recently other hypotheses are also proposed. For instance, Campos et al. (2017) hypothesized that coffee leaf cells don't pass through the process of dedifferentiation but some pools of meristematic cells are already present in leaves and under the right stimulus can proliferate and differentiate to form new viable embryos. More generally, there is currently a debate on the origin of embryogenic cells and more precisely about the possible pre-existence in adult plant tissues of reservoirs of undifferentiated cells that can be mobilized to evolve rapidly towards regenerative processes. The step of dedifferentiation would no longer be the norm!

The hypothesis of ‘pre-existing plant stem cells’ was also discussed by Verdeil (2007) and more recently by Fehér (2019). However, the present study gives us three arguments in favor of the classical concept of totipotency: (i) When a leaf is cut into small squares, an unorganized callus begins to appear on the wounded edges of leaf explants no matter the position of the cutting on the leaf, which refutes the hypothesis of the presence of only a small pool of meristematic cells; (ii) on the other hand, despite numerous histological sections of leaves specifically targeting these wounded edges, we have never observed pre-existing clusters of meristematic cells (pluripotent) or embryogenic cells (totipotent), which is in agreement with Verdeil (2007); (iii) our metabolic and transcriptomic data showed a huge reconfiguration of metabolic pathways before the establishment of the expression of totipotency in embryogenic cells. A total shut-off of genes encoding glycolysis, TCA cycle and secondary metabolites and their respective metabolites characterized the embryogenic cells. Magnani et al. (2017) reported in *Arabidopsis* that embryogenic cells shut off biochemical pathways related to sugar metabolism and activate the transcriptional machinery. This is in agreement with our results where a storage of starch is easily seen in histological studies of embryogenic cells indicating an inhibition in glucose hydrolysis. This shut-off in metabolic pathways suggests that an epigenetic regulation is taking place in order to reprogram the cell function and could be an interesting subject to study (see Perspectives). In addition, a set of regulating genes are switched on during this developmental stage (transcription factor-related genes) which confirms the fact that the transcriptional machinery is coding a new program for the cell. These massive metabolic disruptions occurring during dedifferentiation would not have been necessary under the hypothesis of pre-existing pools of stem cells. A more continuous, spontaneous and easily triggered process would have supported this last hypothesis. Nonetheless, both theories agree on the fact that cells are reprogrammed again at the onset of redifferentiation leading to the re-activation of metabolic pathways (Campos et al., 2017, Magnani et al., 2017).

7. An urgency to optimize somatic embryogenesis in order to disseminate the genetic progress of varieties resulting from breeding programs

In the early 1990s, the strategy of breeding programs was to select varieties with a specific trait (e.g. resistance to leaf rust), then the selection of resistant varieties enters a long breeding cycle of approximately 25 years until the generation of 1 or 2 pure lines having the selected trait in a stable form. This was the case for the “Catimor” cultivars which were selected for their resistance to leaf rust, only the “Costa Rica 95” variety was the result of a long breeding program in Costa Rica and the

“Obata” variety was the result of the one in Honduras. These breeding programs have neglected other traits which systematically resulted in the significant decrease in their cup quality. However, approximately 10 years after their large-scale dissemination, and mainly due to the increasing damages caused by climate change, the Catimor varieties started to lose their resistance to leaf rust. Only Brazil and Colombia can still support such pedigree selection programs of questionable effectiveness. More varieties able to adapt to the abiotic and biotic stresses caused by climate change were urgently needed, which led to the reshaping of the breeding program strategies, and ultimately led to selecting F1 hybrids (Bertrand et al., 2012; Etienne et al., 2012, 2016). This strategy saved a lot of time (from 25 years of selection trials to 8-10 years) and resulted not only in one variety but a whole panel of varieties (about thirty hybrids currently studied in the frame of the Breedcafs H2020 project (<https://www.breedcafs.eu/>) funded by the European Union). These varieties were selected for their adaptability in agroforestry systems (a more sustainable culture system than full-sun systems, particularly in the context of climate change), their vigor and their excellent cup quality. Hybrid varieties are today unanimously recognized by the industry and highly recommended to coffee producers, particularly in the most accessible WCR variety catalogue (<https://varieties.worldcoffeeresearch.org/varieties>). A number of breeding programs to select F1 hybrids have started to emerge in other countries. For example, CATIE (<https://www.catie.ac.cr/>) in Central America, WCR (<https://worldcoffeeresearch.org/>) in Salvador and Nestlé in Colombia also have selected their own panel of F1 hybrids. However, in order to be propagated, F1 hybrids need to be cloned vegetatively. SE was chosen as the most promising tool for the large-scale propagation of elite varieties. Consequently, optimizing SE protocols is urgent and mandatory to be able to disseminate the genetic progress made by breeding programs over the years. For example, the “H1” F1 hybrid variety also commercially called “Centroamericano H1” has proven its quality potential by earning a score of 90 out of 100 points in the world’s leading competition for high quality coffees, the Cup of Excellence, in Nicaragua in 2017 (worldcoffeeresearch.org/news). However, this variety is rather recalcitrant at the different SE stages (the less responsive of a panel including around 20 hybrids). Another variety, the “H14”, a “sibling” of the “H1” coming from the same parents, responds well to the current SE protocols but has the problem of empty cherries when bearing its fruits in the field. Another example is the propagation of sterile male variety “CIR-SM01” by SE before being crossed with a pure line “Marsellesa” to obtain the “Starmaya” variety which is disseminated by seeds (Georget et al., 2019). However, the “CIR-SM01” is recalcitrant at the embryogenic callus stage, this limiting step needs to be optimized to ensure the scaling-up of seed orchards and the further large-scale dissemination of the Starmaya hybrid.

Thus, SE technique has to be scaled-up and democratized in order to be able to sustain the large-scale dissemination of new hybrid varieties and of the genetic progress more generally in all the producing countries in America, Asia and Africa. The transition from an annual production of a few million hybrids to a production of several hundred million hybrids will have to be successfully achieved in the coming years, while solving the problems of plant transport, whether between countries or between different areas of the same country. Of course, the first solutions are to optimize the SE protocols and to carry out numerous technological transfers of the resulting optimized SE process in new production laboratories. However, the production cost of a vitroplant will always remain higher than that of a simple traditional seedling. Complementary horticultural techniques such as rooted mini-cuttings (Georget et al., 2017) have recently been developed allowing the vitroplant batches produced in sophisticated micropropagation laboratories to be re-multiplied in nurseries and therefore at reduced costs. The final price obtained is closer to that of seedling with each re-multiplication cycle with the advantage of producing a plant material with a real added value compared to seedlings of varieties totally unsuited to the current climate change and to the current market. These recent solutions must be multiplied and adapted. For example, these complementary horticultural methods can help to propagate the genetic gain to the most remoted areas where many small coffee growers live. Multiplying the number of nurseries will help to reach more and more remote producers while at the same time solving the problems of long-distance transport of young plants by truck. Local nurseries can also help coffee growers appropriate the new varieties because growers are still reluctant to material directly coming from laboratories. More broadly, we have to help in democratizing the SE technique by granting an open-access to all of our scientific achievements as sharing them with the scientific community to ensure a rapid progress in optimizing SE regardless of the species.

Situations similar to those of coffee are encountered in other woody species and especially in tropical trees that have not been invested in by seed groups or international research centers (e.g. CGIAR, <https://www.cgiar.org/>), and whose scientific communities are small. For these species, we need to think about a new governance between private and public actors within the community working in genetics to allow more collaborations, exchanges of genomic and genetic resources, sharing of biotechnological tools such as SE and joint pre-breeding and breeding programs.

Conclusion

We have successfully demonstrated that the characterization of SE stages cannot be sufficient at the phenotypic level (macroscopic and cellular) and is more efficient at the molecular level through metabolic and transcriptomic profiles. These profiles reflect dramatic changes that occur during transitions of cell fate and are closely linked to environmental conditions (exogenous hormones, cell density, light exposure). Thus, through identifying molecular markers, the SE process could be piloted in a rational way and paves the way, for the first time, to new perspectives after many years marked by very little scientific progress. A number of candidate genes and metabolites have been identified and are awaiting validation. The next step would be to validate the candidate genes by testing them on a panel of more or less recalcitrant genotypes (whose biological efficiencies are known at the different stages of SE) by performing RT-qPCR tests or by comparing different culture conditions with contrasted biological efficiencies (**Figure 19**). It will thus be possible to verify the existing correlation between expression intensity of marker genes and the biological efficiency of the developmental switch that they should pilot. Some candidate metabolites identified as markers of embryogenic potential (embryogenic callus stage) already enter into the composition of the media used in industrial SE protocols; this is the case with myo-inositol and maltose for example. The concentrations of these compounds can therefore be directly modified in the composition of the medium in order to assess their impact on the embryogenic potential, which until now represents the most limiting stage of SE in coffee. For the metabolic markers that do not enter in the media composition, compounds able to modulate the marker activity and possible to be added to media composition should be tested. Once the transcriptomic and metabolomic markers have been validated, the high-throughput approach can be implemented in order to accelerate the optimization of the SE protocols by simultaneously testing hundreds of compounds intervening in the same metabolic pathways in order to select those which are able to modulate the expression levels of these transcriptomic markers or the concentrations of metabolic markers. We have carried out the proof of concept of the high-throughput approach on a pilot scale, the next step will be to carry out this approach on an industrial scale which will certainly require new equipment and new laboratory practices (**Figure 19**). The optimization of SE protocols will thus be able to accompany breeding programs to disseminate the genetic progress on a large scale in order to support coffee growers during the challenges caused by climate change. Our data also show that the mechanisms regulating SE are conserved between species and could therefore be useful for the other coffee cultivated species – Robusta – and for the entire scientific community.

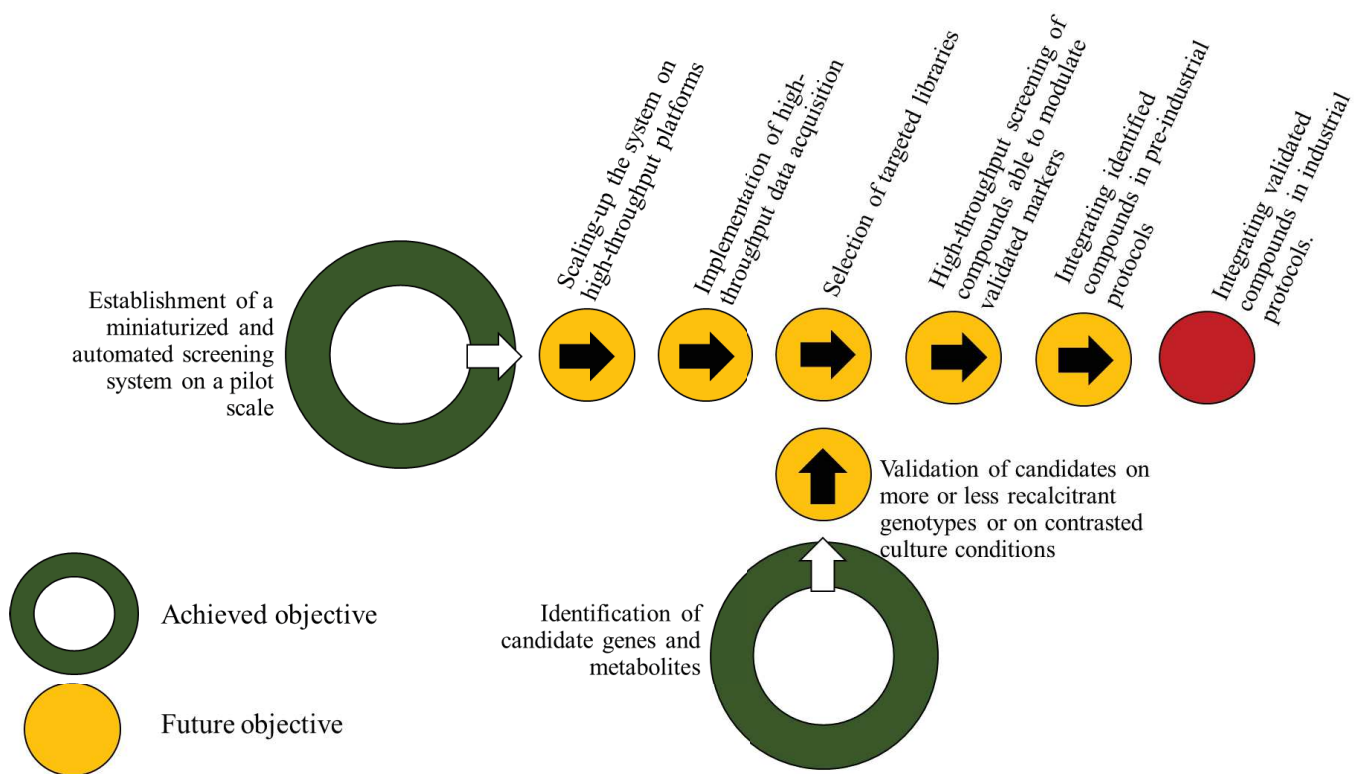


Figure 19. Representation of the objectives set to reach a rational and accelerated optimization of SE protocols. In green, achieved objectives and in orange, future objectives.

