# Impact des éléments transposables sur la structure, la fonction et l'évolution des génomes

Nathalie Picault

Mémoire d'Habilitation à Diriger des Recherches

Délivré par

**UNIVERSITE DE PERPIGNAN VIA DOMITIA**

Unité de Recherche :
Laboratoire Génome et développement des Plantes
UMR5096

Présenté par :

**Nathalie PICAULT**

# Impact des éléments transposables sur la structure, la fonction et l'évolution des génomes

**24 Juin 2015**

**JURY**

| | | |
|---|---|---|
| Cristina Vieira | Rapporteur | Université de Lyon 1 |
| Boulos Chaloub | Rapporteur | INRA Evry |
| Nicolas Gilbert | Rapporteur | INSERM Montpellier |
| Josep Casacuberta | Membre du jury | CRAG Barcelona |
| Christoph Grunau | Président du jury | Université de Perpignan |

# Résumé

Les éléments transposables (ETs) sont des séquences d'ADN mobiles constituant la majeure partie des génomes eucaryotes. Ils ont longtemps été considérés à tort comme de l'ADN « poubelle », car ils ont non seulement un impact sur la structure et la taille des génomes, et de nombreux travaux ont révélé qu'ils ont un impact fonctionnel en régulant l'expression des gènes adjacents. Malgré leur grand nombre dans les génomes des plantes, la plupart des ETs sont inactifs. Ceci est dû aux nombreuses mutations, délétions et autres variations structurales qu'ils subissent au cours de l'évolution, conduisant à leur extinction. De plus, les ETs font l'objet d'un contrôle épigénétique strict qui affecte leur capacité à transposer dans les conditions normales de développement. Cependant, différents types de stress peuvent conduire à la levée de la répression des ETs et induire leur réactivation. Cependant, à mon arrivée au laboratoire, le nombre d'éléments actifs décrits restait très faible par rapport à l'ensemble des ETs caractérisés chez le riz.

Au cours de ces 10 dernières années, mon travail a permis d'identifier de nouveaux éléments actifs au niveau transcriptionnel, mais aussi au niveau transpositionnel chez le riz grâce à l'utilisation de puce à ADN et du séquençage haut débit. En effet, le développement des nouvelles techniques de séquençage à haut débit a ouvert de nouveaux horizons pour étudier l'activité transpositionnelle, à l'échelle du génome entier. Les résultats de mon travail montrent que plusieurs familles d'ETs sont activées transcriptionnellement par différents stress biotiques ou abiotiques comme un excès de fer, une infection par un virus (RYMV) ou lors d'une interaction du riz avec une bactérie (*Azospirillum*). J'ai aussi pu quantifier pour la 1ère fois l'activité transpositionnelle globale chez le riz.

L'ensemble de ce travail a permis de mieux comprendre l'impact des éléments transposables sur la structure, la fonction et l'évolution des génomes des plantes. Toutefois, si ces connaissances forment un socle solide pour la compréhension des ETs, il reste encore beaucoup de choses à découvrir sur ces éléments qui passionnent un grand nombre de chercheurs depuis Barbara McClintock.

Parmi les nombreuses copies d'éléments transposables insérées dans les génomes, certaines ont été « domestiquées » par les génomes et sont devenues des gènes ou des éléments de régulation des gènes. Je vais donc maintenant me concentrer sur cet aspect des ETs afin de voir leur impact évolutif contribuant à la diversité des génomes, et donc à leur évolution.

En effet, les ETs doivent être maintenant regardés comme des éléments-clés dans un grand nombre de processus, refondant les génomes et contrôlant l'activité des gènes. Les études des génomes vont de plus en plus se focaliser sur les séquences répétées et leurs relations avec les gènes. C'est dans ce domaine d'analyse des réseaux d'interactions gènes-éléments transposables, responsables de divers phénotypes que les découvertes les plus stimulantes sont attendues.

# SOMMAIRE

# Dossier Administratif

**Nathalie PICAULT**

Maître de conférences

à l'Université de Perpignan *Via Domitia*

Née le 25 novembre 1975 à Paris, deux enfants

Section CNU : 66

Adresses :

Professionnelle :
Laboratoire Génome et Développement des Plantes
UMR 5096 CNRS, Université de Perpignan,
Bâtiment T
Université de Perpignan
58, avenue Paul Alduy
66860 Perpignan
tel.(+33)468662103 / fax.(+33)468668499
Email : nathalie.picault@univ-perp.fr

Domicile :
6 rue des pâquerettes
66470 Sainte Marie la mer
tel. (+33)609730864

***Titres Universitaires et expériences professionnelles***

1995 **DEUG Biologie**, Université Paris VII

1998 **Magistère de Génétique**, Université Paris VII

1998 **DEA Physiologie Cellulaire et Moléculaire des Plantes**, INA-PG, Université Paris XI, Université Paris VI

1998-2002 **Doctorat en Biologie**, Université Paris XI, Orsay
Travail de recherche effectué à l'Institut de Biotechnologies des Plantes, Université Paris XI, Orsay. Sujet : Caractérisation de transporteurs mitochondriaux d'acides organiques chez les plantes supérieures.

2002-2004 **Stage Post-Doctoral**, CEA de Cadarache, Laboratoire d'Ecophysiologie de la photosynthèse. Sujet : Effets d'une surproduction de peptides chélateurs sur l'accumulation des métaux lourds chez les plantes.

2005- **Maître de conférences**, Université de Perpignan via Domitia (UPVD)

2009-2012 Prime d'excellence scientifique (PES)

***Activités en matière d'enseignement***

J'effectue chaque année 192 heures équivalent TD réparties de la façon suivante :

- En Licence Biologie-Ecologie :
  **Génétique mendélienne, Biologie cellulaire** en 1ère année
  **Génétique eucaryote approfondie** en 3ème année
  **Structure et Dynamique des génomes eucaryotes** en 2ème et 3ème année
  **Bio-informatique** en 3ème année

- En Master recherche BIMoPoDD (Biologie Intégrée des molécules aux populations) :
  **Dynamique du génome** en 1ère année master recherche BIMoPoDD

- En Master professionnel MOBI, GMEA et BDD :
  **Phytoremédiation** en 2ème année

J'ai été responsable du Travail d'étude bibliographique (TEB) en licence Biologie-Ecologie 3ème année au cours de l'année 2007-2008.

Je suis responsable des matières suivantes depuis septembre 2007 :
Génétique mendélienne en Licence Biologie-Ecologie en 1ère année
Génétique eucaryote approfondie, en Licence Biologie-Ecologie 3ème année
Dynamique du génome, en Master recherche BIMoPoDD 1ère année

Je suis responsable de l'enseignement de génétique sur l'Université de Perpignan depuis la rentrée 2009.

***Responsabilités administratives***

***Au sein de l'Université :***

- **Présidente de jury du Master première année spécialité Génomique environnementale**, mention "Biologie intégrée : Molécules, Populations et Développement Durable, (BIMoPoDD)" de l'UPVD depuis septembre 2008.

- Responsable du Travail d'étude bibliographique (TEB) en licence Biologie-Ecologie 3ème année au cours de l'année 2007-2008.

- Participation à l'élaboration des maquettes et constitution du dossier de demande d'habilitation pour le master BIMoPoDD (campagne 2011).

- Membre de jurys restreints :
  Licence Biologie-Ecologie 3ème année 2010-2014
  Master 2ème année BIMoPoDD 2010-2015

***Responsabilités de projets de recherche***

2011-2012 : Responsable d'un projet de collaboration franco-taïwanais (PHC ORCHID). « Identification of retrotransposon/transposon in Phalaenopsis orchid genome » avec notre partenaire Taiwanais, Mme CHEN HONG-Hwa du NCKU.

2009-2012 : Responsable du projet ANR AZORIZ (coord. F. Wisniewski-Dye, Université Lyon1) : « Specificity of the phytostimulatory cooperation between *Azospirullum lipoferum* and rice ».

2008-2011 : Co-responsable d'un projet de collaboration franco-brésilien (EGIDE-COFECUB). « Adaptation aux stress abiotiques chez les plantes : analyse transcriptomique de la réponse au stress métallique (fer et aluminium) chez le riz ».

2006-2009 : Membre du projet ANR ITEGE (coord. M.A. Grandbastien, INRA Versailles) : "Impact of Transposable Elements on Gene Regulation".

***Publications***

***Publications dans des journaux à comité de lecture :***

13- Finatto T., Costa de Oliveira A., Chaparro C., Maia LC., Farias DR., Woyann LG., Mistura CC., Soares-Bresolin AP., Llauro C., Panaud O., and **Picault N.** (2015) Abiotic stress and genome dynamics : specific genes and transposable elements response to iron excess in rice. Accepted in Rice

12- Drogue B., Sanguin H., Chamam A., Mozar M., Llauro C., Panaud O., Pringent-Combaret C., **Picault N.**, Wisniewski-Dyé F. (2014) Plant root transcriptome profiling reveals a strain-dependent response during *Azospirillum*-rice cooperation. *Frontiers in Plant Science*. 5:607

11- El Baidouri, Carpentier MC, Cooke R, Gao D, Lasserre E, Llauro C, Mirouze M, Picault N, Jackson SA, Panaud O. (2014) Widespread and frequent horizontal transfers of transposable elements in plants. *Genome Res*. 24:831-838

10- Calvayrac C, Martin-Laurent F, Faveaux A, **Picault N**, Panaud O, Coste CM, Chaabane H, Cooper JF. (2012) Isolation and characterisation of a bacterial sttrain degrading the herbicide sulcotrione from an agricultural soil. *Pest Manag Sci*. 68: 340-347.

9- Sabot F, **Picault N**, Elbaidouri M, Llauro C, Chaparro C, Piegu B, Roulin A, Guiderdoni E, Delabastide M, McCombie R, Panaud O. (2011). Transpositional landscape of rice genome revealed by Paired-End Mapping of high-throughput resequencing data. *Plant Journal*. 66: 241-246. Contribution équivalente des deux premiers auteurs.

8- **Picault N.**, Chaparro C., Piegu B., Stenger W., Formey D., Llauro C., Descombin J., Sabot F., Lasserre E., Meynard D., Guiderdoni E. and Panaud O. (2009) Identification of an active LTR-retrotransposon in rice. *Plant Journal*. 58: 754-765.

7- Palmieri L., **Picault N**., Arrigoni R., Besin E., Palmieri F. and Hodges M. (2008) Molecular identification of three Arabidopsis thaliana mitochondrial dicarbosylate carrier isoforms: organ distribution, bacterial expression, reconstitution into liposomes and functional characterization. *Biochem J*. 410: 621-629.

6- Piegu B., Guyot R., **Picault N.**, Roulin A., Saniyal A., Kim HR., Collura K., Brar D.S., Jackson S., Wing R.A. and Panaud O. (2006) Doubling genome size without polyploidization : direct evidence of retrotransposition-driven genomic bursts in *Oryza* genus. *Genome Research*. 16: 1262-1269.

5- Gayet L., **Picault N**., Cazale AC., Beyly A., Lucas P., Jacquet H., Suso HP., Vavasseur A., Pelier G. and Forestier C. (2006) Transport of antimony salts by *Arabidopsis thaliana* protoplasts over-expressing the human multidrug resistance-associated protein 1 (MRP1/ABCC1). *FEBS Lett*. 580: 6891-6897.

4- **Picault N**., Cazale AC., Beyly A., Cuine S., Carrier P., Luu DT., Forestier C. and Pelier G. (2006) Chloroplast targeting of phytochelatin synthase in *Arabidopsis*: effects on heavy metal tolerance and accumulation. *Biochimie*. 88: 1743-1750.

3- **Picault N**., Hodges M., Palmieri L. and Palmieri F. (2004) The growing family of mitochondrial carriers in *Arabidopsis*. *Trends Plant Sci.* 9: 138-146.

2- **Picault N**., Palmieri L., Pisano I., Hodges M. and Palmieri F. (2002) Identification of a novel transporter for dicarboxylates and tricarboxylates in plant mitochondria. Bacterial expression, reconstitution, functional characterization, and tissue distribution. *J. Biol. Chem*. 277: 24204-24211.

1- Mourrain P., Beclin C., Elmayan T., Feuerbach F., Godon C., Morel J.B., Jouette D., Lacombe A.M., Nikic S., **Picault N.**, Remoue K., Sanial M., Vo T.A. and Vaucheret H. (2000) *Arabidopsis SGS2* and *SGS3* genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell*. 101: 533-542.

***Thèse de doctorat :***

**Picault N.** (2002) Caractérisation de transporteurs mitochondriaux d'acides organiques chez les plantes supérieures. Thèse de Doctorat en Sciences, Université Paris XI, Orsay.

***Chapitre d'ouvrage :***

Palmieri F., **Picault N**., Palmieri L., and Hodges M. (2004) Plant mitochondrial carriers. In: Advances in Photosynthesis and Respiration (eds., Day D, Millar H and Whelan J).

***Communications***

***Présentations orales*** *(séminaires et congrès) :*

**Picault N.** Transpositional landscape of the rice Genome revealed by paired-End Mapping of high-throughput sequencing. Université de Tainan, Taïwan, 3 décembre 2012. (communication orale sur invitation)

**Picault N**. Update on *O. meridionalis* Genome Sequencing. International Rice Functional Genomics Consortium, 10th ISRFG, Chiang Mai, Thaïlande, 26-29 novembre 2012.

**Picault N**. Transpositional landscape of the rice Genome revealed by paired-End Mapping of

high-throughput sequencing. Université de Pelotas, Brésil, 27 février 2012. (communication orale sur invitation)

Finatto T., Picault N., Chaparro C., Soares-Bresolin AP., Farias DR., Mistura CC., Woyann LG., Maia LC., Panaud O. and Costa de Oliveira A. Cis-regulatory elements occurrence in promoters of antioxidant signaling network genes up-regulated in rice under iron stress toxicity. 8th ISRFG, Bento Gonçalves, Brazil, 18-20 octobre 2010.

**Picault N.** Identification of an active LTR-retrotransposon in rice. Réunion GDR2157: "Evolution des Eléments transposables : du génome aux populations", Colloque GDR éléments transposables, Institut Pasteur, Paris, 9-10 décembre 2009.

**Picault N.** Identification of an active LTR-retrotransposon in rice. Université de Pelotas, Brésil, 15 mai 2009. (communication orale sur invitation)

**Picault N.** The rice LTR-retrotransposons: structure, evolution and function. University of Kobe, Kobe, Japon, mai 2007. (communication orale sur invitation)

**Picault N.** Comment explorer l'activité transcriptionnelle des rétrotransposons chez le riz ? Réunion GDR2157 : "Evolution des Eléments transposables : du génome aux populations", Colloque GDR éléments transposables, Institut Pasteur, Paris, 7-8 décembre 2006.

**Picault N**, Palmieri L, Pisano I, Palmieri F, Besin E, Hodges M. Identification of a novel transporter for dicarboxylates and tricarboxylates in plant mitochondria. 5th General Meeting of the French Societyof Plant Physiologists, Orsay, 9-11 juillet 2003. (communication orale sur invitation)

***Posters :***

Drogue B., Sanguin H., Borland S., **Picault N**., Prigent-Combaret C., Wisniewski-Dyé F. Host specificity of the plant growth-promoting cooperation between *Azospirillum* and rice. 12th Symposium on Bacterial Genetics and Ecology (Ljubljana, Slovénie), 9-13 Juin 2013.

**Picault N**. Drogue B., Sanguin H., Chamam A., Mozar M., Llauro C., Pringent-Combaret C., Panaud O., Wisniewski-Dyé F. Specificity of the phytostimulatory cooperation between *Azospirillum lipoferum* and rice. International Rice Functional Genomics Consortium, Chiang Mai, Thaïlande, 26-29 novembre 2012.

Sanguin H., Chamam A., Drogue B., Progent-Combaret C., Bertrand C., Comte G., Mozar M., Llauro C., **Picault N.**, Panaud O., Wisniewski-Fyé F. AZORIZ: Specificity of the phytostimulatory cooperation between *Azospirillum lipoferum* and rice. 2nd Congress on Transposable Elements, Saint-Malo (France), 21-24 avril 2012.

Sabot F., Picault N., El-Baidouri M., Llauro C., Chaparro C., Piegu B, Roulin A., Guiderdoni E., Delabastide M., McCombie R. and Panaud O. Transpositional landscape of the Rice genome revealed by Paired-end mapping of high-throughput re-sequencing data. 2nd Congress on Transposable Elements, Saint-Malo (France), 21-24 avril 2012.

Chamam A., Sanguin H., Drogue B., Prigent-Combaret C., Bertrand C., Comte G., Mozar M., Picault N., Panaud O., Wisniewski-Dyé F. AZORIZ : Spécificité de la coopération phytostimulatrice *Azospirillum lipoferum* / riz. 10ièmes rencontres plantes-bactéries, Aussois,

30 janvier-3 février 2012.

Finatto T., Picault N., Chaparro C., Soares-Bresolin AP., Farias DR., Mistura CC., Woyann LG., Maia LC., Panaud O. and Costa de Oliveira A. Transcriptome analysis in rice (*oryza sativa* L.) under iron stress toxicity. 57ème congrès de génétique du Brésil, 30 Aout-2 septembre 2011.

Finatto T., Picault N., Chaparro C., Farias DR., Mistura CC., Maia LC., Panaud O. and Costa de Oliveira A. Iron toxicity stress conditions enhances expression of metallothionein genes in rice. Plant and Animal Genome XIX conference, San Diego, California, 15-19 janvier 2011.

Finatto T., **Picault N.**, Maia LC., Panaud O. and Costa de Oliveira A. Analysis of rice transcriptome under iron stress conditions. 30ième anniversaire du Programme CAPES-COFECUB. Salvador de Bahia, Brésil, 24-26 mai 2009.

**Picault N.** Chaparro C, Piégu B, Stenger W, Formey de St Louvent D, Llauro C, Descombin J, Sabot F, Lasserre E, Meynard D, Guiderdoni E and Panaud O. Identification of an active LTR-retrotransposon in rice. International Rice Functional Genomics Consortium, Jeju, South Korea, 10-12 novembre 2008.

**Picault N**, Chaparro C, Piégu B, Stenger W, Formey de St Louvent D, Descombin J, and Panaud O. Transcriptomic Assay of the Rice transposome, Plant and Animal Genome XVI conference, San Diego, California, 12-16 janvier 2008.

**Picault N**, Chaparro C, Piégu B, Stenger W, Formey de St Louvent D, and Panaud O. Transcriptomic Assay of the Rice transposome, 6th Plant Genomics European Meeting, Tenerife, Espagne, 3-6 Octobre 2007.

**Picault N**, Chaparro C, Piégu B, Stenger W, Formey de St Louvent D, Panaud O. The Rice Transposome Project. XV ème congrès national Eléments Transposables, Rennes, France, 3-5 juillet 2007.

C Chaparro, **Picault N**, Piégu B, Guyot R, Zuccolo A et Panaud O. A post-genomic approach to study transposable elements in plants: the Rice (Oryza sativa) case. XIV ème colloque Eléments Transposables, Clermont-Ferrand, France, 3-5 Juillet 2006.

**Picault N**, Beyly A, Cuine S, Carrier P, Forestier C, Peltier G. Effects of overexpressing metal-binding peptides and transporters on heavy metal tolerance and accumulation in *A. thaliana*. 13th International Workshop on Plant Membrane Biology, Montpellier, 6-10 juillet 2004.

**Picault N**, Palmieri L, Pisano I, Palmieri F, Besin E, Hodges M. Identification of a novel transporter for dicarboxylates and tricarboxylates in plant mitochondria. 5th General Meeting of the French Society of Plant Physiologists, Orsay, 9-11 juillet 2003.

**Picault N**, Palmieri L, Pisano I, Hodges M, Palmieri F. Identification of a novel subfamily of plant mitochondrial transporters for dicarboxylates and tricarboxylates. 6th International Symposium on Inorganic Nitrogen Assimilation, Reims, 8-12 juillet 2001.

***Encadrement scientifique***

***Co-direction de thèse :***

- Oct. 2008 - Fév. 2012 : **Taciane Finatto** (étudiante brésilienne en co-tutelle, financement CAPES-COFECUB), Université de Perpignan et Université de Pelotas : Adaptation aux stress abiotiques chez les plantes : analyse transcriptomique de la réponse au stress métallique (fer et aluminium) chez *Oryza sativa* (en France du 15/11/2008 au 31/10/2009 sous ma direction).

Taciane Finatto a obtenu un poste permanent à l'Université Technologique Fédérale de Paraná (« Universidade Tecnológica Federal do Paraná » (UTFPR)) dans l'Etat du Paraná au Brésil en 2013.

- Sept. 2006 - 20 Sept. 2011 : **Christophe Calvayrac** a effectué sa thèse en même temps que son emploi d'enseignant dans un BTS (ce qui explique le durée de la thèse). J'ai co-encadré sa thèse en co-direction avec O. Panaud et JF Cooper, Université de Perpignan : Dégradation biologique de la sulcotrione dans un sol agricole : recherche d'une éventuelle biodégradation accélérée et caractérisation de souches bactériennes potentiellement dégradantes.

Christophe Calvayrac est actuellement enseignant (PRCE) à l'IUT de Perpignan.

***Encadrement d'étudiants en Master :***

- 2013 : Vivien Julia, Master recherche 1ère année BIMoPoDD, Université de Perpignan. Activité des rétrotransposons à LTR chez *Oryza sativa* lors d'un stress salin, analyses de données de RNAseq.
- 2011 : Corinne Marchal, Master recherche 1ère année DINEV, Université de Perpignan. Expression, méthylations et cotranscrits : influence des éléments transposables sur les gènes adjacents lors de l'infection par RYMV chez *O. sativa*.
- 2008 : Damien Formey, Master recherche 2ème année DINEV, Université de Perpignan. Study of the transcriptional activity of LTR retrotransposons and their impact on adjacent genes during viral stress in *Oryza sativa* L. ssp. *japonica*.
- 2007 : Willfried STENGER, Master recherche 2ème année d'Ecologie fonctionnelle, Université de Perpignan. Etude transcriptionnelle des familles de rétrotransposons à LTR du riz grâce à la puce transposome.
- 2007 : Damien Formey, Master recherche 1ère année d'Ecologie fonctionnelle, Université de Perpignan. Etude de l'activité transcriptionnelle et transpositionnelle de familles de rétrotransposons à LTR chez *Oryza sativa* L. ssp *japonica*.

***Encadrement de personnels :***

- 2011 : Michaël Mozar, ingénieur d'étude recruté en CDD d'un an dans le cadre du projet ANR AZORIZ.
- Depuis septembre 2007, je supervise le travail de Christel Llauro qui est assistante ingénieur (CNRS) au sein de l'équipe.

## *Participation à des jurys*

***Jurys de thèse** :*

- Taciane Finatto. Université de Perpignan et de Pelotas (Brésil). Transcriptomic analysis of genes and LTR retrotransposons in rice (*Oryza sativa* ssp. *Japonica*) in response to iron toxicity. 27 février 2012. Co-directrice de thèse.

***Comité de thèse** :*

- Amel Chamam. Université de Lyon 1. Caractérisation de l'interaction entre le riz (*Oryza sativa*) et la bactérie phytostimulatrice *Azospirillum lipoferum*. 8 septembre 2009.

- Benoît Drogue. Université de Lyon 1. Spécificité de la coopération phytostimulatrice Azospirullum-céréales. 18 janvier 2011.

- David Roquis. Université de Perpignan. Possible epigenetic origins of Dauermodifications in the parasite *Schistosoma mansoni* and the coral *Pocillopora damicornis*. 16 juillet 2013.

- David Roquis. Université de Perpignan. Rôle possible des mécanismes épigénétiques dans le phénomène de *Dauermodifikation* chez le corail *Pocillopora damicornis* et le parasite *Schistosoma mansoni*. 20 octobre 2014.

## *Activités éditoriales*

***reviews pour les journaux :***

BMC Plant Biology.
International Journal of Plant Genomics.
Plant Cell Reports.
Planta.
Chromosome research.

***Expertise nationale :***

*ANR blanc.*

# Dossier Scientifique

Concernant la rédaction de mon manuscrit pour l'Habilitation à Diriger des Recherches, je vais me concentrer sur les activités de recherche que j'ai menées depuis mon recrutement au sein de l'Université de Perpignan en février 2005, et plus particulièrement au Laboratoire Génome et Développement des Plantes (UMR5096) dans l'équipe « Analyse du génome et évolution » dirigée par Olivier Panaud. En effet, mes activités de recherche de thèse et de stage post-doctoral étant très éloignées de ce que j'ai pu réalisé ces dernières années, je vais seulement les présenter rapidement.

## ACTIVITÉS DE RECHERCHE ANTÉRIEURES

### *Stage de maîtrise*

Du 1er février au 31 août 1997, j'ai effectué mon stage de maîtrise à l'INRA de Versailles, au laboratoire de Biologie Cellulaire sous la direction d'Hervé Vaucheret. Tout au long de ce stage, j'ai été encadrée par Philippe Mourrain, un étudiant en deuxième année de thèse, pour analyser chez *Arabidopsis thaliana* des mutants affectés dans la mise en place de l'inactivation épigénétique post-transcriptionnelle. L'objectif était d'isoler le ou les gène(s) impliqué(s) dans ce mécanisme par une approche génétique de marche sur le chromosome. Le travail auquel j'ai participé a été publié dans Cell (Mourrain *et al*., 2000) :

Mourrain P., Beclin C., Elmayan T., Feuerbach F., Godon C., Morel J.B., Jouette D., Lacombe A.M., Nikic S., **Picault N.**, Remoue K., Sanial M., Vo T.A. and Vaucheret H. (2000) *Arabidopsis SGS2* and *SGS3* genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell*. 101: 533-542.

### *DEA et Doctorat*

Mon stage de DEA a été effectué à l'Institut de Biotechnologie des Plantes à Orsay sous la direction de Michael HODGES dans l'équipe "Régulation d'enzymes du métabolisme carboné" de Jean VIDAL. A l'issue du DEA et après l'obtention d'une bourse MNRT, j'ai poursuivi ce travail au cours de ma thèse toujours dans le même laboratoire. Mon travail de thèse avait donc pour objectif la caractérisation de transporteurs mitochondriaux d'acides organiques chez les plantes supérieures. J'ai soutenu ma thèse le 11 juin 2002, le jury était composé de Mme Françoise VEDELE et M. Alain VAVASSEUR comme rapporteurs, de Mme Hélène BARBIER-BRYGOO et M. Michael HODGES comme examinateurs et du professeur Pierre GADAL comme président du jury.

**Doctorat en Biologie, Université Paris XI, Orsay**

Travail de recherche effectué à l'Institut de Biotechnologies des Plantes, Université Paris XI, Orsay, sous la direction de Michaël Hodges.

Sujet : Caractérisation de transporteurs mitochondriaux d'acides organiques chez les plantes supérieures.

## RÉSUMÉ

Dans les cellules des eucaryotes, les échanges entre les différents compartiments sont fondamentaux pour le maintient d'une fonction cellulaire efficace et montrent ainsi l'importance de transporteurs pour chacun des organites. Les études biochimiques et moléculaires réalisées pendant ces 20 dernières années ont démontré la présence de plusieurs transporteurs dans la membrane interne mitochondriale des plantes. Les activités de transport mesurées suggèrent qu'il existe des transporteurs identiques à ceux mis en évidence chez les mammifères et la levure, mais aussi des transporteurs spécifiques des plantes. Cependant, des membres de la famille des transporteurs mitochondriaux décrits chez les animaux et la levure n'ont pas encore été identifiés chez les plantes. Grâce au séquençage complet du génome d'*Arabidopsis*, la présence de plusieurs membres putatifs de cette famille a été révélée chez cette plante modèle, mais la fonction et la spécificité des substrats transportés est à déterminer. Il est donc évident que, par rapport aux transporteurs mitochondriaux d'autres organismes, les études sur les transporteurs mitochondriaux végétaux n'en sont qu'à leur début.

Mon travail de thèse avait pour objectif d'identifier des ADNc codant des transporteurs mitochondriaux d'acides dicarboxyliques et tricarboxyliques chez *N. tabacum* et *A. thaliana*, puis de réaliser les caractérisations fonctionnelles des protéines correspondantes. Nous avons décidé d'élucider leur rôle physiologique en examinant l'impact de la modification de leur expression. Les approches choisies pour atteindre ce but ont été la transgénèse soit par sur- ou sous-expression, soit par la recherche de mutants d'insertion chez *A. thaliana*.

Ainsi, trois types de transporteurs mitochondriaux spécifiques des plantes ont été mis en évidence chez les plantes supérieures :

(1) les transporteurs d'acides dicarboxyliques et tricarboxyliques (DTC). La caractérisation des protéines recombinantes a mis en évidence que les ADNc isolés codent des transporteurs mitochondriaux transportant des acides dicarboxyliques comme l'α-cétoglutarate (αKG) et le malate mais aussi des acides tricarboxyliques comme le citrate. Ces transporteurs sont spécifiques des plantes, puisque dans les espèces non végétales, les transporteurs d'αKG ne transportent pas le

citrate. Des expériences de carence en azote suivie d'une ré-alimentation en nitrate chez le tabac ont mis en évidence une réponse coordonnée entre le *DTC* et plusieurs gènes impliqués dans le cycle de Krebs et le métabolisme azoté. Cette observation pourrait indiquer un rôle de cette protéine dans la fourniture en squelette carboné pour l'assimilation de l'azote (Picault *et al.*, 2002).

(2) les transporteurs d'acides dicarboxyliques (DIC). La caractérisation fonctionnelle des protéines montre qu'elles transportent le malate et l'oxaloacétate en échange de phosphate ou de sulfate comme les transporteurs d'acides dicarboxyliques des espèces non végétales, mais elles présentent aussi les caractéristiques du transporteur de l'oxaloacéate de levure. Les transporteurs DIC seraient donc spécifiques des végétaux. Un mutant nul a été isolé et le seul phénotype observé est un taux élevé de graines non germées. Cette observation suggère que le DIC pourrait avoir un rôle dans la gluconéogénèse, mais aussi dans le transfert du pouvoir réducteur (Palmieri *et al*., 2008).

(3) le transporteur de succinate et de malate (SFC). Ce transporteur est capable de complémenter le mutant *sfc* de levure, et les résultats préliminaires de sa caractérisation fonctionnelle indiquent qu'il transporte le succinate, le malate et l'αKG, mais pas le fumarate contrairement à la levure. L'étude d'un mutant d'insertion possédant l'ADN-T dans le gène *SFC* montre que cette mutation est létale à l'état homozygote ; des graines avortées ont pu être observées dans les siliques des plantes hétérozygotes pour la mutation.

Ce travail a donné lieu aux publications suivantes :

**Picault N**., Palmieri L., Pisano I., Hodges M. and Palmieri F. (2002) Identification of a novel transporter for dicarboxylates and tricarboxylates in plant mitochondria. Bacterial expression, reconstitution, functional characterization, and tissue distribution. *J. Biol. Chem*. 277: 24204-24211.

**Picault N**., Hodges M., Palmieri L. and Palmieri F. (2004) The growing family of mitochondrial carriers in *Arabidopsis*. *Trends Plant Sci.* 9: 138-146.

Palmieri L., **Picault N**., Arrigoni R., Besin E., Palmieri F. and Hodges M. (2008) Molecular identification of three *Arabidopsis thaliana* mitochondrial dicarbosylate carrier isoforms: organ distribution, bacterial expression, reconstitution into liposomes and functional characterization. *Biochem J*. 410: 621-629.

J'ai effectué mon stage post-doctoral au CEA de Cadarache au Laboratoire d'Ecophysiologie de la Photosynthèse, sous la direction de Gilles Peltier et au Laboratoire des Echanges Membranaires et Signalisation, sous la direction de Cyrille Forestier, dans l'équipe d'Alain Vavasseur. Mon projet de recherche s'inscrivait dans le programme de Toxicologie Nucléaire du CEA qui avait pour objectif d'étudier l'impact des métaux lourds et les réponses engendrées dans différents organismes et ainsi de proposer des solutions de remédiation pour décontaminer les sols pollués. Mon travail se focalisait sur **l'étude des effets d'une surproduction de chélatants et de transporteurs sur l'accumulation de métaux lourds chez les végétaux, dans l'objectif d'une application à la phytoremédiation**.

Le projet visait à accroître par génie génétique les capacités de stockage de métaux lourds chez les plantes supérieures (le tabac et *Arabidopsis*), dans le but d'utiliser les végétaux transformés dans des procédés de phytoremédiation de sols pollués. La stratégie retenue consistait à sur-exprimer simultanément des peptides (glutathion et phytochélatines) connus pour chélater les métaux lourds comme le cadmium, dans divers compartiments cellulaires (chloroplaste, cytosol) et des transporteurs membranaires connus pour transporter ces métaux dans la vacuole. Cette stratégie devait permettre de lever d'éventuelles barrières à l'accumulation interne de métaux liés au transfert du métal au sein de la plante.

Tout d'abord, deux des enzymes de la voie de biosynthèse des phytochélatines, la gγ-glutamyl cystéine synthase (γECS) d'*E. coli* et la phytochélatine synthase (PCS) d'*A. thaliana* ont été clonées et exprimées indépendamment chez *A. thaliana*. Ces protéines ont été adressées dans le cytosol ou dans le chloroplaste *via* des transformations nucléaires. Des plantes issues de ces transformations ont été obtenues et leur caractérisation (analyses par Western blot, Southern blot et ségrégation) a permis de mettre en évidence des lignées stables (avec une seule insertion du transgène) sur-exprimant les deux enzymes adressées dans le cytosol ou dans le chloroplaste. Un système de culture *in vitro* a été mis au point dans le but de tester les différentes lignées obtenues en présence de cadmium ou d'autres métaux (le plomb, l'arsénite,...) et d'estimer les capacités des plantes à accumuler les métaux lourds et leur résistance vis-à-vis des métaux testés. Des prélèvements sont ainsi réalisés afin de doser le glutathion et les phytochélatines par HPLC, et les métaux accumulés par ICP optique.

Les résultats obtenus lors de la croissance de plantules en présence de concentrations de cadmium croissantes (de 0 à 100 µM) montrent une augmentation des pools de glutathion et de phytochélatines chez les plantes sur-exprimant la γECS et une augmentation du pool de phytochélatines chez celles sur-exprimant la PCS. En parallèle, la longueur des racines a été mesurée à 6, 9 et 13 jours après transfert des plantules sur cadmium. Les premiers résultats indiquent que les lignées sur-exprimant la PCS et la γECS dans le chloroplaste tolèrent mieux le cadmium que les plantes sauvages. Ces résultats suggèrent que ces deux enzymes sont limitantes dans la voie de biosynthèse du glutathion et des phytochélatines.

Par contre, les données obtenues en présence d'arsénite ou d'arséniate sont plus surprenantes. En effet, les lignées sur-exprimant la γECS sont plus résistantes que les plantes sauvages, mais les lignées sur-exprimant la PCS sont plus sensibles. Cette sensibilité pourrait s'expliquer par une diminution du pool de glutathion due à une forte synthèse de phytochélatines. Pour tester cette hypothèse, des croisements entre ces deux lignées ont été réalisés et la caractérisation des descendants montre des résultats intéressants (Picault *et al*., 2006).

Parmi les transporteurs membranaires impliqués dans la résistance aux métaux lourds, on trouve les transporteurs ABC et les ATPases de type P à métaux lourds. Ces transporteurs présentent une "spécificité" biologique intrinsèque pour certains métaux essentiels mais aussi "une fuite" pour d'autres métaux non-biologiquement nécessaires. Ce manque de "sélectivité" conduit à l'entrée de métaux lourds toxiques dans l'organisme, mais peut en contrepartie être exploité dans des stratégies de phytoremédiation. Parmi ces transporteurs, nous avons sélectionné des transporteurs membranaires de type ABC impliqués dans le transport vacuolaire de complexes glutathion-Cd (YCF1 chez *S. cerevisae*, MRP1 chez l'homme) ou de complexes phytochélatines-Cd (HMT1 chez *S. pombe*).

Des fusions YCF1-GFP, MRP1-GFP, et HMT1-GFP ont été réalisées pour obtenir la localisation subcellulaire des différents transporteurs et des plantes d'*Arabidopsis* ont été transformées avec ces constructions. Des transformants ont été obtenus et caractérisés (analyses des transcrits par Northern blot, détermination du nombre d'insertion par Southern blot et ségrégation). Des plantes homozygotes sur-exprimant MRP1-GFP ont été sélectionnées et les premières expériences réalisées en présence de cadmium montrent que ces plantes possèdent un taux d'accumulation de cadmium supérieur aux plantes sauvages. Des protoplastes ont été produits à partir de plantes sur-exprimant la GFP seule et de plantes sur-exprimant MRP1-GFP. Pour le témoin, nous observons la GFP dans le cytosol et dans le noyau des protoplastes. Lorsque MRP1 est

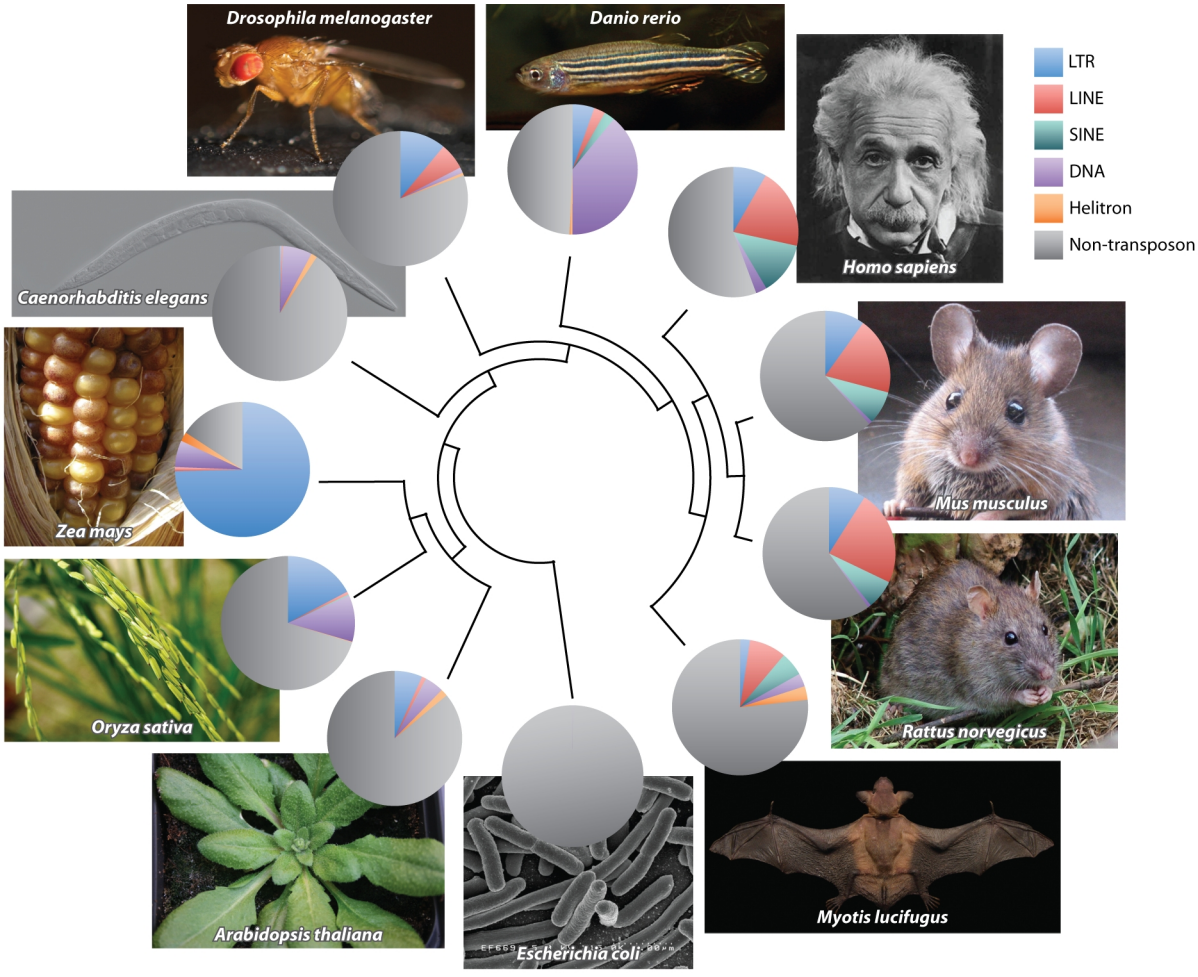
sur-exprimé en fusion avec la GFP, nous observons la fluorescence au niveau de la membrane plasmique, montrant la localisation du transporteur MRP1 chez les plantes. Des expériences similaires, réalisées avec le transporteur YCF1, ont montré qu'il est localisé au niveau de la membrane vacuolaire des protoplastes (Gayet *et al*., 2006).