Perspectives

- **Perspectives that can be directly applied**

- (i) **Scaling-up the high-throughput screening system.** Our results clearly demonstrate the feasibility, in plants, of automated screening for active compounds in miniaturized SE processes. This proof-of-concept represents a key step towards high-throughput optimization of cell culture media in plant micropropagation through SE. For coffee, achieving the intended scaling-up, i.e. high-throughput screening of hundreds of libraries to optimize SE protocols, still requires three important upgrades to our screening system: (1) new equipment, e.g. high-throughput pipetting platforms (on-going collaboration with INRA Versailles); the platforms are coupled with a robotic arm and enable full automation of the screening process; (2) new laboratory strategies, e.g. preparation and maintenance in cryo-banks (conservation in liquid nitrogen) of large amounts of embryogenic tissues; (3) new data acquisition strategies; as embryo development occurs in all 3 dimensions, embryo quantification is still difficult to analyze precisely with an automated 2-dimensional scanning microscope. To that end, early SE-markers would be very helpful.
- (ii) **Validation of transcriptomic and metabolic markers** (see Discussion, section 4)
- (iii) **High-throughput screening can be piloted in a rational way.** A desired cell fate can be enhanced by strictly selecting libraries able to alter the respective molecular markers identified on a metabolic and transcriptional level (after their validation) (**Figure 19**) in order to save time and money. For example, embryogenic cells are characterized by an under-accumulation of their TCA cycle metabolites, a high-throughput screening of libraries targeted to inhibit the TCA cycle in primary callus cells may probably result in obtaining embryogenic cells in a faster way than with current protocols (7 months). We are convinced that non-optimal culture conditions during callus induction could be optimized by using the above described approach leading in a 4-month reduction of the culture time. Similarly, embryogenic cells are characterized by a switch-on of the *BBM* gene, a high-throughput screening of libraries targeted to up-regulate *BBM* may result in increasing the number of embryogenic cells which are the crucial factor for scaling-up the SE process and the reduction of production costs (**Figure 19**). Identified active compounds should be first tested on a pre-industrial scale before adding them in the industrial SE protocols.

- **Additional fundamental studies would help in deciphering the mechanisms underlying somatic embryogenesis**

Five additional fundamental studies will be of interest in decoding the mechanisms underlying SE:

- (i) **Studying gene expression regulation at the Arabica sub-genome level.** One study can be directly carried out using the same data as our transcriptomic study and which consists in analyzing the Arabica sub-genome regulation of gene expression during the SE key developmental stages. *Coffea arabica* is an allotetraploid species ($2n=4x=44$) resulting from a spontaneous hybridization between two diploid species: *C. eugenioides* and *C. canephora* (Lashermes et al., 1999). This additional study is allowed by the fact that the new Arabica reference genome (ACGC, unpublished data) can discriminate the genes belonging to each sub-genome (thanks to the identification of discriminant SNPs) and consequently, allows us to analyze the expression rate of each sub-genome (respective proportions of transcripts for a same gene from each sub-genome). It would be interesting to observe which of those 2 sub-genomes takes the lead during the different developmental phases of the coffee SE process. We hypothesize that since the SE process is established in strict culture conditions with a controlled and rather elevated temperature (26-28°C) which is more favorable for the *C. canephora* species, the sub-genome of this species could probably take the lead at the different SE developmental phases.
- (ii) **Proteomics carried out on the same samples can give new insights on SE.** Since the biological material sampled during the different developmental stages was of great quality, presenting high reliability between biological replicates and sampled in sufficient amounts to perform metabolomic and transcriptomic approaches during this thesis, a sufficient amount of the same samples was saved to be able to carry out a proteomic approach. We believe that proteomics combined with transcriptomics, metabolomics and cell imaging can give a much clearer idea of the mechanisms regulating cell fate. Protein dynamics can be directly correlated to their respective genes in order to evaluate the amount of post-transcriptional modifications. A number of studies have started to perform proteomics on a developmental stage of interest, the coffee embryogenic cells (Tonietto et al., 2012, Campos et al., 2016). Our histo-cytological study also showed a remarkable blue staining of embryogenic cells suggesting high amounts of proteins during this key stage.
- (iii) **Investigating the role of epigenetic regulation in SE.** Since a high majority of the identified differentially expressed genes were down-regulated or even switched-off in

primary callus cells and in embryogenic cells, we hypothesize that an epigenetic regulation may be taking place during these developmental stages. It is interesting to carry out an epigenetic approach on these two stages, and more particularly on the primary callus stage since it represents the major bottleneck causing us to wait 7 long months before obtaining embryogenic callus. We are convinced that a study concerning the amount of methylation of coffee DNA, the amount of acetylation of histones, and the amount of small miRNAs present during these two stages could help us in unravelling gene expression patterns. When 1 μ M of TSA, a histone deacetylase inhibitor, was applied on embryogenic cell clusters during our pilot screening (chapter II), a significant increase in the embryogenic potential was noted confirming our hypothesis of the presence of an epigenetic regulation.

- (iv) **Visualizing global gene networks and metabolic pathways.** On some key developmental stages which have close sampling points, for example the early days of the redifferentiation phase, the MapMan tool (Thimm et al., 2004) could be applied to visualize and compare existing omics data on a transcriptomic and metabolic level in order to visualize the most significantly expressed biological pathways and deduce master regulators of gene networks. This approach was firstly developed for microarray data on Arabidopsis to visualize the effects of different culture conditions on metabolic pathways but is now being extended to RNA-seq data and other crop species (Usadel et al., 2009). In the frame of the Breedcafs H2020 project (<https://www.breedcafs.eu/>) funded by the European Union, MapMan and Robina (RNA-seq data analysis pipelines) were adapted to coffee by the Bjorn Usable team in Germany who participated to the project with CIRAD and IRD.
- (v) **Comparing plant totipotent cells and animal stem cells.** Our study can take a whole other dimension by comparing coffee (and other plant species) totipotent cells (embryogenic cells) to human and other mammalian somatic stem cells. The plasticity of somatic stem cells has stimulated great interest due to their numerous potential applications, in addition to basic and in clinical research, as a regenerative therapy for numerous human diseases or injuries (Grimaldi et al., 2008). Stem cells can be found in three states: quiescent, proliferating or undergoing differentiation to fulfill a specific function (**Figure 19**, Costello & Franklin, 2013). Their differentiation is a result of an environmental signal that causes variations in endogenous growth factors and cytokines followed by redox reactions. The quiescent state is characterized by a shut-off of the majority of metabolic pathways while an increase in glycolysis and oxygen consumption is noted in the proliferating state (Coller, 2019). The TCA cycle is highly active during cell differentiation

(Costello & Franklin, 2013; Ito & Suda, 2014). This is in agreement with our results if we compare the quiescent state to embryogenic callus stage, the proliferating state to embryogenic cell clusters and the differentiation state to the pro-embryogenic masses. Further studies could give first insights about deprogramming specialized adult cells.

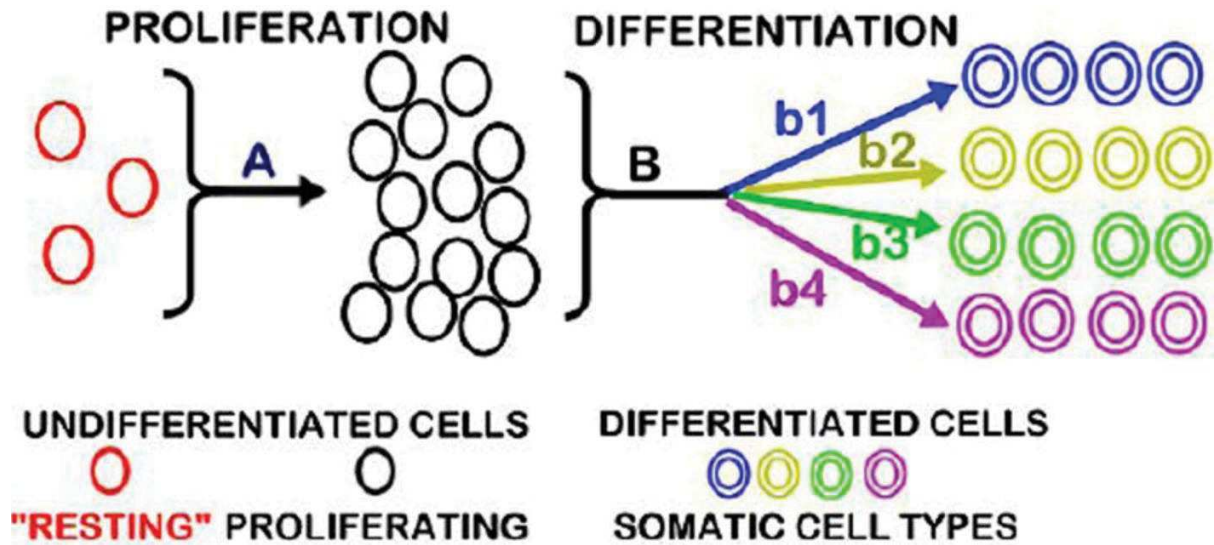


Figure 20. A representation of the stem cell differentiation process. **A**= the proliferation phase for increased population of undifferentiated stem cells. **B**= the differentiation phase leading to specific functional cell types (b1, b2, b3, b4) (from Costello & Franklin, 2013)

References

- Aderkas, Patrick von, Marie-Anne Lelu, and Philippe Label. 2001. "Plant Growth Regulator Levels during Maturation of Larch Somatic Embryos." *Plant Physiology and Biochemistry* 39 (6): 495–502. [https://doi.org/10.1016/S0981-9428\(01\)01271-2](https://doi.org/10.1016/S0981-9428(01)01271-2).
- Alves, Gabriel Sergio Costa, Luana Ferreira Torres, Eveline Déchamp, Jean-Christophe Breitler, Thierry Joët, Frédéric Gatineau, Alan Carvalho Andrade, Benoît Bertrand, Pierre Marraccini, and Hervé Etienne. 2017. "Differential Fine-Tuning of Gene Expression Regulation in Coffee Leaves by CcDREB1D Promoter Haplotypes under Water Deficit." *Journal of Experimental Botany* 68 (11): 3017–31. <https://doi.org/10.1093/jxb/erx166>.
- Arnold, Sara von, and David Clapham. 2008. "Spruce Embryogenesis." *Methods in Molecular Biology (Clifton, N.J.)* 427: 31–47. https://doi.org/10.1007/978-1-59745-273-1_3.
- Arnold, Sara von, Izabela Sabala, Peter Bozhkov, Julia Dyachok, and Lada Filonova. 2002. "Developmental Pathways of Somatic Embryogenesis." *Plant Cell, Tissue and Organ Culture* 69 (3): 233–49. <https://doi.org/10.1023/A:1015673200621>.
- Arroyo-Herrera, A., A. Ku Gonzalez, R. Canche Moo, F. R. Quiroz-Figueroa, V. M. Loyola-Vargas, L. C. Rodriguez-Zapata, C. Burgeff D'Hondt, V. M. Suárez-Solis, and E. Castaño. 2008. "Expression of WUSCHEL in Coffea Canephora Causes Ectopic Morphogenesis and Increases Somatic Embryogenesis." <https://pubag.nal.usda.gov/catalog/732399>.
- Ayil-Gutiérrez, Benajmín, Rosa Galaz-Ávalos, Eduardo Peña-Cabrera, and Victor Loyola-Vargas. 2013. "Dynamics of the Concentration of IAA and Some of Its Conjugates during the Induction of Somatic Embryogenesis in Coffea Canephora." *Plant Signaling & Behavior* 8 (11): e26998. <https://doi.org/10.4161/psb.26998>.
- Barbón, Raúl, Elio Jiménez, and Walter Preil. 2008. "Influence of in Vitro Environment on Somatic Embryogenesis of Coffea Arabica L. Cv. Caturra Rojo: The Effects of Carbon Dioxide on Embryogenic Cell Suspensions." *Plant Cell, Tissue and Organ Culture* 95 (2): 155–61. <https://doi.org/10.1007/s11240-008-9427-4>.
- Baudino, S., S. Hansen, R. Brettschneider, V. F. Hecht, T. Dresselhaus, H. Lörz, C. Dumas, and P. M. Rogowsky. 2001. "Molecular Characterisation of Two Novel Maize LRR Receptor-like Kinases, Which Belong to the SERK Gene Family." *Planta* 213 (1): 1–10. <https://doi.org/10.1007/s004250000471>.
- Berthouly, M., and N. M. Michaux-Ferriere. 1996. "High Frequency Somatic Embryogenesis in Coffea Canephora." *Plant Cell, Tissue and Organ Culture* 44 (2): 169–76. <https://doi.org/10.1007/BF00048196>.
- Bertrand, B., E. Alpizar, L. Lara, R. SantaCreo, M. Hidalgo, J. M. Quijano, C. Montagnon, F. Georget, and H. Etienne. 2011. "Performance of Coffea Arabica F1 Hybrids in Agroforestry and Full-Sun Cropping Systems in Comparison with American Pure Line Cultivars." *Euphytica* 181 (2): 147–58. <https://doi.org/10.1007/s10681-011-0372-7>.
- Bertrand, B., H. Etienne, C. Cilas, A. Charrier, and P. Baradat. 2005. "Coffea Arabica Hybrid Performance for Yield, Fertility and Bean Weight." *Euphytica* 141 (3): 255–62. <https://doi.org/10.1007/s10681-005-7681-7>.
- Bertrand, Benoît, Amélie Bardil, Hélène Baraille, Stéphane Dussert, Sylvie Doubeau, Emeric Dubois, Dany Severac, Alexis Dereeper, and Hervé Etienne. 2015. "The Greater Phenotypic Homeostasis of the Allopolyploid Coffea Arabica Improved the Transcriptional Homeostasis Over That of Both Diploid Parents." *Plant & Cell Physiology* 56 (10): 2035–51. <https://doi.org/10.1093/pcp/pcv117>.
- Bertrand, Benoît, Christophe Montagnon, Frederic Georget, Pierre Charmetant, and Hervé Etienne. 2012. "Création et diffusion de variétés de caféiers Arabica: quelles innovations variétales?" *Cahiers Agricultures* 21: 77–88. [https://doi.org/Création et diffusion de variétés de caféiers Arabica : quelles innovations variétales ?](https://doi.org/Création%20et%20diffusion%20de%20variétés%20de%20caféiers%20Arabica%20:%20quelles%20innovations%20variétales%20?) Bertrand Benoît, Montagnon Christophe, Georget Frederic, Charmetant Pierre, Etienne Hervé. 2012. Cahiers Agricultures, 21 (2-3) : pp. 77-88. <http://dx.doi.org/10.1684/agr.2012.0547> <<http://dx.doi.org/10.1684/agr.2012.0547>>.
- Bhardwaj, Jyoti, Rohit Chauhan, Mohit Kumar Swarnkar, Rakesh Kumar Chahota, Anil Kumar Singh, Ravi Shankar, and Sudesh Kumar Yadav. 2013. "Comprehensive Transcriptomic Study on Horse Gram (Macrotyloma Uniflorum): De Novo Assembly, Functional Characterization and Comparative Analysis in Relation to Drought Stress." *BMC Genomics* 14 (September): 647. <https://doi.org/10.1186/1471-2164-14-647>.