Des croisements ont été réalisés entre des plantes transgéniques sur-exprimant des chélatants dans divers compartiments de la cellule végétale et des plantes sur-exprimant des transporteurs, afin d'associer les capacités accrues de chélation et de transport.

Ce travail a donné lieu aux publications suivantes :

**Picault N**., Cazale AC., Beyly A., Cuine S., Carrier P., Luu DT., Forestier C. and Pelier G. (2006) Chloroplast targeting of phytochelatin synthase in *Arabidopsis*: effects on heavy metal tolerance and accumulation. *Biochimie*. 88: 1743-1750.

Gayet L., **Picault N**., Cazale AC., Beyly A., Lucas P., Jacquet H., Suso HP., Vavasseur A., Pelier G. and Forestier C. (2006) Transport of antimony salts by *Arabidopsis thaliana* protoplasts over-expressing the human multidrug resistance-associated protein 1 (MRP1/ABCC1). *FEBS Lett*. 580: 6891-6897.




Huang CRL, et al. 2012.
Annu. Rev. Genet. 46:651–75

**Figure 1 : Composition en éléments transposables dans différentes espèces (d'après Huang et al., 2012).** Dans le sens des aiguilles d'une montre, les espèces présentées sont *Homo sapiens, Mus musculus, Rattus norvegicus, Myotis lucifigus, Escherichia coli, Arabidospis thaliana, Oryza sativa, Zea mays, Caenorhabditis elegans, Drosophilia melanogaster and Danio rerio*. L'arbre phylogénétique au centre décrit les relations évolutives entre elles. Le graphique circulaire illustre la fraction du génome composée par les différentes classes d'éléments transposables.

## ACTIVITES DE RECHERCHE ACTUELLES

Depuis mon arrivée à l'Université de Perpignan *Via Domitia*, le 1er février 2005, je me suis intégrée dans l'équipe « Analyse du Génome et Evolution » dirigée par O. Panaud (PR) au sein du laboratoire Génome et Développement des Plantes (UMR 5096). L'équipe s'intéresse principalement à l'impact des éléments transposables sur la structure et l'évolution des génomes des plantes et plus particulièrement aux rétrotransposons à LTR.

# Introduction

Les génomes ne sont pas seulement un ensemble de gènes, mais contiennent aussi des séquences mobiles qui peuvent se multiplier de manière plus ou moins autonome et influencer activement leur hôte. Ces éléments mobiles, ou éléments transposables, ont été autrefois considérés comme des parasites génomiques car leur multiplication dans les génomes est purement égoïste et hautement mutagène. L'arrivée de la génomique a mis en évidence que ces éléments composent une grande partie des génomes et que ces derniers sont des entités beaucoup plus dynamiques que ce qui était cru. Cela soulève plusieurs interrogations, mais en particulier : quelle va être l'importance des éléments mobiles sur l'évolution structurale et fonctionnelle des génomes ?

Si les éléments mobiles peuvent être bénéfiques en créant de la variabilité génétique et permettant ainsi l'adaptation de l'hôte à l'environnement ou à différents stress, les insertions de ces éléments peuvent être létales, si elles touchent des gènes impliqués dans le développement ou la reproduction. Bien qu'ils soient présents en grand nombre de copies dans la majorité des génomes, peu d'éléments sont connus pour être actifs. Différents mécanismes de l'hôte sont impliqués dans le contrôle de l'activité de ces éléments. Les éléments transposables sont donc l'un des facteurs majeurs de création de la diversité génétique et l'étude de leur dynamique est essentielle pour comprendre l'évolution des génomes.

## 1. Description des éléments transposables

Les premiers éléments mobiles ont été découverts par Barbara McClintock chez le maïs dans les années 1940. Ses interprétations ont longtemps été controversées jusque dans les années 1970, période ou d'autres éléments ont été découverts chez les bactéries (Shapiro et al., 1969), puis la drosophile (Kidwell et al., 1977). On sait désormais que les éléments transposables (ETs) sont

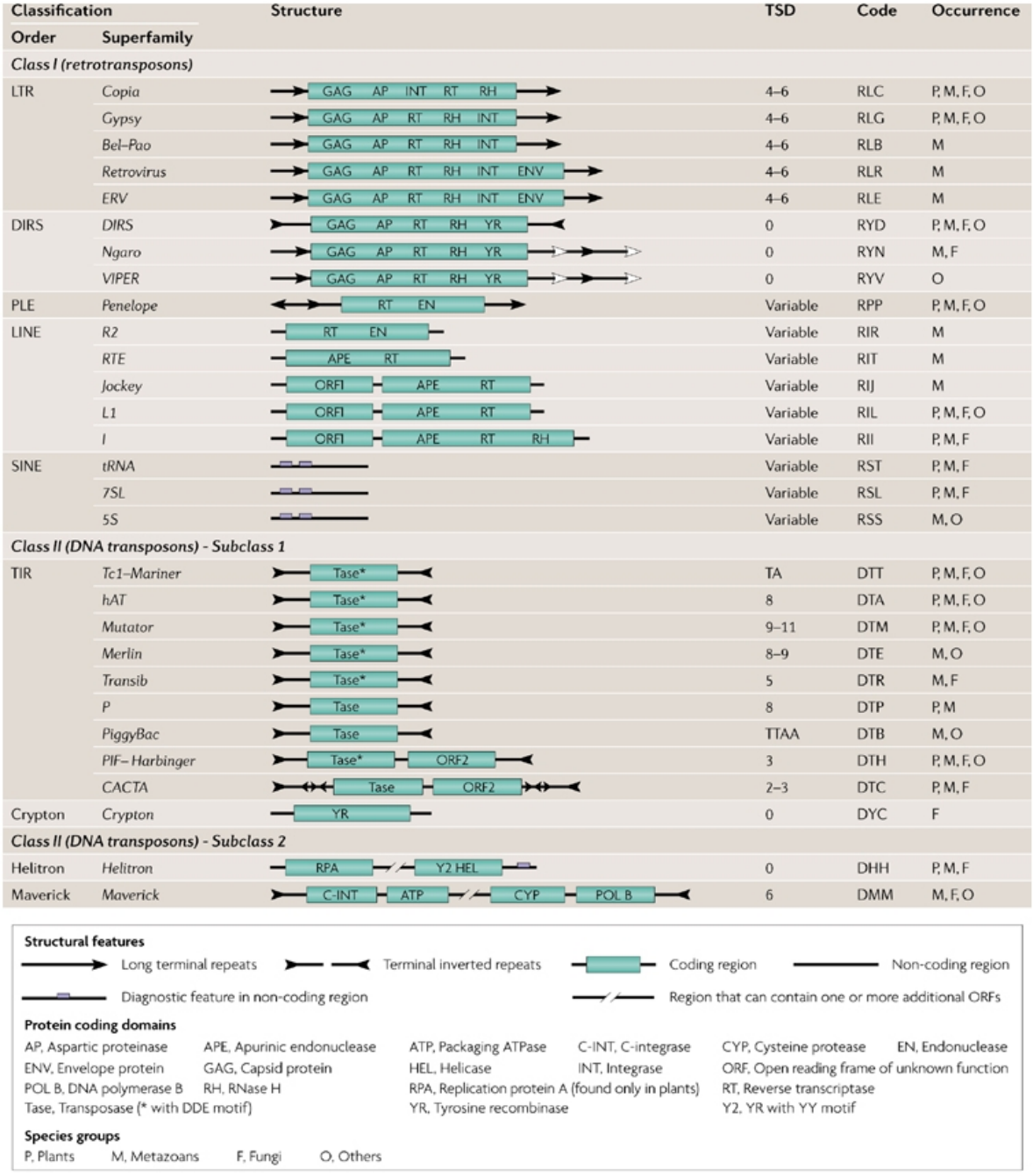



**Figure 2 : Classification complète des éléments transposables de classe I et de classe II (d'après Wicker et al. 2007).**

des composants majeurs des génomes eucaryotes (Figure 1 ; Huang et al., 2012). Ils composent 45 % du génome humain (Lander et al., 2001) et plus de 85 % d'une grande partie des génomes des plantes (Devos, 2010; Estep et al., 2013; Schnable et al., 2009). Les ETs contribuent aux variations de la taille des génomes et il existe une corrélation entre la taille des génomes et le contenu en ET (Bennetzen et al., 2005). De part leur multiplication, les ETs peuvent augmenter la taille des génomes en réponse aux stress ou aux modifications de l'environnement.

Bien qu'ils aient longtemps été considérés comme de « l'ADN poubelle », ce sont probablement des acteurs importants de l'évolution car ils génèrent de la diversité génétique. En effet, les nouvelles insertions d'ETs provoquent de nombreuses mutations et les éléments peuvent ainsi avoir un impact sur l'expression des gènes (Medstrand et al., 2001; Xiao et al., 2008; Zhang et al., 2009). La domestication des ces éléments peut même devenir indispensable au bon fonctionnement de leurs hôtes. Leur rôle dans la protection des extrémités des chromosomes chez la drosophile (Villasante et al., 2007) ou dans l'évolution du système de variabilité immunologique des vertébrés (Jones and Gellert, 2004) sont des exemples de leur impact sur la structure, la fonction et l'évolution des génomes qui les hébergent.

Les éléments transposables sont regroupés selon leur mode de déplacement, puis en fonction de leurs séquences nucléotidiques. La plupart d'entre eux appartiennent à deux grandes classes (Figure 2) : la classe I ou rétrotransposons d'une part, et éléments de classe II, ou transposons, d'autres part (Finnegan, 1989). Les éléments de la classe I transposent par un mécanisme dit de « copier-coller », où après transcription, les ARN sont « reverse transcrit » et l'ADN double brin ainsi formé est intégré à un autre locus dans le génome, dupliquant ainsi la copie initiale (Figure 3 ; Sabot and Schulman, 2006). Les éléments de classe II transposent par un mécanisme de « couper-coller ». Ici, l'élément est excisé et directement réintégré à un autre locus. Cette classification a été complétée après la découverte de nouveaux groupes d'ETs tels que les hélitrons, qui, contrairement aux autres types de classes II « classiques », ne nécessitent pas une coupure double brin du site donneur lors de la transposition (Kapitonov and Jurka, 2001). Cette nouvelle classification est présentée dans la Figure 2 (Wicker et al., 2007).

Les deux classes d'ETs comportent des éléments dits « autonomes » et « non autonomes » selon leur capacité à coder ou non leur propres protéines nécessaires à la transposition. Les éléments autonomes possèdent des ORFs (Open Reading Frame, cadre de lecture ouvert) fonctionnels, alors que les « non autonomes » ont perdu leur capacité à transposer (à cause des mutations, délétions et variations structurales dans les régions codantes) et nécessitent la machinerie transpositionnelle de leurs partenaires autonomes.

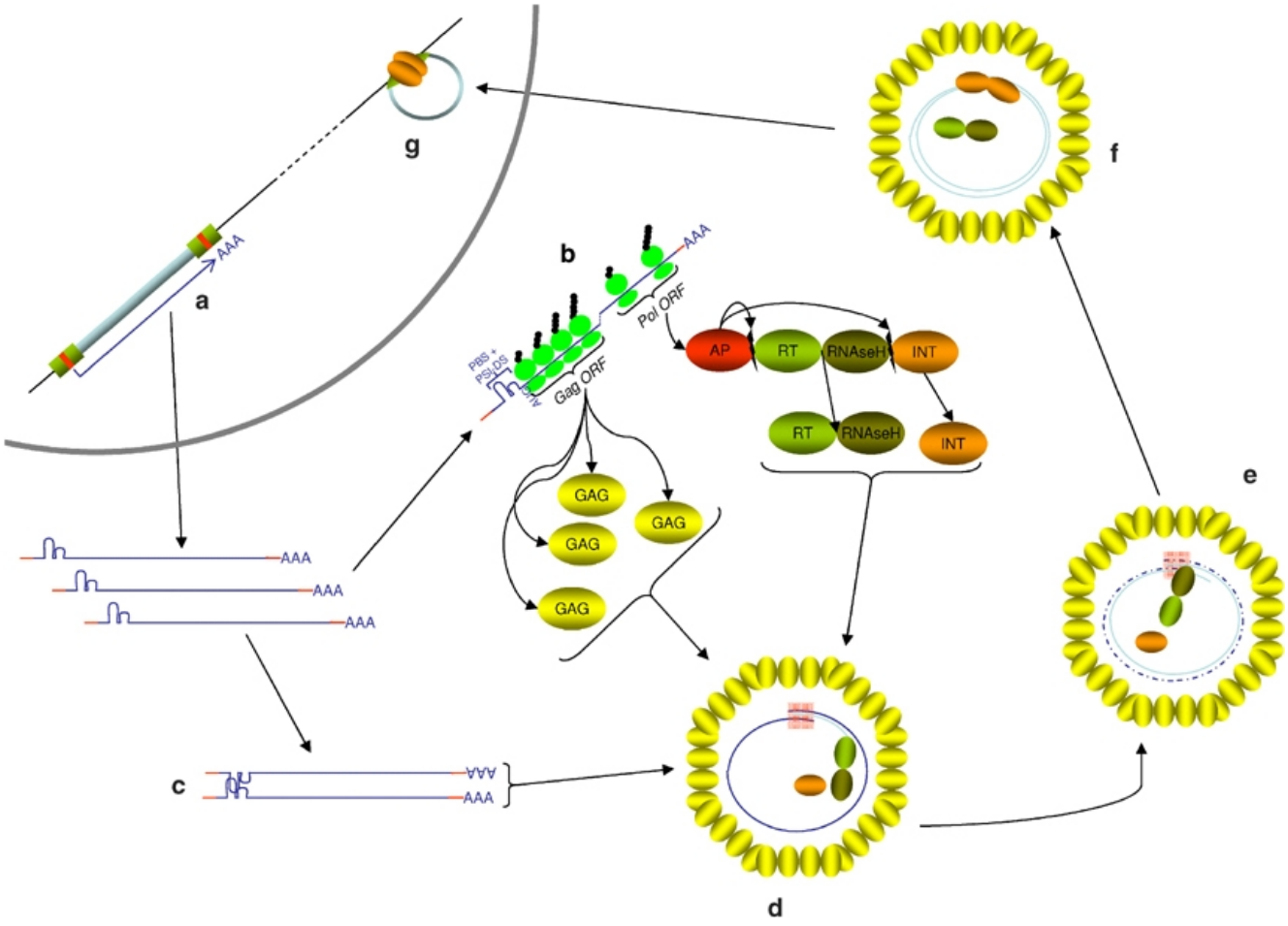



**Figure 3 : Cycle de transposition des rétrotransposons à LTR (d'après Sabot et Schulman, 2006).** Les rétrotransposons autonomes codent une polyprotéine comprenant deux domaines : GAG et POL. Les deux domaines sont transcrits en une fois par la polymérase II (étape a). Le 1er pool d'ARNm est traduit et donne la protéine GAG (protéine capside) et le complexe de protéines POL, divisé en 4 protéines distinctes au niveau post-traductionnel (RT : transcriptase inverse, INT : intégrase, RH : RnaseH, AP : protéase aspartique) (étape b). Les protéines GAG vont se polymériser dans le cytoplasme pour former des VLPs (Virus Like Particules). L'autre pool d'ARNm va entrer dans ces VLPs (étape c), se dimériser (étape d) et, sous l'action de RH et RT, être retro-transcrit en ADN double brin (étapes e et f). L'intégrase (INT) va former une cassure double brin de l'ADN génomique et y intégrer l'ADN du rétrotransposon (étape g). La réparation de cette cassure asymétrique va former les TSDs (Target Site Duplication), signatures caractéristiques laissées par l'insertion de la plupart des ETs (Sabot et Schulman, 2006).

### *1.1. Description des rétrotransposons ou éléments de classe I*

Les éléments de classe I sont divisés en 2 catégories : les rétrotransposons à LTR (Long Terminal Repeat) qui possèdent deux séquences terminales répétées et les rétrotransposons sans LTR. Les rétrotransposons à LTR sont les éléments les plus abondants chez les plantes (Hawkins et al., 2006; Kalendar et al., 2000; Vitte and Panaud, 2005; Zedek et al., 2010) alors que les rétrotransposons sans LTR, les LINEs (Long Interpersed Nuclear Elements) et les SINEs (Short Interpersed Nuclear Elements) sont très répétés chez les mammifères. Chez l'homme, le LINE *L1* (>500 000 copies) et les SINE *Alu* (> 1 million de copies) constituent à eux seuls près de 27,5% du génome humain (Cordaux and Batzer, 2009; Lander et al., 2001).

Du point de vue organisation, les rétrotransposons à LTR sont proches des rétrovirus, à l'exception du gène de l'enveloppe (Capy, 2005). En effet, ils ne possèdent pas la protéine d'enveloppe de glycoprotéines responsables de l'entrée du virus dans une nouvelle cellule (capacité d'infection). Les rétrotransposons à LTR peuvent être autonomes lorsqu'ils possèdent les gènes *GAG* et *POL*. Les protéines Gag ressemblent à la capside des rétrovirus et permettent la formation de l'enveloppe (Virus-like particle, VLP). Le gène *POL* code une polyprotéine qui, après protéolyse, se scinde en protéines requises pour la transposition, notamment une protéase, une reverse transcriptase, une RNAseH et une intégrase qui permet l'insertion dans le génome (Figure 3). Au sein des génomes, le cycle de vie des rétrotransposons à LTR est bien connu. Les LTR contiennent les signaux fonctionnels requis pour l'activation transcriptionnelle de l'élément (Kumar and Bennetzen, 1999). Les rétrotransposons à LTR sont subdivisés en deux groupes principaux suivant l'ordre des domaines protéiques du gène *POL* (Figure 4). Chez les éléments de type *copia*, l'intégrase se trouve du coté 5' de la transcriptase inverse, alors que chez les éléments *gypsy,* elle se trouve du coté 3'.

Ces éléments peuvent donc induire des augmentations importantes de la taille des génomes par des « bursts » de transposition qui peuvent être suivies de phases de diminution. Il a en effet été mis en évidence des mécanismes d'élimination (Vitte and Panaud, 2005) mais aussi des mécanismes capables d'inactiver les rétrotransposons à LTR afin de limiter leur prolifération dans les génomes. Il s'agit ici de mécanismes épigénétiques (Rigal and Mathieu, 2011). Je détaillerai ces mécanismes dans un prochain chapitre, car je vais tout d'abord présenter les éléments de la classe II ainsi que l'impact fonctionnel des ETs mis en évidence dans de nombreux organismes.

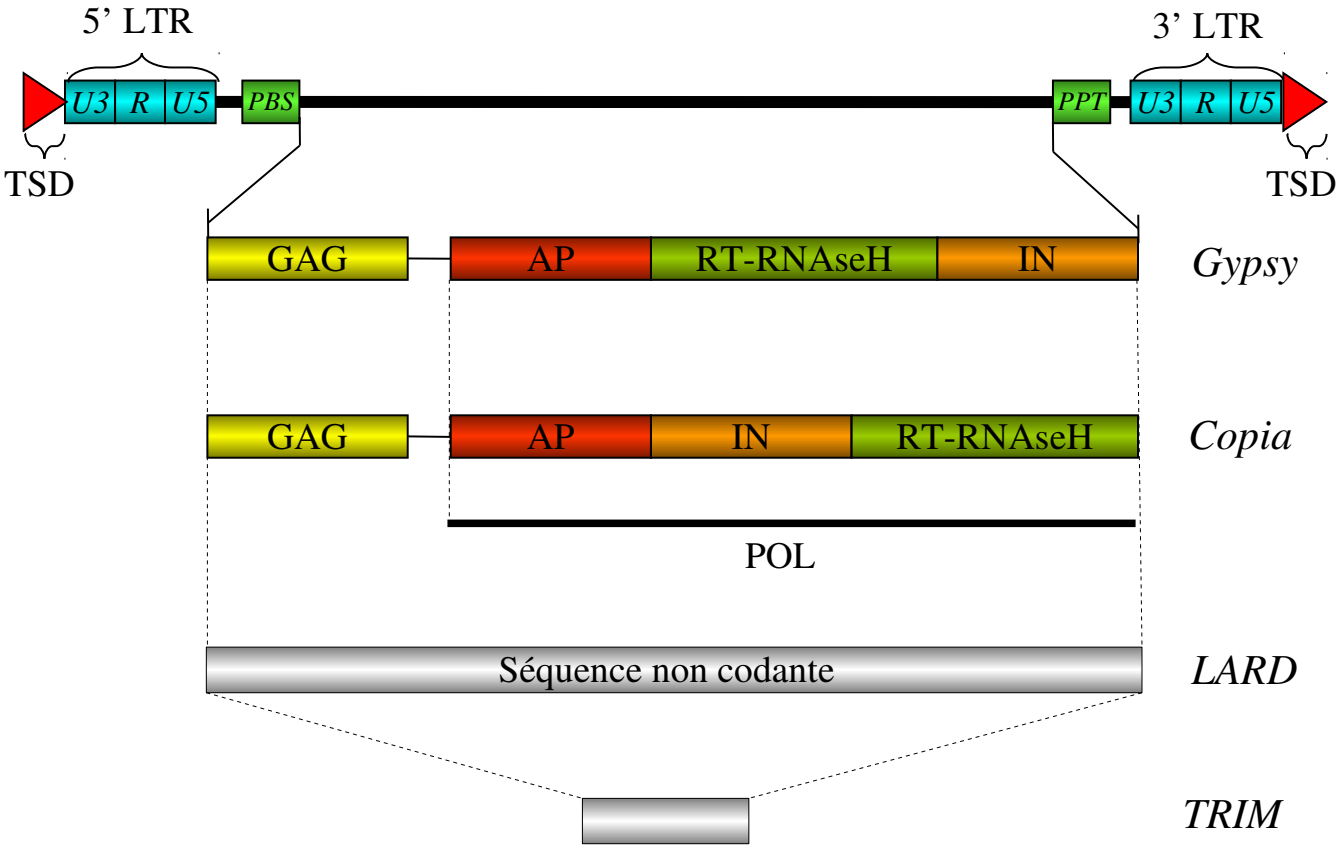



**Figure 4 : Organisation des rétrotransposons à LTR.** Les rétrotransposons sont encadrés par des longues répétitions (Long Terminal Repeats, LTR), qui contiennent le site d'initiation et de terminaison de la transcription. Les séquences correspondantes au PBS (Primer Binding site) et au PPT (PolyPurine Tract) permettent l'initiation de la reverse transcription au cours de la transposition. La région interne de l 'élément code les protéines nécessaires au cycle de transposition : GAG pour les protéines de la capside, et la POL pour les protéines suivantes : aspartic proteinase (AP), intégrase (IN), reverse transcriptase (RT) et RNaseH. Les flèches en rouge indiquent le site dupliqué au niveau de l'insertion de l'élément (TSD : Target Site Duplication) (Sabot et Schulman, 2006).

### *1. 2. Description des transposons ou éléments de classe II*

Les transposons sont également divisés en de nombreuses sous-familles (Figure 2). Ils sont généralement caractérisés par la présence de répétitions terminales inversées TIR (Inverted Terminal Repeat) aux extrémités entourant le gène de la transposase. Certaines familles, telle que celle de l'élément *P* ont été particulièrement étudiées chez la drosophile en raison des mutations génétiques qu'elles provoquent. Cet élément a rapidement envahi le génome de *D. melanogaster*. En effet, jusque dans les années 40 aucune des souches de *D. melanogaster* collectées par les généticiens ne possédaient l'élément *P*. Actuellement, toutes les populations naturelles de cette espèce en contiennent (Bingham et al., 1982). L'élément *P* a depuis été l'objet de nombreuses autres études établissant son implication dans des transferts horizontaux récurrents au sein du genre *Drosophila* (Daniels et al., 1990) et il constitue un des éléments modèles dans l'étude de tels processus.

Les MITEs (Miniature Inverted Transposable Element) constituent l'un des groupes de transposons les plus abondants chez les plantes. Ce sont des transposons non-autonomes de petite taille (< 600 pb) caractérisés par la présence de deux TIRs (10-30 pb) et par l'absence de région interne.

Les éléments transposables sont désormais connus pour avoir un impact fonctionnel chez les organismes, ils peuvent être des modulateurs de l'expression de gènes. Ils sont à ce titre très étudiés et la partie suivante présente quelques exemples caractéristiques de leur l'implication sur la variabilité génétique.

## 2. Description de l'impact fonctionnel des éléments transposables

Dans les premiers temps après leur découverte les éléments transposables ont été considérés comme n'influençant que la taille et la structure des génomes. Toutefois, dès les années 1950, Barbara McClintock a montré que les éléments *Ac/Ds* étaient capables d'entraîner des réarrangements chromosomiques, telles que des inversions, translocations, duplications et délétions chez le maïs (McClintock, 1950, 1951). Depuis, un grand nombre de travaux ont aussi mis en évidence ces phénomènes. Par exemple, chez la tomate, le rétrotransposon à LTR *Rider* est impliqué dans la duplication du locus *Sun* entraînant ainsi l'élongation du fruit (Xiao et al., 2008).

Les éléments transposables peuvent aussi présenter un effet *cis*-régulateurs, en contrôlant finement l'expression des gènes. En effet, les rétrotransposons à LTR peuvent avoir un impact sur l'expression de gènes adjacents grâce au promoteur fort situé dans le LTR. Par exemple, l'insertion

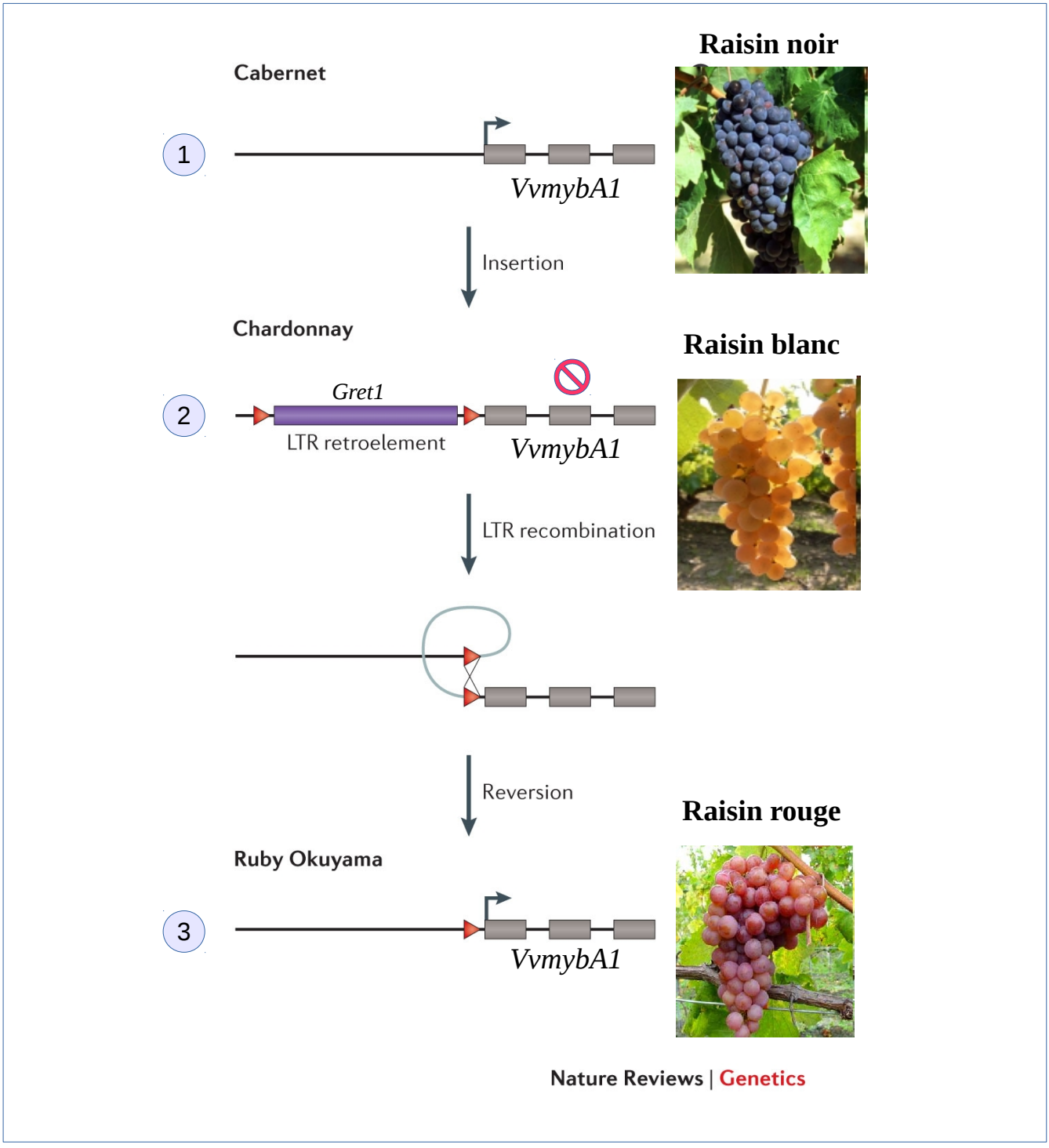



**Figure 5 : Impact du rétrotransposons à LTR *Gret1* sur la coloration de la peau des grains de raisins (d'après Kobayashi et al., 2004 ; Lisch, 2013).** (1) Le gène *VvmybA1* responsable de la pigmentation s'exprime normalement, la peau du grain est noire. (2) Le rétrotransposon *Gret1* est inséré en amont du gène *VvmybA1* qui ne s'exprime plus entraînant une décoloration de la peau du grain. (3) La formation d'un solo LTR à partir de l'élément *Gret1* fait que *VvmybA1* présente un défaut d'expression induisant la coloration rouge de la peau des grains.

du rétrotransposon à LTR *Gret1* en amont du gène *VvmybA1* codant un facteur de transcription régulant la biosynthèse des anthocyanes (pigments) est responsable de la différence de coloration des grains entre trois cultivars de raisins : noir, blanc et rouge (Kobayashi et al., 2004) ; Figure 5).

Les éléments transposables peuvent aussi induire des variations épigénétique, ou épimutations. La grande majorité des ETs dans les génomes des plantes sont sous un contrôle épigénétique qui les inactive efficacement. Il est intéressant de noter plusieurs études qui ont mis en évidence l'impact des ETs sur l'expression des gènes proches à travers la méthylation des cytosines de l'ADN. En effet, l'expression d'un gène peut être affecté par la présence d'un ET dans son voisinage, parce que ce dernier induit des changements de l'état épigénétique de la chromatine dans la région. Dans de tel cas, les ETs agissent comme des médiateurs épigénétiques qui entraînent des changements de l'expression du gène. L'exemple le plus connu est celui concernant la régulation de l'expression du gène *Agouti* par l'élément *IAP* chez la souris. L'insertion de l'élément en amont du gène conduit à une activation ectopique car la transcription est initiée au niveau du LTR de l'élément et se prolonge dans le gène, ce qui augmente le taux de transcrits. En fonction du statut épigénétique de l'élément, le gène *Agouti* est plus ou moins exprimé. Ainsi l'hypométhylation de *IAP* entraîne une augmentation de l'expression du gène et de la coloration jaune du pelage. Au contraire, l'hyperméthylation de l'élément permet de retrouver une expression plus faible du gène et donc d'une coloration plus foncée du pelage (Blewitt et al., 2006; Gaudet et al., 2004; Slotkin and Martienssen, 2007). Il a également été mis en évidence que l'insertion d'un transposon de la famille *hAT* en amont d'un gène *G* (responsable du déterminisme du sexe chez le melon) provoque l'extinction de son expression, expliquant la formation des organes femelles (Martin et al., 2009).

Les éléments transposables peuvent également être domestiqués, c'est-à-dire que leurs gènes peuvent être co-optés par le génome, utilisés comme protéines fonctionnelles et être impliquées dans le développement et la croissance de l'organisme (Feschotte, 2008). Dans la majorité des cas, c'est le gène de la transposase qui est impliqué dans la domestication (Casola et al., 2007; Muehlbauer et al., 2006). Par exemple, chez *Arabidopsis thaliana*, la transposase d'un élément *hAT-like* est indispensable au développement normal de la plante. Les plantes ne possédant pas cet élément présentent un développement anormal des fleurs, une activité photosynthétique réduite et une forte stérilité (Bundock and Hooykaas, 2005).

De nombreux exemples ont été décrits dans la littérature concernant l'impact des éléments transposables sur l'expression des gènes. La Figure 6 expose les principaux mécanismes par lesquels les ETs peuvent influencer la transcription des gènes. Tous ces exemples montrent que les éléments transposables peuvent avoir un impact sur la régulation de l'expression des gènes. Il y a

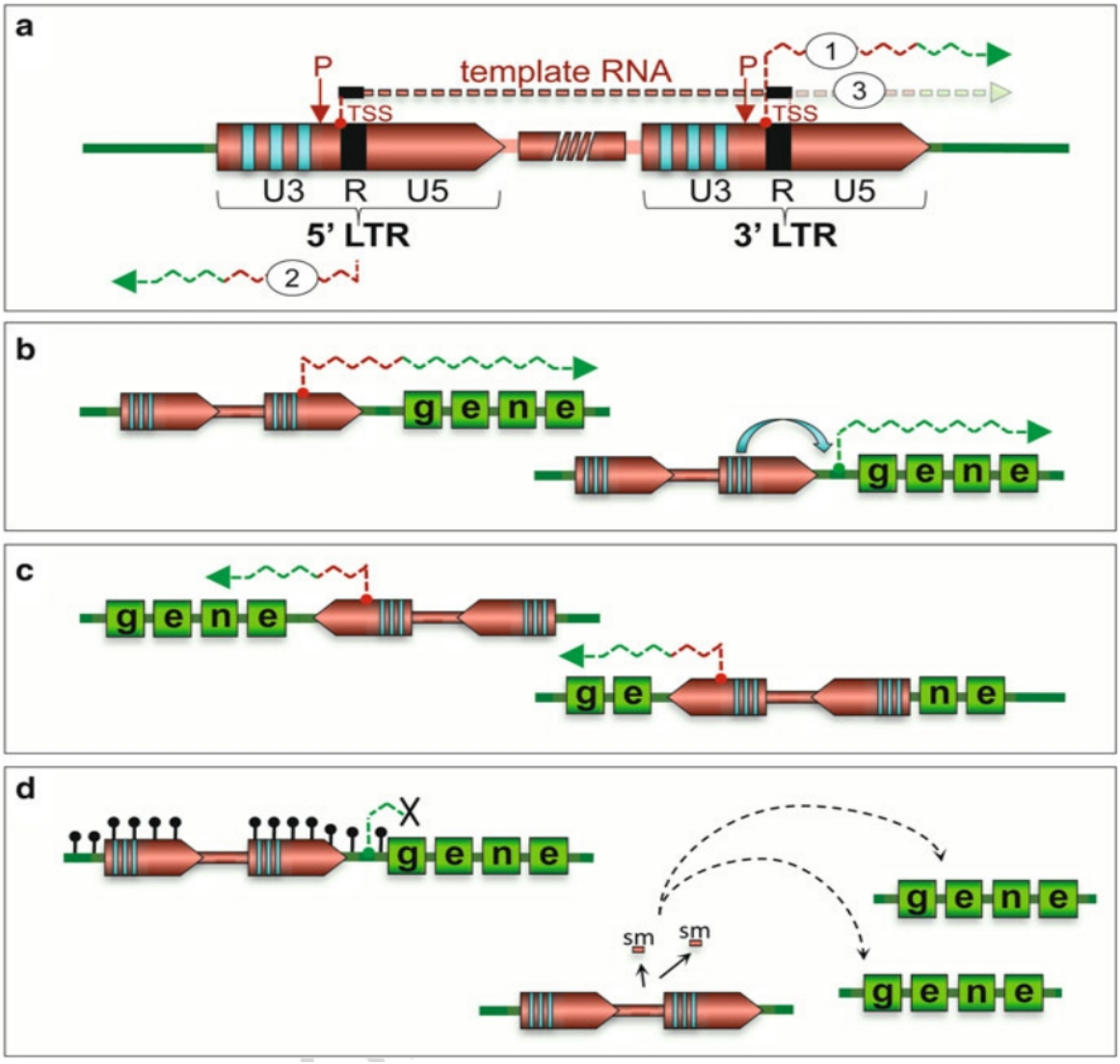



**Figure 6 : Exemples de régulations de l'expression des gènes influencées par la présence de rétrotransposon à LTR (d'après Bui et Grandbastien, 2012 ).** (a) Structure et caractéristiques transcriptionnelles des rétrotransposons à LTR : l'élément est encadré par deux « Long Terminal Repeats » (LTR) qui sont identiques dans les copies qui ont récemment transposé. Les LTR possèdent un promoteur (P), site d'initiation de la transcription (TSS) et des séquences régulatrices (boîtes bleues), et le modèle de l'ARN utilisé pour l'amplification est initié à la limite U3/R dans le LTR 5' et se termine à la limite de R/U5 dans le LTR 3'. (1) Le LTR 3' peut aussi contenir un promoteur, TSS et des séquences régulatrices et peut donc entraîner la co-transcription d'un gène adjacent situé en aval de l'élément. (2) Les LTR peuvent aussi posséder des promoteurs antisens conduisant la co-transcription de gènes adjacents situés en amont du rétrotransposon à LTR. (3) Parfois les ARNs ne s'arrêtent pas au site de terminaison dans le LTR 3' et peuvent ainsi se prolonger dans les séquences en aval de l'élément. (b) Quand l'élément est inséré en amont d'un gène, le LTR 3' peut agir comme un promoteur en initiant la transcription ou en fournissant des séquence cis-régulatrices, comme des sites de fixation pour les facteurs de transcription. (c) Quand l'élément est inséré en orientation antisens du gène (ou en utilisant les promoteurs antisens), les LTR peuvent initier des transcrit antisens qui diminuant l'expression du gène cible. (d) Les rétrotransposons à LTR peuvent aussi entraîner des régulations épigénétiques, comme la méthylation de l'ADN ou des histones des gènes adjacents ou être une source de petits ARNs. Ce sont quelques exemples parmi les multiples possibilités de l'impact fonctionnel des rétrotransposons à LTR qui varie en fonction de leur orientation par rapport aux gènes adjacents et de leur position dans la séquence génique.

donc une nécessité pour l'hôte de contrôler la réactivation de la transposition, dont les effets tant structurels que régulateurs peuvent s'avérer létaux. L'une des possibilités pour ce faire est l'inactivation par modifications épigénétiques.

## 3. Inactivation des éléments transposables par modifications épigénétiques

Je vais maintenant décrire les mécanismes des différentes voies impliquées dans la régulation épigénétique des éléments transposables. Même s'ils sont encore loin d'être entièrement caractérisés, les connaissances dans ce domaine avancent très vite et l'ensemble de la régulation des ETs devient de plus en plus clair (Figure 7).

La répression des ETs se fait à la fois au niveau transcriptionnel (TGS, Transcriptional Gene Silencing) et post-transcriptionnel (PTGS, Post-Transcriptional Gene Silencing). Le TGS intervient au niveau de la répression de la transcription *via* la méthylation de l'ADN et des histones (Vaucheret and Fagard, 2001). Cette méthylation peut être transmise lors des divisions cellulaires (mitoses et méioses) grâce à l'intervention de plusieurs ADN méthyltranférases et d'histones méthyltransférases. Le PTGS est généralement lié à la dégradation de l'ARN messager dans le cytoplasme et cible les transcrits d'ETs qui ont pu échapper au TGS (Voinnet, 2001).