- Bobadilla Landey, Roberto, Alberto Cenci, Romain Guyot, Benoît Bertrand, Frédéric Georget, Eveline Dechamp, Juan-Carlos Herrera, Jamel Aribi, Philippe Lashermes, and Hervé Etienne. 2015. "Assessment of Genetic and Epigenetic Changes during Cell Culture Ageing and Relations with Somaclonal Variation in *Coffea Arabica*." *Plant Cell, Tissue and Organ Culture (PCTOC)* 122 (3): 517–31. <https://doi.org/10.1007/s11240-015-0772-9>.
- Boutillier, Kim, Remko Offringa, Vijay K. Sharma, Henk Kieft, Thérèse Ouellet, Lemin Zhang, Jiro Hattori, et al. 2002. "Ectopic Expression of BABY BOOM Triggers a Conversion from Vegetative to Embryonic Growth." *The Plant Cell* 14 (8): 1737–49. <https://doi.org/10.1105/tpc.001941>.
- Boutillier, Kimberly, Gerrit Cornelis Angenent, Mercedes Soriano Castan, and Li Hui. 2016. Haploid embryogenesis. United States US20160212956A1, filed September 24, 2014, and issued July 28, 2016. <https://patents.google.com/patent/US20160212956/en>.
- Boxtel, Jos van, and Marc Berthouly. 1996. "High Frequency Somatic Embryogenesis from Coffee Leaves." *Plant Cell, Tissue and Organ Culture* 44 (1): 7–17. <https://doi.org/10.1007/BF00045907>.
- Braybrook, Siobhan A., and John J. Harada. 2008. "LECs Go Crazy in Embryo Development." *Trends in Plant Science* 13 (12): 624–30. <https://doi.org/10.1016/j.tplants.2008.09.008>.
- Brown, D. C., and T. A. Thorpe. 1995. "Crop Improvement through Tissue Culture." *World Journal of Microbiology & Biotechnology* 11 (4): 409–15. <https://doi.org/10.1007/BF00364616>.
- Businge, Edward, Klaus Brackmann, Thomas Moritz, and Ulrika Egertsdotter. 2012. "Metabolite Profiling Reveals Clear Metabolic Changes during Somatic Embryo Development of Norway Spruce (*Picea Abies*)." *Tree Physiology* 32 (2): 232–44. <https://doi.org/10.1093/treephys/tp142>.
- Campos, Nádia A., Luciano V. Paiva, Bart Panis, and Sebastien C. Carpentier. 2016. "The Proteome Profile of Embryogenic Cell Suspensions of *Coffea Arabica* L." *PROTEOMICS* 16 (6): 1001–5. <https://doi.org/10.1002/pmic.201500399>.
- Campos, Nádia A., Bart Panis, and Sebastien C. Carpentier. 2017. "Somatic Embryogenesis in Coffee: The Evolution of Biotechnology and the Integration of Omics Technologies Offer Great Opportunities." *Frontiers in Plant Science* 8: 1460. <https://doi.org/10.3389/fpls.2017.01460>.
- Carron, Marc-Philippe, Ludovic Lardet, Antoine Leconte, C. Boko, Goué B. Dea, and J. Kéli. 2003. "Field Growth and Rubber Yield of *Hevea Brasiliensis* (Muëll.-Arg.) from Budded versus in Vitro Micropropagated Plants from Clone IRCA 18." Conference_item. Proceedings of the First International Symposium on Acclimatization and Establishment of Micropropagated Plants, Sani-Halkidiki, Macedonia, Greece, 19-22 September, 2001. 2003. <http://agritrop.cirad.fr/523210/>.
- Carvalho A, Medina Filho HP, Fazuoli LC, Guerreiro Filho O, Lima MMA. 1991. Aspectos genéticos do cafeeiro. *Rev Bras Genet* 14:135–183
- Chen, Yukun, Xiaoping Xu, Zhuanxia Liu, Zihao Zhang, Xu XuHan, Yuling Lin, and Zhongxion Lai. 2020. "Global Scale Transcriptome Analysis Reveals Differentially Expressed Genes Involve in Early Somatic Embryogenesis in *Dimocarpus Longan* Lour." *BMC Genomics* 21 (1): 4. <https://doi.org/10.1186/s12864-019-6393-7>.
- Cheng, Wen-Han, Hua-Guo Zhu, Wen-Gang Tian, Shou-Hong Zhu, Xian-Peng Xiong, Yu-Qiang Sun, Qian-Hao Zhu, and Jie Sun. 2016. "De Novo Transcriptome Analysis Reveals Insights into Dynamic Homeostasis Regulation of Somatic Embryogenesis in Upland Cotton (*G. Hirsutum* L.)." *Plant Molecular Biology* 92 (3): 279–92. <https://doi.org/10.1007/s11103-016-0511-6>.
- Chevalier A. 1947. Les caféiers du globe. III Systématique des caféiers et faux caféiers. Maladies et insectes nuisibles. Lechevalier P ed. Paris.
- Chung, Jen-Ping, Tsu-Liang Chang, Albert Yu-Ming Chi, and Chou Tou Shii. 2006. "Triploid Banana Cell Growth Phases and the Correlation of Medium PH Changes with Somatic Embryogenesis in Embryogenic Cell Suspension Culture." *Plant Cell, Tissue and Organ Culture* 87 (3): 305–14. <https://doi.org/10.1007/s11240-006-9168-1>.
- Coller, Hilary A. 2019. "The Paradox of Metabolism in Quiescent Stem Cells." *FEBS Letters* 593 (20): 2817–39. <https://doi.org/10.1002/1873-3468.13608>.

- Connor, Cornelius J. O', Henning S. G. Beckmann, and David R. Spring. 2012. "Diversity-Oriented Synthesis: Producing Chemical Tools for Dissecting Biology." *Chemical Society Reviews* 41 (12): 4444–56. <https://doi.org/10.1039/C2CS35023H>.
- Corredoira, E., A. Ballester, M. Ibarra, and A. M. Vieitez. 2015. "Induction of Somatic Embryogenesis in Explants of Shoot Cultures Established from Adult Eucalyptus Globulus and E. Saligna × E. Maidenii Trees." *Tree Physiology* 35 (6): 678–90. <https://doi.org/10.1093/treephys/tpv028>.
- Corredoira, Elena, Silvia Valladares, and Ana M. Vieitez. 2006. "Morphohistological Analysis of the Origin and Development of Somatic Embryos from Leaves of Mature Quercus Robur." *In Vitro Cellular & Developmental Biology - Plant* 42 (6): 525–33. <https://doi.org/10.1079/IVP2006827>.
- Costello, Leslie C., and Renty B. Franklin. 2013. "A Review of the Important Central Role of Altered Citrate Metabolism during the Process of Stem Cell Differentiation." *Journal of Regenerative Medicine & Tissue Engineering* 2 (May). <https://doi.org/10.7243/2050-1218-2-1>.
- Coulibaly, I., M. Noirot, M. Lorieux, A. Charrier, S. Hamon, and J. Louarn. 2002. "Introgression of Self-Compatibility from Coffea Heterocalyx to the Cultivated Species Coffea Canephora." *Theoretical and Applied Genetics* 105 (6): 994–99. <https://doi.org/10.1007/s00122-002-1008-z>.
- Cros, J., M. C. Combes, N. Chabrilange, S. Hamon, C. Duperray, and A. Monnot des Angles. 1995. "Nuclear DNA Content in the Subgenus Coffea (Rubiaceae): Inter- and Intra-Specific Variation in African Species." *Canadian Journal of Botany* 73 (1): 14–20. <https://doi.org/10.1139/b95-002>.
- Cros, J., M. C. Combes, P. Trouslot, F. Anthony, S. Hamon, A. Charrier, and P. Lashermes. 1998. "Phylogenetic Analysis of Chloroplast DNA Variation InCoffeaL." *Molecular Phylogenetics and Evolution* 9 (1): 109–17. <https://doi.org/10.1006/mpev.1997.0453>.
- Curaba, Julien, Thomas Moritz, Renaud Blervaque, François Parcy, Vered Raz, Michel Herzog, and Gilles Vachon. 2004. "AtGA3ox2, a Key Gene Responsible for Bioactive Gibberellin Biosynthesis, Is Regulated during Embryogenesis by LEAFY COTYLEDON2 and FUSCA3 in Arabidopsis." *Plant Physiology* 136 (3): 3660–69. <https://doi.org/10.1104/pp.104.047266>.
- Davey, Mark W., Neil S. Graham, Bartel Vanholme, Rony Swennen, Sean T. May, and Johan Keulemans. 2009. "Heterologous Oligonucleotide Microarrays for Transcriptomics in a Non-Model Species; a Proof-of-Concept Study of Drought Stress in Musa." *BMC Genomics* 10 (September): 436. <https://doi.org/10.1186/1471-2164-10-436>.
- Davis, Aaron P. 2011. "Psilanthus Mannii, the Type Species of Psilanthus, Transferred to Coffea." *Nordic Journal of Botany* 29 (4): 471–72. <https://doi.org/10.1111/j.1756-1051.2011.01113.x>.
- Davis, Aaron P., Rafael Govaerts, Diane M. Bridson, and Piet Stoffelen. 2006. "An Annotated Taxonomic Conspectus of the Genus Coffea (Rubiaceae)." *Botanical Journal of the Linnean Society* 152 (4): 465–512. <https://doi.org/10.1111/j.1095-8339.2006.00584.x>.
- Dengler, Nancy G. 1999. "Anisophylly and Dorsiventral Shoot Symmetry." *International Journal of Plant Sciences* 160 (S6): S67–80. <https://doi.org/10.1086/314218>.
- Denoeud, France, Lorenzo Carretero-Paulet, Alexis Dereeper, Gaëtan Droc, Romain Guyot, Marco Pietrella, Chunfang Zheng, et al. 2014. "The Coffee Genome Provides Insight into the Convergent Evolution of Caffeine Biosynthesis." *Science (New York, N.Y.)* 345 (6201): 1181–84. <https://doi.org/10.1126/science.1255274>.
- Deo, Pradeep, Anand Tyagi, Mary Taylor, Rob Harding, and Douglas Becker. 2010. "Factors Affecting Somatic Embryogenesis and Transformation in Modern Plant Breeding." *The South Pacific Journal of Natural and Applied Sciences* 28 (1): 27–40.
- Dethloff, Frederik, Alexander Erban, Isabel Orf, Jessica Alpers, Ines Fehrlé, Olga Beine-Golovchuk, Stefanie Schmidt, Jens Schwachtje, and Joachim Kopka. 2014. "Profiling Methods to Identify Cold-Regulated Primary Metabolites Using Gas Chromatography Coupled to Mass Spectrometry." *Methods in Molecular Biology (Clifton, N.J.)* 1166: 171–97. https://doi.org/10.1007/978-1-4939-0844-8_14.

- Dobrowolska, Izabela, Edward Businge, Ilka N. Abreu, Thomas Moritz, and Ulrika Egertsdotter. 2017. "Metabolome and Transcriptome Profiling Reveal New Insights into Somatic Embryo Germination in Norway Spruce (*Picea Abies*)." *Tree Physiology* 37 (12): 1752–66. <https://doi.org/10.1093/treephys/tpx078>.
- Dodeman, V. L., G. Ducreux, and M. Kreis. 1997. "REVIEW ARTICLE Zygotic Embryogenesis versus Somatic Embryogenesis." *Journal of Experimental Botany* 48 (8): 1493–1509. <https://doi.org/10.1093/jxb/48.8.1493>.
- Dublin P. 1984. Techniques de reproduction végétative in vitro et amélioration génétique chez les cafeiers cultivés. *Café Cacao Thé* 28: 231-244.
- Duclercq, Jérôme, Brigitte Sangwan-Norreel, Manuella Catterou, and Rajbir S. Sangwan. 2011. "De Novo Shoot Organogenesis: From Art to Science." *Trends in Plant Science* 16 (11): 597–606. <https://doi.org/10.1016/j.tplants.2011.08.004>.
- Ducos, J. P., G. Labbe, C. Lambot, and V. Pétiard. 2007. "Pilot Scale Process for the Production of Pre-Germinated Somatic Embryos of Selected Robusta (*Coffea Canephora*) Clones." *In Vitro Cellular & Developmental Biology - Plant* 43 (6): 652–59. <https://doi.org/10.1007/s11627-007-9075-0>.
- Ducos, J.P., R. Alenton, J.F. Reano, C. Kanchanomai, A. Deshayes, and V. Pétiard. 2003. "Agronomic Performance of *Coffea Canephora* P. Trees Derived from Large-Scale Somatic Embryo Production in Liquid Medium." *Euphytica* 131 (2): 215–23. <https://doi.org/10.1023/A:1023915613158>.
- Dussert, Stéphane, Mark W. Davey, Andréina Laffargue, Sylvie Doulbeau, Rony Swennen, and Hervé Etienne. 2006. "Oxidative Stress, Phospholipid Loss and Lipid Hydrolysis during Drying and Storage of Intermediate Seeds." *Physiologia Plantarum* 127 (2): 192–204. <https://doi.org/10.1111/j.1399-3054.2006.00666.x>.
- Etienne, H., and B. Bertrand. 2003. "Somaclonal Variation in *Coffea Arabica*: Effects of Genotype and Embryogenic Cell Suspension Age on Frequency and Phenotype of Variants." *Tree Physiology* 23 (6): 419–26. <https://doi.org/10.1093/treephys/23.6.419>.
- Etienne, H. 2005. "Somatic Embryogenesis Protocol: Coffee (*Coffea Arabica* L. and *C. Canephora* P.)." In *Protocol for Somatic Embryogenesis in Woody Plants*, edited by S. Mohan Jain and Pramod K. Gupta, 167–79. Forestry Sciences. Dordrecht: Springer Netherlands. https://doi.org/10.1007/1-4020-2985-3_14.
- Etienne, Hervé, Benoît Bertrand, Eveline Dechamp, Patrick Maurel, Frederic Georget, Romain Guyot, and Jean-Christophe Breitler. 2016. "Are Genetics and Epigenetic Instabilities of Plant Embryogenic Cells a Fatality? The Experience of Coffee Somatic Embryogenesis." *Human Genetics and Embryology*. <http://agritrop.cirad.fr/580450/>.
- Etienne, Hervé, Benoît Bertrand, Christophe Montagnon, Roberto Bodadilla Landey, Eveline Dechamp, Isabelle Jourdan, Edgardo Alpizar, Eduardo Malo, and Frederic Georget. 2012. "Un exemple de transfert de technologie réussi dans le domaine de la micropropagation : la multiplication de *Coffea arabica* par embryogénèse somatique." *Cahiers Agricultures* 21: 115–24. <http://dx.doi.org/10.1684/agr.2012.0553>
- Etienne, Hervé, David Breton, Jean-Christophe Breitler, Benoît Bertrand, Eveline Déchamp, Rayan Awada, Pierre Marraccini, et al. 2018. "Coffee Somatic Embryogenesis: How Did Research, Experience Gained and Innovations Promote the Commercial Propagation of Elite Clones From the Two Cultivated Species?" *Frontiers in Plant Science* 9: 1630. <https://doi.org/10.3389/fpls.2018.01630>.
- Fehér, Attila. 2019. "Callus, Dedifferentiation, Totipotency, Somatic Embryogenesis: What These Terms Mean in the Era of Molecular Plant Biology?" *Frontiers in Plant Science* 10. <https://doi.org/10.3389/fpls.2019.00536>.
- Fehér, Attila, Krisztina Otvös, Taras P. Pasternak, and Aladór Pettkó Szandtner. 2008. "The Involvement of Reactive Oxygen Species (ROS) in the Cell Cycle Activation (G(0)-to-G(1) Transition) of Plant Cells." *Plant Signaling & Behavior* 3 (10): 823–26. <https://doi.org/10.4161/psb.3.10.5908>.
- Fehér, Attila, Taras P. Pasternak, and Dénes Dudits. 2003. "Transition of Somatic Plant Cells to an Embryogenic State." *Plant Cell, Tissue and Organ Culture* 74 (3): 201–28. <https://doi.org/10.1023/A:1024033216561>.
- Feria, Manuel de, Elio Jiménez, Raúl Barbón, Alina Capote, Maité Chávez, and Elisa Quiala. 2003. "Effect of Dissolved Oxygen Concentration on Differentiation of Somatic Embryos of *Coffea Arabica* Cv. Catimor 9722." *Plant Cell, Tissue and Organ Culture* 72 (1): 1–6. <https://doi.org/10.1023/A:1021202305692>.

- Fernie, L., Greathead, D., Meyer, F. & Monaco, L., Narasimhaswamy, R. 1968. FAO coffee mission to Ethiopia, 1964–65.
- Fischella, Marco, and Stefano Morini. 2003. “Somatic Embryo and Root Regeneration from Quince Leaves Cultured in Ventilated Vessels or under Different Oxygen and Carbon Dioxide Levels.” *In Vitro Cellular & Developmental Biology - Plant* 39 (4): 402–8. <https://doi.org/10.1079/IVP2003429>.
- Florez, Sergio L., Rachel L. Erwin, Siela N. Maximova, Mark J. Gultinan, and Wayne R. Curtis. 2015. “Enhanced Somatic Embryogenesis in Theobroma Cacao Using the Homologous BABY BOOM Transcription Factor.” *BMC Plant Biology* 15 (May): 121. <https://doi.org/10.1186/s12870-015-0479-4>.
- Gaj, Malgorzata D., Shibo Zhang, John J. Harada, and Peggy G. Lemaux. 2005. “Leafy Cotyledon Genes Are Essential for Induction of Somatic Embryogenesis of Arabidopsis.” *Planta* 222 (6): 977–88. <https://doi.org/10.1007/s00425-005-0041-y>.
- Gatica-Arias, Andres M., Griselda Arrieta-Espinoza, and Ana M. Espinoza Esquivel. 2008. “Plant Regeneration via Indirect Somatic Embryogenesis and Optimisation of Genetic Transformation in Coffea Arabica L. Cvs. Caturra and Catuai.” *Electronic Journal of Biotechnology* 11 (1): 0–0. <https://doi.org/10.2225/vol11-issue1-fulltext-9>.
- Gautier, Florian, Philippe Label, Kateřina Eliášová, Jean-Charles Leplé, Václav Motyka, Nathalie Boizot, Zuzana Vondráková, et al. 2019. “Cytological, Biochemical and Molecular Events of the Embryogenic State in Douglas-Fir (Pseudotsuga Menziesii [Mirb.]”). *Frontiers in Plant Science* 10: 118. <https://doi.org/10.3389/fpls.2019.00118>.
- Gazzarrini, Sonia, Yuichiro Tsuchiya, Shelley Lumba, Masanori Okamoto, and Peter McCourt. 2004. “The Transcription Factor FUSCA3 Controls Developmental Timing in Arabidopsis through the Hormones Gibberellin and Abscisic Acid.” *Developmental Cell* 7 (3): 373–85. <https://doi.org/10.1016/j.devcel.2004.06.017>.
- Georget, Frederic, Philippe Courtel, Eduardo Malo Garcia, Martin Hidalgo, Edgardo Alpizar, Jean-Christophe Breitler, Benoît Bertrand, and Hervé Etienne. 2017. “Somatic Embryogenesis-Derived Coffee Plantlets Can Be Efficiently Propagated by Horticultural Rooted Mini-Cuttings: A Boost for Somatic Embryogenesis.” *Scientia Horticulturae* 216: 177–85. <http://dx.doi.org/10.1016/j.scienta.2016.12.017>
- Goevarts, R. M. Ruhsam, L., Andersson, E., Robbrecht, D., Bridson, A., Davis, I., Schanzer, B., Sonké. 2006. World checklist of Rubiaceae. Royal Botanic Gardens, Kew. <http://www.kew.org/wesp/rubiaceae/>.
- Ghini, R., Bettiol, W., Hamada, E. 2011. “Diseases in Tropical and Plantation Crops as Affected by Climate Changes: Current Knowledge and Perspectives.” *Plant Pathology* 60 (1): 122–32. <https://doi.org/10.1111/j.1365-3059.2010.02403.x>.
- Gray, William M. 2004. “Hormonal Regulation of Plant Growth and Development.” *PLoS Biology* 2 (9). <https://doi.org/10.1371/journal.pbio.0020311>.
- Grimaldi, Annalisa, Cristiano Bianchi, Gabriella Greco, Gianluca Tettamanti, Douglas M. Noonan, Roberto Valvassori, and Magda de Eguileor. 2008. “In Vivo Isolation and Characterization of Stem Cells with Diverse Phenotypes Using Growth Factor Impregnated Biomatrices.” *PloS One* 3 (4): e1910. <https://doi.org/10.1371/journal.pone.0001910>.
- Gross, Stephen M., Jeffrey A. Martin, June Simpson, María Jazmín Abraham-Juarez, Zhong Wang, and Axel Visel. 2013. “De Novo Transcriptome Assembly of Drought Tolerant CAM Plants, Agave Deserti and Agave Tequilana.” *BMC Genomics* 14 (August): 563. <https://doi.org/10.1186/1471-2164-14-563>.
- Guillou, Caroline, Audrey Filloudeau, Eric Brulard, David Breton, Simone De Faria Maraschin, Dorothee Verdier, Mathieu Simon, and Jean-Paul Ducos. 2018. “Indirect Somatic Embryogenesis of Theobroma Cacao L. in Liquid Medium and Improvement of Embryo-to-Plantlet Conversion Rate.” *In Vitro Cellular & Developmental Biology - Plant* 54 (4): 377–91. <https://doi.org/10.1007/s11627-018-9909-y>.
- Gutiérrez-Mora, Antonia, Alejandra Guillermina González-Gutiérrez, Benjamín Rodríguez-Garay, Azucena Ascencio-Cabral, and Lin Li-Wei. 2012. “Plant Somatic Embryogenesis: Some Useful Considerations.” *Embryogenesis*, April. <https://doi.org/10.5772/36345>.
- Haarer, A. E. 1958. Modern Coffee production. Leonard Hill.

- Hecht, Valérie, Jean-Philippe Vielle-Calzada, Marijke V. Hartog, Ed D.L. Schmidt, Kim Boutilier, Ueli Grossniklaus, and Sacco C. de Vries. 2001. "The Arabidopsis Somatic Embryogenesis Receptor Kinase 1 Gene Is Expressed in Developing Ovules and Embryos and Enhances Embryogenic Competence in Culture." *Plant Physiology* 127 (3): 803–16.
- Hicks, Glenn R., and Natasha V. Raikhel. 2012. "Small Molecules Present Large Opportunities in Plant Biology." *Annual Review of Plant Biology* 63 (1): 261–82. <https://doi.org/10.1146/annurev-arplant-042811-105456>.
- Honda, H., C. Liu, and T. Kobayashi. 2001. "Large-Scale Plant Micropropagation." *Advances in Biochemical Engineering/Biotechnology* 72: 157–82. https://doi.org/10.1007/3-540-45302-4_6.
- Horstman, Anneke, Marian Bemer, and Kim Boutilier. 2017. "A Transcriptional View on Somatic Embryogenesis." *Regeneration (Oxford, England)* 4 (4): 201–16. <https://doi.org/10.1002/reg2.91>.
- Hughes, JP, S Rees, SB Kalindjian, and KL Philpott. 2011. "Principles of Early Drug Discovery." *British Journal of Pharmacology* 162 (6): 1239–49. <https://doi.org/10.1111/j.1476-5381.2010.01127.x>.
- Indoliya, Yuvraj, Poonam Tiwari, Abhisekh Singh Chauhan, Ridhi Goel, Manju Shri, Sumit Kumar Bag, and Debasis Chakrabarty. 2016. "Decoding Regulatory Landscape of Somatic Embryogenesis Reveals Differential Regulatory Networks between *Japonica* and *Indica* Rice Subspecies." *Scientific Reports* 6 (March): 23050. <https://doi.org/10.1038/srep23050>.
- Ito, Keisuke, and Toshio Suda. 2014. "Metabolic Requirements for the Maintenance of Self-Renewing Stem Cells." *Nature Reviews. Molecular Cell Biology* 15 (4): 243–56. <https://doi.org/10.1038/nrm3772>.
- Ito, Yukihiko, Kazuhiko Takaya, and Nori Kurata. 2005. "Expression of SERK Family Receptor-like Protein Kinase Genes in Rice." *Biochimica Et Biophysica Acta* 1730 (3): 253–58. <https://doi.org/10.1016/j.bbaexp.2005.06.007>.
- Jayasankar, S., Bhaskar R. Bondada, Zhijian Li, and D. J. Gray. 2003. "Comparative Anatomy and Morphology of *Vitis Vinifera* (Vitaceae) Somatic Embryos from Solid- and Liquid-Culture-Derived Proembryogenic Masses." *American Journal of Botany* 90 (7): 973–79.
- Jiménez, Víctor M. 2005. "Involvement of Plant Hormones and Plant Growth Regulators on in Vitro Somatic Embryogenesis." *Plant Growth Regulation* 47 (2): 91–110. <https://doi.org/10.1007/s10725-005-3478-x>.
- Juma, C., Magambo, J.M. and Monteith, H. 1994. "Tissue Culture for Coffee: The case of Uganda." *Biotechnology and Development Monitor*, No. 20, p. 19-20.
- Kobayashi, Toshihiro, Katsumi Highashi, and Hiroshi Kamada. 2003. "4-Hydroxybenzyl Alcohol Accumulates in Flowers and Developing Fruits of Carrot and Inhibits Seed Formation." *Journal of Plant Physiology* 160 (6): 713–16. <https://doi.org/10.1078/0176-1617-00967>.
- Kochko, Alexandre De, Sélastique Akaffou, Alan C. Andrade, Claudine Campa, Dominique Crouzillat, Romain Guyot, Perla Hamon, et al. 2010. "Advances in Coffea Genomics." In *Advances in Botanical Research*, edited by Jean-Claude Kader and Michel Delseny, 53:23–63. Advances in Botanical Research. Academic Press. [https://doi.org/10.1016/S0065-2296\(10\)53002-7](https://doi.org/10.1016/S0065-2296(10)53002-7).
- Komamine, A., R. Kawahara, M. Matsumoto, S. Sunabori, T. Toya, A. Fujiwara, M. Tsukahara, et al. 1992. "Mechanisms of Somatic Embryogenesis in Cell Cultures: Physiology, Biochemistry, and Molecular Biology." *In Vitro Cellular & Developmental Biology - Plant* 28 (1): 11–14. <https://doi.org/10.1007/BF02632185>.
- Kong, Lisheng, and Patrick von Aderkas. 2007. "Genotype Effects on ABA Consumption and Somatic Embryo Maturation in Interior Spruce (*Picea Glaucaengelmanni*)." *Journal of Experimental Botany* 58 (6): 1525–31. <https://doi.org/10.1093/jxb/erm019>.
- Kopka, Joachim, Nicolas Schauer, Stephan Krueger, Claudia Birkemeyer, Björn Usadel, Eveline Bergmüller, Peter Dörmann, et al. 2005. "GMD@CSB.DB: The Golm Metabolome Database." *Bioinformatics (Oxford, England)* 21 (8): 1635–38. <https://doi.org/10.1093/bioinformatics/bti236>.
- Langmead, Ben, and Steven L. Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4): 357–59. <https://doi.org/10.1038/nmeth.1923>.