Il est bien établi que le 'silencing' des ETs au niveau transcriptionnel (TGS) passe par la méthylation de l'ADN et des histones (Lippman et al., 2003). La première évidence de l'implication de la méthylation de l'ADN dans le contrôle de l'activité transpositionnelle provient de l'étude chez le maïs des éléments de classe II *Activator* (*Ac*), *Suppressor-mutator* (*Spm*) et *mutator* (*Mu*) (Banks et al., 1988; Chomet et al., 1987). Dans son état actif, le promoteur du gène de la transposase de ces éléments est hypométhylé. Généralement, le relâchement de la méthylation globale de l'ADN est associé à une réactivation transcriptionnelle et quelques fois transpositionnelle des ETs. Par exemple, chez les mutants d'hypométhylation *ddm1* d'*Arabidopisis thaliana,* affectés dans la protéine de remodelage des nucléosomes appelée DDM1 (Decreased DNA Methylation 1), on observe une réactivation transpositionnelle de plusieurs ETs appartenant à différentes familles (Miura et al., 2001; Singer et al., 2001). La méthylation de l'ADN au niveau des cytosines constitue donc une sorte de verrou et sa levée provoque la réactivation des ETs.

Même si la majorité des cytosines sont méthylées chez les plantes, seules les régions répétées (dont les ETs) connaissent des taux élevés de méthylation, comparées aux régions géniques (Roudier et al., 2009). Cette différence de méthylation entre les gènes et les ETs suscite de nouvelles questions. Par exemple, on ne sait toujours pas comment les génomes hôtes arrivent à distinguer un gène d'un ET (Lisch and Slotkin, 2011). En effet, les différentes classes et familles

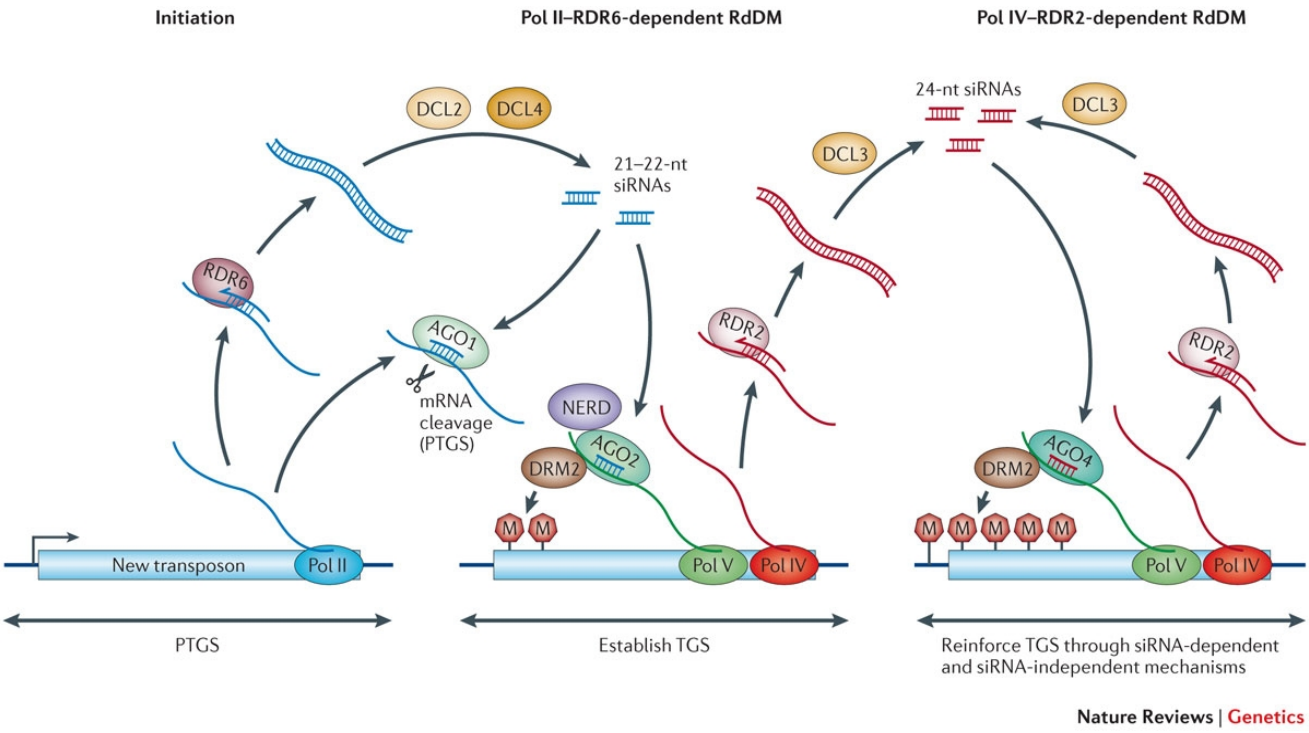



**Figure 7 : La voie RdDM (d'après Matzke and Mosher, 2014).** La voie présentée montre un moyen pour établir le « RNA-directed DNA methylation » (RdDM) et éventuellement assurer le TGS pour un nouvel ET qui était originairement ciblé par le PTGS. Dans le PTGS (à gauche), une nouvelle insertion d'un ET est activée et transcrite par la polymérase II (Pol II). Certains transcrits sont copiés par la RNA-Dependent RNA Polymerase 6 (RDR6) pour produire des ARN double-brins (dsRNAs), qui sont traités par DICER-LIKE 2 (DCL2) et DCL4 en nucléotides de 21–22 appellés small interfering RNAs (siRNAs). Ces siRNAs sont chargés par ARGONAUTE 1 (AGO1) et guident le clivage du transcrit correspondant à l'ET, dans une voie typique du PTGS. Dans une voie dérivée du RdDM (au milieu), des siRNAs de 21-22 nucléotides peuvent aussi cibler à de faibles niveaux la méthylation de l'ADN dans une voie dépendante de DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), Pol V et AGO2, qui interagissent avec NEEDED FOR RDR2-INDEPENDENT DNA METHYLATION (NERD) grâce au motif AGO hook. L'ADN partiellement méthylé recrute la Pol IV, qui initie la transcription d'un ARN simple brin (ssRNA). Le ssRNA est copié par RDR2 en dsRNA qui est clivé par DCL3 en siRNAs de 24 nucléotides. Suite à l'incorporation dans AGO4 (à droite), les siRNAs de 24 nucléotides vont s'hybrider avec les transcrits de Pol V qui entraînent le recrutement de DRM2 et une méthylation dense de l'ET. Les siRNAs sont produits de manière continue à partir du modèle produit par la voie Pol IV, qui renforce le TGS et qui peut être maintenu dans une voie dépendante des siRNA par METHYLTRANSFERASE 1 (MET1), CHROMOMETHYLASE 3 (CMT3) and DECREASED DNA METHYLATION 1 (DDM1) (non montré dans ce schéma).

d'ETs ne montrent aucune homologie de séquence entre elles et présentent une diversité de structure considérable.

La méthylation de l'ADN est souvent associée à la méthylation des histones. Cette dernière se fait sur la lysine (k) des histones H3 (aux positions 9 (di-méthylation) et 27 (mono-méthylation)) par des histones méthyltransférases. De la même manière que les mutants de méthylation de l'ADN, les mutants défectueux pour la méthylation des histones H3 présentent une réactivation des ETs (Ebbs and Bender, 2006; Jacob et al., 2009). Ces deux types de méthylation jouent donc un rôle important dans le maintien de l'état condensé de la chromatine, inactivant l'activité transcriptionnelle et donc transpositionnelle des ETs.

Il faut distinguer la méthylation *de novo* concernant les séquences d'ADN vierges de toute marque épigénétique (c'est le cas par exemple d'une nouvelle insertion d'un ET), et le maintien de la méthylation pré-existante lors de la réplication de l'ADN. Il s'avère que les mécanismes d'inactivation transcriptionnelle et post-transcriptionnelle des éléments transposables interagissent étroitement dans les processus de maintenance et d'inactivation de *novo* des ETs (Chan et al., 2004; Teixeira and Colot, 2009).

La méthylation *de novo* par la voie *RdDM* (Haag and Pikaard, 2011; Matzke and Mosher, 2014) est un processus complexe qui fait intervenir plusieurs protéines qui assurent à la fois la production des siRNA (small interfering RNAs) et leur guidage sur les séquences d'ADN. Le principe de base repose sur la production d'un ARN double brin qui est découpé par des RNAses nommées DICER en siRNA. Un travail conduit par Slotkin et al., en 2009 a permis de montrer l'impact des siRNA sur l'inactivation des éléments transposables dans les lignées germinales. En effet, chez *Arabidopsis thaliana*, la transmission de siRNA du noyau végétatif vers les cellules spermatiques (intervenant dans la formation du zygote) permettrait d'empêcher l'activation des éléments transposables au niveau de ces cellules germinales, limitant ainsi la transmission de copies supplémentaires à la génération suivante (Slotkin et al., 2009). Il apparaît que les mécanismes épigénétique jouent un rôle capital dans le contrôle du nombre de copies par inactivation des éléments au niveau transcriptionnel et post-transcriptionnel (Lisch, 2013; Slotkin and Martienssen, 2007).

## 4. Description de l'impact du stress sur les éléments transposables

Malgré le fait que les ETs soient soumis à une répression épigénétique très efficace, le postulat initial de la régulation des éléments transposables en réponse au stress (McClintock, 1984) a été confirmé dans plusieurs études (Grandbastien, 1998 ; Bui and Grandbastien, 2012). Certains

| Eléments | Stress | Espèces | Références |
|---|---|---|---|
| *Corky* | blessure | chêne | Rocheta et al., 2012 |
| *CLCoy1* | blessure | citron | De Felice et al., 2009 |
| *OARE1* | blessure | avoine | Kimura et al., 2001 |
| *Tnt1* | blessure | tabac | Mhiri et al., 1997 |
| *AtCopeg1* | salinité | *Arabidospis* | Duan et al., 2008 |
| *BARE1* | sécheresse | orge | Kalendar et al., 2000 |
| *CCR* | sécheresse | riz | Neumann et al., 2007 |
| *ONSEN* | chaleur | *Arabidopsis* | Ito et al., 2011 |
| *mPing* | froid | riz | Jiang et al., 2003 |
| *MICRE* | froid | *Medicago sativa* | Ivashuta et al., 2002 |

**Tableau 1 : Quelques exemples d'ETs réactivés en condition de stress chez les plantes.** Pour une liste plus complète voir la revue de Grandbastien, 2015.

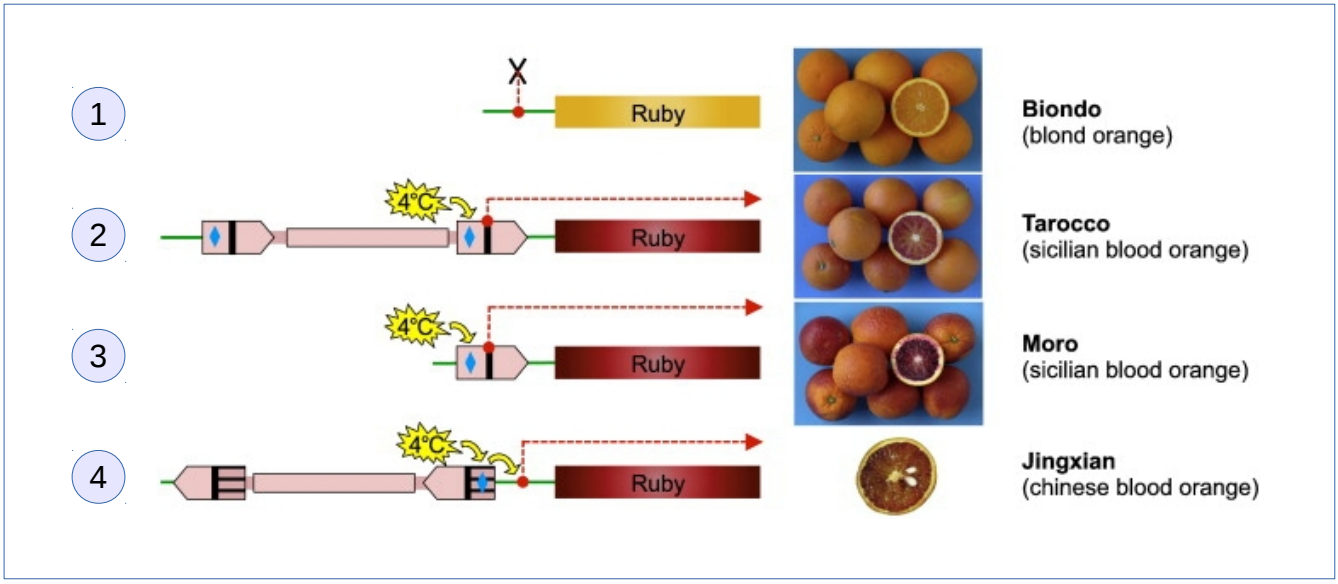



**Figure 8 : Phénotype de l'orange sanguine influencé par la présence d'un rétrotransposon à LTR (d'après Butelli et al., 2012 ; Grandbastien, 2015).** Le gène *Ruby* est inactif dans les oranges « Navel » (1) et il est réactivé spécifiquement dans le fruit et de manière dépendante au froid dans les oranges sanguines du à la présence d'un rétrotransposon à LTR inséré dans le promoteur de ce gène. Dans les oranges sanguines siciliennes, le 3' LTR de l'élément Tcs1 (2) ou le solo LTR dans d'autres accessions (3) fournit le début de la transcription du gène *Ruby* et des séquence régulatrices. Dans une accession d'orange chinoise (4), une insertion d'un élément proche de Tcs2, inséré en orientation opposée, fournit des séquences régulatrices.

stress sont donc capables de réactiver les ETs qui retrouvent leur capacité transpositionnelle et génèrent de nouvelles insertions dans les génomes des plantes stressées.

Il a été mis en évidence que l'activation des ETs peut être induite par des changements environnementaux. En effet, l'expression des ETs peut être activée par une grande diversité de facteurs provenant de stress biotiques ou abiotiques, comme l'attaque par des pathogènes ou des herbivores, la culture cellulaire, la blessure, des substances antifongiques, des stress hydriques, des métaux lourds, les UV ou encore des stress thermiques chauds ou froids (Ebina and Levin, 2007; Grandbastien, 2015; Lesage and Todeschini, 2005). Voici quelques exemples (non exhaustifs) d'éléments induits par différents stress dans le Tableau 1. Certaines de ces études ont aussi démontré que l'activation est du à la présence de sites de fixation de facteurs de transcription dans les régions promotrices des éléments (Mhiri et al., 1997). Ito et al., (2011) ont montré que l'activation transcriptionnelle liée au stress thermique entraîne l'activité transpositionnelle de l'élément *ONSEN* chez *Arabidospsis*. Cependant, pour que *ONSEN* transpose il faut que la plante (en plus du choc thermique) soit déficiente dans la voie RdDM. De plus, la réactivation de ce rétrotransposon n'est observée qu'à la génération suivante, et non directement dans la plante stressée. Ceci suggère que les siRNA transmis par la plante mère sont essentiels pour le contrôle de la transposition (Bucher et al., 2012).

Un exemple, particulièrement intéressant, où un ET est à l'origine d'un nouveau caractère sélectionné en agronomie, est celui de l'orange sanguine, où l'insertion d'un rétrotransposon à LTR en amont du gène d'un facteur de transcription *Myb* est corrélé à une nouvelle coloration du fruit, mais seulement en réponse au froid (Figure 8) (Butelli et al., 2012).

Les promoteurs des rétrotransposons à LTR semblent jouer un rôle essentiel dans la réponse aux différents stress. Plusieurs exemples viennent étayer cette hypothèse, comme l'étude du LTR de l'élément *Tto1* qui a démontré l'existence d'un motif de 13 pb conférant à ce dernier sa capacité à réagir à l'acide jasmonique (Sugimoto et al., 2000; Takeda et al., 1999). Egalement, le promoteur de *Tnt1* possède des séquences régulatrices qui présentent des homologies avec les régions régulatrices de certains gènes de défense contre les pathogènes. Ceci pourrait expliquer l'induction de sa transcription en présence de ce type de stress (Beguiristain et al., 2001; Grandbastien et al., 2005).

Chez les plantes, les interactions entre les rétrotransposons et les différentes étapes des stress biotiques et abiotiques sont encore peu connues. Certaines pistes commencent à être défrichées. Récemment, Anca et al., (2014) ont étudié la réponses de 4 rétrotransposons de tabac à la cryptogéine, un éliciteur mimant un stress biotique. Ils ont montrés que les réponses sont spécifiques à chaque élément et qu'elles sont complexes et seraient donc contrôlées par différentes

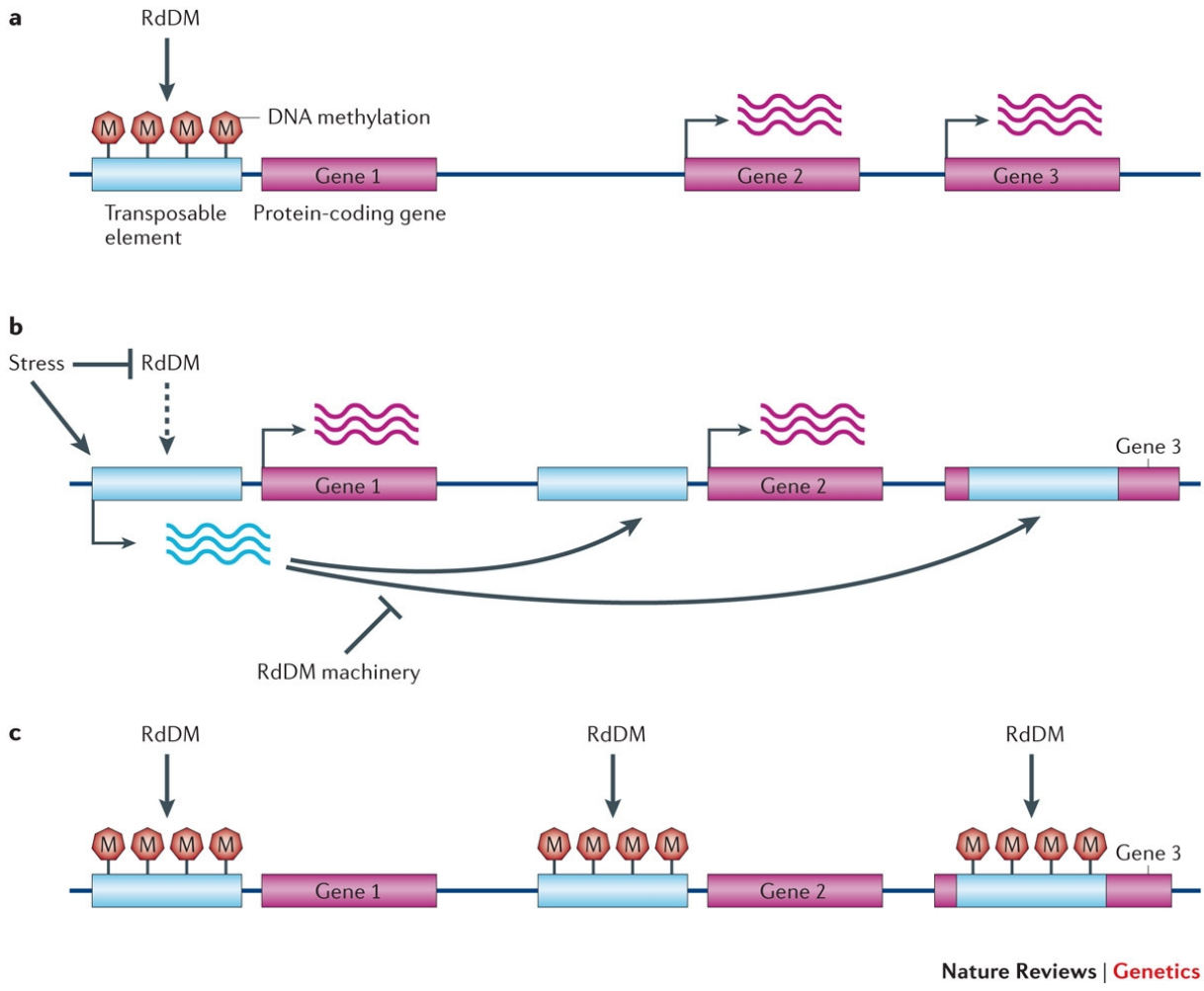



**Figure 9 : Représentation schématique du relâchement du RdDM au cours d'un stress (d'après Matzke and Mosher, 2014).** (a) Un locus est schématisé représentant un ET inactivé par méthylation entraînant l'inactivation du gène adjacent (Gène 1). Au contraire, les gènes 2 et 3 sont transcrits et ne sont pas sous le contrôle de l'ET. (b) L'ET est activé par un stress biotique ou abiotique par une combinaison de perte de RdDM et une réponse transcriptionnelle au stress. Par conséquence, le gène 1 est aussi activé par perte de méthylation dans sa région promotrice. L'ET ré-activé peut s'insérer dans le génome à de nouveaux loci, bien que la machinerie RdDM inhibe l'insertion de certains éléments par un mécanisme encore inconnu (Ito et al., 2011). Les nouveaux ETs insérés peuvent établir une transcription en réponse au stress pour de nouveaux gènes (Gène 2) ou ils peuvent entrainer la perte de fonction permanente (Gène 3).

voies encore inconnues.

Pour résumer, les stress environnementaux peuvent causer l'induction de la transcription et parfois la transposition de certaines familles d'ETs (voir revue Grandbastien, 2015). Cette induction de l'activité des ETs en réponse aux stress pourrait donc jouer un rôle important dans l'adaptation des plantes à leur environnement. Dans les populations naturelles, plusieurs stress peuvent fortement perturber les mécanismes de silencing et donc induire la réactivation des ETs. Ceci va leur permettre d'envahir assez rapidement les génomes en multipliant leur nombre de copies. Si certains stress sont capables de réactiver certaines familles d'ETs, les causes exactes de la sensibilité des ETs vis-à-vis du stress ne sont pas très claires. Toutefois, plusieurs facteurs peuvent être impliqués, comme les promoteurs de ces éléments activés par des facteurs de transcription des gènes impliqués dans la réponse à certains types de stress, mais aussi les stress qui vont perturber le paysage épigénétique de la cellule et causer la levée du TGS et/ou du PTGS (Figure 9).

On ignore la fréquence et l'importance de ces événements *in planta*, car toutes ces études ont été faites dans des conditions bien précises et contrôlées en laboratoire. Par ailleurs, faut-il encore que les nouvelles insertions générées en condition de stress soient transmises aux générations futures, car d'une part, les réponses des plantes aux attaques des pathogènes sont souvent très localisées (au niveau des feuilles par exemple) et d'autre part, le contrôle des ETs dans les cellules germinales est très renforcé (Slotkin et al., 2009). Un exemple de l'impact des changements environnementaux sur l'activité transpositionnelle *in planta* a été apporté par le rétrotransposon *BARE-1* chez l'orge. Cet élément montre un polymorphisme d'insertion dans les populations naturelles vivants dans différents micro-climat dans une vallée d'Israël (Kalendar et al., 2000). Les exemples sont assez rares car jusqu'à maintenant seul un autre cas a été identifié dans des hybrides de tournesol ayant évolué dans des conditions extrêmes (le désert) (Ungerer et al., 2006).

Ces exemples illustrent bien que les éléments transposables constituent des régulateurs de l'expression des gènes et que les modifications qu'ils induisent sont à l'origine de variabilités génétiques permettant une adaptation des organismes en fonction des facteurs environnementaux. La compréhension des mécanismes impliqués dans ces processus est donc indispensable.

D'autres études sont donc nécessaires pour établir le rôle exact des stress biotiques ou abiotiques dans l'induction de l'activité des ETs, ainsi que son impact potentiel direct ou indirect sur le paysage épigénétique. Pour essayer d'apporter des éléments de réponses concernant le rôle des différents stress sur l'activité des rétrotransposons à LTR, j'ai concentré mon travail sur une espèce modèle dans le domaine végétal, le riz.

## 5. Description du riz en génomique et post-génomique

Le riz asiatique, *Oryza sativa* L., est la base alimentaire de plus des deux tiers de la population mondiale et est la 2ème céréale la plus cultivée au monde. En plus de son intérêt agronomique, le génome du riz cultivé est relativement petit : 389 Mb, contre 4700 Mb pour le blé et 2160 Mb pour le maïs. Il constitue donc un bon modèle en génomique. Le génome du riz (*Oryza sativa* L.*, japonica* cv Nipponbare) a été entièrement séquencé (International Rice Genome Sequencing Project (IRGSP), 2005) et totalement annoté (Rice Annotation Project et al., 2008). Ainsi, 33% de la fraction génomique du riz (ce qui représente 130 Mb) correspondent aux ETs, toutes classes confondues. Les éléments majoritaires sont les rétrotransposons à LTR qui représentent 23 % du génome, soit plus de 90 Mb. Des travaux récents menés dans l'équipe ont mis en évidence dans le génome plus de 400 familles de rétrotransposons à LTR possédant de une à plusieurs centaines de copies par famille. Nous disposons d'une base de données exhaustive des séquences génomiques de rétrotransposons à LTR (El Baidouri and Panaud, 2013).

Malgré leur propension à envahir les génomes, très peu d'éléments transpositionnellement actifs ont été identifiés chez les plantes. A mon arrivée au laboratoire, seuls les éléments suivants étaient connus comme actif chez le riz : le rétrotransposon *Tos17* (Hirochika et al., 1996), le LINE *Karma* (Komatsu et al., 2003) et les transposons de classe II *dTok* (Moon et al., 2006), *nDart* (Tsugane et al., 2006) et *mPing/Pong* (Jiang et al., 2003). Toutes ces familles réunies ne représentent pas plus de 100 kpb dans le génome de Nipponbare, soit seulement 0,1 % des ETs connus chez le riz.

Une des grandes difficultés pour identifier les éléments actifs est l'utilisation d'un crible approprié pour la détection de nouvelles insertions. Les transposons *dTok* et *nDart* chez le riz ont été découverts car leur insertion dans les gènes entraînait un changement de phénotype (Moon et al., 2006; Tsugane et al., 2006). Leur identification a été possible grâce au clonage positionnel de ces gènes. Dans le cas des éléments *mPing/Pong*, les auteurs ont combinés une étude *in silico* des insertions récentes de MITES dans le génome du riz avec une confirmation de la mobilité de ces éléments dans des cultures cellulaires en utilisant la technique du transposon display (Jiang et al., 2003). Cette approche n'est toutefois pas toujours efficace. Par exemple, elle ne le sera pas dans le cas où les éléments insérés récemment ne sont plus actifs à ce jour. Les éléments *Tos17* et *Karma* ont été mis en évidence par clonage et séquençage des ADNc en utilisant des amorces choisies au niveau du gène de la reverse transcriptase (Hirochika et al., 1996; Komatsu et al., 2003). Cette méthode est robuste mais non applicable pour une étude exhaustive de tous les rétrotransposons à LTR.

Le fil conducteur de mes activités de recherche est d'étudier l'impact structural et fonctionnel des éléments transposables sur la plasticité des génomes au cours de variations environnementales. Il y a un réel intérêt à mieux comprendre les mécanismes moléculaires régulant les éléments transposables à la vue de leurs rôles important dans l'évolution et la dynamique des génomes.

De ce fait, je me suis jusqu'à présent plus particulièrement intéressée à l'activation transcriptionnelle et transpositionnelle des rétrotransposons à LTR par des facteurs de stress chez le riz. Pour cela, j'ai étudié ces éléments à l'échelle du génome pour avoir une vue d'ensemble la plus complète possible.

# Résultats

## 1. Impact des rétrotransposons à LTR sur la structure des génomes

A mon arrivée au LGDP, j'ai intégré l'équipe d'Olivier Panaud composée à cette époque exclusivement de bio-informaticiens. De part mon expérience et afin de valider des résultats *in silico*, j'ai tout naturellement participé à l'étude de l'impact des éléments transposables sur la structure des génomes.

Le genre *Oryza* est composé d'espèces diploïdes et tétraploïdes. Au sein des espèces diploïdes, nous pouvons observer de grandes variations de taille du génome, allant de 380 MB pour le riz cultivé *O. sativa* à 960 Mb pour l'espèce sauvage *O. australiensis*. Nous nous sommes demandés si ces variations de taille de génome chez *O. australiensis* étaient dues à certains éléments transposables qui auraient connu des événements massifs de transposition (« bursts »). L'objectif a été d'identifier les séquences à l'origine de cette augmentation de taille du génome et de caractériser leur dynamique au sein du génome d'*O. australiensis*.

A partir de séquences d'extrémités de BAC disponibles pour *O. australiensis*, l'analyse *in silico* a permis d'identifier trois rétrotransposons à LTR, *RIRE1*, *Wallabi* et *Kangourou.* En combinant l'utilisation des séquences d'extrémités de BAC à une analyse par « dot-blot », le nombre de copies complètes et de solo LTR de chaque rétrotransposon a été estimé. Enfin, une approche de paléogénomique, consistant à transformer les divergences observées entre copies paralogues en dates, a permis d'estimer l'âge des bursts de transposition des trois familles.

Nous avons ainsi pu caractériser l'activation transpositionnelle de trois familles de rétrotransposons à LTR chez *O. australiensis* et expliquer l'origine du doublement de la taille du génome par rapport à *O. sativa* durant les deux derniers millions d'années. Les résultats sont présentés dans l'article suivant publié en 2006 dans Genome Research. Mon travail dans cet article a consisté à réaliser toutes les expériences permettant l'obtention des séquences complètes de chacune des trois familles (par clonage et séquençage), ainsi que l'estimation du nombre de copies par dot-blot et leur représentation dans le génome. J'ai aussi déterminé l'origine des trois rétrotransposons dans le genre *Oryza* par Southern blot.

Letter

# Doubling genome size without polyploidization: Dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice

Benoit Piegu,[1] Romain Guyot,[1] Nathalie Picault,[1] Anne Roulin,[1] Abhijit Saniyal,[3] Hyeran Kim,[4] Kristi Collura,[4] Darshan S. Brar,[2] Scott Jackson,[3] Rod A. Wing,[4] and Olivier Panaud[1,5]

[1]*Laboratoire Génome et Développement des Plantes, UMR 5096 CNRS-IRD, Université de Perpignan, Perpignan 66860, France;* [2]*Plant Breeding Genetics and Biochemistry Division, International Rice Research Institute, Manila 1099, Philippines, USA;* [3]*Agricultural Genomics, Purdue University, West Lafayette, Indiana 47907, USA;* [4]*Arizona Genomics Institute, Department of Plant Sciences, University of Arizona, Tucson, Arizona 85721, USA*

**Retrotransposons are the main components of eukaryotic genomes, representing up to 80% of some large plant genomes. These mobile elements transpose via a "copy and paste" mechanism, thus increasing their copy number while active. Their accumulation is now accepted as the main factor of genome size increase in higher eukaryotes, besides polyploidy. However, the dynamics of this process are poorly understood. In this study, we show that *Oryza australiensis*, a wild relative of the Asian cultivated rice *O. sativa*, has undergone recent bursts of three LTR-retrotransposon families. This genome has accumulated more than 90,000 retrotransposon copies during the last three million years, leading to a rapid twofold increase of its size. In addition, phenetic analyses of these retrotransposons clearly confirm that the genomic bursts occurred posterior to the radiation of the species. This provides direct evidence of retrotransposon-mediated variation of genome size within a plant genus.**

**[Supplemental material is available online at www.genome.org.]**

The considerable diversity of genome sizes in higher eukaryotes, in particular the lack of correlation between genome size and biological complexity, first formulated as the *c-value paradox* (Thomas 1971), was one of the most disturbing discoveries in the early days of structural genomics. It has since been established for many organisms that Transposable Elements (TEs) are the main components of complex genomes and that transposition can be regarded as the predominant force driving their structural changes, besides polyploidy (Bennetzen et al. 2005; Vitte and Panaud 2005). In this regard, a particular class of TEs, the retrotransposons, is considered an important factor of genomic inflation in both plants and animals because of their propensity to increase their copy number during transposition (Kumar and Bennetzen 1999). This is well documented in grasses, where Long Terminal Repeat (LTR)-retrotransposons can compose more than half of the genome of some species (SanMiguel et al. 1996; Vicient et al. 1999; Kalendar et al. 2000; Schulman and Kalendar 2005). Much less is known, however, about the dynamics of this process, i.e., the timing of the genomic expansions caused by the activity of LTR-retrotransposons. Do genomes expand gradually through slow accumulation of retrotransposons, or are the expansions a saltatory process caused by large, sudden bursts of retrotransposition? Many studies have shown that in large plant genomes, LTR-retrotransposon families often contain thousands (or tens of thousands) of copies with high sequence identity, which suggests that they originate from a recent massive retrotransposition event (SanMiguel et al. 1998; Vicient et al. 1999). However, none of these studies have so far provided direct evidence at a genome scale that the activity of LTR-retrotransposons could increase genome size to such a large extent over a short evolutionary time.

The genus *Oryza*, to which the Asian-cultivated rice species *O. sativa* belongs, contains 24 species (17 diploids and seven tetraploids) distributed throughout the world (Ge et al. 1999). This genus has been extensively studied because of the economical importance of Asian rice and because the wild relatives of the cultivated species constitute a useful reservoir of genetic diversity which is exploited for rice breeding (Brar and Khush 1997). The taxonomic status and phylogenetic relationships among the 24 species of the genus have been established using phenotypic, cytogenetic, and molecular data (Ge et al. 1999). In addition, their estimated genome size ranges from 357 Mbp for *O. glaberrima* (diploid, genome type AA) to 1283 Mbp for *O. coarctata* (tetraploid, genomes HHKK) (Ammiraju et al. 2006). It is 390 Mbp for the model species *O. sativa* (diploid, genome AA), whose genome has been sequenced (International Rice Genome Sequencing Project 2005). The four largest genome sizes are found in the tetraploid species for which such data is available, which illustrates well that polyploidy in plants is an important factor of genome size variation. However, there is also a significant variation of genome size within the diploid *Oryza* species, with a 2.7-fold variation between the smallest (i.e., *O. glaberrima* 357 Mbp) and the largest genome (i.e., 965 Mbp for *O. australiensis*, genome EE) (Ammiraju et al. 2006). Moreover, as shown in Figure 1, this large difference in size between the genome of *O. australiensis* compared with that of the most closely related diploid species (i.e., *O. sativa*, *O. glaberrima*, *O. officinalis*, and *O. punctata*) suggests that dramatic and recent structural genomic changes have occurred specifically in the lineage of *O. australiensis*.

**[5]Corresponding author.**
**E-mail panaud@univ-perp.fr; fax 33-04-468664899.**
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.5290206.


16:1262–1269 ©2006 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/06; www.genome.org

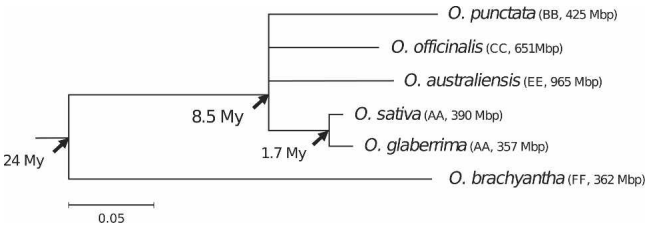



**Figure 1.** Phylogenetic tree of six diploid *Oryza* species, established with the *ADH2* gene. Only nodes with a bootstrap value >70% are shown. The tentative dates of the radiation event are given for each node. The computational methods are given in the Methods section.

By combining a molecular cloning approach and in silico analyses, we have characterized three LTR-retrotransposon families, whose transcriptional activity has generated over 500 Mbp of genomic sequence for the species. Moreover, detailed phenetic studies of these three families in the genus *Oryza* allowed us to characterize the dynamics of this process by providing evidence for several distinct and successive retrotranspositional bursts which have occurred within the last three million years, after the speciation of *O. australiensis*.

## Results and Discussion

### Three LTR-retrotransposon families compose 60% of the *O. australiensis* genome

In order to investigate the cause of the genomic expansion in *O. australiensis*, we first applied Representational Difference Analysis (RDA). RDA (Lisitsyn et al. 1993; Panaud et al. 2002) is a PCR-based cloning procedure that allows isolation of sequences which are specific to one genome (the tester) compared with another (the blocker). It is based on subtractive hybridization of genomic fractions (the representation) of both the tester and the blocker. These representations are obtained after digestion of total genomic DNA, followed by ligation with adapters and PCR amplification of the ligated products using a primer homologous to the adapter. Prior to the subtraction, a new set of adapters is ligated to the representation of the tester DNA only, thus allowing the amplification of specific sequences. Using *O. australiensis* genomic DNA as tester and *O. sativa* as blocker we obtained a library primarily composed of a single 359-bp *O. australiensis*-specific fragment (data not shown). This suggested that, at least in the representation we obtained, the difference in composition between these two genomes could be explained by the presence of a sequence which is highly repeated in the genome of *O. australiensis* but which is absent (or present at a much lower copy number) from the *O. sativa* genome. Sequencing of this RDA sequence revealed that it is part of the LTR of *RIRE1*, a previously characterized *TY1/Copia* type LTR-retrotransposon (Uozu et al. 1997). In order to estimate the copy number of *RIRE1*, dot-blot assays were performed using either LTR or internal region probes (Supplemental data #1). We found that there are 30,000 ± 3000 complete *RIRE1* copies and 10,000 ± 1000 apparent single LTRs, considered as recombinational variants called solo-LTRs (Shirasu et al. 2000), which makes this element one of the most highly repeated within a plant genome. *RIRE1* therefore appears to contribute a total of about 265 Mbp, i.e., 27% of the genome of *O. australiensis* (Table 1). This observation led us to use another strategy to identify other repeated sequences that could have contributed to the genomic expansion of the species in addition to *RIRE1*.

The *Oryza* Map Alignment Project (OMAP, http://www.omap.org) has generated a large amount of genomic resources for 12 *Oryza* species, including 137,000 BAC end sequences (BES) of *O. australiensis* (Ammiraju et al. 2006). In silico analyses of these *O. australiensis* BES allowed identification of 25 highly redundant sequences which were not homologous to *RIRE1*. These were successively extended and merged using the BES data, allowing reconstruction of the putative sequence of three *TY3/Gypsy* type LTR-retrotransposons, named *Kangourou*, *Wallabi*, and *Dingo* (Fig. 2; Supplemental data #2). Three BAC clones from the *O. australiensis* genomic libraries were then fully sequenced to validate the structure of these three new elements. One clone harbors the *HD1* locus (and was chosen because it also harbors at least one copy of *RIRE1*). The other two clones were chosen based on the homology of one of their BES with either *Wallabi* or *Kangourou*. Comparative sequence analysis with the *O. sativa* genome suggests that none of the three clones are located in a pericentromeric region. Overall, 350 kb (350,792 bp) of genomic sequence were generated and analyzed (Fig. 3). Seven complete copies and five solo-LTRs of *RIRE1*, *Kangourou*, and *Wallabi* were identified from these genomic clones, thus validating the sequences inferred from the in silico approach. Moreover, *Kangourou* and *Wallabi* were searched for homology with already known rice retrotransposons. We found that *Kangourou* exhibits a low but significant sequence identity with the *Retrosat1/RIRE2* retrotransposon family (65% overall identity in the internal region with 75% in the GAG-POL region). This is an indication that these two families are homologous. However, given the low sequence identity between *Kangourou* and *Retrosat1*, we kept a distinct name for each.

Based on dot-blot assays, *Kangourou* and *Wallabi* were estimated to contribute a total of 90 ± 9 Mbp and 250 ± 25 Mbp, i.e., 9% and 26% of the genome of *O. australiensis*, respectively (Table 1). The copy number of *Dingo* was estimated to be between 3700 and 4300 and could thus clearly be considered as highly

**Table 1.** Description of the three retrotransposons, *RIRE1*, *Kangourou*, and *Wallabi*, in the genome of *O. australiensis*

| | | Size in bp | Number of copies | Size in the genome | Total |
|---|---|---|---|---|---|
| *RIRE1* | Full element | 8300 | 30,000 ± 3000 | 250 ± 25 Mbp | 265 ± 26.5 Mbp |
| | Apparent single LTR | 1500 | 10,000 ± 1000 | 15 ± 1.5 Mbp | |
| *Kangourou* | Full element | 9200 | 9500 ± 1000 | 87 ± 9 Mbp | 90 ± 9 Mbp |
| | Apparent single LTR | 3500 | 1000 ± 100 | 3.5 ± 0.5 Mbp | |
| *Wallabi* | Full element | 9000 | 27,000 ± 3000 | 240 ± 24 Mbp | 250 ± 25 Mbp |
| | Apparent single LTR | 500 | 12,000 ± 1000 | 6 ± 0.5 Mbp | |
| | | | | | 605 ± 60 Mbp |

The number of copies is estimated based on dot-blot hybridizations. Mean and standard deviation (based on eight repetitions, see Supplemental data #1) are given for each element (either for the full element or for the apparent single LTR).

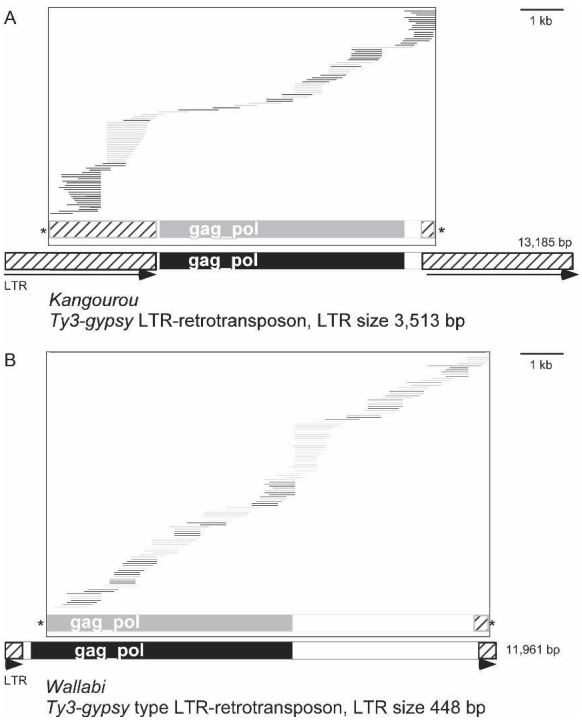



**Figure 2.** In silico reconstruction of *Kangourou* (*A*) and *Wallabi* (*B*) retrotransposons. The copies of the BES contigs are shown as horizontal lines (alternate black and gray according to their final position on the element). The schematic representation of the element assembled from *O. australiensis* BES is represented in gray. The schematic representation of the elements, as it is found in the *O. australiensis* sequenced BAC clones, is given in black at the *bottom* of the figures. The size scale in bp is given at the *bottom*.

repeated, but its contribution to the genomic expansion of *O. australiensis* (i.e., <5% of the present size) was considered negligible compared with the other three elements. *Dingo* was therefore not included in further analyses. Altogether, *RIRE1*, *Kangourou*, and *Wallabi* contribute ~60% (605 ± 40 Mbp) of the *O. australiensis* genome (Table 1). In contrast, BLAST searches of the genomic sequence of *O. sativa* (*cv*. Nipponbare) revealed that it contains only two, 10, and 16 complete copies of *RIRE1*, *Kangourou*, and *Wallabi* elements, respectively. The retrotransposition bursts of these three elements alone could thus account for the increase in the genome size of *O. australiensis*, compared with that of *O. sativa*. Our data therefore show that, at least in the case of the genus *Oryza*, retrotransposition has contributed to genome size variation to an extent which is comparable with that of polyploidization.

We first anticipated that BES may not be a random representation of the genome of *O. australiensis*, mainly because the BAC library was constructed using the HindIII restriction enzyme, but our results show a posteriori that this approach was nevertheless successful for retrieving the most highly repeated elements from the genome regardless of the representational bias that may have been caused by the restriction enzyme. In fact, it is clear from Figure 2 that some regions of the retrotransposon are overrepresented in the BES database, probably because most of the paralogs harbor a HindIII site at this location. Contrastingly, it is expected that BES corresponding to the regions of the element where the majority of the paralogs do not harbor a cleavage site should be far less frequent. However, in our case, we were able to retrieve at least one BES in all the regions of the three elements, probably because of the high copy number of each family, thus allowing their in silico reconstruction.

Deep annotation of the sequences of the three BAC clones revealed the presence of many TEs which are distinct from *RIRE1*, *Kangourou*, and *Wallabi* (Fig. 3). In all, 61 putative transposable elements were identified, accounting for 195,277 bp of sequence and representing 55.6% of the complete BAC sequences. Among these, a majority belong to the Class I retrotransposon group (accounting for 65% of all the TE and 42.9% of the BAC sequences). The 39 LTR-retrotransposons identified in these sequences belong to 11 distinct families. The LTR-retrotransposons, analyzed and annotated in the three sequenced BAC clones, were found mainly clustered within intergenic regions (Fig. 3). These elements are frequently disrupted by successive insertions of other LTR-retrotransposons, leading to the formation of nested structures as often observed in larger genomes of other cereal species (SanMiguel et al. 1996; Wicker et al. 2001, 2003). Distal parts of BACs OA_59I14 and OA_AB10104J14 and the central part of BAC OA_ABa0008H03 showed a high density of clustered and nested LTR-retrotransposons, with respectively five, four, and eight elements clustered within a distance of 41, 30, and 50 kb. Altogether, *Kangourou*, *Wallabi*, and *RIRE1* account for ~29% of all the BAC sequences which is much less than their overall genome contribution (60%). However, FISH experiments previously revealed that *RIRE1* elements are more abundant in pericentromeric than in distal regions of *O. australiensis* chromosome arms (Uozu et al. 1997), a characteristic shared with other LTR-retrotransposon families in the *O. sativa* genome (Jiang et al. 2002; Vitte and Panaud 2003). Consequently, the distribution of the elements in the three BAC sequences may not reflect their actual distribution in the genome.

## The genomic amplification of *O. australiensis* occurred after its speciation

In order to trace the origin of the three elements *RIRE1*, *Kangourou*, and *Wallabi*, we surveyed their presence in nine different genome types of the genus *Oryza* by Southern hybridization using probes corresponding to either the LTR or the internal region (Fig. 4). The results clearly show that all three elements are present in at least one other wild *Oryza* species, indicating an ancient origin in the genus. Moreover, the strong hybridization signals obtained for some species distantly related to *O. australiensis* (e.g., in *O. granulata* [GG] for *Wallabi*) suggest that independent transposition bursts of *RIRE1* and *Wallabi* have occurred in distinct genome types of the genus, although to a lesser extent than in the genome of *O. australiensis*. In order to tentatively characterize and date the transposition bursts of the three elements in the *O. australiensis* genome, we conducted phenetic analyses of the three elements based on the OMAP BES data of the 12 *Oryza* species (Fig. 5; Supplemental data #3): For all three elements, the paralogs found in *O. australiensis* form a cluster which is distinct from those found in other *Oryza* species (i.e., supported by a bootstrap value which is >70%), suggesting that the retrotransposition bursts occurred concomitantly or after the speciation of *O. australiensis*.

In order to test this hypothesis, the dates of retrotransposition of *RIRE1*, *Kangourou*, and *Wallabi* were estimated in the *O. australiensis* genome (Fig. 6). We applied an approach of genomic paleontology, which consists of translating the nucleotide divergence observed between the paralogs mined out from the BES into a radiation date. This approach relies on the estimation of

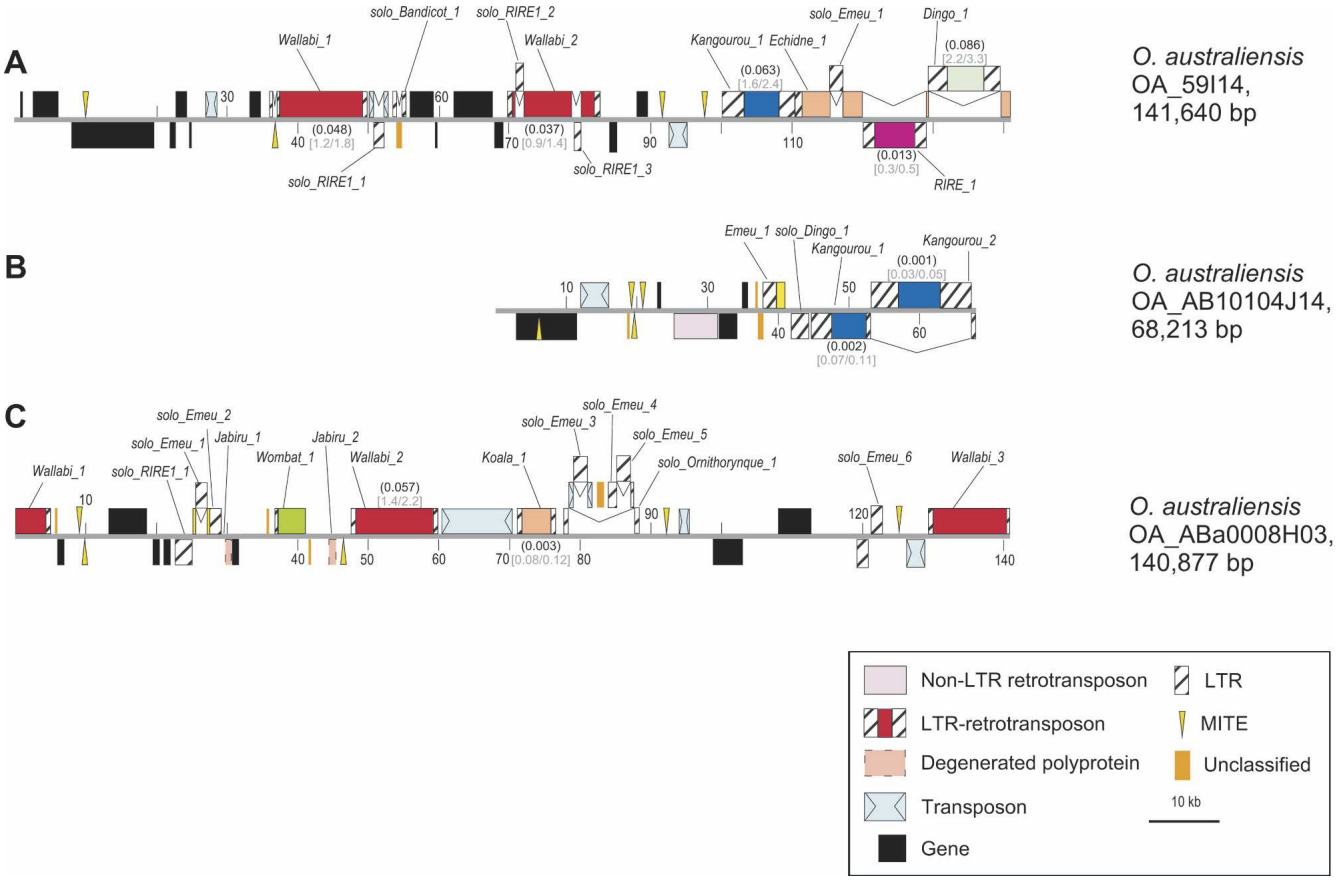



**Figure 3.** Physical map of three sequenced *O. australiensis* BAC clones. Black boxes represent predicted coding regions. Colored boxes represent different types of TEs as indicated on the figure. Numbers in parentheses indicate the estimated date of LTR-retrotransposon insertions (in million years) using the two molecular clocks MC1 and MC2 (see text). *A,B,C* correspond to the sequence of the BAC clones OA_59114, OA_AB10104J14, and OA_ABa0008H03, respectively.

the rate of the molecular clock (MC) of retrotransposon sequences once they are inserted in the genome. The first examples of such studies in plants were conducted using an MC of $6.5 \times 10^{-9}$ synonymous substitutions/site/year (SanMiguel et al. 1998), an estimation based on the MC of the *ADH2* gene in the *Poaceae* family (Gaut et al. 1996). Several subsequent studies have led to a re-estimation of the MC of retrotransposons in rice, i.e., $2 \times 10^{-8}$ subst/site/year (Vitte et al. 2004), referred to as MC1, and $1.3 \times 10^{-8}$ subst/site/year (Ma et al. 2004), referred to as MC2. The data provided in the present paper are given using both these new MC. In any case, the translated dates can only be considered as rough estimates and only large differences should be retained as putatively significant. The figure clearly shows that the transpositional activity of the three elements has not been continuous during the last 3 to 4 Myr: A peak of activity (defined here as a burst) is indeed observed at ~0.5–0.75, 1.2–1.8, and 2–3 Mya for *RIRE1*, *Wallabi*, and *Kangourou*, respectively. Moreover, the size of the peaks shown in Figure 6 is proportional to the

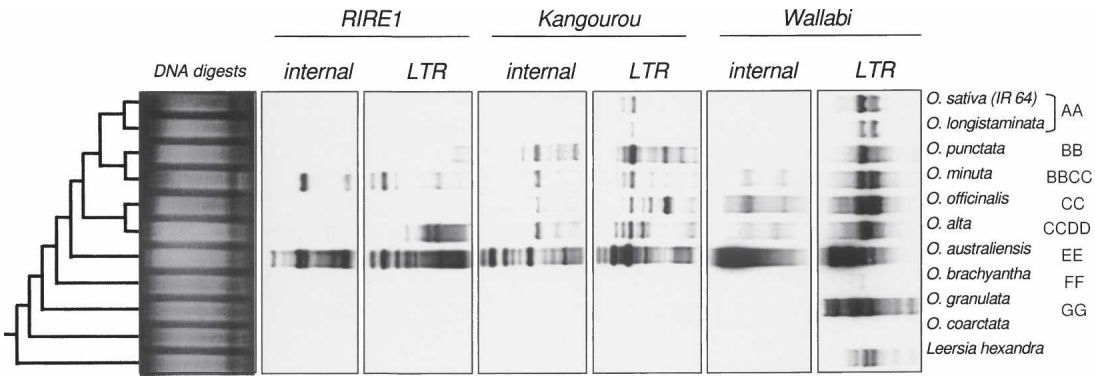



**Figure 4.** Southern hybridization of the three retrotransposons, *RIRE1*, *Kangourou*, and *Wallabi* on total genomic DNA of *Oryza* species digested with RsaI. The phylogenetic tree given on the figure is extrapolated from Ge et al. (1999). The direction of migration is from *left* to *right*.

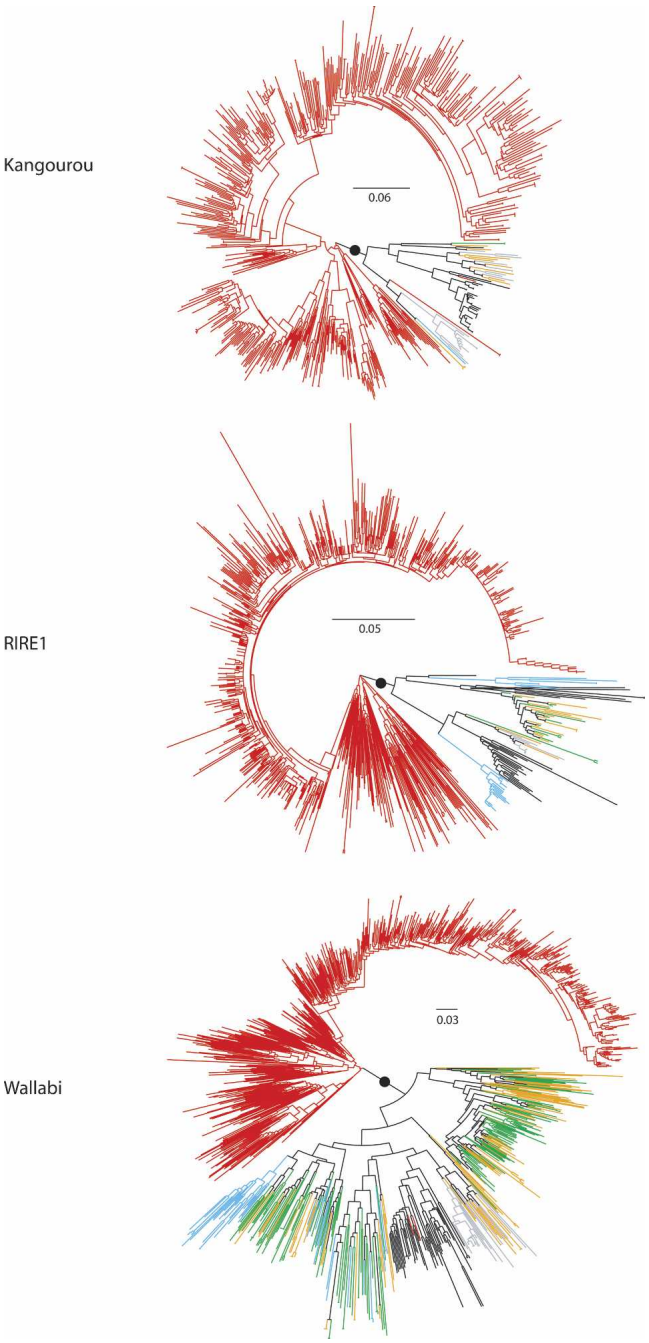



**Figure 5.** Phenetic relationships of *RIRE1*, *Kangourou*, and *Wallabi* in the genus *Oryza*: The neighbor-joining tree was constructed based on the alignments given in Supplemental data #3. For each tree, the dot shows the branch separating the *O. australiensis* sequences from the others. The number given near the dot corresponds to the bootstrap value. Color coding: black for the A-genome species; gray for *O. punctata*; orange for *O. minuta*; green for *O. officinalis*; blue for *O. alta*, and red for *O. australiensis*. The numbers of aligned sequences used to build the tree were as follows: for *RIRE1*: 752 *O. australiensis* sequences and 113 other *Oryza* sequences; for *Kangourou*: 570 *O. australiensis* sequences and 67 others; for *Wallabi*: 757 *O. australiensis* sequences and 422 others.

number of complete copies of the corresponding elements in the *O. australiensis* genome. This representation clearly shows that the largest bursts are the most recent and, therefore, that most of the genomic expansions that led to the doubling of the genome size of *O. australiensis* are of recent origin (i.e., within the last 3 or 4 Myr). This is further supported by dating of the insertions of *RIRE1*, *Kangourou*, and *Wallabi* elements found in the three BAC sequences (Fig. 3). In the case of full-length LTR-retrotransposons, the date of insertion can be estimated based on the divergence between their two LTRs (SanMiguel et al. 1998; Vitte et al. 2004). The estimated insertion time of these elements ranges from 1.6 to 2.4 Mya for *Kangourou_1* (OA_59I14) to 0.03–0.05 Mya for *Kangourou_2* (OA_AB10104J14). Because the date of the radiation of *O. australiensis* species is estimated at 8.5 Myr (Fig. 1), we conclude from all these lines of evidence that the genomic expansion is posterior to and not concomitant with the speciation. These results also suggest that the strong hybridization signals observed in some other *Oryza* species in the Southern hybridization experiments (e.g., for *Wallabi* in *O. granulata*, Fig. 4) reflect distinct bursts of the corresponding elements in these lineages. In this regard, the large size of the *O. granulata* genome (i.e., 880 Mbp) (Ammiraju et al. 2006), compared with that of *O. sativa*, may be partly accounted for by the retrotranspositional activity of *Wallabi*, although more detailed analyses are needed to quantify precisely this contribution.

The phenetic analyses also provide interesting insights into the dynamics of the genomic expansions observed in *O. australiensis*. The peaks shown in Figure 6 are not overlapping, thus suggesting that the maximum transpositional activity of the three elements did not occur concomitantly. The cause of these successive waves of retrotransposition could be regulatory, i.e., the result of an activation (triggered by external stimuli, such as biotic or abiotic stress) and/or by the repression of silencing of

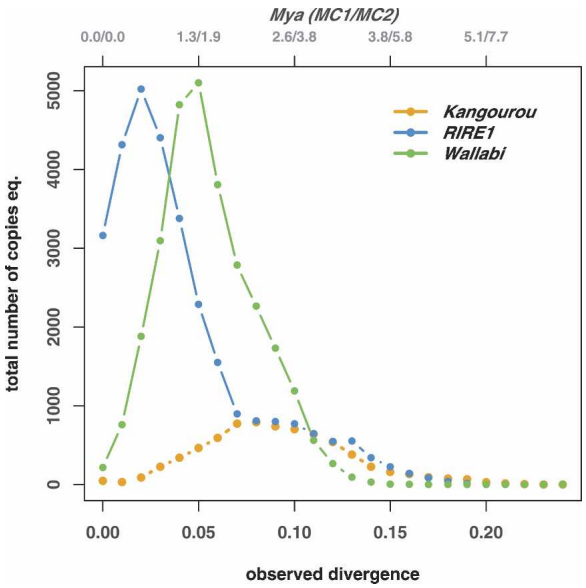



**Figure 6.** Timing of the bursts of the three retrotransposons, *RIRE1*, *Kangourou*, and *Wallabi*: For each element, the curves represent the distribution of the observed divergence between each paralog (given at the *bottom x*-axis). *Top x*-axis represents the date of divergence in Mya translated from the observed divergence, using the two molecular clocks MC1 and MC2 (see Methods section). The groups of paralogs used to compute the pairwise distances are defined within the phenetic subgroups shown in the phenogrammes given in Supplemental data #4. The *y*-axis represents the total number of copy equivalent, i.e., (the frequency at which the divergence time occurred) × (the number of paralogs in the genome of *O. australiensis*, based on the dot-blot experiments, Table 1).

the corresponding elements. Alternatively, these distinct bursts may be explained by the presence of active elements in the genome only during a short period (corresponding to the peaks). These active elements may have arisen from ectopic recombinations between two defective copies, a mechanism known for retroviruses (Bartosch et al. 2004). The presence of *RIRE1*, *Kangourou*, and *Wallabi* in the genome of many other *Oryza* species (Fig. 4) suggests, however, that functional copies of these three elements have probably been present in the genus since its origin. In order to test whether active copies may still be present in the genome of *O. australiensis*, we determined for each paralog the shortest distance found among all the pairwise distances computed with all the other copies (at the nucleotide level). The distributions of these shortest distances are given in Figure 7. Interestingly, for *RIRE1*, *Kangourou*, and *Wallabi*, several pairs of very closely related paralogs can be found (the first bar of the histogram), suggesting recent transposition (<200,000 yr ago) of active elements. Consequently, transpositional bursts observed in this species may have their origin in a regulatory process, rather than a structural mechanism, although this remains a hypothesis that should be further tested. As a first step one should conduct expression studies of the three elements in *O. australiensis* in order to assess whether its genome still harbors transcriptionally active copies.

Several reports on both plants and animals have shown that transposable elements are efficiently eliminated from eukaryotic genomes, either by recombination or deletion (Petrov et al. 1996; Shirasu et al. 2000; Ma et al. 2004; Chantret et al. 2005). In particular, LTR-retrotransposons tend to be partially eliminated through ectopic recombinations between their two LTRs, leading to the formation of solo-LTRs (Shirasu et al. 2000). This was taken into account in our estimation of the total contribution of the three elements to the genome size increase of *O. australiensis* by using both LTR and internal region probes in the dot-blot assay

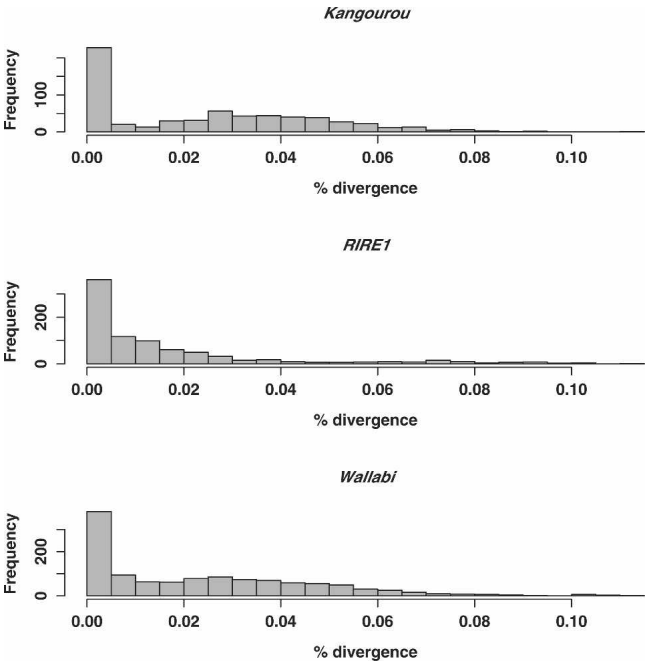



**Figure 7.** Histograms of the most recent among all observed divergence computed for each paralog (compared with all others) of the three retrotransposons *RIRE1*, *Kangourou*, and *Wallabi*.

(Table 1; Supplemental data #1). Interestingly, the apparent single LTRs represent a significant percentage of the total number of copies that we estimated (i.e., 25%, 9.5%, and 30% for *RIRE1*, *Kangourou*, and *Wallabi*, respectively), regardless of the age of the bursts (that of *RIRE1* being the most recent). This corroborates earlier reports suggesting that the process of partial removal of LTR-retrotransposons through ectopic recombinations leading to solo-LTRs is concomitant with (or occurs shortly after) retrotransposition (Vitte and Panaud 2003). Illegitimate recombination mechanisms targeting LTR-retrotransposons have been identified as inducing considerable loss of DNA and contributing to genome size reduction in *Arabidopsis* and rice (Petrov 2002; Ma et al. 2004). Analysis by sequence alignment of *Wallabi*, *Kangourou*, and *RIRE1* LTR-retrotransposon families from the three fully sequenced BAC clones (Fig. 3) revealed the presence of limited small deletions, corresponding to 2.4%, 6.7%, and 2.6% of the total length of the *Wallabi*, *Kangourou*, and *RIRE1* elements, respectively (data not shown). Altogether, these results indicate that processes leading to the elimination of retrotransposon sequences such as unequal and illegitimate recombinations occurred in the genome of *O. australiensis* similarly to *O. sativa*. Both the extent and timing of these DNA losses need to be investigated in order to clarify their overall contribution to the *O. australiensis* genome size variation following the bursts of the three retrotransposon families. Nevertheless, we anticipate that the overall DNA loss of the three LTR-retrotransposons through small illegitimate recombinations is negligible and did not lead to overestimation of their overall contribution in the genome of the species, based on our dot-blot assay.

The current model of eukaryotic genome evolution in relation to the activity of TEs posits that genome size should result from two balanced forces: increase, induced by retrotransposition, and decrease, caused by recombinations and deletions (Petrov 2002; Vitte and Panaud 2005). The evolutionary dynamics of *RIRE1*, *Kangourou*, and *Wallabi* in the genus *Oryza* provides an opportunity to test this hypothesis. Our Southern hybridization and phenetic data suggest that the three elements were present in the genome of the ancestor of the genus. This is further supported by the presence of homologs of these elements in BES of nearly all the *Oryza* species of the OMAP project. We also show that they have undergone independent amplification in distinct lineages in the genus, leading to one case of genomic obesity (i.e., in *O. australiensis*). In other lineages, their strong regulation and elimination may have led to a decrease in genome size (e.g., in *O. glaberrima*), but this still remains purely speculative. In this regard one should point out that, if all the copies of the three retrotransposons families *RIRE1*, *Kangourou*, and *Wallabi* were to be removed from the genome of *O. australiensis*, about 360 Mbp of genomic DNA would remain, i.e., a size comparable with that of the smallest diploid genomes of the *Oryza* genus. Further examination of complete sequences of high copy number LTR-retrotransposons in *O. australiensis* will provide better insights into the dynamics of the elimination process. The model also predicts that the successive events of TE insertions followed by their elimination should cause a fast turnover of intergenic regions, leading to their rapid divergence among distinct evolutionary lineages. This has been confirmed in several comparative studies between the genomes of maize, sorghum, and rice (Tikhonov et al. 1999; Ma et al. 2005). Comparative genomics studies within the *Oryza* genus, and in particular between closely-related species, would allow us to test this hypothesis and give insight into the dynamics of the process.

Maize, wheat, and barley are large genomes that contain 50%–80% of LTR-retrotransposons. It is now commonly accepted that retrotranspositions have played a crucial role in genomic expansion and architecture and could also have an impact on the transcriptional regulation of genes of these major crop species (Kashkush et al. 2003). However, the relatively older burst of amplification of these elements and the limited genomic sequence information from these large genomes make it difficult to reconstruct the history and study the impact of retrotransposon amplification at the whole genome level. The present study provides the first direct evidence that active LTR-retrotransposons can contribute to large variations in genome size over short periods of time, i.e., at the species level. As in the case of maize, wheat, and barley, LTR-retrotransposons are the main component of the *O. australiensis* genome. Cytologically, the mitotic chromosomes of *O. australiensis* show a twofold size increase compared with the *Oryza sativa* species (Uozu et al. 1997), indicating their dramatic impact on chromosome morphology and suggesting major events of genome reshaping. The availability of comprehensive genomic resources for many species in the genus *Oryza* makes possible physical comparison between closely related genomes contrasting for their size and will allow studies on the evolutionary history of the LTR-retrotransposons at the genus scale. This provides a unique and promising opportunity to unlock our knowledge on the causes of retrotransposition bursts as well as their impact on plant genome architecture and gene expression.

## Methods

### Research of highly repeated sequences

The 137,000 *O. australiensis* BES were used as query for a BLASTN search (Altschul et al. 1990) against themselves (all-by-all search). Only BES with 200 or more "complete" matches with at least 95% identity were kept. A "complete" match is a match with subject or query completely aligned. A total of 1132 BES were obtained and then used as query for a BLASTN search against known LTR-retrotransposon sequences. Six hundred and forty-nine BES matched perfectly with *RIRE1* (>95% identity). The remaining 483 BES not matching with *RIRE1* were assembled using the Sequencher software (Gene Codes Corporation) with a minimum match of 95%. A total of 25 assembled sequences (seeds) were obtained following this procedure. All these steps were automated with Perl scripts (available on request).

### In silico reconstruction of the retrotransposons

The 25 highly redundant seed contigs were submitted to a BLAST search against the 137,000 BES of *O. australiensis*. Seeds and conserved overlapping BES (*E*-value < e-100) were assembled using the Sequencher software with a minimum match of 90%. This rather low threshold was especially necessary for the reconstruction of *Wallabi*, given that most of the paralogs found in the BES are more ancient than those of both *RIRE1* and *Kangourou*. The contigs identified de novo were extended by reiterative comparisons and assemblies with overlapping BES. At each step of the contig extension, the accuracy of the assembly was checked by comparative analysis with the corresponding *O. sativa* homologous LTR-retrotransposons.

### Southern hybridizations

Blots were prepared using 2µg of total genomic DNA digested with RsaI transferred onto Hybond-N+ membrane. A nonradioactive procedure for probe labeling and detection signal was used (Panaud et al. 1993). Stringency washes were performed at 65°C in 0.5 × SSC. The probes were obtained by cloning the amplification products of PCR reactions on total genomic DNA of *O. australiensis* using primer pairs corresponding to the selected regions of the three elements. The accessions used for all the species were *cv.IR 64* for *O. sativa*; acc. 110,404 for *O. longistaminata*; acc. 105,690 for *O. punctata*; acc. 101,089 for *O. minuta*; acc. 101,116 for *O. officinalis*; acc. 105,143 for *O. alta*; acc. 100,882 for *O. australiensis*; acc. 101,232 for *O. brachyantha*; acc. 102,118 for *O. granulata*, and acc. 104,502 for *O. coarctata*. No accession number is available for *Leersia hexandra*. The DNA was provided by the International Rice Research Institute, Manila, Philippines.

### BAC sequence analysis

BAC sequences were analyzed using BLASTN algorithms (Altschul et al. 1990) against public and local nucleotide databases. Detailed analysis was performed with the EMBOSS package (Rice et al. 2000) and by dot-plot (using DOTTER software; Sonnhammer and Durbin 1995). Putative genes were determined by a combination of coding region prediction software available through the RiceGAAS Web site (http://ricegaas.dna.affrc.go.jp) and similarity searches. Final annotation was performed with the Artemis tool (Rutherford et al. 2000). Putative TEs were both identified and annotated by similarity searches against local databases of plant TEs and by investigating structural properties of the elements. Dating of intact LTR-retrotransposon insertions was conducted according to SanMiguel et al. (1998) with the EMBOSS package using two distinct molecular clocks referred to as MC1 and MC2 (respectively $2 \times 10^{-8}$ subst/site/year from Vitte et al. 2004 and $1.3 \times 10^{-8}$ subst/site/year from Ma et al. 2004). Putative TEs were classified according to their mobility mechanisms.

### Phenetic analyses

For each of the three elements, *RIRE1*, *Kangourou*, and *Wallabi*, the total sequence was split into subsequences of either 400 bp (for *Wallabi*) or 200 bp (for *RIRE1* and *Kangourou*). Each of these subsequences was used as query for a BLASTN search against all the OMAP BES released in GenBank. Only hits showing homology over at least 90% of the total length of the query were kept. For each element, only one subregion, for which the highest number of *Oryza* species were represented in the data set, was chosen for further analyses. All the subsequences corresponding to the match were aligned using ClustalX (Thompson et al. 1997) and the alignments were modified by hand using SEAVIEW software (Galtier et al. 1996). Final alignments were used to construct a Neighbor-Joining dendrogram using the ClustalX software, using the observed divergence distance and performing 1000 bootstraps. A circular classification tree was drawn using the Treedyn package (http://www.treedyn.org/). The age of the retrotransposition bursts was estimated at the peaks of the distribution of the pairwise nucleotide distances obtained within phenetic groups for each element (Fig. 5; Supplemental data #4). The observed divergence was translated into an insertion date following the method first described by SanMiguel et al. (1998), but using the two molecular clocks MC1 and MC2 (see above).

### Phylogenetic analyses of *Oryza* species

The sequence of *ADH2* gene was retrieved from GenBank for six diploid *Oryza* species. We used the GenBank accession numbers AF148623 for *O. australiensis*, AF148632 for *O. brachyantha*, AF148606 for *O. glaberrima*, AF148613 for *O. officinalis*, AF148611 for *O. punctata*, and AF148602 for *O. sativa* (Ge et al. 1999). For each accession, the coding sequence (CDS) was extracted. These

six CDS were then aligned using ClustalX, and a pairwise distance matrix was computed using the Nei and Gojobori method (Nei and Gojobori 1986). Only synonymous substitutions were thus taken into account. A neighbor-joining tree was built using 500 bootstrap replicates. We estimated the date of the nodes using a rate of $6.5 \times 10^{-9}$ synonymous substitutions/site/year (Gaut et al. 1996). The Mega3 software was used for this work (Kumar et al. 2004).

## Acknowledgments

BP, RG, NP, AR, and OP were supported by CNRS funding. HK, RAW, AS, HK, KC, and SJ were supported by NSF grants IOB 0208329 and DBI 0321678. We thank R. Cooke, T. Wicker and V. Colot for their useful comments on the manuscript. The GenBank accession numbers of the LTR-retrotransposons described in this article are DQ365821 for *Kangourou*, DQ365824 for *Wallabi*, and DQ365822 for *Dingo*.

## References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Ammiraju, J.S., Luo, M., Goicoechea, J.L., Wang, W., Kudrna, D., Mueller, C., Talag, J., Kim, H., Sisneros, N.B., Blackmon, B., et al. 2006. The *Oryza* bacterial artificial chromosome library resource: Construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus *Oryza*. *Genome Res.* **16:** 140–147.

Bartosch, B., Stefanidis, D., Myers, R., Weiss, R., Patience, C., and Takeuchi, Y. 2004. Evidence and consequence of porcine endogenous retrovirus recombination. *J. Virol.* **78:** 13880–13890.

Bennetzen, J.L., Ma, J., and Devos, K.M. 2005. Mechanisms of recent genome size variation in flowering plants. *Ann. Bot. (Lond.)* **95:** 127–132.

Brar, D.S. and Khush, G.S. 1997. Alien introgression in rice. *Plant Mol. Biol.* **35:** 35–47.

Chantret, N., Salse, J., Sabot, F., Rahman, S., Bellec, A., Laubin, B., Dubois, I., Dossat, C., Sourdille, P., Joudrier, P., et al. 2005. Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). *Plant Cell* **17:** 1033–1045.

Galtier, N., Gouy, M., and Gautier, C. 1996. SEAVIEW and PHYLO_WIN: Two graphic tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci.* **12:** 543–548.

Gaut, B.S., Morton, B.R., McCaig, B.C., and Clegg, M.T. 1996. Substitution rate comparisons between grasses and palms: Synonymous rate differences at the nuclear gen *Adh* parallel rate differences at the plastid gene *rbcL*. *Proc. Natl. Acad. Sci.* **93:** 10274–10279.

Ge, S., Sang, T., Lu, B.R., and Hong, D.Y. 1999. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proc. Natl. Acad. Sci.* **96:** 14400–14405.

International Rice Genome Sequencing Project. 2005. The map-based sequence of the rice genome. *Nature* **436:** 793–800.

Jiang, N., Bao, Z., Temnykh, S., Cheng, Z., Jiang, J., Wing, R.A., McCouch, S.R., and Wessler, S.R. 2002. Dasheng: A recently amplified nonautonomous long terminal repeat element that is a major component of pericentromeric regions in rice. *Genetics* **161:** 1293–1305.

Kalendar, R., Tanskanen, J., Immonen, S., Nevo, E., and Schulman, A.H. 2000. Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc. Natl. Acad. Sci.* **97:** 6603–6607.

Kashkush, K., Feldman, M., and Levy, A. 2003. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat. Genet.* **33:** 102–106.

Kumar, A. and Bennetzen, J.L. 1999. Plant retrotransposons. *Annu. Rev. Genet.* **33:** 479–532.

Kumar, S., Tamura, K., and Nei, M. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* **5:** 150–163.

Lisitsyn, N., Lisitsyn, N., and Wigler, M. 1993. Cloning the differences between two complex genomes. *Science* **259:** 946–951.

Ma, J., Devos, K.M., and Bennetzen, J.L. 2004. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res.* **14:** 860–869.

Ma, J., SanMiguel, P., Lai, J., Messing, J., and Bennetzen, J.L. 2005. DNA rearrangement in orthologous orp regions of the maize, rice and sorghum genomes. *Genetics* **170:** 1209–1220.

Nei, M. and Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3:** 418–426.

Panaud, O., Magpantay, G., and McCouch, S.R. 1993. A protocol for non-radioactive DNA labelling and detection in the RFLP analysis of rice and tomato using single-copy probes. *Plant Mol. Biol. Rep.* **11:** 54–59.

Panaud, O., Vitte, C., Hivert, J., Muzlak, S., Talag, J., Brar, D.S., and Sarr, A. 2002. Characterization of transposable elements in the genome of rice (*Oryza sativa* L.) using Representational Difference Analysis (RDA). *Mol. Genet. Genomics* **268:** 113–121.

Petrov, D.A. 2002. Mutational equilibrium model of genome size evolution. *Theor. Popul. Biol.* **61:** 531–544.

Petrov, D.A., Lozovskaya, E.R., and Hartl, D.L. 1996. High intrinsic rate of DNA loss in *Drosophila*. *Nature* **384:** 346–349.

Rice, P., Longden, I., and Bleasby, A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **16:** 276–277.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. 2000. Artemis: Sequence visualization and annotation. *Bioinformatics* **16:** 944–945.

SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., et al. 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274:** 765–768.

SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., and Bennetzen, J.L. 1998. The paleontology of intergene retrotransposons of maize. *Nat. Genet.* **20:** 43–45.

Schulman, A.H. and Kalendar, R. 2005. A movable feast: Diverse retrotransposons and their contribution to barley genome dynamics. *Cytogenet. Genome Res.* **110:** 598–605.

Shirasu, K., Schulman, A.H., Lahaye, T., and Schulze-Lefert, P. 2000. A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. *Genome Res.* **10:** 908–915.

Sonnhammer, E.L. and Durbin, R. 1995. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **167:** 1–10.

Thomas, C.A. 1971. The genetic organization of chromosomes. *Annu. Rev. Genet.* **5:** 237–256.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25:** 4876–4882.