- Larkin, P. J., and W. R. Scowcroft. 1981. "Somaclonal Variation - a Novel Source of Variability from Cell Cultures for Plant Improvement." *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 60 (4): 197–214. <https://doi.org/10.1007/BF02342540>.
- Lashermes, P., M. C. Combes, J. Robert, P. Trouslot, A. D'Hont, F. Anthony, and A. Charrier. 1999. "Molecular Characterisation and Origin of the Coffea Arabica L. Genome." *Molecular & General Genetics: MGG* 261 (2): 259–66. <https://doi.org/10.1007/s004380050965>.
- Lashermes, Philippe, Alan Carvalho Andrade, and Hervé Etienne. 2008. "Genomics of Coffee One of the World's Largest Traded Commodities." In *Genomics of Tropical Crop Plants*, edited by Paul H. Moore and Ray Ming, 1:203–26. New York, NY: Springer New York. https://doi.org/10.1007/978-0-387-71219-2_9.
- Lelu-Walter, Marie-Anne, David Thompson, Luc Harvengt, Leopoldo Sanchez, Mariano Toribio, and Luc E. Pâques. 2013. "Somatic Embryogenesis in Forestry with a Focus on Europe: State-of-the-Art, Benefits, Challenges and Future Direction." *Tree Genetics & Genomes* 9 (4): 883–99. <https://doi.org/10.1007/s11295-013-0620-1>.
- Lin, Brenda B. 2011. "Resilience in Agriculture through Crop Diversification: Adaptive Management for Environmental Change." *BioScience* 61 (3): 183–93. <https://doi.org/10.1525/bio.2011.61.3.4>.
- Lotan, T., M. Ohto, K. M. Yee, M. A. West, R. Lo, R. W. Kwong, K. Yamagishi, R. L. Fischer, R. B. Goldberg, and J. J. Harada. 1998. "Arabidopsis LEAFY COTYLEDON1 Is Sufficient to Induce Embryo Development in Vegetative Cells." *Cell* 93 (7): 1195–1205. [https://doi.org/10.1016/s0092-8674\(00\)81463-4](https://doi.org/10.1016/s0092-8674(00)81463-4).
- Ma, Shisong, Shawn Bachan, Matthew Porto, Hans J. Bohnert, Michael Snyder, and Savithramma P. Dinesh-Kumar. 2012. "Discovery of Stress Responsive DNA Regulatory Motifs in Arabidopsis." *PLoS One* 7 (8): e43198. <https://doi.org/10.1371/journal.pone.0043198>.
- Maciel, Anna Lygia de Rezende, Filipe Almendagna Rodrigues, Moacir Pasqual, Carlos Henrique Siqueira de Carvalho, Anna Lygia de Rezende Maciel, Filipe Almendagna Rodrigues, Moacir Pasqual, and Carlos Henrique Siqueira de Carvalho. 2016. "Large-Scale, High-Efficiency Production of Coffee Somatic Embryos." *Crop Breeding and Applied Biotechnology* 16 (2): 102–7. <https://doi.org/10.1590/1984-70332016v16n2a16>.
- Magnani, E., J. M. Jiménez-Gómez, L. Soubigou-Taconnat, L. Lepiniec, and E. Fiume. 2017. "Profiling the Onset of Somatic Embryogenesis in Arabidopsis." *BMC Genomics* 18 (1): 998. <https://doi.org/10.1186/s12864-017-4391-1>.
- Martínez-Gómez, Pedro, Raquel Sánchez-Pérez, and Manuel Rubio. 2012. "Clarifying Omics Concepts, Challenges, and Opportunities for Prunus Breeding in the Postgenomic Era." *Omics: A Journal of Integrative Biology* 16 (5): 268–83. <https://doi.org/10.1089/omi.2011.0133>.
- Masson, Aurélien, Jean-Marc Julien, and Luc Boedt. 2013. "Multiplication Industrielle Par Bouturage de Clones Matures d' Hevea Brasiliensis." *BOIS & FORETS DES TROPIQUES* 317 (317): 51. <https://doi.org/10.19182/bft2013.317.a20523>.
- Matsunaga, Fabio Takeshi, Jonas Barbosa Tosti, Armando Androcio-Filho, Jacques Duílio Brancher, Evelyne Costes, and Miroslava Rakocevic. 2016. "Strategies to Reconstruct 3D Coffea Arabica L. Plant Structure." *SpringerPlus* 5 (1): 2075. <https://doi.org/10.1186/s40064-016-3762-4>.
- Maurin, Olivier, Aaron P. Davis, Michael Chester, Esther F. Mvungi, Yasmina Jaufeerally-Fakim, and Michael F. Fay. 2007. "Towards a Phylogeny for Coffea (Rubiaceae): Identifying Well-Supported Lineages Based on Nuclear and Plastid DNA Sequences." *Annals of Botany* 100 (7): 1565–83. <https://doi.org/10.1093/aob/mcm257>.
- Maximova, Siela N., Larence Alemanno, Ann Young, Nicole Ferriere, Abdoulaye Traore, and Mark J. Gultinan. 2002. "Efficiency, Genotypic Variability, and Cellular Origin of Primary and Secondary Somatic Embryogenesis of Theobroma Cacao L." In *In Vitro Cellular & Developmental Biology - Plant* 38 (3): 252–59. <https://doi.org/10.1079/IVP2001257>.
- Mayr, Lorenz M., and Peter Fuerst. 2008. "The Future of High-Throughput Screening." *Journal of Biomolecular Screening* 13 (6): 443–48. <https://doi.org/10.1177/1087057108319644>.
- Méndez-Hernández, Hugo A., Maharshi Ledezma-Rodríguez, Randy N. Avilez-Montalvo, Yary L. Juárez-Gómez, Analesa Skeete, Johny Avilez-Montalvo, Clelia De-la-Peña, and Víctor M. Loyola-Vargas. 2019. "Signaling Overview of Plant Somatic Embryogenesis." *Frontiers in Plant Science* 10: 77. <https://doi.org/10.3389/fpls.2019.00077>.

- Menéndez-Yuffá, Andrea, Dominique Barry-Etienne, Benoît Bertrand, Frédéric Georget, and Hervé Etienne. 2010. “A Comparative Analysis of the Development and Quality of Nursery Plants Derived from Somatic Embryogenesis and from Seedlings for Large-Scale Propagation of Coffee (*Coffea Arabica* L.)” *Plant Cell, Tissue and Organ Culture (PCTOC)* 102 (3): 297–307. <https://doi.org/10.1007/s11240-010-9734-4>.
- Mira, Mohamed M., Owen S. D. Wally, Mohamed Elhiti, Adel El-Shanshory, Dhadi S. Reddy, Robert D. Hill, and Claudio Stasolla. 2016. “Jasmonic Acid Is a Downstream Component in the Modulation of Somatic Embryogenesis by Arabidopsis Class 2 Phytoglobin.” *Journal of Experimental Botany* 67 (8): 2231–46. <https://doi.org/10.1093/jxb/erw022>.
- Molina, Diana M., María E. Aponte, Hernando Cortina, and Germán Moreno. 2002. “The Effect of Genotype and Explant Age on Somatic Embryogenesis of Coffee.” *Plant Cell, Tissue and Organ Culture* 71 (2): 117–23. <https://doi.org/10.1023/A:1019965621041>.
- Moon, Heung-Kyu, Yong-Wook Kim, Yong-Pyo Hong, and So-Young Park. 2013. “Improvement of Somatic Embryogenesis and Plantlet Conversion in *Oplopanax Elatus*, an Endangered Medicinal Woody Plant.” *SpringerPlus* 2 (August). <https://doi.org/10.1186/2193-1801-2-428>.
- Mueller, L., Strickler, S. R., Domingues, D. S., Pereira, L. F. P., Andrade, A. A., Marraccini, P., et al. 2015. “Towards a better understanding of the coffee arabica genome structure,” in *Proceedings of the 25th International Conference on Coffee Science ASIC*. (Armenia, CO), 42–45.
- Mukul-López, Héctor Gabriel, Clelia De-la-Peña, Rosa María Galaz-Ávalos, and Víctor Manuel Loyola-Vargas. 2012. “Evaluation of the Extracellular Proteome Profile During the Somatic Embryogenesis Process of *Coffea* Spp.” *Journal of the Mexican Chemical Society* 56 (1): 72–79.
- Murashige, Toshio, and Folke Skoog. 1962. “A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures.” *Physiologia Plantarum* 15 (3): 473–97. <https://doi.org/10.1111/j.1399-3054.1962.tb08052.x>.
- Na, Haeyoung, Ki Woo Kim, Yurina Kwack, Sung Kyeom Kim, and Changhoo Chun. 2007. “Comparative Anatomy of Embryogenic and Non-Embryogenic Calli From *Pimpinella Brachycarpa*.” *Journal of Plant Biology* 50 (3): 344–50. <https://doi.org/10.1007/BF03030665>.
- Namasivayam, Parameswari. 2007. “Acquisition of Embryogenic Competence during Somatic Embryogenesis.” *Plant Cell, Tissue and Organ Culture* 90 (1): 1–8. <https://doi.org/10.1007/s11240-007-9249-9>.
- Neelakandan, Anjanasree K., and Kan Wang. 2012. “Recent Progress in the Understanding of Tissue Culture-Induced Genome Level Changes in Plants and Potential Applications.” *Plant Cell Reports* 31 (4): 597–620. <https://doi.org/10.1007/s00299-011-1202-z>.
- Nic-Can, Geovanny I., Rosa M. Galaz-Ávalos, Clelia De-la-Peña, Armando Alcazar-Magaña, Kazimierz Wrobel, and Víctor M. Loyola-Vargas. 2015. “Somatic Embryogenesis: Identified Factors That Lead to Embryogenic Repression. A Case of Species of the Same Genus.” *PloS One* 10 (6): e0126414. <https://doi.org/10.1371/journal.pone.0126414>.
- Nic-Can, Geovanny I., Adolfo López-Torres, Felipe Barredo-Pool, Kazimierz Wrobel, Víctor M. Loyola-Vargas, Rafael Rojas-Herrera, and Clelia De-la-Peña. 2013. “New Insights into Somatic Embryogenesis: LEAFY COTYLEDON1, BABY BOOM1 and WUSCHEL-RELATED HOMEODOMAIN4 Are Epigenetically Regulated in *Coffea Canephora*.” *PLOS ONE* 8 (8): e72160. <https://doi.org/10.1371/journal.pone.0072160>.
- Oono, Youko, Fuminori Kobayashi, Yoshihiro Kawahara, Takayuki Yazawa, Hirokazu Handa, Takeshi Itoh, and Takashi Matsumoto. 2013. “Characterisation of the Wheat (*Triticum Aestivum* L.) Transcriptome by de Novo Assembly for the Discovery of Phosphate Starvation-Responsive Genes: Gene Expression in Pi-Stressed Wheat.” *BMC Genomics* 14 (February): 77. <https://doi.org/10.1186/1471-2164-14-77>.
- Page, Justin T., Alan R. Gingle, and Joshua A. Udall. 2013. “PolyCat: A Resource for Genome Categorization of Sequencing Reads from Allopolyploid Organisms.” *G3 (Bethesda, Md.)* 3 (3): 517–25. <https://doi.org/10.1534/g3.112.005298>.
- Pais, Maria Salomé. 2019. “Somatic Embryogenesis Induction in Woody Species: The Future After OMICs Data Assessment.” *Frontiers in Plant Science* 10: 240. <https://doi.org/10.3389/fpls.2019.00240>.