Tikhonov, A.P., SanMiguel, P.J., Nakajima, Y., Gorenstein, N.D., Bennetzen, J.L., and Avramova, Z. 1999. Colinearity and its exceptions in orthologous ADH regions of maize and sorghum. *Proc. Natl. Acad. Sci.* **96:** 7409–7414.

Uozu, S., Ikehashi, H., Ohmido, N., Ohtsubo, H., Ohtsubo, E., and Fukui, K. 1997. Repetitive sequences: Cause for variation in genome size and chromosome morphology in the genus *Oryza*. *Plant Mol. Biol.* **35:** 791–799.

Vicient, C.M., Suoniemi, A., Anamthawat-Jonsson, K., Tanskanen, J., Beharav, A., Nevo, E., and Schulman, A.H. 1999. Retrotransposon BARE-1 and its role in genome evolution in the genus hordeum. *Plant Cell* **11:** 1769–1784.

Vitte, C. and Panaud, O. 2003. Formation of Solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice *Oryza sativa* L. *Mol. Biol. Evol.* **20:** 528–540.

———. 2005. LTR retrotransposons and plant genome size: Emergence of the increase/decrease model. *Cytogenet. Genome Res.* **110:** 91–107.

Vitte, C., Ishii, T., Lamy, F., Brar, D.S., and Panaud, O. 2004. Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa*, L.). *Mol. Genet. Genomics* **272:** 504–511.

Wicker, T., Stein, N., Albar, L., Feuillet, C., Schlagenhauf, E., and Keller, B. 2001. Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. *Plant J.* **26:** 307–316.

Wicker, T., Yahiaoui, N., Guyot, R., Schlagenhauf, E., Liu, Z.D., Dubcovsky, J., and Keller, B. 2003. Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and Am genomes of wheat. *Plant Cell* **15:** 1186–1197.

*Received May 9, 2006; accepted in revised form August 2, 2006.*

Au cours de ce travail, nous avons mis en évidence que *RIRE1, Wallabi et Kangourou* comptent respectivement 30 000, 27 000 et 10 000 copies expliquant ainsi à eux seuls le doublement de la taille du génome d'*O. australiensis* par rapport à celui d'*O. sativa*. L'ensemble de ces copies représentent environ 600 Mb sur les 960 Mb soit 60 % du génome d'*O. australiensis*. La datation des bursts a montré que les événements de transposition se sont déroulés durant les 2 derniers millions d'années, bien après la spéciation d'*O. australiensis*, ce qui explique pourquoi *RIRE1, Wallabi et Kangourou* sont en faible nombre de copies chez les autres espèces étudiées. Ceci démontre que la morphologie des chromosomes peut être modifiée de manière drastique et sur un temps évolutif court. De tels processus auront donc sûrement un impact sur l'expression des gènes et donc sur l'évolution de l'espèce. Néanmoins, les phénomènes permettant aux rétrotransposons d'échapper aux mécanismes d'inactivation restent encore peu connus, même s'ils sont de plus en plus étudiés ces dernières années.

## 2. Impact fonctionnel des rétrotransposons à LTR sur les génomes

Comme je l'ai expliqué dans l'introduction, la recherche d'ETs actifs est nécessaire pour l'étude de leur biologie et de leur dynamique au sein des génomes. Malgré leur grand nombre dans les génomes des plantes, peu d'ETs ont été décrits comme actifs. Ceci est dû à plusieurs raisons. (1) L'activité transcriptionnelle des ETs est strictement contrôlée dans les génomes hôtes *via* le TGS et seulement une infime partie est exprimée. (2) Dans leur grande majorité les ETs ont perdu leur capacité à transposer à cause des différentes mutations qu'ils subissent. En conséquence, la plupart des transcrits sont aberrants et donc incapables de coder la machinerie indispensable à la transposition. (3) Même s'ils produisent des transcrits « viables », les ETs doivent échapper au système de régulation post-transcriptionnel (PTGS) pour réaliser un cycle complet de transposition. Malgré ce strict contrôle dans les cellules, les ETs sont capables, dans certaines conditions, de retrouver leur capacité transpositionnelle et générer de nouvelles insertions dans les génomes hôtes.

Historiquement, les insertions d'ETs ont été détectées parce qu'elles causaient un phénotype visible. Le clonage positionnel de l'allèle mutant permettait de mettre en évidence le lien causal entre le phénotype et l'insertion d'un ET (Chopra et al., 1999). Encore très récemment, le clonage positionnel du locus responsable du déterminisme du sexe chez le melon par Martin et al., (2009) a permis de mettre en évidence le rôle de l'insertion d'un transposon de type *hAT* dans l'apparition de l'andromonoécie chez cette espèce (la présence de fleurs hermaphrodites). La découverte de telles insertions est souvent fortuite, car l'objectif de départ des études n'était pas centré sur les ETs.

L'analyse de l'activité transcriptionnelle des ETs peut s'avérer utile dans la recherche d'éléments actifs. Même si la transcription n'est pas suffisante pour l'accomplissement d'un cycle de transposition, elle constitue un indice fort quant à la capacité d'un élément à transposer. L'utilisation de la RT-PCR (Hirochika et al., 1996), des bases de données transcriptomiques (collections d'EST), des banques d'ADNc ont permis de mettre en évidence l'activité transcriptionnelle de plusieurs familles d'ETs qui se sont révélées par la suite transpositionnellement actives (Grandbastien, 2015).

Pour comprendre la dynamique de tous les ETs au niveau du génome entier, il faut utiliser des techniques qui permettent une quantification globale de l'activité transcriptionnelle ou transpositionnelle, comme l'utilisation de puce et le re-séquencage. A mon arrivée en 2005, le re-séquencage n'étant pas encore développé, nous avons utilisé les puces pour approfondir nos connaissances.

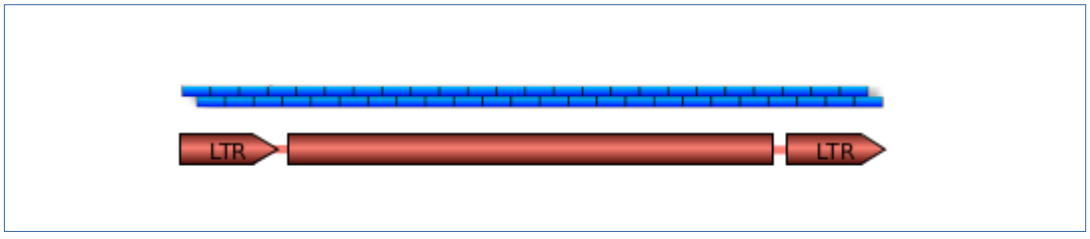



**Figure 10 : Représentation schématique du « design » de la puce version 1.0 sur un rétrotransposon à LTR.** En rouge est représenté le rétrotransposon à LTR et les blocs bleus indiquent les oligonucléotides utilisés pour la puce transposome version 1.0.

## *2.1. Mise en évidence de rétrotransposons à LTR transpositionnellement actifs chez le riz*

### *2.1.1. Développement d'une puce à façon*

A la vue de mes connaissances sur l'aspect fonctionnel, je me suis chargée de développer cette thématique au sein de l'équipe. Cet aspect a été financé par un projet ANR ITEGE (coord. M.A. Grandbastien, INRA Versailles) :"Impact of Transposable Elements on Gene Regulation" (2006-2009). Comme nous l'avons décrit, les ETs sont des facteurs centraux de la création de variation fonctionnelle. Ils peuvent modifier la fonction des gènes et engendrer des promoteurs alternatifs et des spécificités tissulaires. De plus, ils répondent à des stimuli particuliers, et cette réponse modifie à son tour l'expression de gènes adjacents. Cette nouvelle image suggère que les ETs se placent au cœur de réseaux régulateurs complexes permettant la modulation fine de l'expression génique. L'objectif du projet ITEGE fut donc de réaliser une première évaluation de la contribution des ETs à la diversité régulatrice des gènes végétaux. Il visait à établir une revue globale des associations entre gènes et ETs chez quatre plantes modèles (arabidopsis, riz, tomate et blé), et à étudier les modifications d'expression géniques qui en résultaient, ainsi que les mécanismes moléculaires utilisés. Ce projet regroupait quatre des principaux laboratoires français (INRA Versailles, INRA/URGV Evry, Université de Clermont-Ferrand, LGDP Perpignan) reconnus pour leurs compétences sur les ETs de plantes, et combinait des approches bioinformatiques et transcriptomiques. Il s'agissait de la première analyse à grande échelle de ce type chez les plantes, visant à démontrer un impact massif des ETs sur l'expression des génomes végétaux.

La première étape de ce projet consistait en la mise au point des outils nécessaire à nos analyses, et pour cela nous avons développé une puce « transposome » (technologie Nimblegen, 400 000 oligos), permettant de suivre l'expression de tous les rétrotransposons à LTR répertoriés dans le génome du riz. Ceci a été possible grâce au développement d'une base de données : Retroryza (www.retroryza.org), contenant toutes les copies des rétrotransposons à LTR du riz (Chaparro et al., 2007).

Cette approche post-génomique ne nécessite pas d'avoir un phénotype mutant associé à la transposition de l'élément et elle permet l'étude transcriptionnelle de tous les rétrotransposons à LTR. Ce type d'expérience n'ayant jamais été réalisée, nous avons choisi de faire un design des oligonucléotides correspondants à un tiling (chevauchement des oligonucléotides) des rétrotransposons à LTR, permettant ainsi une couverture complète d'au moins un élément de chaque famille (Figure 10).

Afin de valider l'utilisation de cette puce, j'ai choisi comme condition témoin la culture *in vitro*. Les cellules végétales sont capables de se dédifférencier, c'est-à-dire de revenir à un état non spécialisé. On parle de totipotence cellulaire. Il est donc possible de régénérer une plante entière à partir d'une seule cellule, en dirigeant son développement grâce aux hormones de croissance. La culture *in vitro* exploite cette caractéristique propre aux plantes. L'utilisation d'explants permet ainsi de produire tout d'abord des cals (amas de cellules indifférenciées) grâce à un milieu de culture adapté. Les cals vont ensuite se différencier en nouveaux tissus jusqu'à régénérer une plante entière.

Ces différentes étapes de la culture *in vitro* qui permettent de reprogrammer les cellules constituent un stress physiologique pour ces dernières. Les cultures de cals et les plantes régénérées sont connues pour être le lieu de réactivation transcriptionnelle et transpositionnelle de plusieurs famille d'ETs. Par exemple, les rétrotransposons à LTR *Tos17* chez le riz (Hirochika et al., 1996), *BARE-1* chez l'orge (Kalendar et al., 2000) ou encore *Tnt1* chez le tabac (Grandbastien et al., 1989) ont été caractérisés comme étant actifs dans les cultures de cals. Ces conditions permettent aussi la réactivation d'autres types d'ETs, comme le LINE *Karma* (Komatsu et al., 2003) ou le MITE *mPing* (dans les cals déviés d'explants d'anthères) chez le riz (Jiang et al., 2003). La culture *in vitro* est largement utilisée pour générer des collections de mutants d'insertion chez plusieurs plantes. Chez le riz, par exemple, *Tos17* est utilisé comme agent mutagène grâce à sa capacité à générer de nouvelles insertions durant la culture *in vitro*. Ainsi, une importante collection de lignées mutantes de riz a été générée pour des études de génomique fonctionnelle (Larmande et al., 2008; Miyao et al., 2007).

J'ai utilisé des cultures de cals de 14 semaines pour tester mon approche. Comme dans mon cas les cals ont été générés à partir d'embryons, j'ai utilisé des embryons comme témoins négatifs de l'analyse. Grâce à cette expérience d'hybridation sur la puce « transposome », j'ai identifié 13 familles de rétrotransposons à LTR transcriptionnellement actifs ainsi qu'un nouvel élément transpositionnellement actif chez le riz : *Lullaby* (Picault et al., 2009). Les résultats sont présentés dans l'article suivant publié en 2009 dans Plant Journal.

the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

*The Plant Journal* (2009) **58**, 754–765 doi: 10.1111/j.1365-313X.2009.03813.x

# Identification of an active LTR retrotransposon in rice

**Nathalie Picault[1,*], Christian Chaparro[1], Benoit Piegu[1], Willfried Stenger[1], Damien Formey[1], Cristel Llauro[1], Julie Descombin[1], Francois Sabot[1], Eric Lasserre[1], Donaldo Meynard[2], Emmanuel Guiderdoni[2] and Olivier Panaud[1]**

[1]*Laboratoire Génome et Développement des Plantes, UMR CNRS/IRD/UPVD, Université de Perpignan Via Domitia, 52 avenue Paul Alduy, 66860 Perpignan cedex, France, and*

[2]*CIRAD, Biological Systems Department, UMR DAP 1098, TAA96/03, 2477 Avenue Agropolis, 34398 Montpellier cedex 5, France*

Received 4 September 2008; revised 7 January 2009; accepted 15 January 2009; published online 19 March 2009.
*For correspondence (fax +33 468668499; e-mail nathalie.picault@univ-perp.fr).

## SUMMARY

**Transposable elements are ubiquitous components of plant genomes. When active, these mobile elements can induce changes in the genome at both the structural and functional levels. Availability of the complete genome sequence for several model plant species provides the opportunity to study TEs in plants at an unprecedented scale. In the case of rice, annotation of the genomic sequence of the variety Nipponbare has revealed that TE-related sequences form more than 25% of its genome. However, most of the elements found are inactive, either because of structural alterations or because they are the target of various silencing pathways. In this paper, we propose a new post-genomic strategy aimed at identifying active TEs. Our approach relies on transcript profiling of TE-related sequences using a tiling microarray. We applied it to a particular class of TEs, the LTR retrotransposons. A transcript profiling assay of rice calli led to identification of a new transpositionally active family, named *Lullaby*. We provide a complete structural description of this element. We also show that it has recently been active *in planta* in rice, and discuss its phylogenetic relationships with *Tos17*, the only other active LTR retrotransposon described so far in the species.**

**Keywords: transposable elements, transcript profiling, rice, LTR retrotransposon, *Lullaby, Tos17.***

## INTRODUCTION

Recent advances in plant genomics have yielded considerable resources and information on the nature and structure of the genome of several model angiosperms. These include the full genome sequence of *Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000), rice (*Oryza sativa*, International Rice Genome Sequencing Project, 2005), poplar (*Populus trichocarpa*, Tuskan *et al.*, 2006) and grapevine (*Vitis vinifera*, Jaillon *et al.*, 2007), as well as a large proportion of the genomic sequences for many other species. One of the major discoveries of the past decade in the field of plant structural genomics is that plant genomes are mostly composed of transposable elements (TEs, Feschotte *et al.*, 2002). In some species, such as wheat, these mobile elements constitute over 90% of the nuclear genome (Devos *et al.*, 2005). All the classes of TEs identified so far in eukaryotes have been found in flowering plants; these include both class I elements (LTR retrotransposons, Long Interspersed Elements (LINEs) and Short Interspersed Elements (SINEs) and class II elements (Terminal Inverted Repeat (TIR) transposons and *Helitrons*) (Wicker *et al.*, 2007). In addition, plant genomes harbour pararetrovirus-related sequences that show some homology with class I elements, particularly LTR retrotransposons (Harper *et al.*, 2002). Nevertheless, the most predominant type of TEs found in flowering plants (in terms of the percentage of genome they represent compared to other types) are LTR retrotransposons (Vitte and Bennetzen, 2006). As an example, the first analysis of the rice sequence showed that they represent 55 Mbp of genomic sequence and 42% of all TEs (International Rice Genome Sequencing Project, 2005). As a consequence, TEs have been a major focus of genomicists in recent years, with the main objectives being to better understand their impact on the structure, evolution and function of plant genomes (Bennetzen, 2007).

At the structural level, most TEs are found in inter-genic regions, although small elements, such as Miniature Inverted repeat Transposable Elements (MITEs), are often found in genes, either in UTRs or introns (Wessler *et al.*, 1995). In addition, Ma *et al.* (2007) have recently shown that LTR retrotransposons are ubiquitous in plant centromeres,

although their role in the function of these chromosomal regions is not yet established. The impact of TEs in genome evolution has been particularly well studied in the case of LTR retrotransposons (Bennetzen *et al.*, 2005). These elements transpose via a copy and paste mechanism, and therefore increase their copy number while active, causing genomic expansion. This process can reach such an extent that such elements are now considered to be the primary factor in genome size variation in plants, other than polyploidization (Piegu *et al.*, 2006). Moreover, several studies have shown that LTR retrotransposons are quickly eliminated from the genome through deletions and recombination, leading to rapid turnover of inter-genic regions (Vitte *et al.*, 2007). These studies have demonstrated the central role played by these elements in genomic differentiation in plants (Bennetzen, 2007).

Despite their propensity to invade and densely populate plant genomes, only a few transpositionally active TEs have been identified so far in plants. As an example, in rice, the only known active elements are the LTR retrotransposon *Tos17* (Hirochika *et al.*, 1996), the LINE *Karma* (Komatsu *et al.*, 2003), and the transposons *dTok* (Moon *et al.*, 2006), *nDart* (Tsugane *et al.*, 2006) and *mPing/Pong* (Jiang *et al.*, 2003). Altogether, these five TE families represent less than 100 kbp of the Nipponbare genome, which contrasts with the fact that the rice genome harbours several hundreds of TE families, totalling 250 000 copies, making up 130 Mbp (International Rice Genome Sequencing Project, 2005). Most of the TEs found in complex eukaryotic genomes are not functional, either because they are the target of silencing processes that impede their proliferation in the genome (Slotkin and Martienssen, 2007) or because they have been structurally altered by mutations (Vitte *et al.*, 2007). It is therefore expected that few elements would be active at a given time in a plant species. Moreover, the presence of more than five TE families still active in the rice genome today might be also anticipated. Many studies have shown insertion polymorphisms of several class I and class II elements among rice varieties, and have demonstrated their transpositional activity in the recent past (i.e. after rice domestication 10 000 years ago; Vitte *et al.*, 2004).

One of the major difficulties in identifying active TEs is a suitable screen for the detection of new insertions. The *dTok* and *nDart* transposons were discovered as a result of mutation of the genes into which they were inserted and the consequent change in the plant phenotype. Their identification was then made possible by positional cloning of these genes. Although there is no doubt that the genetic study of more rice mutants will lead to the discovery of new active transposons, this approach remains tedious because of the cloning procedure. In the case of *mPing*/*Pong* elements, an *in silico* survey of recent MITE insertions in the rice genome sequence was combined with confirmation of the mobility of these TEs in cell culture using the transposon display procedure (Jiang *et al.*, 2003). The efficiency of this double approach demonstrates that availability of the full genomic sequence of rice considerably facilitates the discovery of new active TEs. However, not all recently inserted TEs are still transpositionally active. This is particularly the case for rice LTR retrotransposons, as is shown here. The discovery of both *Tos17* and *Karma* elements was achieved through cloning and sequencing of cDNAs that were amplified by PCR using primers for the conserved domains of the reverse transcriptase domain. Although this method is robust and straightforward, it may not be suitable for an exhaustive survey of all the transcriptionally active LTR retrotransposons because of the bias associated with the PCR amplification.

TEs require transcription and translation of the genetic information they encode in order to be transpositionally active. In fact, class I elements transpose via their mRNA, which is synthesized by RNA polymerase II. Some of these mRNAs are processed and translated into protein complexes, the functions of which are formation of a virus-like particle (VLP) in the cytoplasm on one hand and provision of the enzymatic activities necessary to complete the transposition cycle of the element on the other. Most of the transcripts of a transpositionally active LTR retrotransposon are thus imported into the VLP where they are reverse-transcribed into DNA. The new copies of the element are then re-inserted into the nuclear genome (Kumar and Bennetzen, 1999).

It is therefore expected that all transpositionally active LTR retrotransposons are initially present in the form of transcripts in the tissues in which they are activated. Consequently, their transposition should yield new copies of the original element. The strategy that we followed to identify new active LTR retrotransposons in rice consists of two steps. First we surveyed the transcriptional activity of a large repertoire of rice LTR retrotransposons in rice calli (which are known to induce transposition of the LTR retrotransposon *Tos17*) using a microarray-based hybridization procedure. We then searched the genome of rice lines derived from callus culture for new copies of transcriptionally activated retrotransposons. This strategy allowed us to identify *Lullaby*, a new active LTR retrotransposon in rice, for which we provide a full characterization.

## RESULTS

### Development of the rice LTR retrotransposon array

We developed a microarray to study the transcriptional activity of rice LTR retrotransposons on a genome-wide scale. The first step of our work consisted of gathering a comprehensive dataset of the rice LTR retrotransposon sequences, which was based on the retrOryza database (http://www.retroryza.org) that was developed by our group

(Chaparro *et al.*, 2007). This database contains over 300 families of LTR retrotransposons, the majority of which are not included in either the Rice-repeat or Repbase databases because they correspond to single- or low-copy elements. These sequence data were directly used for construction of the array. We anticipated that the repeat nature of LTR retrotransposons could complicate analysis of the transcript profiling assays, mainly because a given hybridization hit corresponds to several genomic loci. We therefore chose to represent the nucleotidic diversity of the repeated families on the array where possible by increasing the number of oligomers specific to various members of these families. This is the case for the reference elements *Osr42*/*RN156*, *Osr8*/*copia1_os* and *Osr34*/*retrosat2*, which belong to the same family but differ slightly in their nucleotidic composition.

Finally, the array that we used in this study consisted of 113 747 oligomers providing a tiling of a total of 305 LTR retrotransposon families. Altogether, these oligomers represent 24 982 loci and 52 Mbp (12%) of the rice genome (Table 1). These values are similar to the estimation of the rice LTR retrotransposon content given elsewhere (International Rice Genome Sequencing Project, 2005).

### Transcript profiling of rice LTR retrotransposons in callus

Using the microarray described above, we conducted genome-wide analysis of the transcriptional activity of rice LTR retrotransposons in calli, using imbibed embryos as a control. First, we examined the transcript profile of the LTR retrotransposon *Tos17*, because this element is known to be transcriptionally activated and is therefore considered in our assay as a positive control. *Tos17* is represented by a tiling of 138 oligomers. Figure 1(a) shows that all probes except three show a significant increase in transcription of *Tos17* in calli. It should however be noted that the intensity of the hybridization signals is not homogenous throughout the whole element. A similar observation was made for the transcript profile of the actin gene (Figure 1b). However, we performed semi-quantitative RT-PCR, which confirmed the expression level of these two loci (Figure 2). Therefore, these results show that our tiling array can be successfully used to detect over-expression of active LTR retrotransposons. The technical aspects of the use of our tiling array for transcript profiling of TEs are discussed in detail in Appendix S1.

We examined the transcript profiles of the other rice LTR retrotransposon families that are represented on the array. The 290 families showed significant differential expression between embryo and callus for at least one oligomer. In most cases, we observed heterogeneity in the level of the hybridization signals throughout the tiling of the elements, as in the case of *Tos17* and the actin gene. Moreover, for several elements, only a few oligomers showed a significant differential expression while the remaining probes did not (Figure 1c). We therefore determined a threshold of detection of differential expression in our screen for the entire set of LTR retrotransposon families. In order to be considered differentially expressed in callus, 30% of the oligomers tiling the entire sequence should have expression above the background (based on normalization of the biological replicates), and, among these, 30% should be either under- or over-expressed (see Appendix S1 for a detailed explanation of the method). Using these criteria, 15 and 13 families were found to be under- or over-expressed, respectively (Table 2).

**Table 1** Number of elements and size of genomic sequence represented by each superfamily

| Superfamily | Number of loci | Genomic representation (bp) |
|---|---|---|
| *Copia* | 5335 | 12 238 915 |
| *Gypsy* | 11 201 | 28 864 280 |
| Large Retrotransposon Derivative (LARD) | 801 | 1 879 202 |
| Unclassified[a] | 7645 | 9 471 564 |
| Total LTR-retrotransposons | 801 | 52 453 961 |

[a]LTR retrotransposons for which no reverse transcriptase could be identified were considered as unclassified.
For each superfamily of LTR retrotransposons, the number of loci, including complete, truncated and solo LTR, are shown, as well as the genomic size represented by each superfamily.

### Experimental confirmation of the transcript profiling data

The 13 families that were considered over-expressed in calli in our transcript profiling experiment were further tested using semi-quantitative RT-PCR experiments. The actin gene was used as a control. We also tested differential expression of the LTR retrotransposon *Houba* that was shown to be under-expressed using our criteria (Figures 1d and 2) and two LTR retrotransposons (*RN265* and *RN552*) that showed the same expression under the two conditions. In all cases, the RT-PCR experiments confirmed the tiling array hybridization results (Figure 2). This shows that the threshold parameters that we chose were correct, and validates use of the array for a genome-wide study of the transcriptional activity of rice LTR retrotransposons.

We anticipated that, in the case of repeated TE families, several paralogues may be transcriptionally activated in calli. As a consequence, the transcriptome of this family will consist of many different mRNA molecules showing nucleotidic diversity that could lead to heterogeneity of the hybridization signal. In order to test this hypothesis, we cloned and sequenced the RT-PCR products (about 20 per element) obtained for validation of the array hybridization (see above). The sequences were then mapped *in silico*

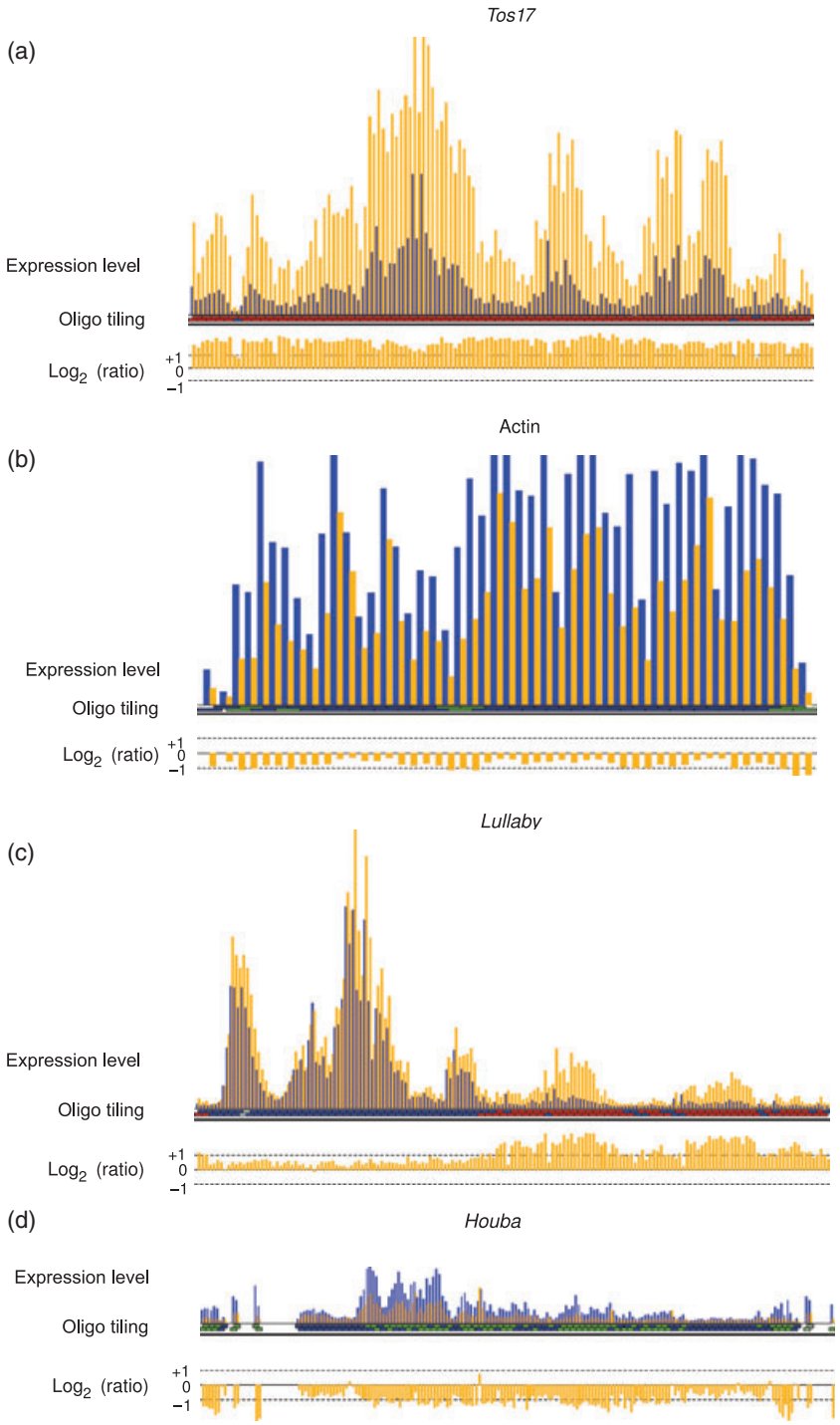



**Figure 1.** Graphical representation of the hybridization signals and their relative expression for selected retro-element families.
(a) *Tos17,* (c) *Lullaby*, (d) *Houba* and control gene, (b) actin gene.
In each case, the upper part of the figure shows the position of each oligomer in the tiling, and above each oligomer two bars corresponding to the expression levels in embryos and calli. Oligomers that did not show differential expression between the two conditions are shown in blue, those that were over-expressed are shown in red, and those that were under-expressed are shown in green. When the oligomers are not unique, they are shown in light blue. The expression level for each oligomer is shown as a vertical bar for each condition. Blue bars represents the expression level for the embryo and yellow bars correspond to calli. The lower part of each figure represents the $\log_2$ ratio for each oligomer, with the +1 and −1 limits that show a twofold increase or decrease in expression when comparing embryos and calli.

onto the rice pseudomolecules (International Rice Genome Sequencing Project, 2005) in order to identify the transcriptionally active copies for each family (Table 3). For two elements, *RN16* and *RN364*, seven and three different transcripts were isolated in calli, respectively. However, for the remaining 11 families, only one copy was found to be transcriptionally active. This shows that, for most of the LTR retrotransposon families, only a few copies (if not only one) are transcriptionally active in the genome of rice. Moreover, *in silico* mapping showed that these active copies are evenly spread across the 12 rice chromosomes. Among the 21 transcripts that were examined, three were located near the centromere, at less than 0.5 Mb. In addition, the 13 families found to be transcriptionally active in calli were present in only a few copies in the rice genome. This result clearly suggests that the repeated

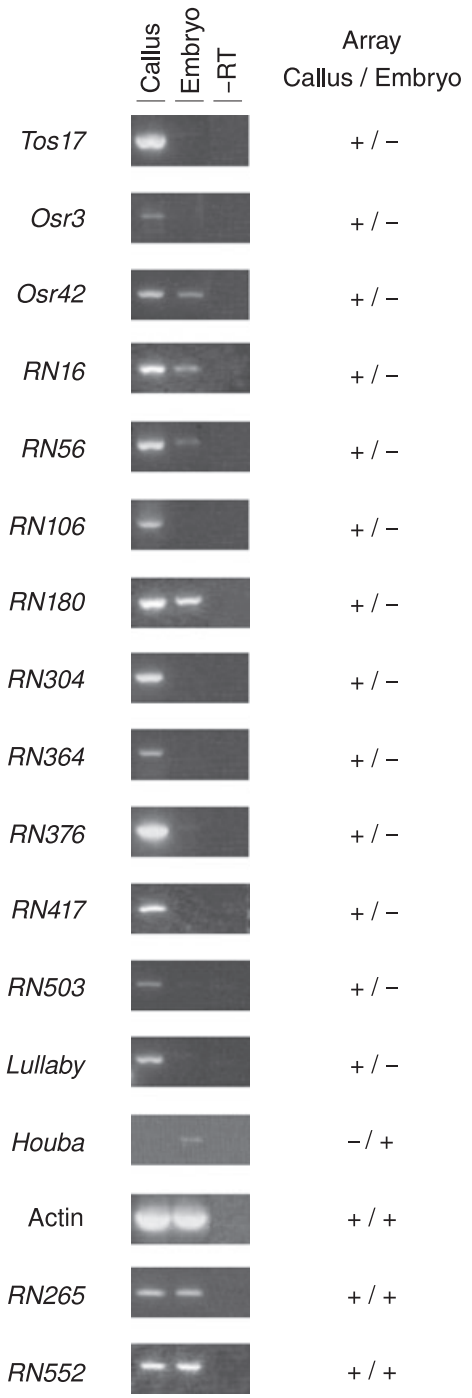



**Figure 2.** Transcription levels of LTR retrotransposons were examined by semi-quantitative RT-PCR in callus and embryo.
'-RT' indicates negative control experiments in which reverse transcriptase was not added before the final PCR amplification step. The actin gene was used as positive control. Each RT-PCR experiment was repeated three times on three batches of independent tissues, and identical results were obtained.

families are either not trancriptionally activated in calli, or that they are silenced by a copy number-dependent mechanism.

**Table 2** Number of over- and under-expressed families classified by superfamily

| | Number of families on the microarray | Number of families differentially expressed | |
|---|---|---|---|
| | | Under | Over |
| *Copia* | 123 | 4 | 6 |
| *Gypsy* | 119 | 8 | 5 |
| LARDs | 3 | 2 | 0 |
| Unclassified[a] | 60 | 1 | 2 |
| Total | 305 | 15 | 13 |

[a]LTR retrotransposons for which no reverse transcriptase could be identified were considered as unclassified.
The number of families in each superfamily are shown, as well as LARDs and those unclassified as either the *Copia* or *Gypsy* superfamilies. For each classification, the number of families showing over- or under-expression when comparing embryos and calli is shown.

### Identification of a new transpositionally active LTR retrotransposon

The low copy number of the 13 LTR retrotransposon families found to be over-expressed in calli enabled us to test their transpositional activity by Southern hybridization. This was achieved by testing several lines derived from cultured calli (Figure 3). Among 62 lines tested, 41 showed the presence of a neocopy of the retrotransposon *Tos17*. This was expected as rice callus culture is used to generate insertional mutant lines carrying transposed copies of this retrotransposon family. In addition, when the other 12 LTR retrotransposon families were tested using the same material, *Lullaby* also showed the presence of additional bands in the Southern hybridization for two lines (23.4 and 35.1, Figure 4). This suggests that this family also transposed in the callus from which these lines were regenerated, as in the case of *Tos17*. We confirmed insertion of a new copy of *Lullaby* in both lines by inverse PCR (iPCR) and retrotransposon-based insertion polymorphism assay (Figure 4). Line 23.4 carries a neocopy of *Lullaby* on chromosome 9 (position 20 648 712), and line 35.1 carries a neocopy of *Lullaby* on chromosome 2 (position 32 078 848). This element is the second active LTR retrotransposon so far described in rice.

The Nipponbare genome harbours two native copies of *Lullaby* (on chromosomes 6 and 9, positions 23 681 894–23 687 035 and 13 521 300–13 524 921, respectively). We fully characterized the newly inserted elements for both lines by cloning and sequencing (Appendix S2). *Lullaby* is 5142 bp and has 125 bp LTRs. The two insertions generated a 5 bp target site duplication. In both cases, the transposed element is identical in sequence to the native copy from chromosome 6, except for three and four Single Nucleotide Polymorphisms (SNPs) for the neocopies of chromosomes 2 and 9, respectively. These results

**Table 3** Description of the 13 LTR retrotransposon families that are transcriptionally active in calli, showing the number of complete copies, truncated copies and solo LTRs found in the complete genome

| Element name | Family | Size of the full element (bp) | Size of the LTR (bp) | Number of complete copies | Number of truncated copies | Number of solo LTRs | Number of different transcripts (by RT-PCR) in calli | Localization of transcriptionally active copies in calli |
|---|---|---|---|---|---|---|---|---|
| *Tos17* | *Copia* | 3986 | 138 | 2 | 0 | 0 | 1 | Chr 7 (27 354 953–27 359 066) |
| *Osr3* | *Copia* | 5200 | 146 | 2 | 6 | 3 | 1 | Chr 3 (15 281 287–15 286 487) |
| *osr42* | *Gypsy* | 5605 | 358 | 1 | 1 | 0 | 1 | Chr 4 (7 922 608–7 928 212) |
| *RN16* | *Copia* | 5144 | 118 | 5 | 1 | 0 | 1 | Chr 7 (5 227 105–5 232 635) |
| *RN56* | Unclassified | 6704 | 289 | 2 | 19 | 3 | 7 | Chr 8 (22 188 461), Chr 10 (9 465 101), Chr 7 (17 367 865),Chr 6 (9 898 000), Chr 2 (25 269 946), Chr 6 (28 198 863), Chr 7 (11 609 306) |
| *RN106* | *Gypsy* | 8967 | 224 | 1 | 7 | 2 | 1 | Chr 9 (5 762 526–5 771 510) |
| *RN180* | *Gypsy* | 13842 | 216 | 1 | 3 | 0 | 1 | Chr 4 (16 880 687–16 891 179) |
| *RN304* | *Copia* | 4137 | 133 | 1 | 0 | 0 | 1 | Chr 7 (19 474 119–19 478 300) |
| *RN364* | *Gypsy* | 10566 | 1138 | 8 | 16 | 25 | 3 | Chr 1 (16 545 753–16 547 872), Chr 7 (16 308 810–16 309 856), Chr 8 (13 577 638–13 588 203) |
| *RN376* | Unclassified | 4969 | 117 | 1 | 1 | 0 | 1 | Chr 8 (21 614 190–21 619 285) |
| *RN417* | *Gypsy* | 7710 | 1051 | 2 | 2 | 3 | 1 | Chr 5 (10 263 360–10 271 341) |
| *RN503* | *Copia* | 5187 | 406 | 2 | 3 | 3 | 1 | Chr 11 (3 440 583–3 435 397) |
| *Lullaby* | *Copia* | 5142 | 125 | 2 | 1 | 0 | 1 | Chr 6 (23 681 894–23 687 035) |

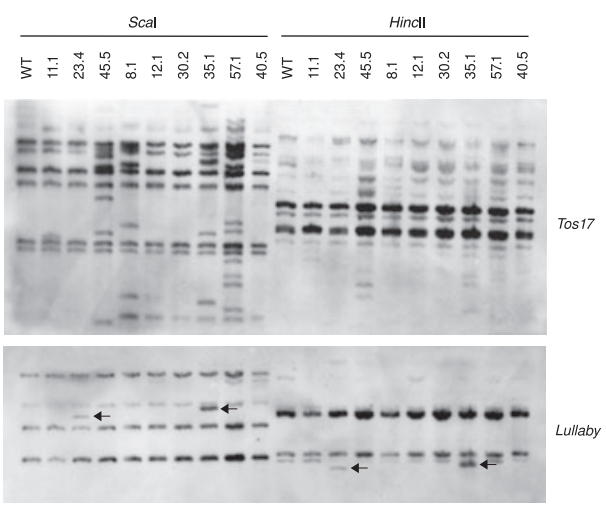



**Figure 3.** Transposition analyses of *Tos17* and *Lullaby* in regenerated plants by Southern hybridizations.
Total genomic DNA was digested by *Sca*I or *Hin*cII.

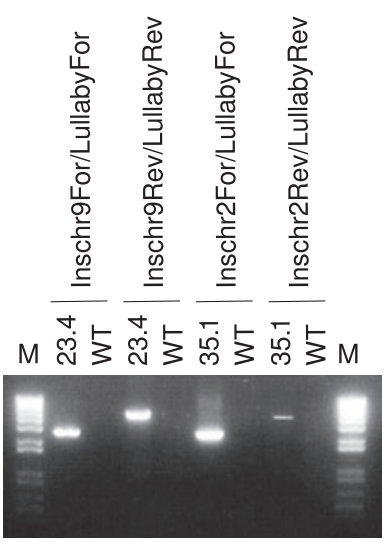



**Figure 4.** Validation of the insertion of new copies of *Lullaby* in two lines by a retrotransposon-based insertion polymorphism assay.
The primers Inschr9For/Inschr9Rev and Inschr2For/Inschr2Rev were localized on chromosomes 9 and 2, respectively. Primers specific to *Lullaby* were used.

suggest that, as in the case of *Tos17*, only one of the two copies of *Lullaby* is transpositionally active, which is in accordance with the transcript profiling analyses. As shown on Figure 5, *Lullaby* harbours all the structural features of a complete and autonomous *Copia* element, such as a Primer Binding Site (PBS), a double ORF (showing all *gag* and *pol* domains: AntiGen (GAG), Zn finger, Intergrase (INT), Reverse Transcriptase RnaseH (RT-RH)), and a PolyPurine Tract (PPT).