- Park, Y. S., M. A. Lelu-Walter, L. Harvengt, J. F. Trontin, I. MacEacheron, K. Klimaszewska, and J. M. Bonga. 2006. "Initiation of Somatic Embryogenesis in *Pinus Banksiana*, P. Strobilus, P. Pinaster, and P. Sylvestris at Three Laboratories in Canada and France." *Plant Cell, Tissue and Organ Culture* 86 (1): 87–101. <https://doi.org/10.1007/s11240-006-9101-7>.
- Phillips, Gregory C. 2004. "In Vitro Morphogenesis in Plants-Recent Advances." *In Vitro Cellular & Developmental Biology - Plant* 40 (4): 342–45. <https://doi.org/10.1079/IVP2004555>.
- Quint, Marcel, and William M Gray. 2006. "Auxin Signaling." *Current Opinion in Plant Biology* 9 (5): 448–53. <https://doi.org/10.1016/j.pbi.2006.07.006>.
- Quiroz-Figueroa, Francisco, Marcela Méndez-Zeel, Felipe Sánchez-Teyer, Rafael Rojas-Herrera, and Víctor M. Loyola-Vargas. 2002. "Differential Gene Expression in Embryogenic and Non-Embryogenic Clusters from Cell Suspension Cultures of *Coffea Arabica*." *Journal of Plant Physiology* 159 (11): 1267–70. <https://doi.org/10.1078/0176-1617-00878>.
- Quiroz-Figueroa, Francisco R., Rafael Rojas-Herrera, Rosa M. Galaz-Avalos, and Víctor M. Loyola-Vargas. 2006. "Embryo Production through Somatic Embryogenesis Can Be Used to Study Cell Differentiation in Plants." *Plant Cell, Tissue and Organ Culture* 86 (3): 285. <https://doi.org/10.1007/s11240-006-9139-6>.
- Radzun, Khairul Adzfa, Juliane Wolf, Gisela Jakob, Eugene Zhang, Evan Stephens, Ian Ross, and Ben Hankamer. 2015. "Automated Nutrient Screening System Enables High-Throughput Optimisation of Microalgae Production Conditions." *Biotechnology for Biofuels* 8 (1): 65. <https://doi.org/10.1186/s13068-015-0238-7>.
- Rasmussen, Susanne, Anthony J. Parsons, and Christopher S. Jones. 2012. "Metabolomics of Forage Plants: A Review." *Annals of Botany* 110 (6): 1281–90. <https://doi.org/10.1093/aob/mcs023>.
- Reinert, J. 1958. "Morphogenese und ihre Kontrolle an Gewebekulturen aus Carotten." *Naturwissenschaften* 45 (14): 344–45. <https://doi.org/10.1007/BF00640240>.
- Rezende, Juliana Costa de, Carlos Henrique Siqueira de Carvalho, Moacir Pasqual, Ana Carolina Ramia Santos, and Stephan Malfitano de Carvalho. 2011. "Calli Induction in Leaf Explants of Coffee Elite Genotypes." *Ciência Rural* 41 (5): 0–0. <https://doi.org/10.1590/S0103-84782011000500031>.
- Robert, Stéphanie, Natasha V. Raikhel, and Glenn R. Hicks. 2009. "Powerful Partners: Arabidopsis and Chemical Genomics." *The Arabidopsis Book* 2009 (7). <https://doi.org/10.1199/tab.0109>.
- Robinson, Andrew R., Rebecca Dauwe, Nicholas K. Ukrainetz, Ian F. Cullis, Rick White, and Shawn D. Mansfield. 2009. "Predicting the Regenerative Capacity of Conifer Somatic Embryogenic Cultures by Metabolomics." *Plant Biotechnology Journal* 7 (9): 952–63. <https://doi.org/10.1111/j.1467-7652.2009.00456.x>.
- Rodríguez-Furlán, Cecilia, Giovanna Miranda, Martín Reggiardo, Glenn R. Hicks, and Lorena Norambuena. 2016. "High Throughput Selection of Novel Plant Growth Regulators: Assessing the Translatability of Small Bioactive Molecules from Arabidopsis to Crops." *Plant Science* 245 (April): 50–60. <https://doi.org/10.1016/j.plantsci.2016.01.001>.
- Rojas-Herrera, R., F. Quiroz-Figueroa, L. Sánchez-Teyer, and V. M. Loyola-Vargas. 2002. "Molecular analysis of somatic embryogenesis: An overview." *Physiology and Molecular Biology of Plants* 8 (2): 171–84.
- Rozo, Yomara, Carolina Escobar, Álvaro Gaitán, and Marco Cristancho. 2012. "Aggressiveness and Genetic Diversity of *Hemileia Vastatrix* During an Epidemic in Colombia." *Journal of Phytopathology* 160 (11–12): 732–40. <https://doi.org/10.1111/jph.12024>.
- Ruduś, Izabela, Ewa Kępczyńska, and Jan Kępczyński. 2006. "Comparative Efficacy of Abscisic Acid and Methyl Jasmonate for Indirect Somatic Embryogenesis in *Medicago Sativa* L." *Plant Growth Regulation* 48 (1): 1. <https://doi.org/10.1007/s10725-005-5136-8>.
- Santana, N., M. E. González, M. Valcárcel, A. Canto-Flick, M. M. Hernández, F. J. Fuentes-Cerda, F. Barahona, J. Mijangos-Cortés, and V. M. Loyola-Vargas. 2004. "Somatic Embryogenesis: A Valuable Alternative for Propagating Selected Robusta Coffee (*Coffea Canephora*) Clones." *In Vitro Cellular & Developmental Biology - Plant* 40 (1): 95–101. <https://doi.org/10.1079/IVP2003486>.
- Santana-Buzzy, Nancy, Rafael Rojas-Herrera, Rosa M. Galaz-Avalos, José R. Ku-Cauch, Javier Mijangos-Cortés, Luis C. Gutiérrez-Pacheco, Adriana Canto, Francisco Quiroz-Figueroa, and Víctor M. Loyola-Vargas. 2007. "Advances in

Coffee Tissue Culture and Its Practical Applications.” *In Vitro Cellular & Developmental Biology - Plant* 43 (6): 507–20. <https://doi.org/10.1007/s11627-007-9074-1>.

Scalabrin, Simone, Lucile Toniutti, Gabriele Di Gaspero, Davide Scaglione, Gabriele Magris, Michele Vidotto, Sara Pinosio, et al. 2020. “A Single Polyploidization Event at the Origin of the Tetraploid Genome of *Coffea Arabica* Is Responsible for the Extremely Low Genetic Variation in Wild and Cultivated Germplasm.” *Scientific Reports* 10 (1): 1–13. <https://doi.org/10.1038/s41598-020-61216-7>.

Scherling, Christian, Christiane Roscher, Patrick Giavalisco, Ernst-Detlef Schulze, and Wolfram Weckwerth. 2010. “Metabolomics Unravel Contrasting Effects of Biodiversity on the Performance of Individual Plant Species.” *PloS One* 5 (9): e12569. <https://doi.org/10.1371/journal.pone.0012569>.

Schmidt, E. D., F. Guzzo, M. A. Toonen, and S. C. de Vries. 1997. “A Leucine-Rich Repeat Containing Receptor-like Kinase Marks Somatic Plant Cells Competent to Form Embryos.” *Development (Cambridge, England)* 124 (10): 2049–62.

Serrano, Mario, Erich Kombrink, and Christian Meesters. 2015. “Considerations for Designing Chemical Screening Strategies in Plant Biology.” *Frontiers in Plant Science* 6 (April). <https://doi.org/10.3389/fpls.2015.00131>.

Shimada, Takehiko, Toshiro Hirabayashi, Tomoko Endo, Hiroshi Fujii, Masayuki Kita, and Mitsuo Omura. 2005. “Isolation and Characterization of the Somatic Embryogenesis Receptor-like Kinase Gene Homologue (CitSERK1) from Citrus Unshiu Marc.” *Scientia Horticulturae* 103 (2): 233–38. <https://doi.org/10.1016/j.scienta.2004.07.005>.

Silva, Anderson Tadeu, Douglas Barduche, Kalynka Gabriella do Livramento, Wilco Ligterink, and Luciano Vilela Paiva. 2014. “Characterization of a Putative Serk-Like Ortholog in Embryogenic Cell Suspension Cultures of *Coffea Arabica* L.” *Plant Molecular Biology Reporter* 32 (1): 176–84. <https://doi.org/10.1007/s11105-013-0632-x>.

Silva, Anderson Tadeu, Douglas Barduche, Kalynka Gabriella do Livramento, and Luciano Vilela Paiva. 2015. “A Putative BABY BOOM-like Gene (CaBBM) Is Expressed in Embryogenic Calli and Embryogenic Cell Suspension Culture of *Coffea Arabica* L.” *In Vitro Cellular & Developmental Biology - Plant* 51 (1): 93–101. <https://doi.org/10.1007/s11627-014-9643-z>.

Simões-Costa, M.C., E. Carapuça, and I.R. Moura. 2009. “SOMATIC EMBRYOGENESIS INDUCTION IN DIFFERENT GENOTYPES OF *COFFEA* SPP.” *Acta Horticulturae*, no. 812 (February): 295–300. <https://doi.org/10.17660/ActaHortic.2009.812.40>.

Steward, F. C., Marion O. Mapes, and Kathryn Mears. 1958. “Growth and Organized Development of Cultured Cells. II. Organization in Cultures Grown from Freely Suspended Cells.” *American Journal of Botany* 45 (10): 705–8. <https://doi.org/10.2307/2439728>.

Stoffelen, Piet, Michel Noirot, Emmanuel Couturon, Sylvain Bontems, Petra De Block, and François Anthony. 2009. “*Coffea Anthonyi*, a New Self-Compatible Central African Coffee Species, Closely Related to an Ancestor of *Coffea Arabica*.” *TAXON* 58 (1): 133–40. <https://doi.org/10.1002/tax.581014>.

Su, Ying-Hua, Yu-Bo Liu, and Xian-Sheng Zhang. 2011. “Auxin–Cytokinin Interaction Regulates Meristem Development.” *Molecular Plant* 4 (4): 616–25. <https://doi.org/10.1093/mp/ssr007>.

Sugimoto, Keiko, Lin Xu, Uta Paszkowski, and Makoto Hayashi. 2018. “Multifaceted Cellular Reprogramming at the Crossroads Between Plant Development and Biotic Interactions.” *Plant & Cell Physiology* 59 (4): 651–55. <https://doi.org/10.1093/pcp/pcy066>.

Sylvain, P. G. 1955. Some observations on *Coffea arabica* L. in Ethiopia. *Turrialba*. 5, 37–53.

Szymański, Paweł, Magdalena Markowicz, and Elżbieta Mikiciuk-Olasik. 2011. “Adaptation of High-Throughput Screening in Drug Discovery—Toxicological Screening Tests.” *International Journal of Molecular Sciences* 13 (1): 427–52. <https://doi.org/10.3390/ijms13010427>.

Thimm, Oliver, Oliver Blasing, Yves Gibon, Axel Nagel, Svenja Meyer, Peter Krüger, Joachim Selbig, Lukas A. Müller, Seung Y. Rhee, and Mark Stitt. 2004. “MAPMAN: A User-Driven Tool to Display Genomics Data Sets onto Diagrams of Metabolic Pathways and Other Biological Processes.” *The Plant Journal: For Cell and Molecular Biology* 37 (6): 914–39. <https://doi.org/10.1111/j.1365-313x.2004.02016.x>.