### Activity of the retrotransposon *Lullaby* in cultivated rice

We surveyed the presence of *Lullaby* in several rice varieties by Southern hybridization using *Lullaby* as a probe. Figure 6 shows that this element is present in all accessions, which suggests that it is present in the genome of the ancestor of the cultivated rice species *O. sativa*. Moreover, there is a clear

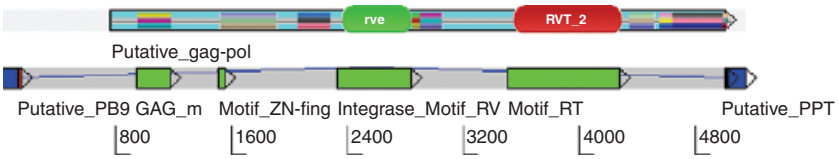



**Figure 5.** *Lullaby* structure and functional annotation.
The structure of the complete element is shown. The LTRs are shown in dark blue, the PBS and PPT in red, and the structural motifs in green (extrapolated from Pfam and ProSite analyses), and the coding DNA sequence (CDS) in light blue. The Pfam analysis for the CDS is shown; *rve* and *RVT_2* (plain blocks) are Pfam-A motifs, and multi-coloured blocks are Pfam-B motifs. The *rve* motif is the *Copia* integrase and *RVT_2* is the *Copia* reverse transcriptase.

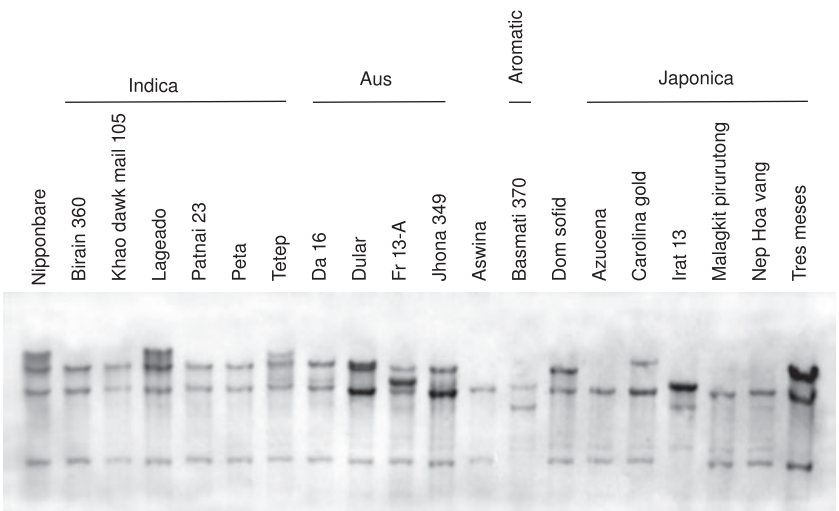



**Figure 6.** Activity of the retrotransposon *Lullaby* in cultivated rice analysed by Southern hybridization.
These accessions represent a sub-sample of a representative collection of the diversity, consisting of 246 accessions established to best represent *Oryza sativa* species based on various parameters such as geographic origin, type of culture and position in the isozyme classification (Glaszmann *et al.*, 1996). Total genomic DNA was isolated from leaf tissue using the MATAB method (Murray and Thompson, 1980, modified by using MATAB in place of CTAB) and digested with *Eco*RI.

polymorphism of the hybridization patterns among the various accessions tested. This is a strong indication that *Lullaby* could have been active *in planta* in the recent past, although the exact process of its activation remains unknown.

### Phylogenetic relationships among LTR retrotransposons with respect to their expression in callus

We examined the phylogenetic relationships among the LTR retrotransposons found to be either over- or under-expressed in calli. Figure 7 shows two phenetic trees corresponding to either *Copia*- or *Gypsy*-type elements. These trees were constructed using the Reverse Transcriptase (RT) domain sequence of the elements mined from the rice pseudomolecules. Only 74 *Copia*- and 89 *Gypsy*-type families are represented on Figure 7, although 123 *Copia*- and 119 *Gypsy*- type families were included on the array. This is due to the fact that several families (50 *Copia*-type and 30 *Gypsy*-type) do not harbour any RT domain and could therefore not be included in the multiple sequence alignments that were used for phylogenetic reconstruction. As observed from the two phenetic trees, both over- and under-expressed elements are found in all phenetic groups, suggesting that there is no clear correlation between their transcription level and their phylogenetic relatedness. Nevertheless, *Tos17* and *Lullaby*, which are both transpositionally active, belong to the same cluster, together with *RN304*, which is also over-expressed in calli.

## DISCUSSION

The microarray-based strategy that we developed enabled us to conduct genome-wide transcript profiling of rice LTR retrotransposons. Among the 305 families that are represented on the array, our results show that 28 families are differentially expressed in calli compared to imbibed embryos. These include *Tos17*, an element that was known to be over-expressed in these tissues. Li *et al.* (2006) used a similar microarray platform as the one used in the present study in order to perform transcript profiling of the rice genome. The only difference was that the tiling array that they constructed concerned only the non-repeated fraction of the rice genome. Moreover, they did not provide any data on the variation in the level of hybridization signals between the oligomers that constitute the tiling of a single locus, as was observed in our experiments. Therefore, we cannot directly compare the results from both studies. Nevertheless, the present study confirms that tiling arrays can be successfully used in rice for genome-wide transcript profiling experiments, even in the case of repeated sequences.

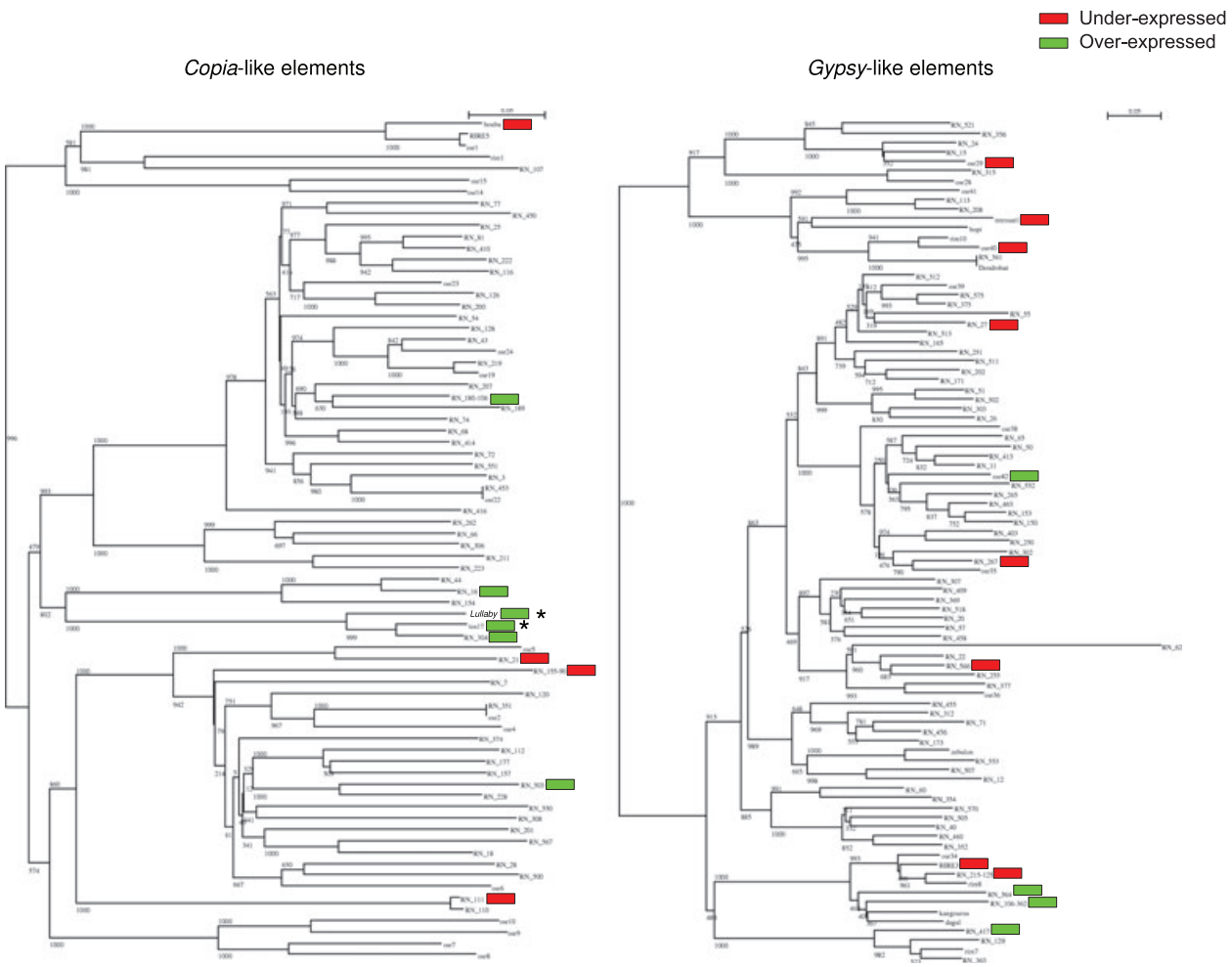



**Figure 7.** Phylogenetic relationships among rice LTR retrotransposons.
Both phenetic trees were constructed using the protein sequence of the RT domain of the polyprotein (POL) regions of the reference copy of each family (see Experimental procedures). Green and red rectangles indicate the elements that were over- and under-expressed, respectively, in the transcript profiling experiment. Bootstrap values (based on 1000 replicates) are given on the branches of the tree. The asterisks indicate the two transpositionally active LTR retrotransposons in rice (*Tos17* and *Lullaby*).

Our results show that 13 LTR retrotransposon families are over-expressed in rice calli. These 13 families were further tested for their transpositional activity by Southern analyses of several lines regenerated from these cultured tissues. One family, named *Lullaby*, was found to be transpositionally active. Overall, the strategy that we used to detect active LTR retrotransposons in rice was successful, and could be applied to other physiological states. In particular, it could ultimately lead to establishment of a transpositional landscape of the rice LTR retrotransposons in response to various stresses (either biotic or abiotic).

The recent development of high-throughput sequencing technologies, such as the Illumina (http://illumina.com) or GS FLX (http://www.454.com) platforms provides a new opportunity to conduct highly efficient studies such as genome re-sequencing, metagenomics or transcript profiling experiments. Recently, the FLX technology was used for large-scale characterization of transposon flanking sequence tags (FSTs) in *Petunia* (Vandenbussche *et al.*, 2008). These new cost-efficient strategies could be applied to study the TE transcriptome in plants as an alternative to the use of array-based approaches such as that used in the present study. So far, the main limitation of these techniques is the small size of the sequence reads (i.e. 36 bases for the Illumina technology). This is not a problem for establishing the expression profile of genes, because each read can unambiguously be assigned to a single locus. However, in the case of repeated TE families, this could impede identification and localization of the differentially expressed loci. Our microarray-based strategy relies upon the Nimblegen technology (http://www.nimblegen.com) which uses a tiling of several 60-mer oligoprobes and can circumvent this problem. The newest version of the FLX technology generates longer reads, and could therefore allow more efficient

identification of active copies of a given family, and should therefore be tested for transcript profiling of repeated TEs.

Seed embryo-derived callus culture has been used extensively to generate *Tos17*, T-DNA, Ac/Ds or En/Spm insertion libraries for functional genomics studies in rice. Several mutant collections are available worldwide (Hirochika *et al.*, 2004), totalling more than 560 000 lines. Until now, the only known transpositionally active LTR retrotransposon was *Tos17.* In order to exploit these lines for reverse genetics studies, FSTs of *Tos17* insertions have been produced from cell culture-derived (Miyao *et al.*, 2003) and T-DNA (Piffanelli *et al.*, 2007) plants. Our results show that at least a second LTR retrotransposon family is active in callus. Some lines of these collections may harbour mutations caused by the insertion of *Lullaby* for which corresponding FSTs should be obtained in order to further exploit this material. Moreover, the mutant lines have been phenotyped for forward-genetics studies. Only a small proportion (5–10%) of the mutant phenotypes have been successfully tagged by *Tos17* (Hirochika, 2001). Analysis of *Lullaby* insertions in these lines may help to increase this proportion.

Only 13 out of the 305 LTR retrotransposon families that were analysed for their transcriptional activity (i.e. 4.3%) are over-expressed in calli. It has been demonstrated that activation of LTR retrotransposons can be triggered by both biotic and abiotic stresses (such as *in vitro* culture) in plants (Lucas *et al.*, 1995; Wessler, 1996; Hirochika, 1997; Grandbastien, 1998; Liu and Wendel, 2000; Kashkush *et al.*, 2003). These elements harbour regulatory regions in their promoters that enable their activation under specific conditions. For example, activation of the tobacco *Tnt1* element is regulated by *cis*-acting elements that are similar to motifs found in defence genes (Grandbastien *et al.*, 2005). The low percentage of over-expressed LTR retrotransposons that we observed in calli suggests that only a few families may harbour such regulatory signals for their transcriptional activation in calli. We analysed the phylogenetic relationships among 11 of the 13 over-expressed elements (Figure 7). For the remaining two elements, this was not possible because of the absence of an RT domain. No clear clustering of the five *Gypsy*-like elements was observed, suggesting that they have no common origin that could explain that they share any common structural features such as the presence of regulatory signals involved in transcriptional response. In the case of the sixover-expressed *Copia*-like elements, three (*Tos17*, *Lullaby* and *RN304*) form a distinct cluster. Interestingly, this cluster contains the only two transpositionally active rice LTR retrotransposons. We therefore looked for the presence of common structural features in the LTR region of these elements. Several sequences with putative regulatory function were identified, but none of them were conserved in the whole set of retrotransposons that are transcriptionally activated during callus culture. This indicates that this activation can be achieved through distinct mechanistic pathways that remain to be elucidated. However, one element, a tandem repeat of seven base pairs (Figure 8), was found to be highly conserved in the 5′ half of the LTR of *Lullaby*, *Tos17* and *RN304*. Moreover, this element was found to be present in the LTR of another retrotransposon that is activated in calli, *Osr3*, and in the LTR of *RN73*, *RN154* and *RN172*, for which few oligomers showed differential expression between embryo and callus. Therefore, this repeat, which is not present in the LTR of previously characterized plant retrotransposons, is an

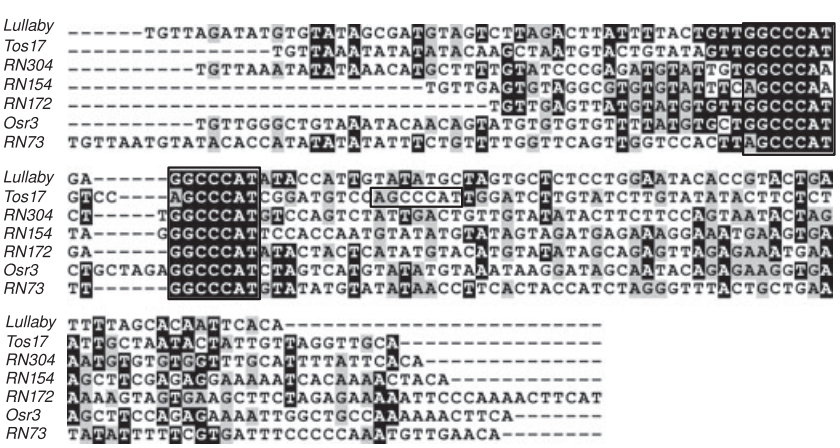



**Figure 8.** Identification of a conserved tandem repeat in the LTR of several retrotransposons transcriptionally activated in calli. Conserved motifs were identified by MEME (Multiple Em for Motif Elicitation) on a training set composed of the LTR from *Lullaby*, *Tos17* and *RN304*. LTRs containing this motif were identified using MAST (Motif Alignment and Search Tool) on the whole set of LTRs from the retrOryza database. An HMM (Hidden Markov Model) profile was built based on a multiple alignment of this motif using HMMBUILD of the EMBOSS suite. The six LTRs that contained this motif were then aligned to this profile using HMMALIGN from the EMBOSS suite. The alignment was printed out using BOXSHADE. Black and grey shading indicate identical and similar residues, respectively. The conserved element is boxed. MEME and MAST programs can be found at http://meme.sdsc.edu/meme/, HMMBUILD, HMMALIGN and BOXSHADE at http://mobyle.pasteur.fr/. Default parameters were used for each program. Note that the LTR of *Tos17* has an additional copy of the repeat.

interesting candidate for a *cis*-element that could account for the transcriptional activation in calli.

In addition, the 305 families that were included in our analysis differ in terms of their copy number in the rice genome, e.g. one copy for *RN304* to several hundred copies for *Hopi* or *Houba*. Our results show that the over-expressed elements are have a low-copy number, while the highly repeated ones are either down-regulated (e.g. *Houba*) or not differentially expressed in rice calli. Several mechanisms, in two major pathways, can control the activity of TEs in plants (Feschotte *et al.*, 2002). These are methylation of both DNA and histones, as well as post-transcriptional silencing through the siRNA pathway (Slotkin and Martienssen, 2007). In addition, RNA-directed DNA methylation (RdDM), a process that leads to chromatin modification through siRNA, is also dependent on the number of copies of a given element and the level of its transcription (Perez-Hormaeche *et al.*, 2008). Our results suggest that the highly repeated rice LTR retrotransposons may be the target of such silencing pathways, in contrast with the 13 low-copy families that are over-expressed in calli.

The array-based strategy that we used in the present study was successful in identifying a second active LTR retrotransposon in rice. This approach can be used for other types of rice TEs, as expertly curated databases of these elements are available and may be used to construct a suitable tiling array. Moreover, use of such a tool under various conditions would allow establishment of both the transcriptional and transpositional landscapes of rice with respect to developmental stages and various stress responses. Finally, this type of transcript profiling data can be exploited to study the functional impact of TEs in rice, by integrating the data with those for rice genes under similar experimental conditions.

## EXPERIMENTAL PROCEDURES

### Microarray design

We designed an isothermal 60-mer custom microarray, which was synthesized at the Nimblegen expression facility (Madison, WI, USA). We used a tiling design in which probes were chosen at intervals of approximately 30 nucleotides. Oligomers were designed with GC content between 40 and 60%, and melting temperature ($T_m$) between 80 and 110°C. Another restriction was the number of cycles needed to synthesize the oligomers, which was limited to 180 cycles for a 60-mer design. When adding a new batch of oligomers, homology searches were performed using megaBLAST (Zhang *et al.*, 2000, http://www.ncbi.nlm.nih.gov) against the selected oligomers, and those that had fewer than four bases of difference with any of the oligomers already selected were eliminated and the selected oligomer was marked as 'shared'. The consensus sequence for the 305 LTR retro-element families is represented by 113 747 probes. In addition, 40 000 random probes were designed by Nimblegen as internal controls.

### Microarray data analysis

We used the SVN package that is included in the Bioconductor suite of the R statistical software (http://www.R-project.org) to normalize the raw data of the four hybridizations. We used the probes homologous to genes for normalization between the various treatments, and applied the normalization parameters to the entire array. This approach eliminates deviation due to the non-independent nature of the retro-element probes, as well as the fact that these retro-elements represent families of elements for which it is not known beforehand whether they account for the expression of one or several molecules. Normalization was confirmed by determining the expression rates of housekeeping genes such as ubiquitin and actin. We used the random 60-mers to calculate a background cut-off value for each array. Outliers of these random 60-mers were eliminated by considering only the probes that fell into the range of the mean of the intensity distribution $\pm$ one standard deviation. The background cut-off value was defined as the mean of the filtered random probes plus twice the standard deviation.

We developed an in-house software solution using the scripting language Tcl/Tk to represent the results of the tiling, in which the position of the oligomers and their raw expression values as well as the $\log_2$ ratio are represented.

### Plant materials

*Oryza sativa* ssp. *Japonica* cv. Nipponbare was used for all experiments. De-hulled seeds were imbibed in distilled water overnight, and embryos were isolated from the endosperm and deep frozen in liquid nitrogen. Regenerated rice plants were obtained from calli induced from seed embryo as previously described (Sallaud *et al.*, 2003). Briefly, scutellum primary callus yielded secondary embryogenic units 5 weeks after seed inoculation onto the callus induction medium. The units were sub-cultured twice for a four-week period on a callus maintenance medium before transfer to a maturation medium for 1 week, and then to regeneration conditions for 9 weeks, totalling 14 weeks culture under a de-differentiated state.

### RNA isolation, cDNA synthesis and RT-PCR analysis

Total RNA was isolated from calli that had been cultured for 14 weeks, and embryos were imbibed for 24 h using the TRIzol method (Invitrogen, http://www.invitrogen.com/). RNAs were purified using an RNeasy plant mini kit (Qiagen, http://www.qiagen.com/) according to the manufacturer's recommendations. RNAs were reverse-transcribed using a Superscript double-stranded cDNA synthesis kit (Invitrogen) using the recommendations of Nimblegen. For RT-PCR, amplification of cDNAs was performed for 30 cycles at an annealing temperature of 58°C. Products were separated by 1% agarose gel electrophoresis. The primers used for RT-PCR are listed in Appendix S3.

### Southern blot analysis

Blots were prepared using 2 μg of total genomic DNA digested with *Sca*I or *Hin*cII and transferred onto Hybond-$N^+$ membrane (http://www5.gelifesciences.com). A non-radioactive procedure and detection signal (Dig system, Roche, http://www.roche.com) was used in accordance with the manufacturer's protocol. Stringency washes were performed at 65°C in 0.5 x SSC. Probes were amplified by PCR using the following primers: for *Tos17*, 5′-CTAATGTACTG-TATAGTTGGCCC-3′ and 5′-AACCAAAGATCACTAGCAACTC-3′; for

*Lullaby*, 5′-CACAATTCACATGGTATCAGAG-3′ and 5′-GTGACTG-TAGACGAGTTTCCT-3′.

### Inverse PCR experiment

Total genomic DNA (1 μg) for the two rice lines 23.4 and 35.1 was digested using *Hin*fI, then precipitated by adding 1/10th of the volume of NaOAc (3 M, pH 5.2) and two volumes of 100% EtOH. After resuspension in distilled water, the DNA was self-ligated using 1 unit of T4 ligase (Promega, http://www.promega.com/). The following primer combinations were used for the PCR amplification and cloning of the flanking insertions of *Lullaby*: 5′-CAC-AATTCACATGGTATCAGAG-3′ and 5′-AAATCAGTACGGTGTATT-CC-3′ for line 35.1, and 5′-CCCTTCAAACTCAAGGTGGA-3′ and 5′-TGTTAGATATGTGTATAGCGATG-3′ for line 23.4. Standard PCR conditions were used for the amplification, i.e. 35 cycles of 1 min at 94°C, 1 min at 55°C and 90 sec at 72°C, using 0.5 units of Taq polymerase (Promega). PCR products were cloned in a PGEM-T Easy vector (Promega) according the manufacturer's instructions. Each clone was sequenced using the T7 and SP6 primers in an ABI377 sequencer (Applied Biosystems, http://www.appliedbiosystems.com/).

### Retrotransposon-based insertion polymorphism assay

We tested for the presence/absence of *Lullaby* insertions in lines 35.1, 23.4 and Nipponbare. PCR conditions were identical to those given by Vitte *et al.* (2004). We used the following primers: Inschr9For (5′-GCTCTGCATATTTCTGATTAGCTC-3′) and Inschr9Rev (5′-CAGTGTCCTCTGTCAAACGA-3′); Inschr2For (5′-TTACATGTTC-TGCTGATGCA-3′) and Inschr2Rev (5′-ACATCGTTTCCAAAGTACCC-3′); *Lullaby* forward (5′-CGACAACTTCTTTCTTCCGA-3′) and *Lullaby* reverse (5′-ATCTGGTCTGGTTATAGTGAG-3′).

### Reconstruction of complete sequence of Lullaby

Transgenic lines 35.1 and 23.4 were sequenced using primers for *Lullaby*, and the sequences obtained were cleaned using Phred-Phrap (Ewing *et al.*, 1998; Ewing and Green, 1998; http://www.phrap.org) and assembled using Consed (Gordon *et al.*, 1998) software.

### Structural annotations

Annotation was performed using Artemis software (Rutherford *et al.*, 2000), and ORFs longer than 100 residues were automatically extracted from the element sequences. The ORFs were then scanned online using a combination of Pfam, ProSite and BLASTp analyses, using standard parameters. The results were then uploaded in Artemis, in order to reconstruct the complete structure of each element. LTRs were identified using Dotter (Sonnhammer and Durbin, 1995), and the PBS and Polypurine tract (PPT) were manually determined.

### Phylogenetic analyses

The amino acid sequence of the RT domain of the LTR retrotransposons used for the transcript profiling experiment was extrapolated from the reference copies deposited in the retrOryza database. Sequences were aligned using CLUSTALX and modified by hand using Seaview software (Galtier *et al.*, 1996). A neighbour-joining tree was constructed using 1000 bootstrap replicates.

## ACKNOWLEDGEMENTS

This project was supported by Agence National de la Recherche grant number ANR-05-BLAN-0244-03). We thank E. Galinato and S.A. Jackson (Department of Agronomy, Perdue University) for their useful comments on the manuscript.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

**Appendix S1.** Technical aspects of the microarray hybridization and exploitation of transcript profiling data.

**Appendix S2.** Sequence of neocopies of *Lullaby* on chromosomes 2 and 9.

**Appendix S3.** List of primers used for RT-PCR.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

## REFERENCES

**Arabidopsis Genome Initiative** (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.

**Bennetzen, J.L.** (2007) Patterns in grass genome evolution. *Curr. Opin. Plant Biol.* **10**, 176–181.

**Bennetzen, J.L., Ma, J. and Devos, K.M.** (2005) Mechanisms of recent genome size variation in flowering plants. *Ann. Bot.* **95**, 127–132.

**Chaparro, C., Guyot, R., Zuccolo, A., Piegu, B. and Panaud, O.** (2007) RetrOryza: a database of the rice LTR-retrotransposons. *Nucleic Acids Res.* **35**, D66–D77.

**Devos, K.M., Ma, J., Pontaroli, A.C., Pratt, L.H. and Bennetzen, J.L.** (2005) Analysis and mapping of randomly chosen bacterial artificial chromosome clones from hexaploid bread wheat. *Proc. Natl Acad. Sci. USA*, **102**, 19243–19248.

**Ewing, B. and Green, P.** (1998) Base-calling of automated sequencer traces using phred II. Error probabilities. *Genome Res.* **8**, 186–194.

**Ewing, B., Hillier, L., Wendl, M.C. and Green, P.** (1998) Base-calling of automated sequencer traces using phred I. Accuracy assessment. *Genome Res.* **8**, 175–185.

**Feschotte, C., Jiang, N. and Wessler, S.R.** (2002) Plant transposable elements: where genetics meets genomics. *Nat. Rev. Genet.* **3**, 329–341.

**Galtier, N., Gouy, M. and Gautier, C.** (1996) SEAVIEW and PHYLO_WIN: two graphics tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci.*, **12**, 543–548.

**Glaszmann, J.C., Mew, T., Hibino, H., Kim, C.K., Vergel de Dios-Mew, T.J., Vera Cruz, C.M., Notteghem, J.L. and Bonman, J.M.** (1996) Molecular variation as a diverse source of disease resistance in cultivated rice. In *Rice Genetics III. Proceedings of the Third International Rice Genetics Symposium, Manila, Philippines* (Kush, G.S., ed.) Los Baños, Laguna, The Philippines: International Rice Research Institute, pp. 460–465.

**Gordon, D., Abajian, C. and Green, P.** (1998) Consed: a graphical tool for sequence finishing. *Genome Res.* **8**, 195–202.

**Grandbastien, M.A.** (1998) Activation of plant retrotransposons under stress conditions. *Trends Plant Sci.* **3**, 181–187.

**Grandbastien, M.A., Audeon, C., Bonnivard, E. *et al.*** (2005) Stress activation and genomic impact of *Tnt1* retrotransposons in Solanaceae. *Cytogenet. Genome Res.* **110**, 229–241.

**Harper, G., Hull, R., Lockhart, B. and Olszewski, N.** (2002) Viral sequences integrated into plant genomes. *Annu. Rev. Phytopathol.* **40**, 119–136.

**Hirochika, H.** (1997) Retrotranspons of rice: their regulation and use for genome analysis. *Plant Mol. Biol.* **35**, 231–240.

**Hirochika, H.** (2001) Contribution of the *Tos17* retrotransposon to rice functional genomics. *Curr. Opin. Plant Biol.* **4**, 118–122.

**Hirochika, H., Sugimoto, K., Otsuki, Y., Tsugawa, H. and Kanda, M.** (1996) Retrotransposons of rice involved in mutations induced by tissue culture. *Proc. Natl Acad. Sci. USA*, **93**, 7783–7788.

**Hirochika, H., Guiderdoni, E., An, G. *et al.*** (2004) Rice mutant resources for gene discovery. *Plant Mol. Biol.* **54**, 325–334.

**International Rice Genome Sequencing Project** (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793–800.

**Jaillon, O., Aury, J.M., Noël, B. *et al.*** (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, **449**, 463–467.

**Jiang, N., Bao, Z., Zhang, X., Hirochika, H., Eddy, S.R., McCouch, S.R. and Wessler, S.R.** (2003) An active DNA transposon family in rice. *Nature*, **421**, 163–167.

**Kashkush, K., Feldman, M. and Levy, A.A.** (2003) Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat. Genet.* **33**, 102–106.

**Komatsu, M., Shimamoto, K. and Kyozuka, J.** (2003) Two-step regulation and continuous retrotransposition of the rice LINE-type retrotransposon *Karma*. *Plant Cell*, **15**, 1934–1944.

**Kumar, A. and Bennetzen, J.L.** (1999) Plant retrotransposons. *Annu. Rev. Genet.* **33**, 479–532.

**Li, L., Wang, X., Stolc, V. *et al.*** (2006) Genome-wide transcription analyses in rice using tiling microarrays. *Nat. Genet.* **38**, 124–129.

**Liu, B. and Wendel, J.F.** (2000) Retrotransposon activation followed by rapid repression in introgressed rice plants. *Genome*, **43**, 874–880.

**Lucas, H., Feuerbach, F., Grandbastien, M.A. and Caboche, M.** (1995) The tobacco retrotransposon *Tnt1* transposes in *Arabidopsis thaliana*. *EMBO J.* **14**, 2364–2373.

**Ma, J., Wing, R.A., Bennetzen, J.L. and Jackson, S.A.** (2007) Plant centromere organization: a dynamic structure with conserved functions. *Trends Genet.* **23**, 134–139.

**Miyao, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K., Shinozuka, Y., Onosato, K. and Hirochika, H.** (2003) Target site specificity of the *Tos17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *Plant Cell*, **15**, 1771–1780.

**Moon, S., Jung, K.H., Lee, D.E., Jiang, W.Z., Koh, H.J., Heu, M.H., Lee, D.S., Suh, H.S. and An, G.** (2006) Identification of active transposon *dTok*, a member of the hAT family, in rice. *Plant Cell Physiol.* **47**, 1473–1483.

**Murray, M.G. and Thompson, W.F.** (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acid Res.* **8**, 4321–4325.

**Perez-Hormaeche, J., Potet, F., Beauclair, L., Le Masson, I., Courtial, B., Bouche, N. and Lucas, H.** (2008) Invasion of the Arabidopsis genome by the tobacco retrotransposon *Tnt1* is controlled by reversible transcriptional gene silencing. *Plant Physiol.* **147**, 1264–1278.

**Piegu, B., Guyot, R., Picault, N. *et al.*** (2006) Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res.* **16**, 1262–1269.

**Piffanelli, P., Droc, G., Mieulet, D. *et al.*** (2007) Large-scale characterization of *Tos17* insertion sites in a rice T-DNA mutant library. *Plant Mol. Biol.* **65**, 587–601.

**Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A. and Barrell, B.** (2000) Artemis: sequence visualization and annotation. *Bioinformatics*, **16**, 944–945.

**Sallaud, C., Meynard, D., Van Boxtel, J. *et al.*** (2003) Highly efficient production and characterization of T-DNA plants for rice (*Oryza sativa* L.) functional genomics. *Theor. Appl. Genet.* **106**, 1396–1408.

**Slotkin, R.K. and Martienssen, R.** (2007) Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **8**, 272–285.

**Sonnhammer, E.L. and Durbin, R.** (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene*, **167**, 1–10.

**Tsugane, K., Maekawa, M., Takagi, K., Takahara, H., Qian, Q., Eun, C.H. and Iida, S.** (2006) An active DNA transposon *nDart* causing leaf variegation and mutable dwarfism and its related elements in rice. *Plant J.* **45**, 46–57.

**Tuskan, G.A., Difazio, S., Jansson, S. *et al.*** (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr& Gray). *Science*, **313**, 1596–1604.

**Vandenbussche, M., Janssen, A., Zethof, J. *et al.*** (2008) Generation of a 3D indexed Petunia insertion database for reverse genetics. *Plant J.* **54**, 1105–1114.

**Vitte, C. and Bennetzen, J.F.** (2006) Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. *Proc. Natl Acad. Sci. USA*, **103**, 17638–17643.

**Vitte, C., Lamy, F., Ishii, T., Brar, D.S. and Panaud, O.** (2004) Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). *Mol. Genet. Genomics.* **272**, 504–511.

**Vitte, C., Panaud, O. and Quesneville, H.** (2007) LTR retrotransposons in rice (*Oryza sativa* L.): recent burst amplifications followed by rapid DNA loss. *BMC Genomics*, **8**, 218.

**Wessler, S.R.** (1996) Plant retrotransposons: turned on by stress. *Curr. Biol.* **6**, 959–961.

**Wessler, S.R., Bureau, T.E. and White, S.E.** (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Curr. Opin. Genet. Dev.* **5**, 814–821.

**Wicker, T., Sabot, F., Hua-Van, A. *et al.*** (2007) A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982.

**Zhang, Z., Schwartz, S., Wagner, L. and Miller, W.** (2000) A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7**, 203–214.

La nouvelle stratégie, basée sur l'utilisation de la puce « transposome », que nous avons développé au sein de l'équipe, nous a permis de suivre l'activité transcriptionnelle de tous les rétrotransposons à LTR au sein du génome du riz. Parmi les 305 familles représentées sur la puce, nous avons identifié 13 familles de rétrotransposons à LTR transcriptionnellement actifs dans les cals. Parmi ces éléments, nous avons trouvé l'élément *Tos17*, déjà connu pour être actif dans ces tissus, et l'avons utilisé comme témoin positif. Ces 13 familles ont ensuite été testées pour leur activité transpositionnelle et outre *Tos17*, seul un nouvel élément, *Lullaby,* était actif dans les cals.

Cette stratégie a permis de mesurer l'activité transcriptionnelle de différentes familles d'ETs et de mettre en évidence un nouveau rétrotransposon à LTR actif chez le riz. La puce ainsi développée est un investissement technique permettant de tester l'activité transcriptionnelle des ETs dans différents stress, et sera améliorée en ajoutant des oligonucléotides correspondants à tous les gènes du riz, de manière à suivre les variations d'expression génique dans sa globalité. En combinant toutes les données de la puce, nous aurons une vue d'ensemble de l'impact fonctionnel des rétrotransposons à LTR sur l'expression des gènes dans différentes conditions de stress.

Par contre, il est important de souligner qu'avec cette étude, nous avons pu mettre en évidence que la plupart des éléments transcriptionnellement actifs ne transposent pas forcément *in planta*. Cela montre que seul le séquençage complet du génome pourra nous fournir une étude exhaustive du paysage transpositionnel.

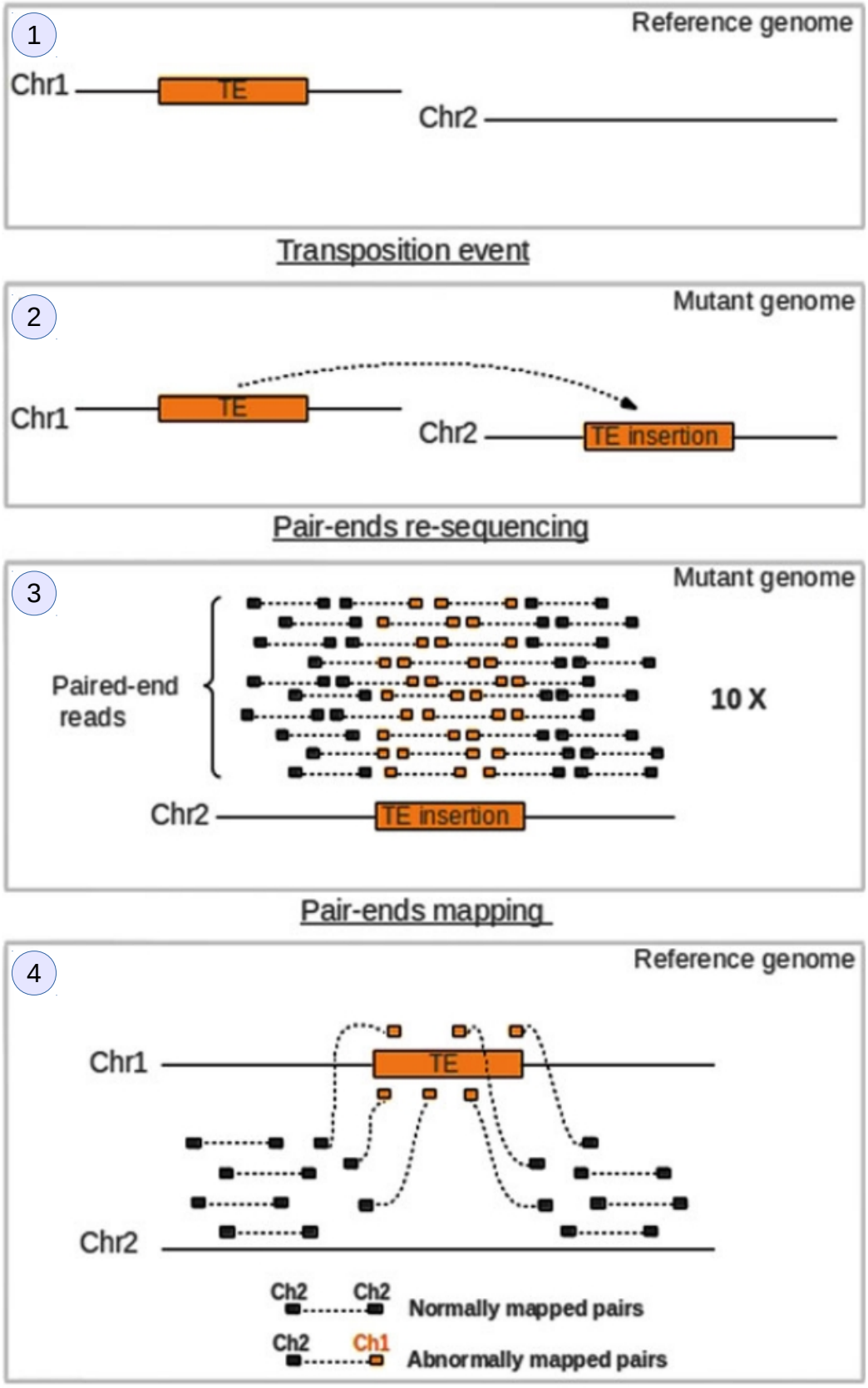



**Figure 11 : Représentation schématique du «paired-end-mapping» pour la détection de variants structuraux associés aux ETs.** (1) Une vue de deux localisations génomiques, l'une hébergeant un ET (chromosome 1) et l'autre un site vide (chromosome 2). (2) L'ET localisé sur le chromosome 1 transpose sur le chromosome 2. (3) Séquençage du génome mutant. Les amplicons illumina sont séquencés aux deux extrémités (les petits blocs représentent les lectures des « paired-ends »). (4) Cartographie des lectures illumina. La cartographie des lectures est considérée comme « normale » si les 2 lectures d'un amplicon se retrouvent proches sur la séquence de référence et comme « anormale » lorsque l'une des deux lectures s'aligne sur l'élément qui a transposé et donc absent à ce locus dans la séquence de référence.