- Thomas, Clément, Denise Meyer, Christophe Himber, and André Steinmetz. 2004. "Spatial Expression of a Sunflower SERK Gene during Induction of Somatic Embryogenesis and Shoot Organogenesis." *Plant Physiology and Biochemistry: PPB* 42 (1): 35–42. <https://doi.org/10.1016/j.plaphy.2003.10.008>.
- Tonietto, Ângela, Juliana Hiromi Sato, João Batista Teixeira, Emanuel M. de Souza, Fabio O. Pedrosa, Octávio Luiz Franco, and Angela Mehta. 2012. "Proteomic Analysis of Developing Somatic Embryos of Coffea Arabica." *Plant Molecular Biology Reporter* 30 (6): 1393–99. <https://doi.org/10.1007/s11105-012-0425-7>.
- Trontin, Jean-François, Krystyna Klimaszewska, Alexandre Morel, Catherine Hargreaves, and Marie-Anne Lelu-Walter. 2016. "Molecular Aspects of Conifer Zygotic and Somatic Embryo Development: A Review of Genome-Wide Approaches and Recent Insights." *Methods in Molecular Biology (Clifton, N.J.)* 1359: 167–207. https://doi.org/10.1007/978-1-4939-3061-6_8.
- Usadel, Björn, Fabien Poree, Axel Nagel, Marc Lohse, Angelika Czedik-Eysenberg, and Mark Stitt. 2009. "A Guide to Using MapMan to Visualize and Compare Omics Data in Plants: A Case Study in the Crop Species, Maize." *Plant, Cell & Environment* 32 (9): 1211–29. <https://doi.org/10.1111/j.1365-3040.2009.01978.x>.
- Vahdati, Kourosh, Shima Bayat, Hassan Ebrahimpzadeh, Maryam Jariteh, and Masoud Mirmasoumi. 2008. "Effect of Exogenous ABA on Somatic Embryo Maturation and Germination in Persian Walnut (Juglans Regia L.)." *Plant Cell, Tissue, and Organ Culture*. <http://agris.fao.org/agris-search/search.do?recordID=US201300883926>.
- Velculescu, V. E., L. Zhang, B. Vogelstein, and K. W. Kinzler. 1995. "Serial Analysis of Gene Expression." *Science (New York, N.Y.)* 270 (5235): 484–87. <https://doi.org/10.1126/science.270.5235.484>.
- Verdeil, J. 2001. "Ultrastructural Changes in Coconut Calli Associated with the Acquisition of Embryogenic Competence." *Annals of Botany* 88 (1): 9–18. <https://doi.org/10.1006/anbo.2001.1408>.
- Verdeil, Jean-Luc, Laurence Alemanno, Nicolas Niemenak, and Timothy John Tranbarger. 2007. "Pluripotent versus Totipotent Plant Stem Cells: Dependence versus Autonomy?" *Trends in Plant Science* 12 (6): 245–52. <https://doi.org/10.1016/j.tplants.2007.04.002>.
- Vondrakova, Zuzana, Petre I. Dobrev, Bedrich Pesek, Lucie Fischerova, Martin Vagner, and Vaclav Motyka. 2018. "Profiles of Endogenous Phytohormones Over the Course of Norway Spruce Somatic Embryogenesis." *Frontiers in Plant Science* 9: 1283. <https://doi.org/10.3389/fpls.2018.01283>.
- Vossen, Herbert van der. 2001. Coffee breeding practices. In *Coffee: Recent Developments*, 184–201 (Eds R. J. Clarke and O. G. Vitzthum) Oxford: Blackwell Science Ltd.
- Vossen, Herbert van der, Benoît Bertrand, and André Charrier. 2015. "Next Generation Variety Development for Sustainable Production of Arabica Coffee (Coffea Arabica L.): A Review." *Euphytica* 204 (2): 243–56. <https://doi.org/10.1007/s10681-015-1398-z>.
- Wang, Zhong, Mark Gerstein, and Michael Snyder. 2009. "RNA-Seq: A Revolutionary Tool for Transcriptomics." *Nature Reviews. Genetics* 10 (1): 57–63. <https://doi.org/10.1038/nrg2484>.
- Wienkoop, Stefanie, Julia Weiss, Patrick May, Stefan Kempa, Susann Irgang, Luis Recuenco-Munoz, Matthias Pietzke, et al. 2010. "Targeted Proteomics for Chlamydomonas Reinhardtii Combined with Rapid Subcellular Protein Fractionation, Metabolomics and Metabolic Flux Analyses." *Molecular BioSystems* 6 (6): 1018–31. <https://doi.org/10.1039/b920913a>.
- Wintgens, Jean Nicholas. 2012. 'The coffee plant', in Nicolas, W. J. (ed.) *Coffee: Growing, processing, sustainable production*. Wiley-VCH, pp. 3–24.
- Yang, Xiyang, and Xianlong Zhang. 2010. "Regulation of Somatic Embryogenesis in Higher Plants." *Critical Reviews in Plant Sciences* 29 (1): 36–57. <https://doi.org/10.1080/07352680903436291>.
- Yang, Xiyang, Xianlong Zhang, Daojun Yuan, Fangyan Jin, Yunchao Zhang, and Jiao Xu. 2012. "Transcript Profiling Reveals Complex Auxin Signalling Pathway and Transcription Regulation Involved in Dedifferentiation and Redifferentiation during Somatic Embryogenesis in Cotton." *BMC Plant Biology* 12 (July): 110. <https://doi.org/10.1186/1471-2229-12-110>.

- Yang, Yongil, and Dale Karlson. 2012. "Effects of Mutations in the Arabidopsis Cold Shock Domain Protein 3 (AtCSP3) Gene on Leaf Cell Expansion." *Journal of Experimental Botany* 63 (13): 4861–73. <https://doi.org/10.1093/jxb/ers160>.
- Zimmerman, J. Lynn. 1993. "Somatic Embryogenesis: A Model for Early Development in Higher Plants." *The Plant Cell* 5 (10): 1411–23. <https://doi.org/10.2307/3869792>.
- Zivy, Michel, Stefanie Wienkoop, Jenny Renaut, Carla Pinheiro, Estelle Goulas, and Sebastien Carpentier. 2015. "The Quest for Tolerant Varieties: The Importance of Integrating 'Omics' Techniques to Phenotyping." *Frontiers in Plant Science* 6. <https://doi.org/10.3389/fpls.2015.00448>.
- Zuo, Jianru, Qi-Wen Niu, Giovanna Frugis, and Nam-Hai Chua. 2002. "The WUSCHEL Gene Promotes Vegetative-to-Embryonic Transition in Arabidopsis." *The Plant Journal: For Cell and Molecular Biology* 30 (3): 349–59. <https://doi.org/10.1046/j.1365-313x.2002.01289.x>.

ANNEXES

An innovative automated active compound screening system allows high-throughput optimization of somatic embryogenesis in *Coffea arabica*

Rayan Awada^{1,2,3*}, Dorothée Verdier¹, Solène Froger¹, Eric Brulard¹, Simone de Faria Maraschin⁴, Hervé Etienne^{2,3} & David Breton¹

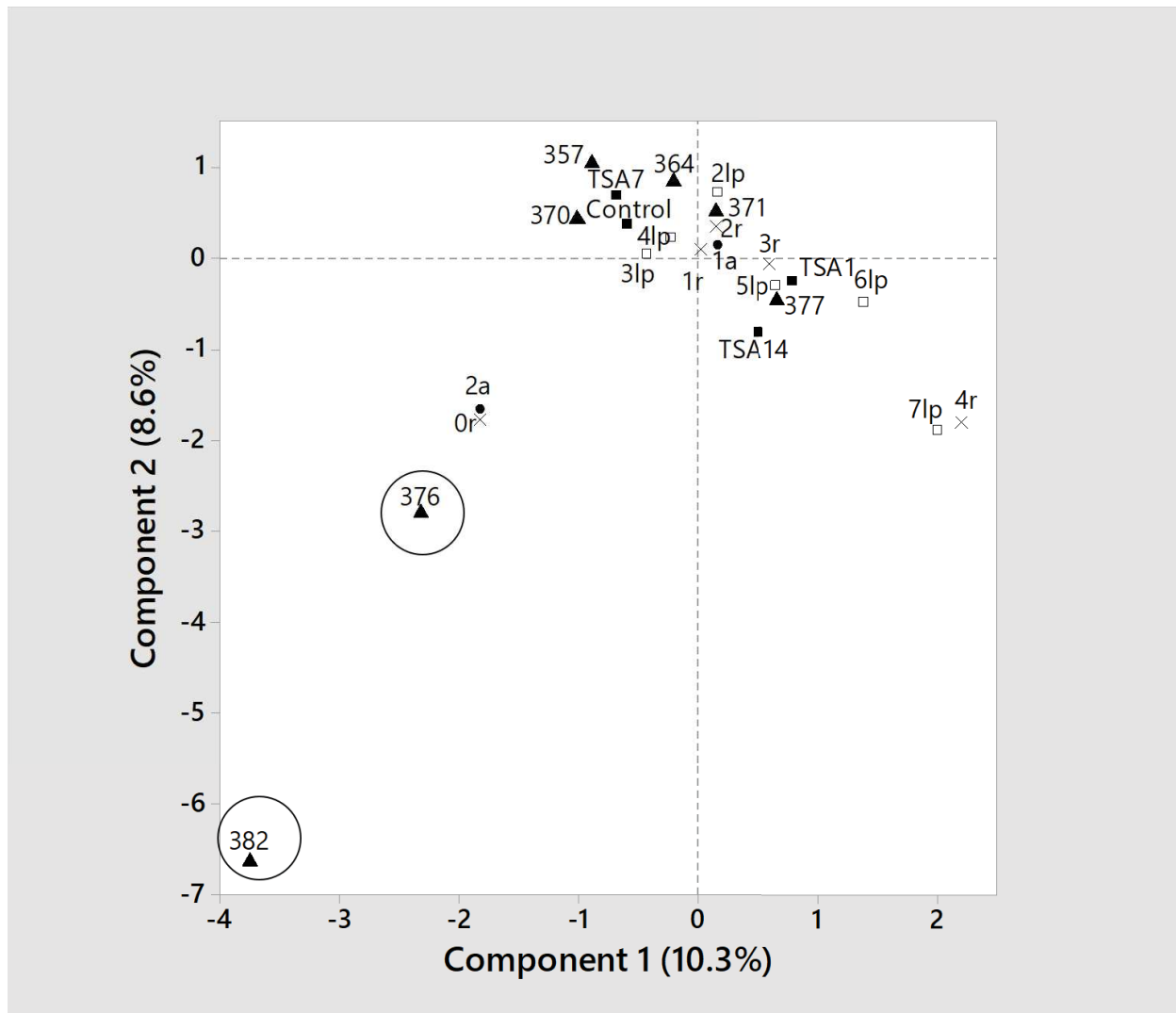
¹ *Nestlé R&D Center Tours, 101 avenue Gustave Eiffel, 37097 Tours Cedex 2, France*

² *CIRAD, UMR IPME, F-34398 Montpellier, France*

³ *UMR IPME, Université de Montpellier, CIRAD, IRD, F-34398 Montpellier, France*

⁴ *Bioalternatives SAS, 1 bis Rue des Plantes, 86160 Gençay, France*

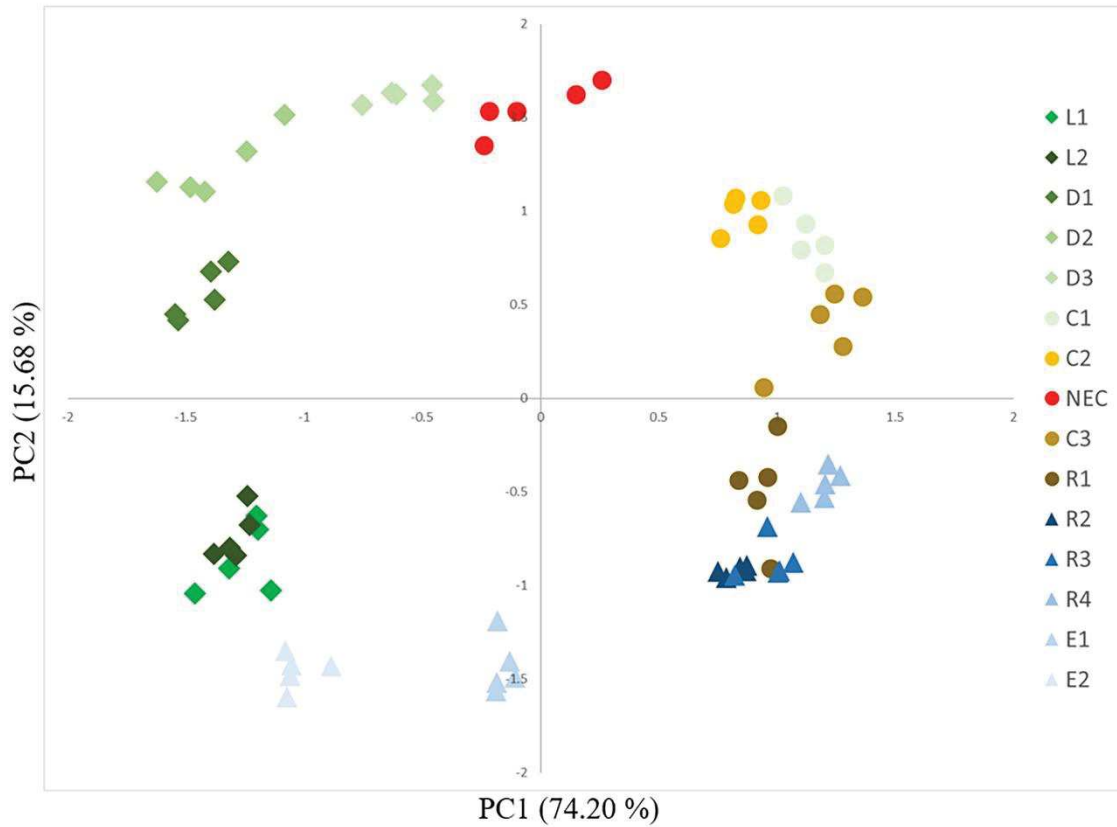
* Correspondence: Rayan Awada, rayan.awada@cirad.fr



Supplementary data. *In vitro* plant phenotypic characterization following 1 μ M TSA treatment during the coffee SE process. A graphic representation of a Multiple Component Analysis (MCA) was performed on plant phenotypic variables. Data from 160 plants were plotted according to the first 2 axes of the MCA. The number of pairs of leaves are represented by white squares, the number of roots by crosses, the number of stems by black spots, the leaf colour index by triangles and the *in vitro* treatment prior to regeneration by black squares. Plotting categorical data according to axes 1 and 2 revealed a strong contribution of 3 atypical plants scored with a single colour index. Therefore, no distinction was detected between TSA treatments. We decided to discard those data and run another MCA (Fig. 6).