### *2.1.2. Utilisation du séquençage haut débit*

En parallèle de cette précédente approche, et afin d'étudier le paysage transpositionnel d'un génome, nous avons développé une nouvelle stratégie basée sur l'utilisation de la technologie du séquençage à haut débit (NGS). L'essor de nouvelles technologies de séquençage à haut débit et le développement d'outils d'analyse bio-informatique puissants ont permis pour la première fois dans l'histoire de la biologie de séquencer des génomes entiers à très bas coût, et en un temps relativement court comparé au séquençage Sanger classique. Ceci a ouvert de nouveaux horizons quant à l'étude de la dynamique des génomes à l'échelle individuelle, ce qui permet de suivre l'évolution des génomes en temps réel et en particulier l'activité des ETs.

La première analyse de variants structuraux à l'échelle du génome entier avec le NGS a été développée chez l'homme par Korbel et al., 2007. Cette stratégie nommée le « Paired-end mapping » (PEM, Figure 11) a par la suite été améliorée avec succès par Stewart et al., (2011) et a permis la mise en évidence de l'activité des ETs.

Le re-séquençage de lignées mutantes (ou de plusieurs individus) permet ainsi de détecter le polymorphisme d'insertion des ETs sur l'ensemble du génome grâce aux séquences de référence de l'espèce disponibles publiquement. L'utilisation du re-séquençage pour l'analyse de l'activité transpositionnelle comprend plusieurs étapes : (1) Le séquençage du génome mutant (comportant potentiellement de nouvelles insertions) à l'aide des nouvelles technologies de séquençage à haut débit. (2) La cartographie des lectures générées par le séquençage à haut débit sur le génome de référence. (3) L'utilisation d'outils bio-informatique appropriés qui permettent de détecter les sites polymorphes en comparant les données de séquençage avec la séquence de référence de l'espèce étudiée.

Nous avons re-séquencé entièrement, à l'aide de la technologie *Illumina*, une lignée mutante de riz (AB156635), issue de la culture de cals de la variété Nipponbare. Cette lignée mutante *Tos17* présentait 11 nouvelles insertions de cet élément (mises en évidence par des expériences de Southern blot). L'objectif d'utiliser cette lignée était de tester si dans ce mutant, en plus de l'élément *Tos17,* d'autres éléments étaient actifs. Nous avons utilisé le paired-end-mapping afin de détecter les nouvelles insertions d'ETs. Pour cela, un pipeline informatique a été développé au sein de notre équipe afin de traiter automatiquement les millions de lectures générées par le re-séquençage. Ce pipeline nous a fourni une liste d'insertions d'ETs candidats du mutant AB156365. Par une approche expérimentale (Dotter, PCR et séquençage) j'ai pu valider toutes les insertions détectées *in silico*.

Le re-séquençage d'une lignée mutante de riz (provenant de culture de cals) a révélé une vingtaine de nouvelles insertions d'éléments transposables, permettant de caractériser de nouveaux éléments actifs dans le génome du riz (Sabot, Picault *et al*., 2011, où je suis 1ère co-auteur). Les résultats sont présentés dans la publication suivante acceptée en 2011 dans Plant Journal.

the plant journal

SEB Society for Experimental Biology

*The Plant Journal* (2011) **66**, 241–246 doi: 10.1111/j.1365-313X.2011.04492.x

# Transpositional landscape of the rice genome revealed by paired-end mapping of high-throughput re-sequencing data

**François Sabot[1,†], Nathalie Picault[1,†], Moaine El-Baidouri[1], Christel Llauro[1], Cristian Chaparro[1], Benoit Piegu[1], Anne Roulin[1], Emmanuel Guiderdoni[2], Mélissa Delabastide[3], Richard McCombie[3] and Olivier Panaud[1,*]**

[1]*Laboratoire Génome et Développement des Plantes, UMR CNRS/IRD/UPVD, Université de Perpignan Via Domitia, 52 Avenue Paul Alduy, F-66860 Perpignan Cedex, France,*

[2]*CIRAD, UMR DAP, Biological Systems Department, 2477 Avenue Agropolis, F-34398 Montpellier Cedex 5, France, and*

[3]*Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA*

Received 10 August 2010; revised 13 December 2010; accepted 4 January 2011; published online 1 March 2011.

*For correspondence (fax +33 468668499; e-mail panaud@univ-perp.fr).

†These authors contributed equally to this work.

## SUMMARY

**Transposable elements (TEs) are mobile entities that densely populate most eukaryotic genomes and contribute to both their structural and functional dynamics. However, most TE-related sequences in both plant and animal genomes correspond to inactive, degenerated elements, due to the combined effect of silencing pathways and elimination through deletions. One of the major difficulties in fully characterizing the molecular basis of genetic diversity of a given species lies in establishing its genome-wide transpositional activity. Here, we provide an extensive survey of the transpositional landscape of a plant genome using a deep sequencing strategy. This was achieved through paired-end mapping of a fourfold coverage of the genome of rice mutant line derived from an *in vitro* callus culture using Illumina technology. Our study shows that at least 13 TE families are active in this genotype, causing 34 new insertions. This next-generation sequencing-based strategy provides new opportunities to quantify the impact of TEs on the genome dynamics of the species.**

**Keywords: rice, transposable elements, next-generation sequencing, genomics, mutant line, paired-end mapping.**

## INTRODUCTION

Transposable elements (TEs) are mobile entities that are the major constituents of most eukaryotic genomes. Several reports have shown that TEs have a significant impact on the structure (SanMiguel *et al.*, 1996; Morgante, 2006; Feschotte and Pritham, 2007; Feschotte, 2008; Cordaux and Batzer, 2009) and function (Kobayashi *et al.*, 2004; Xiao *et al.*, 2008) of both plant and animal genomes. TE-associated structural variations have been widely documented by recent genome-level studies (Korbel *et al.*, 2007; Huang *et al.*, 2008). These variations have been detected based on comparative genomic studies among individuals in natural populations (Korbel *et al.*, 2007) or, in the case of crops, between varieties from the same species (Huang *et al.*, 2008). Most TE families are under the strict control of silencing pathways at both the transcriptional and post-transcriptional levels (Slotkin and Martienssen, 2007), and subsequently become non-functional due to accumulation of mutations (Vitte *et al.*, 2007). Moreover, TE insertion polymorphisms can be retained within gene pools for long periods, either through lineage sorting (particularly in the case of domesticated species) or because they are neutral (in the case of large populations). Therefore, the polymorphisms observed today only reflect the past transpositional activity of TE families that may since have become extinct. However, identification of active families in a given genome is necessary in order to fully understand the molecular basis of biodiversity. Moreover, in the case of crops, active TEs not only represent a source of such variation, but can also be exploited to develop genetic markers for genetic mapping and in marker-assisted breeding (Flavell *et al.*, 1998).

Asian rice (*Oryza sativa*) is grown worldwide and provides the basic food supply for billions of people in Asia, Africa and America. Completion of its genomic sequence (International Rice Genome Sequencing Project, 2005), which consists of 12 chromosomes with a total size of 430 Mbp, has enabled it to become a model species in plant genomics. Despite its relatively small size compared to other cereal crops such as sorghum (*Sorghum bicolor*; 900 Mbp), maize

(*Zea mays*; 2500 Mbp) or wheat (*Triticum aestivum*; 18 000 Mbp), the rice genome harbours a significant proportion of TE-related sequences (at least 33%) that belong to most of the classes described in eukaryotes (International Rice Genome Sequencing Project, 2005; Wicker *et al.*, 2007). The most predominant are the long terminal repeat (LTR) retrotransposons, with nearly 400 families ranging from one to several hundred copies, which constitute over 90 Mbp of the genome. Other types of TEs (long interspersed repetitive elements (LINEs), short interspersed repetitive elements (SINEs), transposons, miniatures inverted repeat transposable element (MITEs) and helitrons) represent an additional 40 Mbp. Altogether, several hundred TE families have been characterized in the rice genome, some of which are curated in dedicated databases (Jurka *et al.*, 2005; Chaparro *et al.*, 2007). At present, only eight of these TE families are known to exhibit transpositional activity *in planta*. These are the LTR retrotransposons *Tos17* (Hirochika *et al.*, 1996), *Lullaby* (Picault *et al.*, 2009), *osr7* and *osr23* (Wang *et al.*, 2009), the LINE *Karma* (Komatsu *et al.*, 2003), the MITEs *dTok* (Moon *et al.*, 2006), *nDart* (Tsugane *et al.*, 2006) and *mPing* (Jiang *et al.*, 2003), and the transposon *Pong* (Jiang *et al.*, 2003). Altogether, these eight TE families represent <100 kbp, i.e. 0.1% of the TE-related sequences in the rice genome.

A number of rice mutant collections have been developed for functional genomics studies (Rice Annotation Project, 2008). These have been produced either through *in vitro* callus culture or transformation using a T-DNA vector (which also requires *in vitro* cultivation for regeneration of the plant). Four of the eight families mentioned above are activated during *in vitro* culture of rice calli (*Tos17*, *Lullaby*, *mPing* and *Karma*). Consequently, it is expected that rice mutant lines harbour several TE-associated mutations that remain to be characterized in order to fully exploit these resources. Here we provide an extensive survey of the transpositional landscape of a plant genome using next-generation sequencing (NGS) technology. This was achieved through paired-end mapping (PEM) of a fourfold genome coverage of a rice mutant derived from *in vitro* callus culture using Illumina technology. Our study reveals that at least 13 TE families are active in a single mutant plant, and thus constitutes the first step towards understanding of the impact of these mobile elements on the genome dynamics of the species.

## RESULTS AND DISCUSSION

A genome-wide survey of TE activity was performed using NGS technology. As a proof of concept, we selected a line showing a high number of *Tos17* insertions in order to use this element as a positive control of TE activity in the analysis. The line AB156365 was thus selected from those available in the mutant collections because it harbours at least 11 such insertions, as indicated by a Southern hybridization experiment using *Tos17* as a probe (Figure S1). A total of 44 061 780 36-mer paired-end reads, representing a 4.17-fold genome coverage (base count), were generated using Illumina technology. These data are freely available upon request. PEM (Korbel *et al.*, 2007) was performed in order to identify TE-associated homology breakpoints compared to the reference genome sequence (International Rice Genome Sequencing Project rice pseudomolecules version 4; http://rapdb.dna.affrc.go.jp/): amplicons for which one end mapped to an unambiguously unique, non-TE-related sequence and the other corresponded to a known rice TE located at a distant site in the reference pseudomolecules were selected (see Figure 1 and Experimental procedures). In this process, many of the homology breakpoints that were identified in the first screening of the paired-end reads (i.e. abnormally mapped pairs) were not fully characterized, in particular those caused by insertions in repeated sequences. In addition, $R_3$ plants are expected to harbour a significant proportion of the mutations that originated from the callus culture at the heterozygous state. For such loci, half of the amplicons (corresponding to the wild-type) would exhibit no abnormal pairing, while the other half (corresponding to the mutated allele) would exhibit abnormal paired-end mapping. This leads to a discrepancy in the mapping phase of the analysis pipeline and subsequently to their elimination from TE insertion candidate loci. For these reasons, the present study provides a partial, non-exhaustive view of the overall transpositional activity of the rice genome. Nevertheless, our method allowed identification of 34 TE insertions caused by the activity of 13 TE families in the mutant line (Tables 1 and S1). A set of 27 randomly chosen breakpoints, representing the 13 active families, were confirmed using PCR amplification, cloning and sequencing of the region for a set of ten selfing progenies of the original mutant line (Figure 2 and Appendix S1). Our

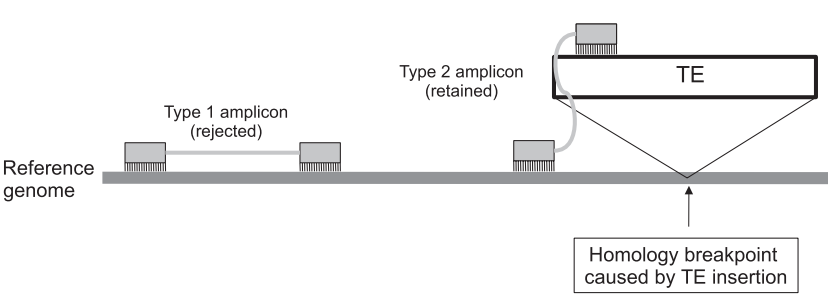



**Figure 1.** Schematic view of the procedure followed for PEM.

**Table 1** TE insertions identified in the mutant line

| TE name | TE type | Number of insertions | Number of copies | Expression |
|---|---|---|---|---|
| *BAJIE* | LTR retrotransposon | 1 | 663 | 0 |
| *osr10* | LTR retrotransposon | 1 | 248 | 0 |
| *osr37* | LTR retrotransposon | 3 | 128 | 0 |
| *RIRE2* | LTR retrotransposon | 2 | 451 | – |
| *RIRE3* | LTR retrotransposon | 2 | 587 | – |
| *RN363* | LTR retrotransposon | 1 | 1 | 0 |
| *RN216* | LTR retrotransposon | 1 | 135 | 0 |
| *Tos17* | LTR retrotransposon | 11 | 2 | + |
| *Mite#1* | MITE | 3 | 1696 | |
| *Mite#2* | MITE | 1 | 2554 | |
| *Mite#3* | MITE | 2 | 59 | |
| *mPing* | MITE | 3 | 8 | |
| *Tami2* | MITE | 3 | 42 | |

The expression column refers to previously published data (Picault *et al.*, 2009). 0, no difference in expression between callus and imbibed embryos; –, down-regulation in callus; +, over-expression in callus. Expression data are not available for the MITEs.

strategy allowed us to confirm and map *in silico* the 11 insertions of the LTR retrotransposon *Tos17* that were first identified based on the Southern hybridization experiment (Figure S1). The 23 additional insertions were caused by the activity of 12 other distinct TE families (seven LTR retrotransposons and five MITEs). Among these, only the MITE *mPing* was previously shown to be active in the rice genome (Jiang *et al.*, 2003), while the remaining 11 families were only identified through annotation of the rice pseudomolecules (Rice Annotation Project, 2008). Therefore, the PEM strategy implemented here allowed identification of active TEs in the rice genome at an unprecedented scale, increasing the number of known active families from eight to 19 in a single experiment. As indicated by the number of insertions that we identified, the most active type of TEs in rice callus appears to be the LTR retrotransposons, with eight distinct families causing 22 mutations. The second most active type of TEs in our rice mutant line are MITEs (Wicker *et al.*, 2007), which are short degenerated forms of transposons. Our study did not show transposition of two of the four TE families previously shown to be active in callus, i.e. the LTR retrotransposon *Lullaby* and the LINE *Karma*. As mentioned above, our screening for TE insertions is biassed towards gene-rich regions, and we cannot conclude that these two families are not active in the re-sequenced mutant. In addition, transposition is a stochastic process, and analysis of a single genotype is not sufficient to draw conclusions on the activity of all rice TE families. Additional studies of this type may further extend the list of active TEs in the species.

*Tos17* has been used to develop large mutant collections derived from *in vitro* culture of rice calli (Miyao *et al.*, 2007). Over 100 000 mutant lines are now available to the scientific community for both forward and reverse genetics studies in rice. Until now, identification of the insertion sites of *Tos17* was achieved through direct sequencing of the

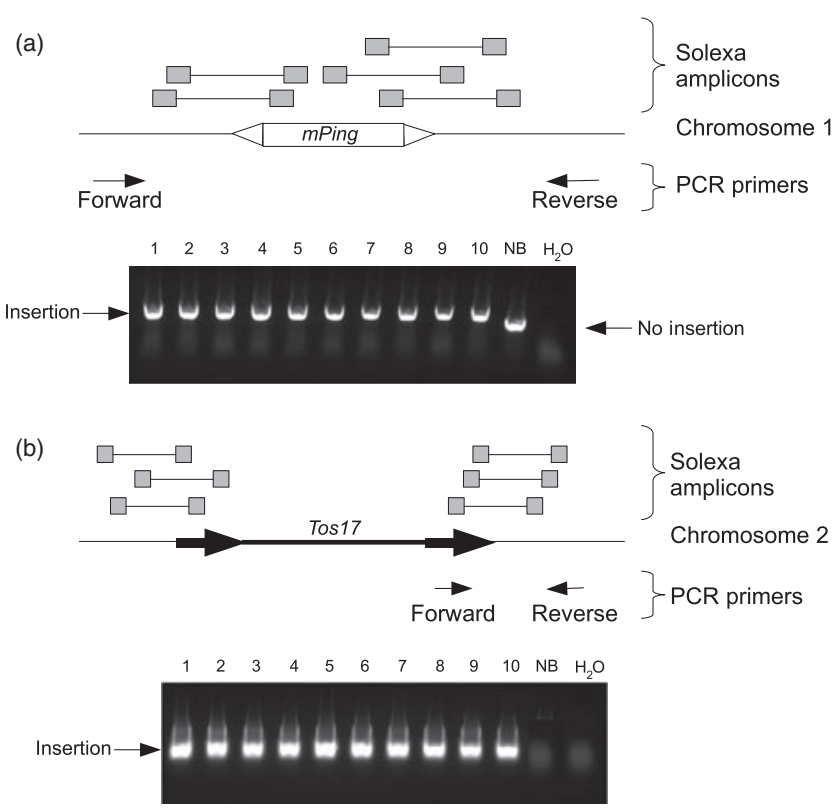



**Figure 2.** Validation of the insertions of *mPing* on chromosome 1 (a) and *Tos17* on chromosome 2 (b).
The Illumina amplicons mapped *in silico* are indicated above the genome sequence. The primers used for the PCR are shown below. The gels show PCR amplifications for 10 selfing progenies.

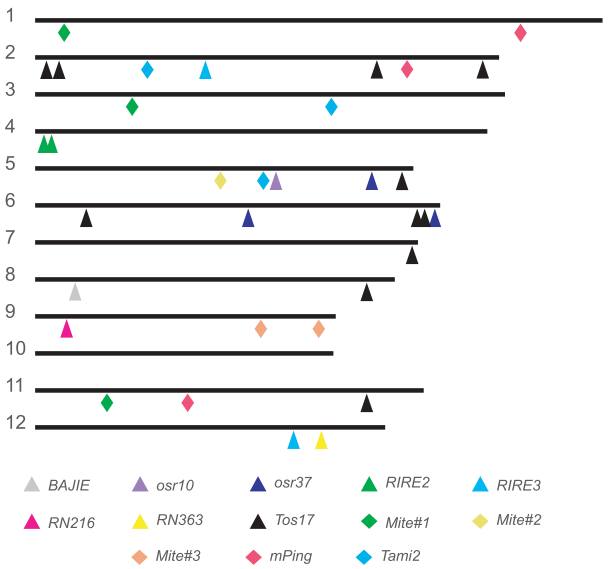



**Figure 3.** *In silico* mapping of the 34 TE insertions detected using our method on each chromosome (1–12, solid black lines).
Arrows and diamonds represent LTR retrotransposons and MITEs, respectively.

genomic region flanking the element (i.e. the flanking site tag; Miyao *et al.*, 2007). However, in the case of mutant lines harbouring multiple insertions of this element, flanking site tags cannot be obtained in a single sequencing experiment. The PEM strategy that we used allowed precise and unambiguous localization of all the *Tos17* insertions present in the line AB156365. Moreover, it allowed mapping of the 23 other mutations caused by the insertion of TEs that were not known to be active and for which flanking site tag mapping was therefore not applicable. These mutations are evenly distributed in the genome of the mutant line (Figure 3). Moreover, the *in silico* mapping of all the TE insertions was examined with regard to the position of the nearest rice gene. A complete list of the genes for which a TE insertion was identified within 1 kb in both the upstream and downstream directions in line AB156365 is given in Table 2. Fifteen of the 34 identified insertions occurred in the vicinity of genes, with the majority (ten) being caused by the activity of LTR retrotransposons. Eight of the 15 gene sequences were modified by insertion in either an exon, an intron or one UTR, which is twice the number caused by insertion of a *Tos17* element. Only 10% of the phenotypes observed in the rice mutant collections are tagged by flanking site tags of this element (Miyao *et al.*, 2007). As shown in the present study, PEM enables fast characterization of all genes mutated by TE insertions, and therefore increases the probability of gene discovery through forward genetics in mutants exhibiting altered phenotypes.

The LTR retrotransposon *Tos17* is present in two copies in the genome of the rice cultivar Nipponbare, on chromosomes 7 and 10. However, only the copy on chromosome 7 is active, the other being silenced by DNA methylation (Cheng *et al.*, 2006). No such information is available for the other seven LTR retrotransposon families that we identified in this study, as this report is the first evidence of their transpositional activity. Except for *RN363* (of which there is only one copy in the Nipponbare genome), these elements are present at a high copy number, which prevents identification of the active copy. As mentioned above, TEs are under the strict control of epigenetic silencing (Slotkin and Martienssen, 2007). Several reports have described the role of major pathways in plants, such as DNA methylation, histone methylation, RNA polymerase IV/V and siRNAs, by

**Table 2** Genes for which a TE insertion was found within 1 kbp

| Chromosome | Position | Gene ID | Gene name | Relative position | Location of insertion | TE type |
|---|---|---|---|---|---|---|
| 2 | 1001426 | Os02g0118900 | NBS-LRR disease resistance protein, putative | 0 | Exon | *Tos17* |
| 2 | 1657933 | Os02g0131800 | Similar to root-specific metal transporter | 1000 | | *Tos17* |
| 2 | 29511802 | Os02g0696500 | Glycosyl hydrolases family 16, putative, expressed | 0 | 3′ UTR | *mPing* |
| 2 | 35472538 | Os02g0809401 | Conserved hypothetical protein | −400 | | *Tos17* |
| 4 | 909838 | Os04g0115500 | Zinc finger, C2H2-type containing domain protein | −1000 | | *RIRE2* |
| 5 | 14832678 | Os05g0319900 | Similar to (1.4)-β-xylan endohydrolase, isoenzyme X-II (EC 3.2.1.8) (fragment) | 0 | Exon | *Mite#2* |
| 5 | 18323796 | Os05g0378800 | Methyltransferase, putative, expressed | 0 | Exon–intron | *Tami2* |
| 5 | 29177343 | Os05g0584600 | AAA family ATPase, putative, expressed | 0 | Exon–intron | *Tos17* |
| 6 | 31913095 | Os06g0728766 | Glycosyl transferase, family 48 protein | −900 | | *osr37* |
| 7 | 30050157 | Os07g0691100 | Similar to pectin methyltransferase 6 (fragment) | 300 | | *Tos17* |
| 8 | 3226371 | Os08g0155700 | Similar to RNA polymerase II largest subunit (fragment) | 0 | Exon | *BAJIE* |
| 8 | 26383373 | Os08g0528100 | Conserved hypothetical protein | 0 | Exon–intron | *Tos17* |
| 9 | 18074709 | Os09g0460500 | Gibberellin receptor GID1L2, putative, expressed | 200 | | *Mite#3* |
| 11 | 5770880 | Os11g0211300 | Similar to NBS-LRR disease resistance protein homologue (fragment) | −100 | | *Mite#1* |
| 11 | 26064634 | Os11g0621300 | DUF1399-containing protein, putative, expressed | 0 | Exon | *Tos17* |

The relative position is given in bp.

analysis of mutants impeding one or several of the enzymatic activities involved in these pathways (e.g. *met1*, *kyp* or *nrpd1a*; Mirouze *et al.*, 2009). The line AB156365 analyzed here was regenerated from *in vitro* culture of calli obtained from the wild-type cv. Nipponbare. The exact mechanisms of transpositional activation in cultured tissues are not fully understood, except for the LTR retrotransposon *Tos17*. Thus, whether demethylation is involved in activation of the other 12 TE families identified in this study remains to be elucidated. PEM of high-throughput sequence data could help to decipher the silencing pathways in rice on a genome-wide scale by analysis of the above-mentioned mutants.

In a previous study, we performed a genome-wide analysis of the transcriptional activity of rice LTR retrotransposons in calli using a microarray hybridization strategy (Picault *et al.*, 2009). All transpositionally active LTR retrotransposons are expected to be present in the form of transcripts in the tissues in which they are activated. The results of this transcriptomic survey are given in Table 1. Surprisingly, of the eight LTR retrotransposon families found to be active transpositionally, only *Tos17* showed over-expression in calli, while two (*RIRE2* and *RIRE3*) were down-regulated in these tissues. The remaining five did not show any significant difference in expression level between calli and the control. *RIRE2* and *RIRE3* contrast with *Tos17* in terms of copy number. There are only two copies of *Tos17*, but *RIRE2* and *RIRE3* are among the most highly repeated families in the rice genome. As suggested previously (Picault *et al.*, 2009), the expression level of LTR retrotransposons in rice may be controlled by a copy number-dependent pathway. However, the present results show a lack of correlation between expression level and transpositional activity, which suggests that retrotransposition occurs in calli regardless of transcriptional or post-transcriptional control, although this requires confirmation in further studies.

The development of high-throughput sequencing techniques has opened new perspectives in the field of genomics. They offer a fast and low-cost alternative to the classical Sanger-based technology. PEM has been used to detect structural variants in various organisms, such as human (Korbel *et al.*, 2007), *Bombyx* (Xia *et al.*, 2009) and Arabidopsis (Mirouze *et al.*, 2009). It was also used recently to perform high-density mapping in rice (Huang *et al.*, 2008). In that study, we used PEM to determine the transpositional activity in the rice genome, and showed that this approach allows fast identification of TE insertions in a mutant line, a task that until now required production of several generations of back-crossing, followed by positional cloning of the genes of interest. More generally, PEM could be used to detect TE activity in any species for which the complete genome sequence is available, and in any genetic background or physiological state. Use of this technology should clarify many aspects of TE biology in complex genomes, such as their epigenetic control and their genomic impact on stress responses (whether biotic or abiotic).

## EXPERIMENTAL PROCEDURES

### Sequencing

Rice mutant line AB156365 (from the Tos17 library at the National Institute of Agrobiological Sciences, Tsukuba, Japan) was selected for this study based on the number of *Tos17* insertions that its genome harbours (see Results and discussion). This line was obtained by A. Miyao and collaborators through regeneration of callus culture. Callus from strain NF7023 of the cultivar Nipponbare (from which the reference rice genome sequence was obtained) was obtained by induction in MS medium supplemented with 2 mg $L^{-1}$ 2,4-dichlorophenoxyacetic acid (2,4-D) and 0.3% gellan gum, followed by transfer into liquid medium (N6 + 1 mg $L^{-1}$ 2,4-D) after 21 days. After 19 weeks, calli were transferred into pre-regeneration medium (N6 + 0.3% casamino acid + 0.1% proline + 1 mg $L^{-1}$ 2,4-D) for 12 days, followed by 10 days in regeneration medium (N6 + 0.1 mg $L^{-1}$ benzyladenine (BA) + 0.01 mg $L^{-1}$ naphtaleneacetic acid (NAA) + 1% agarose). Plants ($R_0$) were then transferred into the field. Total genomic DNA was extracted from a single $R_2$ progeny of the regenerated plants (corresponding to two generations of selfing). Sequencing was performed in two lanes of the Illumina GA (http://www.illumina.com/) for 36-cycle paired-end sequencing.

### Data analyses

The 36 bp paired-end reads were mapped against the International Rice Genome Sequencing Project rice genome sequence (IRGSP version 4, http://rgp.dna.affrc.go.jp/E/IRGSP/Build4/build4.html), using bowtie software (Langmead *et al.*, 2009) with non-stringent parameters authorizing 3 mismatches (option -v 3) and an insertion size for pair ends of 300 bases (option -X 300) as amplicons had a theoretical size of 250 bases. We discarded all matching pairs and continued the analysis with the non-matching pairs using bowtie software to map them against a collection of transposable elements. The unmatched pairs from the previous step were mapped individually against the transposable elements database, and those not producing hits were BLAST searched against the genome. A list was constructed, indexed by chromosome position, for all non-mapped pairs, containing the position and best match of each member of the pair. Then the chromosome matching reads were grouped using a window of 600 bp, and those that presented more than 60% (from a minimum of three) of the corresponding pair reads pertaining to the same element or to the same chromosomal location were filtered. From this filtered set, a 1 kb region surrounding the candidate locations was subjected to a low-stringency BLAST search against the transposable elements database in order to eliminate candidates that showed resemblance to their corresponding pair hit. The remaining candidates were subjected to a manual curation step whereby dot plots were generated from the chromosomic regions, the inserted element and the paired ends. Those candidates that showed a clear pattern of paired-end mapping on the chromosomal region and the transposable element were retained for further analysis. This step allowed elimination of false positives that passed our previous filters.

### Experimental validation

Validation of the insertions was achieved through PCR amplification followed by sequencing of the PCR products on an Applied Biosystems 3130xl genetic analyzer (http://www.appliedbiosystems.com/). Amplification was performed for 35 cycles (2 min at

94°C, 35 cycles of 45 sec at 94°C, 45 sec at 58°C, 1 min at 72°C) with an annealing temperature of 58°C. PCR products were separated by 1% agarose gel electrophoresis. After purification using a Geneclean® Turbo kit (MP Biomedicals, http://www.mpbio.com/) according the manufacturer's instructions, each product was sequenced using corresponding primers in an ABI377 sequencer (Applied Biosystems). The sequences of the primers used for PCR and sequencing are available upon request.

## ACKNOWLEDGEMENTS

This work was supported by CNRS, IRD and the University of Perpignan, France. The authors thank Giuseppe Casacuberta (CSIC Barcelona, Spain) and Scott Jackson (Purdue University, USA) for their critical review of the manuscript. The authors thank A. Miyao and collaborators (NIAS, Tsukuba, Japan) for providing AB156365 seeds and information related to the tissue culture pedigree of the line.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

**Figure S1.** Southern hybridization of mutant line AB156365 progeny using the *Tos17* probe.

**Table S1.** Detailed description of the TE insertions.

**Appendix S1.** Validation of the TE insertions.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

## REFERENCES

**Chaparro, C., Guyot, R., Zuccolo, A., Piégu, B. and Panaud, O.** (2007) RetrOryza: a database of the rice LTR-retrotransposons. *Nucleic Acids Res.* **35**, D66–D70.

**Cheng, C., Daigen, M. and Hirochika, H.** (2006) Epigenetic regulation of the rice retrotransposon Tos17. *Mol. Genet. Genomics*, **276**, 378–390.

**Cordaux, R. and Batzer, M.A.** (2009) The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.* **10**, 691–703.

**Feschotte, C.** (2008) Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **9**, 397–405.

**Feschotte, C. and Pritham, E.J.** (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**, 331–368.

**Flavell, A.J., Knox, M.R., Pearce, S.R. and Ellis, T.H.** (1998) Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. *Plant J.* **16**, 643–650.

**Hirochika, H., Sugimoto, K., Otsuki, Y., Tsugawa, H. and Kanda, M.** (1996) Retrotransposons of rice involved in mutations induced by tissue culture. *Proc. Natl. Acad. Sci. USA*, **93**, 7783–7788.

**Huang, X., Lu, G., Zhao, Q., Liu, X. and Han, B.** (2008) Genome-wide analysis of transposon insertion polymorphisms reveals intraspecific variation in cultivated rice. *Plant Physiol.* **148**, 25–40.

**International Rice Genome Sequencing Project** (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793–800.

**Jiang, N., Bao, Z., Zhang, X., Hirochika, H., Eddy, S.R., McCouch, S.R. and Wessler, S.R.** (2003) An active DNA transposon family in rice. *Nature*, **421**, 163–167.

**Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O. and Walichiewicz, J.** (2005) Repbase update: a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467.

**Kobayashi, S., Goto-Yamamoto, N. and Hirochika, H.** (2004) Retrotransposon-induced mutations in grape skin color. *Science*, **304**, 982.

**Komatsu, M., Shimamoto, K. and Kyozuka, J.** (2003) Two-step regulation and continuous retrotransposition of the rice LINE-type retrotransposon Karma. *Plant Cell*, **15**, 1934–1944.

**Korbel, J.O., Urban, A.E., Affourtit, J.P. *et al.*** (2007) Paired-end mapping reveals extensive structural variation in the human genome. *Science*, **318**, 420–426.

**Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L.** (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25.

**Mirouze, M., Reinders, J., Bucher, E., Nishimura, T., Schneeberger, K., Ossowski, S., Cao, J., Weigel, D., Paszkowski, J. and Mathieu, O.** (2009) Selective epigenetic control of retrotransposition in Arabidopsis. *Nature*, **461**, 427–430.

**Miyao, A., Iwasaki, Y., Kitano, H., Itoh, J.I., Maekawa, M., Murata, K., Yatou, O., Nagato, Y. and Hirochika, H.** (2007) A large-scale collection of phenotypic data describing an insertional mutant population to facilitate functional analysis of rice genes. *Plant Mol. Biol.* **63**, 625–635.

**Moon, S., Jung, K.H., Lee, D.E., Jiang, W.Z., Koh, H.J., Heu, M.H., Lee, D.S., Suh, H.S. and An, G.** (2006) Identification of active transposon dTok, a member of the hAT family, in rice. *Plant Cell Physiol.* **47**, 1473–1483.

**Morgante, M.** (2006) Plant genome organisation and diversity: the year of the junk!. *Curr. Opin. Biotechnol.* **17**, 168–173.

**Picault, N., Chaparro, C., Piegu, B. *et al.*** (2009) Identification of an active LTR retrotransposon in rice. *Plant J.* **58**, 754–765.

**Rice Annotation Project** (2008) The Rice Annotation Project Database (RAP-DB): 2008 update. *Nucleic Acids Res.* **36**, D1028–D1033.

**SanMiguel, P., Tikhonov, A., Jin, Y.K. *et al.*** (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science*, **274**, 765–768.

**Slotkin, R.K. and Martienssen, R.** (2007) Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **8**, 272–285.

**Tsugane, K., Maekawa, M., Takagi, K., Takahara, H., Qian, Q., Eun, C.H. and Iida, S.** (2006) An active DNA transposon nDart causing leaf variegation and mutable dwarfism and its related elements in rice. *Plant J.* **45**, 46–57.

**Vitte, C., Panaud, O. and Quesneville, H.** (2007) LTR retrotransposons in rice (*Oryza sativa* L.): recent burst amplifications followed by rapid DNA loss. *BMC Genomics*, **8**, 218.

**Wang, H., Chai, Y., Chu, X., Zhao, Y., Wu, Y., Zhao, J., Ngezahayo, F., Xu, C. and Liu, B.** (2009) Molecular characterization of a rice mutator-phenotype derived from an incompatible cross-pollination reveals transgenerational mobilization of multiple transposable elements and extensive epigenetic instability. *BMC Plant Biol.* **9**, 63.

**Wicker, T., Sabot, F., Hua-Van, A. *et al.*** (2007) A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982.

**Xia, Q., Guo, Y., Zhang, Z. *et al.*** (2009) Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*). *Science*, **326**, 433–436.

**Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J. and van der Knaap, E.** (2008) A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science*, **319**, 1527–1530.

Dans cette étude, nous avons tout d'abord validé notre approche avec la localisation des 11 nouvelles insertions de *Tos17*. De plus, nous avons identifié 23 nouvelles insertions d'ETs, dont 11 sont des rétrotransposons à LTR répartis dans 7 familles distinctes et 12 sont des MITES (Miniature Inverted Transposable Elements) répartis dans 5 familles différentes. Il est intéressant de noter que parmi ces 12 familles d'ETs, seulement une (le MITE *mPing*) avait été préalablement décrite comme active (Jiang et al., 2003). Ce travail suggère que les études globales de la transposition au niveau des génomes, en utilisant le NGS, permettent la découverte rapide des éléments actifs dans un génome. De plus, cette approche par le re-séquençage nous a permis d'identifier les gènes mutés par des insertions d'ETs.

Nos travaux montrent que plusieurs familles d'ETs peuvent être réactivées et possèdent donc encore leur capacité à transposer. De telles réactivations restent néanmoins stochastiques et seuls quelques éléments d'une famille sont activés. Pour avoir une liste exhaustive de toutes les familles d'ETs actives, il est essentiel de tester plusieurs individus partageant un même fond génétique sous différentes conditions physiologiques.

Ces analyses permettent de détecter les ETs actifs dans les génomes des plantes et des animaux et offrent une meilleure compréhension de la dynamique des génomes en terme de variants structuraux, soit dans des mutants, soit dans des populations naturelles. Elles peuvent aussi ouvrir de nouvelles voies pour établir un lien entre ces variants structuraux et la diversité fonctionnelle chez les eucaryotes.

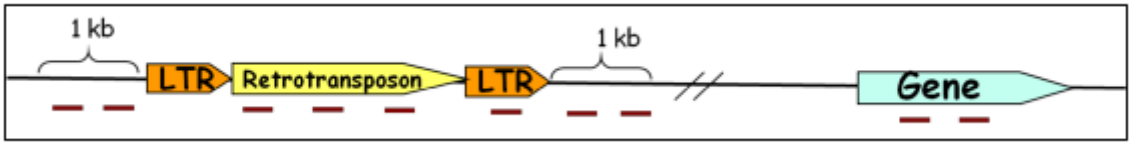



**Figure 12 : Représentation schématique du « design » de la puce version 2.0.** Les blocs rouges indiquent les oligonucléotides de 60 mers dans les régions flanquantes (1kb en amont et en aval des rétrotransposons à LTR) et dans chaque gène du riz (*Oryza sativa* ssp. Japonica) d'après la version 3 de TIGR, mais ensuite ré-annoté sur RAPDB 5.0.

## *2.2. Mise en évidence de rétrotransposons à LTR transcriptionnellement actifs chez le riz et impact sur l'expression des gènes*

Ce volet du travail a consisté en la mise au point de stratégies destinées, à échelle génomique, à associer des modifications globales d'expression génique avec la présence et l'expression d'ETs. Le principe de base est l'utilisation des outils transcriptomiques (hybridation sur puces de cDNAs extraits de différents tissus et conditions d'expression) permettant de révéler des modifications d'expression génique. La stratégie mise au point a consisté en l'amélioration de la puce "transposome", en associant des oligonucléotides spécifiques des ETs et des oligonucléotides correspondant aux gènes, mais aussi aux régions flanquant les insertions. En effet, dans certains cas, des oligonucléotides chevauchant la jonction entre l'élément et sa région flanquante ont été inclus. Cette 2ème version de la puce permet ainsi de corréler directement, pour une condition d'expression donnée, la production de transcrits correspondants aux ETs et celle de transcrits correspondants aux gènes, aux régions flanquantes et/ou aux co-transcrits hybrides.

Par rapport à la puce décrite précédemment, nous avons ajouté 90 000 oligonucléotides représentant tous les gènes du riz (45 000 gènes) avec 2 oligos par gène, 2 oligonucléotides dans les deux régions flanquantes des rétrotransposons à LTR et diminué la couverture des rétrotransposons à LTR avec au moins 6 oligonucléotides par élément (car dans la première puce nous avions réalisé du tiling), soit 290 000 oligonucléotides pour les rétrotransposons à LTR (Figure 12).

Afin de réaliser une étude globale des modifications d'expression génique associées aux ETs, nous avons réalisé ces expériences dans le cadre de stress biotiques ou abiotiques : sur des plantes soumises à un stress ferrique (Projet avec le Brésil EGIDE/COFECIB), au cours d'une interaction entre le riz et une bactérie (projet ANR-AZORIZ) et sur des plantes infectées par le virus RYMV (Rice Yellow Mottle Virus).

### *2.2.1. Stress ferrique*

Nous avons utilisé cette approche transcriptomique lors d'un stress ferrique en collaboration avec Antonio Costa de Oliveira de l'Université de Pelotas au Brésil. De 2008 à 2011, j'ai été co-responsable du projet de collaboration franco-brésilien (EGIDE-COFECUB) : « Adaptation aux stress abiotiques chez les plantes : analyse transcriptomique de la réponse au stress métallique (fer) chez le riz ».

Le fer est essentiel pour la croissance des plantes. Il est indispensable pour la photosynthèse, le transport d'électrons et les réactions redox. Il est soumis à une régulation très fine pour éviter la toxicité cellulaire qu'il cause lorsque présent en excès. La toxicité du fer est l'une des plus importantes contraintes pour la production du riz. Son entrée excessive par les racines, puis son transport dans les feuilles peut entraîner la formation d'espèces réactives d'oxygènes (ROS), entraînant les symptômes de jaunissement des feuilles et une diminution des rendements de 10 à 100 %.