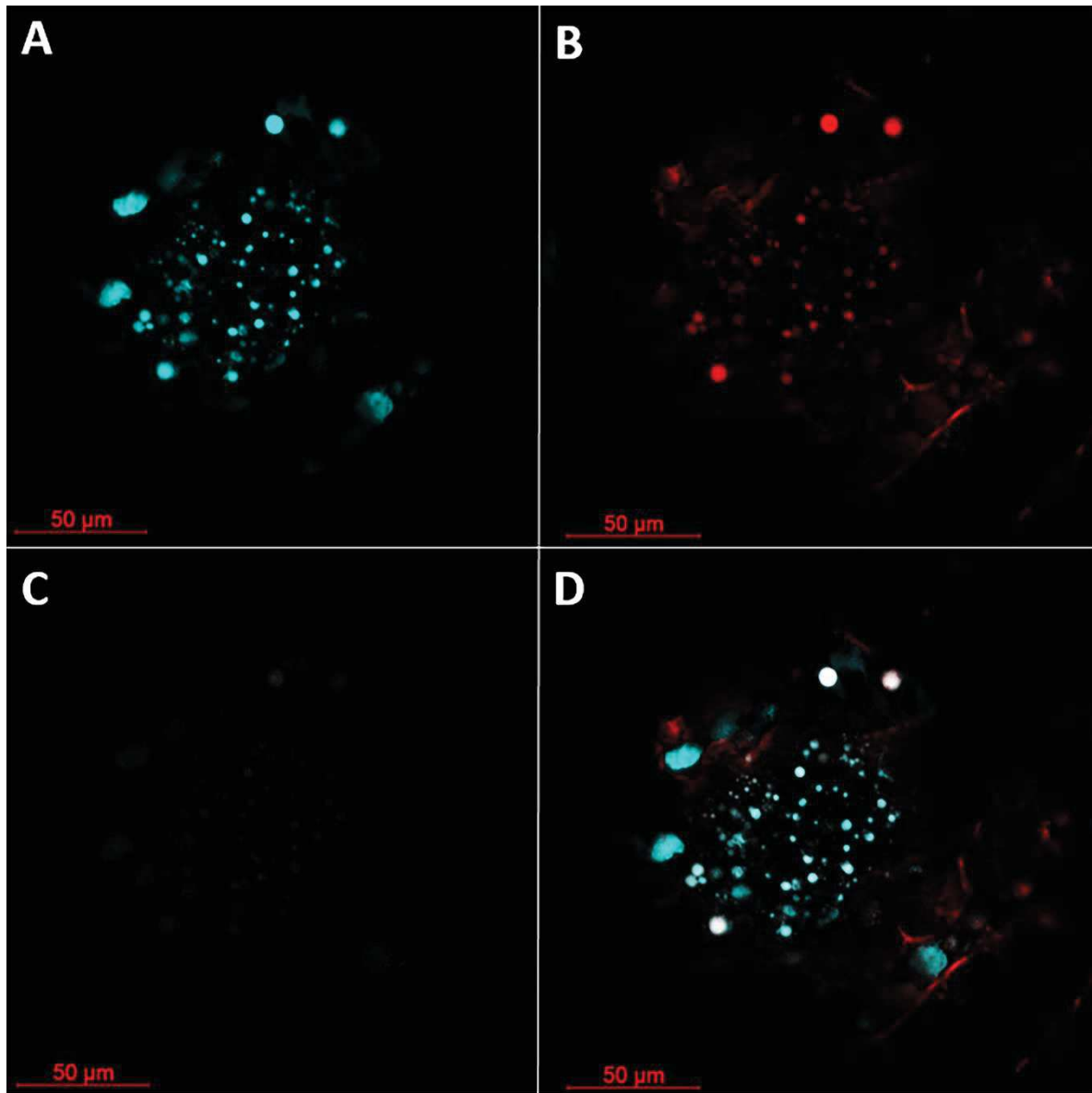
2 **Unravelling the Metabolic and Hormonal Machinery During Key Steps of**
3 **Somatic Embryogenesis: A Case Study in Coffee**

4



5

6 **Figure S1.** Assessment of homogeneity of metabolic replicates in 15 key sampled stages of the somatic
7 embryogenesis process. A Principal Component Analysis (PCA) was carried out to verify
8 homogeneity of 5 independent replicates; i.e. 5 independent leaf introductions in February, April,
9 June, October and December 2016. Samples correspond to : Leaves from greenhouse plants (L1),
10 explants during dedifferentiation [0 h (L2), 1 week (D1), 2 weeks (D2), 5 weeks (D3)], compact primary
11 callus obtained 3 months after induction (C1), embryogenic callus obtained 7 months after induction
12 (C2), established cell clusters obtained after 4 months in proliferation medium (C3), early regeneration
13 of embryos from cell clusters [1 week in DIF medium without reducing cell density (R1), 24 h in DIF
14 medium after reducing cell density (R2), 72 h (R3), 10 d (R4)], globular embryos (E1) and torpedo-
15 shaped embryos (E2). Non-embryogenic callus (NEC) was also sampled at the same time as the
16 embryogenic callus (C2).



17

18 **Figure S2.** Validation of the presence of total chlorogenic acids in Arabica globular-shaped embryos
19 by multiphoton microscopy combined with emission spectral analysis. After obtaining the spectral
20 acquisitions, the LinearUnmixing function (Zen software) was executed to separate, pixel by pixel,
21 the mixed signals of defined pure autofluorescent compounds (chlorogenic acids (A) and chlorophyll
22 (B)), using the entire emission spectrum of each compound, plus a residual channel (C). An overlay
23 image of the three channels was finally generated (D).

24

25

26

27

28

29

30

31 **Table S1.** List of primary and secondary metabolites detected over the 15 sampled stages of the Arabica somatic
 32 embryogenesis process.

Metabolite classification	Metabolites detected
Primary metabolites	2-Imidazolidone-4-carboxylic acid
	Aconitic acid, cis-
	Adenosine-5-monophosphate/Guanosine-5-monophosphate
	Alanine
	Alanine, beta-
	Arabinonic acid
	Arabinose
	Asparagine
	Aspartic acid
	Benzoic acid
	Benzoic acid, 4-hydroxy-
	Benzylalcohol
	Boric acid
	Butanoic acid, 4-hydroxy-
	Caffeic acid, trans-
	Citric acid
	Erythritol
	Erythronic acid
	Ethanolamine
	Ethanolaminephosphate
	Fructose
	Fructose-1-phosphate
	Fructose-6-phosphate
	Fumaric acid
	Furan-2-carboxylic acid
	Galactaric acid
	Galactinol
	Galactonic acid
	Gentiobiose
	Gluconic acid
	Glucose
	Glucose-6-phosphate
	Glutaric acid, 2-oxo-
	Glutaric acid, 3-hydroxy-3-methyl-
	Glyceric acid
	Glyceric acid-3-phosphate
	Glycerol
	Glycerol-3-phosphate
	Glycerophosphoglycerol
	Glycine
	Glycolic acid
	Guanosine
	Hexadecanoic acid
	Hydroquinone
	Inosine
	Inositol, myo-

Inositol-phosphate, myo-
Isocaproic acid, 2-oxo-
Lactic acid
Leucine
Lysine
Malic acid
Malic acid, 2-methyl-
Malonic acid
Maltose
Maltotriose
Mannitol
Mannose-6-phosphate
Methionine
Nicotinic acid
Octadecanoic acid
Orotic acid
Phenylalanine
Phosphoric acid
Phosphoric acid monomethyl ester
Piperidine-2-carboxylic acid
Proline
Proline, 4-hydroxy-, cis-
Propanoic acid, 2-amino-2-methyl-3-hydroxy-
Purine, 6-benzylamino-, 9-beta-D-glucopyranosyl-
Putrescine
Pyroglutamic acid + Glutamic acid + Glutamine
Quinic acid
Raffinose
Ribitol
Ribonic acid
Ribose
Saccharic acid
Salicylic acid
Shikimic acid
Sorbitol
Starch
Succinic acid
Sucrose
Thiazole, 4-methyl-5-hydroxyethyl-
Threonic acid
Trigonelline
Tryptophan
Uridine
Valine
Xanthosine
Xylose

Secondary metabolites

3-Caffeoylquinic acid
3,4-diCaffeoylquinic acid

3,5-diCaffeoylquinic acid
4-Caffeoylquinic acid
4,5-diCaffeoylquinic acid
5-Caffeoylquinic acid
Caffeine
Catechin
Epicatechin
Feruloylquinic acid
Mangiferin
Theobromine

33

34

35

Résumé

Contexte

L'embryogenèse somatique (ES) est l'un des processus de propagation les plus prometteurs pour la diffusion à grande échelle de variétés élités. Cependant, quelle que soit l'espèce considérée, les recherches sur l'ES restent essentiellement empiriques et à bas débit, entraînant une difficulté à répondre à la demande croissante en matériel végétal, notamment en raison des faibles progrès techniques au cours des dernières décennies. Cette thèse présente un travail pionnier où deux groupes leader - Nestlé et le Cirad - ont décidé d'unir leurs forces pour démêler les mécanismes moléculaires impliqués dans les étapes clés de l'ES du caféier Arabica, l'un des procédés les plus avancés, en y appliquant les dernières technologies de métabolomique et de transcriptomique. En parallèle, la mise au point d'un système de criblage à haut débit de composés actifs sur l'ES, qui permettrait d'accélérer l'optimisation des protocoles, n'a toujours pas été proposée pour la micropropagation végétale.

Résultats

Soixante-dix échantillons (14 stades de développement x 5 répétitions biologiques) couvrant l'ensemble du processus d'ES ont été étudiés. Une approche statistique robuste appliquée aux données RNA-seq a permis d'identifier 6 transitions ('switches') de développement clés pour le succès biologique de l'ES c.à.d. la production en masse d'embryons somatiques, et a révélé l'influence directe des facteurs ('drivers') environnementaux. La métabolomique a mis en évidence d'énormes reconfigurations des contenus hormonaux, des voies métaboliques primaires et des composés phénoliques lors des événements précoces de la dédifférenciation et de la redifférenciation. La totipotence atteint son expression la plus visible au stade embryogène lorsque sont totalement inhibées les voies biochimiques liées au métabolisme du sucre et, au contraire, lorsqu'est activée l'expression des gènes codant pour des facteurs de transcription. L'étude a également montré que les profils de transcrits et de métabolites représentent des signatures discriminantes du destin cellulaire. Une liste de 23 gènes et 17 métabolites marqueurs du succès des différentes phases de développement a été identifiée, y compris le stade callos embryogène si stratégique car à l'interface Dédifférenciation-Redifférenciation. En parallèle, nous avons réussi à reproduire à l'échelle pilote la différenciation d'embryons somatiques à partir de suspensions dans un système miniaturisé compatible avec une plateforme automatisée et ainsi permis le criblage de 4 régulateurs épigénétiques (HDACi) à différentes concentrations, apportant ainsi une preuve de concept de la faisabilité d'une optimisation haut-débit des milieux nutritifs.

Conclusion

Pour toutes les étapes de développement clés de l'ES, le décryptage des principaux mécanismes moléculaires ainsi que l'identification de marqueurs moléculaires spécifiques offre de nouvelles perspectives pour la compréhension fine de ce phénomène de régénération et son optimisation raisonnée. Des stratégies sont proposées pour la validation des marqueurs candidats et l'industrialisation du système de criblage haut-débit. Il est proposé une combinaison des deux approches -omiques et automatisation pour une optimisation des protocoles rationnelle et plus efficace, permettant aux procédés d'ES de soutenir la diffusion à très grande échelle du progrès génétique chez le caféier.

Mots clés : criblage haut-débit, destin cellulaire, embryogenèse somatique, métabolomique, totipotence, transcriptomique.

Abstract

Background

Somatic embryogenesis (SE) is one of the most promising propagation processes for the large-scale dissemination of elite varieties. However, whatever the species considered, research on SE remains essentially empirical and low-throughput, making it difficult to meet the growing demand for plant material, in particular due to the slow technical progress in recent decades. This thesis presents a pioneering work in which two leading groups - Nestlé and CIRAD - have decided to join forces to unravel the molecular mechanisms involved in the key stages of SE in Arabica coffee, one of the most advanced processes, by applying the latest metabolomics and transcriptomics technologies. In parallel, the implementation of a high-throughput screening system for active compounds on SE, that would speed up the optimization of protocols, has not yet been proposed for plant micropropagation.

Results

Seventy samples (14 developmental stages x 5 biological replicates) covering the whole SE process were studied. A robust statistical approach applied to RNA-seq data led to the identification of 6 key developmental switches for the biological success of SE i.e. the mass production of somatic embryos, and revealed the direct influence of environmental drivers. Metabolomics highlighted huge reconfigurations of hormonal contents, primary metabolic pathways and phenolic compounds during the early events of dedifferentiation and redifferentiation. Totipotency reaches its most visible expression at the embryogenic stage when the biochemical pathways related to sugar metabolism are inhibited and, on the contrary, when the expression of genes coding for transcription factors is activated. The study also showed that transcript and metabolite profiles represent discriminant signatures of cell fate. A list of 23 genes and 17 metabolites that mark the success of the different developmental phases has been identified, including the embryogenic callus stage, so strategic due to its position at the Dedifferentiation-Redifferentiation interface. In parallel, we succeeded in reproducing on a pilot scale the differentiation of somatic embryos from liquid cell suspensions in a miniaturized system compatible with an automated platform and thus allowed the screening of 4 epigenetic regulators (HDACi) at different concentrations, bringing a proof of concept of the feasibility of nutrient media high-throughput optimization.

Conclusion

For all SE key developmental stages, deciphering the main molecular mechanisms as well as the identification of specific molecular markers offers new perspectives for a detailed understanding of this regeneration phenomenon and its rational optimization. Strategies are proposed for the validation of candidate markers and the industrialization of the high-throughput screening system. A combination of both -omics and automation approaches is proposed for rational and more efficient protocol optimization, allowing the SE processes to support the very large-scale dissemination of genetic progress in coffee.

Key words: cell fate, high-throughput screening, metabolomics, somatic embryogenesis, totipotency, transcriptomics.