Les plantes supérieures ont développé deux stratégies pour l'acquisition du fer à partir de la rhizosphère : la réduction du fer (stratégie I) ou sa chélation (stratégie II). La stratégie I est utilisée par toutes les plantes, sauf par les graminées, qui ont l'exclusivité de la stratégie II. Il est important de souligner que l'une des particularité du riz est qu'en plus de réaliser la stratégie II, cette espèce est capable d'absorber le fer *via* la stratégie I (Ishimaru et al., 2006). Ceci est un très grand avantage pour croître en condition de submergence, car dans ces conditions, le fer est sous forme réduite et donc plus toxique pour les plantes.

La toxicité du fer est un stress abiotique qui a lieu fréquemment dans les sols immergés du Brésil et principalement dans les régions du sud (Rio Grande Do Sul). L'état de Rio Grande do Sul est responsable de 63% de la production du pays, où la plupart (95%) des superficies cultivées sont irriguées par inondation (IRGA, 2008). Nous avons développé cette collaboration afin d'essayer de découvrir les mécanismes impliqués dans la réponse globale du riz lors d'un excès de fer. Nous avons réalisé l'analyse complète de l'expression des gènes et des rétrotransposons à LTR et identifié les voies métaboliques impliquées dans la réponse à ce stress.

Dans le cadre de ce projet, j'ai co-encadré une étudiante en thèse qui est venue en France de novembre 2008 à octobre 2009, et j'ai effectué deux missions de 15 jours à l'Université Fédérale de Pelotas, où j'ai réalisé des expériences et formé des étudiants en thèse et en master. J'ai rédigé avec l'étudiante en thèse un article dont je suis dernier auteur qui vient d'être accepté dans Rice (Finatto et al., 2015).

Finatto et al. Rice (2015) 8:13
DOI 10.1186/s12284-015-0045-6

Rice
a SpringerOpen Journal

RESEARCH Open Access

# Abiotic stress and genome dynamics: specific genes and transposable elements response to iron excess in rice

Taciane Finatto[1,4], Antonio Costa de Oliveira[1], Cristian Chaparro[2,3,5], Luciano C da Maia[1], Daniel R Farias[1], Leomar G Woyann[1], Claudete C Mistura[1], Adriana P Soares-Bresolin[1], Christel Llauro[2,3], Olivier Panaud[2,3] and Nathalie Picault[2,3*]

## Abstract

**Background:** Iron toxicity is a root related abiotic stress, occurring frequently in flooded soils. It can affect the yield of rice in lowland production systems. This toxicity is associated with high concentrations of reduced iron ($Fe^{2+}$) in the soil solution. Although the first interface of the element is in the roots, the consequences of an excessive uptake can be observed in several rice tissues. In an original attempt to find both genes and transposable elements involved in the response to an iron toxicity stress, we used a microarray approach to study the transcriptional responses of rice leaves of cv. Nipponbare (*Oryza sativa* L. ssp. *japonica*) to iron excess in nutrient solution.

**Results:** A large number of genes were significantly up- or down-regulated in leaves under the treatment. We analyzed the gene ontology and metabolic pathways of genes involved in the response to this stress and the *cis*-regulatory elements (CREs) present in the promoter region of up-regulated genes. The majority of genes act in the pathways of lipid metabolic process, carbohydrate metabolism, biosynthesis of secondary metabolites and plant hormones. We also found genes involved in iron acquisition and mobilization, transport of cations and regulatory mechanisms for iron responses, and in oxidative stress and reactive oxygen species detoxification. Promoter regions of 27% of genes up-regulated present at least one significant occurrence of an ABA-responsive CRE. Furthermore, and for the first time, we were able to show that iron stress triggers the up-regulation of many LTR-retrotransposons. We have established a complete inventory of transposable elements transcriptionally activated under iron excess and the CREs which are present in their LTRs.

**Conclusion:** The short-term response of Nipponbare seedlings to iron excess, includes activation of genes involved in iron homeostasis, in particular transporters, transcription factors and ROS detoxification in the leaves, but also many transposable elements. Our data led to the identification of CREs which are associated with both genes and LTR-retrotransposons up-regulated under iron excess. Our results strengthen the idea that LTR-retrotransposons participate in the transcriptional response to stress and could thus confer an adaptive advantage for the plant.

**Keywords:** Rice; Microarray; Iron toxicity; LTR-retrotransposon; *cis*-regulatory elements

* Correspondence: nathalie.picault@univ-perp.fr
[2]Laboratoire Génome et Développement des Plantes, UMR 5096, Université de Perpignan Via Domitia, F-66860 Perpignan, France
[3]CNRS, Laboratoire Génome et Développement des Plantes, UMR 5096, F-66860 Perpignan, France
Full list of author information is available at the end of the article

## Background

Rice, *Oryza sativa* L., is the staple food for more than two thirds of the world's population and is the second most widely grown cereal. Beyond its social and economic importance, rice is a model plant among the monocots because it has a small genome (~389 Mbp), which has been sequenced for the cultivars Nipponbare (*japonica* rice) (International Rice Genome Sequencing Project 2005) and 93–11 (*indica* rice) (Yu et al. 2002). Iron (Fe) is essential to mineral nutrition of plants. It is necessary for photosynthesis, electron transport and other redox reactions (Marschner 1995). Fe is subjected to tight control to avoid cellular toxicity (Connolly and Guerinot 2002). Although essential, it can be toxic when in excess (Stein et al. 2009a). Iron toxicity is one of the most important constraints to rice production on acid soils. The uptake of excessive Fe(II) by rice roots and its xylem transport *via* the transpiration stream into the leaves can lead to the generation of reactive oxygen species (Thongbai and Goodman 2000), causing the typical leaf-bronzing symptoms and entailing yield losses in the range of 10%-100% (Audebert and Fofana 2009).

Higher plants have evolved two main strategies for Fe acquisition from the rhizosphere: Fe reduction (Strategy I) or Fe chelation (Strategy II). Strategy I is employed by all plant species, with the exception of graminaceous plants, and involves pumping protons by $H^+$-ATPases to acidify the rhizosphere and increase Fe solubility in the soil. A ferric chelate reductase (FRO) reduces $Fe^{3+}$ to $Fe^{2+}$, and $Fe^{2+}$ transporters (IRTs - Iron Reduced Transporters) carry Fe into cells (Kim and Guerinot 2007). The Strategy II Fe-uptake system is limited to graminaceous plants, which release mugineic acid family phytosiderophores (MAs) from their roots to solubilize the sparingly soluble $Fe^{3+}$ in the soil (Takagi 1976; Takagi et al. 1984; Ishimaru et al. 2006). Exceptionally, in addition to a chelation strategy, rice possesses an Fe-uptake system that directly absorbs $Fe^{2+}$ (strategy I) (Ishimaru et al. 2006). This is advantageous for growth in submerged conditions because, unlike other grasses, rice is well adapted to grow in flooded conditions where $Fe^{2+}$ is more abundant than $Fe^{3+}$ (Ishimaru et al. 2006).

Rice plants have developed morphological and physiological mechanisms for avoiding adverse iron-toxic soil conditions and large amounts of iron in the plant (Becker and Asch 2005) and/or tolerance mechanisms to cope with and survive such conditions. Three major types of adaptation strategies have been proposed by Becker and Asch 2005. Firstly, an Fe-exclusion/avoidance strategy where the plants exclude $Fe^{2+}$ at the root level and hence avoid $Fe^{2+}$ damage to the shoot tissue (rhizospheric oxidation and root ion selectivity). Secondly, the Fe-inclusion/avoidance strategy where $Fe^{2+}$ is taken up into the rice root, but tissue damage may be avoided by either compartmentalization (immobilization of active iron in "dumping sites", *e.g.*, old leaves or photosynthetically less active leaf sheath tissue) or exclusion from the symplast (immobilization in the leaf apoplast). Finally, the Fe-inclusion/tolerance strategy where plants tolerate elevated levels of $Fe^{2+}$ within leaf cells, probably *via* enzymatic "detoxification" in the symplast (Becker and Asch 2005; Stein et al. 2009a). However, at present, these mechanisms are not very well characterized.

Several studies aiming to identify genes involved in response of rice to iron have been described, although most published studies concern iron deficiency (Kobayashi and Nishizawa 2012). Iron homeostasis has mainly been studied through differential expression of target genes, for example genes related to iron transport such as *OsIRT* (Lee and An 2009), *OsNRAMP1 and OsNRAMP2* (Zhou and Yang 2004), storage proteins like ferritin (*OsFER*) (Silveira et al. 2009; Stein et al. 2009b) and transcription factor like *OsWRKY*80 (Ricachenevsky et al. 2010). In fact, several transporters potentially involved in metal ion homeostasis have been identified in the rice genome (Kobayashi and Nishizawa 2012). Most of these metal transporters are capable of transporting one or several divalent cations including $Fe^{2+}$, $Zn^{2+}$, $Mn^{2+}$ and $Cu^{2+}$ (Narayanan et al. 2007). However, the contribution of each transporter and the precise iron flux still need to be clarified for each step involved in iron translocation.

In addition, the responses of plants to stress are complex and its perception requires interaction between multiple sensors. After initial recognition of stress by cells, a signal transduction cascade is triggered through secondary messengers that transmit the signal, activating responsive genes and generating an initial response. The products of the induced genes may be involved in response to stress and signal transduction. The stress genes enable plants to support these adverse conditions through short and long term responses (Grennan 2006). Transcription factors (TFs) regulate the first step of gene expression and are usually defined as proteins containing a DNA-binding domain that recognize specific DNA sequences, *cis*-acting regulatory elements (CREs), located in gene promoter regions (Mitsuda and Ohme-Takagi 2009). In order to understand mechanisms controlling gene expression in response to iron excess, it is important to know if specific CREs are present in the promoters of differentially-expressed genes.

The rice genome harbors a significant proportion of transposable element-related sequences (at least 33%) (IRGSP 2005). The most predominant type of transposable elements (TEs) are long terminal repeat (LTR) retrotransposons, with nearly 400 families ranging from one to several hundred copies, which constitute over 90 Mbp of the genome (*i.e.* 23%) (El Baidouri and Panaud 2013). However, the general impact of TEs on the

structure, evolution and function of plant genomes are not yet fully understood (Ma et al. 2004; Bennetzen 2007). LTR-retrotransposons require transcription and translation of the genetic information they encode in order to be transpositionally active. In fact, class I elements transpose *via* their mRNA, which is synthesized by RNA polymerase II (Kumar and Bennetzen 1999). The existence of transpositionally active LTR-retrotransposons in rice has only been demonstrated for 11 elements (Hirochika et al. 1996; Picault et al. 2009; Wang et al. 2009; Sabot et al. 2011). Although most TEs are transpositionally inactive, they can sometimes be activated by stress (Grandbastien 1998), although the exact process of transcription induction remains unknown. Transcriptional activation of LTR-retrotransposons can also significantly alter the expression of adjacent genes like in the case of blood orange (Butelli et al. 2012).

Recently, a transcriptomic approach was used (Quinet et al. 2012) and revealed modifications in gene expression after 3 days of exposure to iron excess concerning genes involved in hormonal signaling but also those involved in C-compound and carbohydrate metabolism, oxygen and electron transfer, oxidative stress, and iron homeostasis and transport in *indica* rice (cv. Kong Pao). This cultivar was classified as tolerant to iron toxicity by Engel et al. (2012) whereas Nipponbare cultivar was classified as susceptible. Thus, we decided to observe if the response to iron toxicity is a feature that could depend on genotypes presenting different degrees of tolerance/susceptibility. In addition, as we have the complete sequence of the Nipponbare cultivar, we designed a microarray that contained oligomers corresponding to all genes and LTR-retrotransposons of Nipponbare with the aim of discovering the mechanisms involved in the global response of rice in iron excess condition. In the present study, we performed a complete analysis of the expression of genes and LTR-retrotransposons, identifying the metabolic pathways and genes involved in response to this stress. Furthermore, we analyzed CREs in LTR of LTR-retrotransposons and gene promoters to potentially highlight common mechanisms in response to an abiotic stress.

## Results

### Phenotyping of rice plants

After four days of exposure to iron excess in nutrient solution, 18-day-old stressed rice plants showed necrotic spots on the leaves, which is a typical symptom of direct toxicity due to the accumulation of iron in the leaves (Figure 1). In the plants that grew in the control condition (without iron excess) no symptoms were observed and the leaves showed a uniform green color. In addition, no difference was observed in the fresh and dry matter between the plants with or without exposure to iron excess.

### Quantification of micronutrients in the tissue of rice leaves

The quantification of micronutrients was realized in shoots and significant differences ($p = 0.00374$) were found between iron content in leaves of 18-day-old rice seedlings after four days of iron excess exposure compared with control plants (Table 1). These results indicate that leaves of cv. Nipponbare cultivated in iron excess solution absorbed more than twice as much iron as seedlings in control conditions (optimal amount of iron). In contrast, for micronutrients such as manganese, copper and zinc, no significant differences ($p > 0.05$) between control and iron excess conditions were observed, indicating that the excess of iron had no influence on absorption of these cations in our conditions.

### Differentially expressed genes in microarray

Transcriptomic profiles were obtained using microarrays on Nipponbare rice plants under iron excess condition compared to non-stressed condition (three independent biological replicates for each condition). Genes significantly regulated were selected using a $P_{adjusted}$-value threshold of 0.05 and a fold change cutoff of 2 ($|Log_2(FC)| \geq 1$). A total of 2,525 genes are differentially expressed in rice under iron excess, which represent about 5.5% of the entire set of rice genes. Among these,

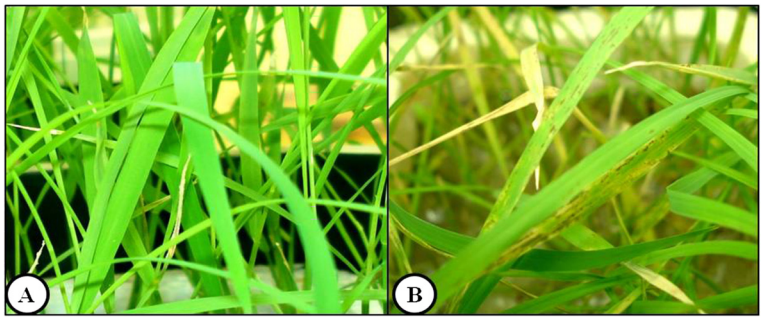



**Figure 1 Leaves of 18-day-old rice seedlings (*Oryza sativa* L. ssp. *japonica* cv. Nipponbare). A)** after development in a complete nutrient solution (control treatment, 10 μM Fe-EDTA) or **B)** after four days of iron excess exposure (7000 μM Fe-EDTA).

**Table 1 Means of micronutrient content in leaves, plant height and root length and shoot and dry matter in 18-day-old rice seedlings (*Oryza sativa* ssp. *japonica* cv. Nipponbare) after four days of iron excess exposure**

| Condition | Micronutrients content (mg $kg^{-1}$) | | | | Plant height (cm) | Root lenght (cm) | Shoot dry matter (g) | Root dry matter (g) |
|---|---|---|---|---|---|---|---|---|
| | Manganese | Cooper | Zinc | Iron | | | | |
| Control | 500.50 | 27.72 | 91.68 | 795.75* | 19.5 | 14.2 | 0.086 | 0.042 |
| Iron toxicity | 527.87 | 26.38 | 105.04 | 1927.12 | 18.8 | 13.8 | 0.092 | 0.039 |

*Indicate significant differences in means between control and iron toxicity conditions by ANOVA F-test.

2,457 were up-regulated and 68 down-regulated (Additional file 1: Table S1). Thus, in our experimental conditions, more than 97% of differentially expressed genes are up-regulated.

### Microarray validation by RT-qPCR

Microarray data were confirmed by analyzing expression levels of 17 representative genes using reverse transcription quantitative polymerase chain reaction (RT-qPCR). The correlation coefficient between microarray and RT-qPCR data was 0.7856 ($P < 0.01$), which can be considered as a good positive correlation. Table 2 shows the relative quantification of gene expression obtained by RT-qPCR, and the gene expression (fold-change) obtained by microarray for 17 genes. These results show that the array data are in accordance with the RT-qPCR data.

### Gene ontology

The cellular component classification after GO analysis (Figure 2) shows the greatest number of gene products are located in an intracellular membrane bound organelle, and most up-regulated gene products are localized in plastids (558) or in mitochondria (428), followed by comparable numbers in the cytosol (101) and vacuole (97). The down-regulated genes have the same distribution, with the exception of the cytosol. Concerning molecular functions, the binding and catalytic activity are the most highly represented with, respectively, 1159 and 1085 genes for those which are up-regulated and 31 and 26 for down-regulated genes. For transporter activity, only up-regulated genes (111) are observed under iron excess. For biological processes, the most frequent are metabolic process with 1187 up-regulated genes which represent more than one-third of genes affected by iron excess, followed by cellular process (996) and response to stimulus (450). The majority of up-regulated genes in the metabolic pathway are involved in lipid metabolic process (184), carbohydrate metabolism (175), biosynthesis of secondary metabolites (113) and biosynthesis of plant hormones (60) (Additional file 2: Table S2). Among

**Table 2 Differential expression of genes in the microarray ($Log_2FC$) and relative quantification (RQ expressed in $Log_2$) by RT-qPCR in leaves of 18-day-old rice seedlings after four days of iron excess exposure**

| *Locus* | $Log_2FC$ | RQ ($log_2$) | Description |
|---|---|---|---|
| *Os02g0121700* | 2.14 | 2.3 | Terpenoid synthase domain containing protein |
| *Os02g0594800* | 2.48 | 3.0 | OsNAC50, No apical meristem (NAM) protein |
| *Os02g0740700* | 1.68 | 2.5 | Peptidase M10A and M12B |
| *Os05g0162000* | 2.25 | 1.5 | Similar to Peroxidase |
| *Os06g0257450* | 2.26 | 3.7 | Similar to Ribonucleoside-diphosphate reductase |
| *Os08g0467400* | 1.87 | 0.8 | OsZIP14, Zinc/iron permease family protein. |
| *Os08g0508000* | 2.58 | 2.0 | Cytochrome P450 family proten |
| *Os12g0601800* | 1.67 | 1.2 | OsbZIP88, Similar to BZIP transcription factor family |
| *Os06g0141200* | −2.54 | −2.0 | OsZFP1, Similar to RNA-binding protein EWS. |
| *Os05g0506000* | −1.02 | −1.8 | Similar to MS5-like protein |
| *Os10g0521900* | −1.21 | −1.6 | Similar to Membrane protein |
| *Os06g0649000* | 1.41 | 2.5 | OsWRKY28 |
| *Os12g0567800* | 2.01 | 1.2 | OsMIT1f, Plant metallothionein, family 15 protein. |
| *Os12g0567800* | 2.01 | 0.5 | OsMIT1f, Plant metallothionein, family 15 protein. |
| *Os03g0288000* | 1.12 | 2.0 | OsMIT1b, Similar to Metallothionein |
| *Os05g0399300* | 3.29 | 3.0 | OsChi1b, Similar to Chitinase. |
| *Os12g0571000* | 2.43 | 2.5 | OsMIT1g, Metallothionein-like protein type 1. |
| *Os04g0486600* | - | - | OsGAPDH, Similar to Glyceraldehyde-3-phosphate dehydrogenase, cytosolic 3 |

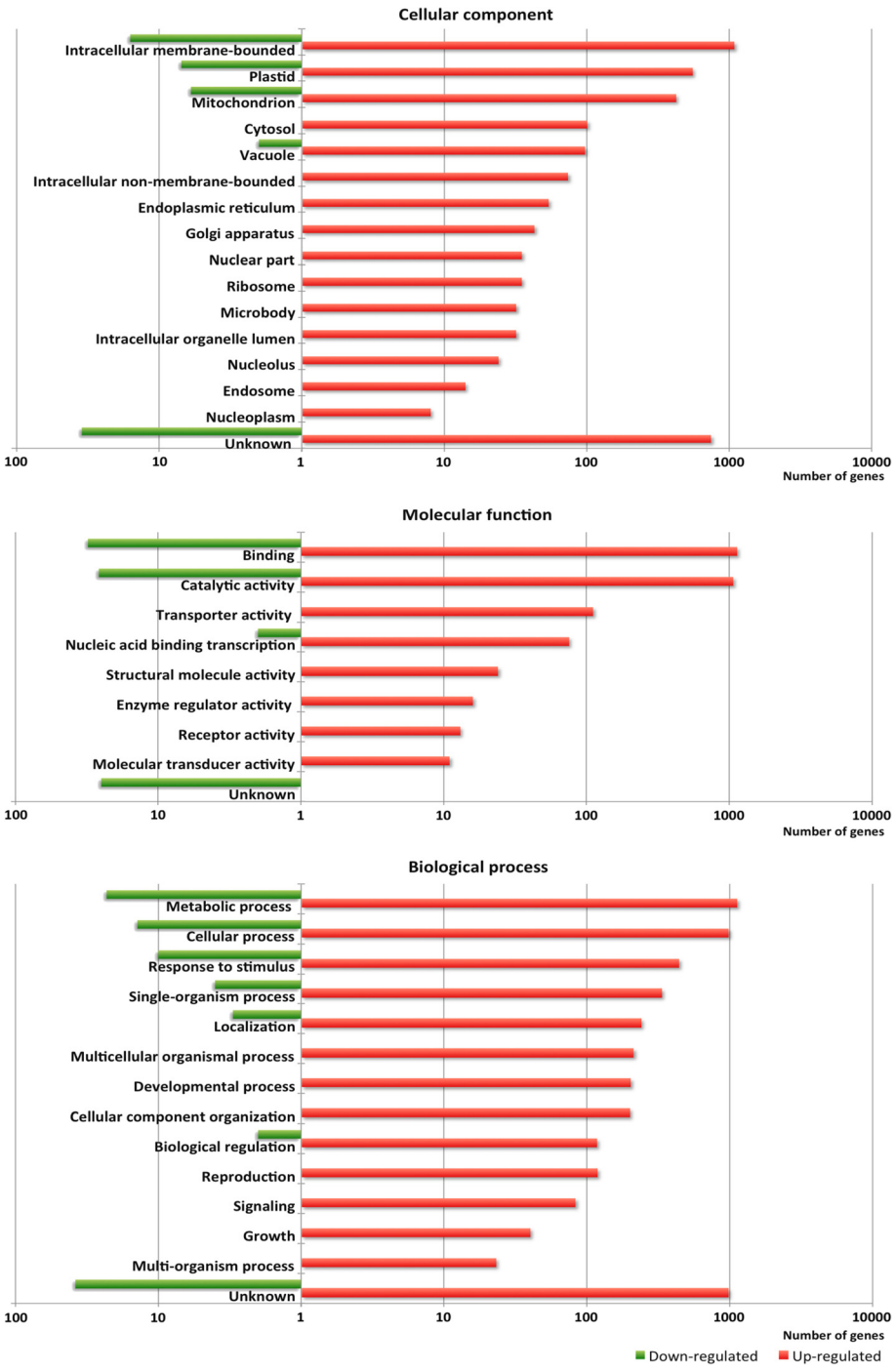



**Figure 2 Number of up-regulated and down-regulated genes in leaves of 18-day-old rice seedlings (*Oryza sativa* ssp. *japonica* cv. Nipponbare) after four days of iron excess exposure.** Gene ontology was generated by GO Slim with Biological Process and Molecular Function level 2 and Cellular Component level 6. A logarithmic scale was used for the number of genes.

the down-regulated genes, methane metabolism (5 genes) is particularly affected. In biological processes, another function with a large number of up-regulated genes is transport with 239 up-regulated genes.

## Expression profile of genes

Among the 2525 differentially expressed genes, 1773 correspond to genes with known functions (1720 up-regulated and 53 down-regulated) and 752 are hypothetical genes or

proteins (Additional file 1: Table S1). Thus, for the following analysis we concentrated on those with a known function. We compared these differentially expressed genes with genes or families of genes already described in the literature as encoding proteins involved in iron metabolism (*i.e.* transporters and enzymes involved in iron homeostasis) (Gross et al. 2003; Zheng et al. 2009; Ricachenevsky et al. 2010; Ricachenevsky et al. 2011; Kobayashi and Nishizawa 2012; Quinet et al. 2012; Victoria et al. 2012; Wang et al. 2013; Dabarni et al. 2013). In our study, we found 61 genes up-regulated in shoots in response to iron excess in common with the 392 candidate genes (Table 3). Among these, we found genes encoding proteins involved in iron uptake, distribution, redistribution and storage. These were classified into 3 classes according to their putative functions in iron homeostasis: (1) iron acquisition and mobilization (2) iron transport (3) regulatory mechanisms for iron responses.

Concerning genes involved in iron acquisition and mobilization, few of them were induced by iron excess in shoots. Transcription levels of only some genes involved in strategy I and II varied under iron excess. Among them, the ferric-chelate reductase oxidase gene, *OsFRO2* (*Os04g0578600*), involved in strategy I for the reduction of $Fe^{3+}$ and also described as playing a role in shoots (Sperotto et al. 2010) was up-regulated under iron excess in our condition. Expression of genes implicated in strategy II, the nicotianamine aminotransferase, *OsNAAT1* (*Os02g0306401*) (Inoue et al. 2008), and a receptor like protein RMC, *OsRMC* (*Os04g0659300*), recently shown to be involved in regulation of iron acquisition in rice (Yang et al. 2013), were up-regulated under iron excess. We found genes coding for known or potential metal transporters. The *YSL* (*yellow stripe-like*) genes are known as components of strategy II and are believed to transport NA-metal chelates across plant cell membranes. Experimental evidence points to a role of the YSL proteins in the long-distance and intra-cellular transport of metals (Curie et al. 2009; Ishimaru et al. 2010). Of the 18 rice genes that belong to the *OsYSL* family (Koike et al. 2004), two were induced in our conditions (*OsYSL1, Os01g0238700* and *OsYSL17, Os08g0280300*).

Vacuolar sequestration is an important mechanism in regulating iron homeostasis, and could serve as a safe iron storage strategy. Several genes have been shown to regulate iron trafficking between the cytosol and vacuoles, among which *OsVIT1* (*Os09g0396900*), localized to the vacuolar membrane and able to transport $Fe^{2+}$ across the tonoplast into the vacuole (Zhang et al. 2012), was induced under our conditions. Among the 8 genes in the family of natural resistance-associated macrophage protein (NRAMP) metal cation transporters shown to be involved in metal uptake and transport in plants (Belouchi et al. 1997; Narayanan et al. 2007; Takahashi et al. 2011), only *OsNRAMP6* (*Os01g0503400*) was up-regulated in our conditions. It has been shown that zinc transporters may play important roles in iron homeostasis in plants (Ricachenevsky et al. 2011). In the *ZIP* (*Zinc Iron Permease*) family genes, known to participate in divalent metal transport in plant (Guerinot et al. 2000; Lee et al. 2010), twelve homologs of the transporter OsIRT1 are present in the rice genome (Ishimaru et al. 2005). Among them, *OsZIP8* (*Os07g0232800*), *OsZIP10* (*Os06g0566300*) and *OsZIP14* (*Os08g0467400*) were up-regulated in iron toxicity. For the *Zinc-induced facilitator-like* (*ZIFL*) family genes, among the 13 genes of this family in rice genome (Ricachenevsky et al. 2011), *OsZIFL2* (*Os01g0279400*) was down-regulated and *OsZIFL7* (*Os11g0135900*) and *OsZIFL13* (*Os12g0133300*) up-regulated in our conditions.

The *Phenolic Efflux Zero1* (*OsPEZ1*, *Os03g0571900*) gene, shown to load protocatechuic acid (PCA) and caffeic acid into the rice xylem facilitating remobilization of precipitated apoplasmic iron inside the plant (Ishimaru et al. 2011), was induced under iron excess. Three up-regulated genes (*Os01g0684900, Os10g0345100, Os06g0495500*), belonging to the family of citrate transporters (Multi antimicrobial extrusion protein MATE family protein), were observed under iron excess. Some members of this family were found to be involved in the efficient translocation of iron from roots to shoots in rice (Yokosho et al. 2009).

Concerning the regulatory mechanisms for iron responses, we found four up-regulated genes (*Os03g0288000, Os05g0202800, Os12g0567800, Os12g0571000*) coding for metallothioneins (MTs) that have already been shown to have a significant role in maintaining intracellular metal homeostasis, metal detoxification and protection against intracellular oxidative damage (Zhou et al. 2006; Nath et al. 2014). MTs participate in controlling the concentration of "free" metals and reactive oxygen species that would activate defenses, *e.g. via* the MAPK cascade. These responses would help to regain cellular oxidant and metal homeostasis (Polle and Schützendübel 2003).

A large number of up-regulated genes (130) encoding transcription factors (TFs) was observed, mostly zinc finger (46 up-regulated genes) and WRKY (21 up-regulated genes) transcription factors. Other families of TFs encoded by up-regulated genes were bHLH (helix-loop-helix) (9 genes), ERFs (5 genes) and MYB (8 genes). Among the transcription factors, four were previously reported to be associated with iron response (Table 3): *OsNAC4* (*Os01g0816100*) and *OsEBP1* (*Os03g0860100*) by Zheng et al. (2009) and Wu et al. (2011), *OsNAC6* (*Os01g0884300*) by Nakashima et al. (2007) and Todaka et al. (2012), and *OsPRL1* (*Os03g0339100*) by Duc et al. (2009) and (Dabarni et al. 2013).

Antioxidant/scavenger enzymes are involved in detoxification of reactive oxygen species (ROS) produced

**Table 3 List of genes found both in our experiments and which are known to be involved in iron homeostasis**

| Gene-ID | Description | $log_2FC$ |
|---|---|---|
| **Iron uptake** | | |
| Os04g0578600 | OsFRO2, ferric-chelate reductase/oxidase protein | 1.40 |
| Os02g0306401 | OsNAAT1, Similar to Nicotianamine aminotransferase A | 1.34 |
| Os04g0659300 | OsRMC, receptor like protein | 3.04 |
| Os06g0486800 | OsFDH, Similar to Formate dehydrogenase, mitochondrial precursor | 1.46 |
| **Transport** | | |
| Os01g0238700 | OsYSL1 | 1.45 |
| Os08g0280300 | OsYSL17 | 1.78 |
| Os09g0396900 | OsVIT1, putative vacuolar iron/manganese transporter | 1.47 |
| Os01g0503400 | OSNRAMP6 | 1.43 |
| Os07g0232800 | OsZIP8 | 1.53 |
| Os06g0566300 | OsZIP10 | 1.25 |
| Os08g0467400 | OsZIP14 | 1.87 |
| Os01g0279400 | OsZIFL2, Major facilitator superfamily antiporter | −1.82 |
| Os11g0135900 | OsZIFL7, Major facilitator superfamily protein | 1.09 |
| Os12g0133300 | OsZIFL13, Similar to Carbohydrate transporter/sugar porter/transporter | 1.13 |
| Os03g0571900 | OsPEZ1, phenolics efflux transporter | 1.61 |
| Os01g0684900 | Multi antimicrobial extrusion protein MatE family protein | 1.07 |
| Os10g0345100 | Multi antimicrobial extrusion protein MatE family protein | 1.16 |
| Os06g0495500 | Multi antimicrobial extrusion protein MatE family protein | 1.55 |
| Os09g0440700 | OsCOPT7, putative copper cation transporter | 1.14 |
| Os01g0304100 | OsCCC2, putative cation:chloride co-transporter | 1.29 |
| **Homeostasis** | | |
| Os03g0288000 | OsMT1b, Similar to Metallothionein | 1.12 |
| Os05g0202800 | OsMT3b, Similar to Metallothionein-like protein 3B | 1.55 |
| Os12g0567800 | OsMT1f, Plant metallothionein, family 15 protein | 2.01 |
| Os12g0571000 | OsMT1g, Metallothionein-like protein type 1 | 2.53 |
| **Transcription factors** | | |
| Os01g0816100 | OsNAC4, Similar to NAC domain protein | 1.60 |
| Os01g0884300 | OsNAC6, No apical meristem (NAM) protein domain containing protein | 1.20 |
| Os03g0339100 | OsPRL1 | 1.01 |
| Os03g0860100 | OsEBP1 (ethylene responsive element binding protein encoding gene) | 1.94 |
| **Stress oxydatif** | | |
| Os08g0561700 | OsSOD4, OsCSD4, Similar to Superoxide dismutase | 1.23 |
| Os12g0613250 | Similar to BTB/POZ; Superoxide dismutase, copper/zinc binding; NPH3 | 1.27 |
| Os07g0665300 | Similar to Superoxide dismutase | 1.69 |
| Os03g0285700 | OsAPX1, APXa, OsAPx01, Similar to L-ascorbate peroxidase | 1.20 |
| Os10g0415300 | OsGR3, GR3, Similar to Chloroplastic glutathione reductase | 1.02 |
| Os02g0280700 | Similar to Iron/ascorbate-dependent oxidoreductase | 1.42 |
| Os07g0531400 | Similar to Peroxidase 27 precursor (EC 1.11.1.7) (PRXR7) | 1.20 |
| Os11g0112200 | Similar to Cationic peroxidase 1 precursor (EC 1.11.1.7) (PNPC1) | 2.09 |
| Os01g0327100 | Haem peroxidase family protein | 1.81 |
| Os10g0527400 | Similar to Tau class GST protein 3 | 1.54 |
| Os10g0530900 | Similar to Glutathione S-transferase GST 30 (EC 2.5.1.18) | 1.85 |
| **Hormonal regulation** | | |
| Os03g0860100 | OsEBP1 (ethylene responsive element binding protein encoding gene) | 1.94 |
| Os02g0201900 | DNA/RNA helicase, C-terminal domain containing protein | 1.54 |
| Os04g0475600 | 2OG-Fe(II) oxygenase domain containing protein. (ethylene) | 1.35 |
| Os05g0178600 | Similar to Auxin-responsive protein (Aux/IAA) (Fragment) | 1.01 |
| Os02g0643800 | Auxin responsive SAUR protein family protein | 1.02 |
| Os09g0491740 | Auxin efflux carrier domain containing protein | 1.14 |
| Os09g0554300 | Auxin efflux carrier domain containing protein | 1.20 |
| Os10g0147400 | Similar to Auxin influx carrier protein | 1.25 |
| Os12g0529300 | Similar to Auxin-binding protein | 1.51 |
| Os04g0288100 | Similar to Auxin-binding protein ABP20 | 1.76 |
| Os07g0164900 | Similar to ABA aldehyde oxidase | 1.29 |
| Os12g0555200 | Similar to Probenazole-inducible protein PBZ1. Gibberellins | 1.13 |
| **Cytochrome P450** | | |
| Os03g0760200 | Cytochrome P450 family protein | 1.71 |
| Os02g0570700 | Cytochrome P450 family protein | 1.21 |
| Os07g0418500 | Similar to Cytochrome P450 | 2.77 |
| Os07g0635500 | Similar to Cytochrome P450 | 1.89 |
| Os10g0515900 | Cytochrome P450 family protein | 1.26 |
| **Senescence and stress marker** | | |
| Os04g0650000 | ORYZAIN, Similar to Oryzain alpha chain precursor (EC 3.4.22.-) | 1.32 |
| Os04g0670200 | Similar to Oryzain beta chain precursor (EC 3.4.22.-) | 1.80 |
| Os02g0709800 | OsGAP1. RabGAP/TBC domain containing protein | 1.33 |

**Table 3 List of genes found both in our experiments and which are known to be involved in iron homeostasis** *(Continued)*

| | | |
|---|---|---|
| Os09g0532000 | OsSGR, TonB box, conserved site domain containing protein | 1.62 |
| Os04g0600300 | OsAOX1b, Homodimeric diiron-carboxylate protein, Cyanide-resistant respiration in mitochondria | 1.79 |

under iron stress toxicity (Becker and Asch 2005). Most of the genes involved in oxidative stress and detoxification of ROS were up-regulated in response to iron excess. Among the up-regulated genes, we found 9 genes coding for glutathione S-transferases (GSTs) which catalyze the conjugation of the reduced form of glutathione (GSH) to xenobiotic substrates for the purpose of detoxification in the cells. We also found genes encoding antioxidant enzymes, for example three superoxide dismutase (SOD) genes, two L-ascorbate peroxidase (APX) genes, seven genes encoding proteins similar to peroxidase (PRX) and one encoding chloroplastic glutathione reductase (GR).

Sixty up-regulated genes are involved in the biosynthesis of plant hormones (Additional file 3: Table S3). Plant hormones are implicated in the regulation of assimilate metabolism and growth. Furthermore, some of them, like abscisic acid (ABA), auxin and ethylene play a major role in controlling iron homeostasis in plants (Chen et al. 2010; Lingam et al. 2011; Wu et al. 2011). Among the up-regulated genes in our condition, we found genes implicated in hormone responses (Table 3) like the ABA signaling pathway (1 gene), auxin response (7 genes), ethylene response (3 genes, among them *OsEBP1*) and gibberellin response (1 gene).

Among genes involved in oxygen and electron transfers, the Cytochrome P450 family was represented by 44 up-regulated genes, most of which have a molecular function of monooxygenase activity or iron ion binding. Cytochrome P450 monooxygenases (P450s) play an essential role in the synthesis of lignin, pigments, defense compounds, fatty acids, hormones and signaling molecules in all plant species (Schuler 1996; Werck-Reichhart et al. 2002; Nielsen and Moller 2005, Pan et al. 2009). Recently, Quinet et al. (2012) found 11 up-regulated cytochrome P450 genes in the roots of *indica* rice (cv. Kong Pao) after three days of iron exposure. Among these genes, six (*Os08g0105700, Os03g0760200, Os02g0570700, Os07g0418500, Os07g0635500* and *Os10g0515900*) are also up-regulated in our conditions (Table 4), suggesting that these enzymes are important for iron homeostasis.

It is interesting to note that some differentially expressed genes are members of gene families already described as related to stress response or senescence. For example, oryzain-alpha and OsGAP1 were already characterized as related to iron excess in rice (Ricachenevsky et al. 2010). Oryzain-alpha (*Os04g0650000*), which is a cysteine proteinase found to be up-regulated by various stresses in rice leaves and which has been suggested to be involved in the final steps of leaf senescence (Fu et al. 2007), is up-regulated in our conditions, as is oryzain-beta (*Os04g0670200*). The C2 domain protein OsGAP1 (*Os02g0709800*) is also up-regulated. Among the other genes known as senescence and stress markers, we observed over-expression of *OsSGR* (*Os09g0532000*) (Park et al. 2007) and *OsAOX1b* (*Os04g0600300*) which codes for, an alternative oxidase involved in reactive oxygen species (ROS) scavenging (Saika et al. 2002; Feng et al. 2013).

### Occurrence of cis-Regulatory Elements (CREs) in putative promoters of up-regulated genes

For the identification of known plant regulatory promoter elements, 1 kbp upstream from transcription start sites of up-regulated genes were analyzed for potential consensus sequences using the 469 CREs experimentally validated in the literature (Mangeon et al. 2010; Tsutsui et al. 2011) and in the PLACE Database. We found 338 predicted CREs with a significant occurrence ($p \leq 0.05$). The number of different predicted CREs in 1 kb upstream regions ranged from zero to 40 (Figure 3A). Individual CREs were found in zero to 495 different putative promoter regions (Figure 3B).

Among the most represented CREs, ACGTABOX, GT1CORE, ARFAT, RYREPEATGMGY2, TATABOX3 and AMYBOX1 are each present in more than 400 up-regulated genes. Some of them have been identified in many plant genes regulated by diverse environmental and physiological conditions (Green et al. 1988; Huang et al. 1990; Lelievre et al. 1992; Foster et al. 1994). Interestingly, ARFAT has been shown to be an auxin-responsive element (Ulmasov et al. 1999). Additionally, among the 29 ABA responsive CREs (Table 5), we identified 23 that have at least one significant occurrence in the 1 kb region upstream from 665 up-regulated genes in response to iron toxicity (this represents *ca*. 30% of up-regulated genes). The most frequent CREs were ABREOSRAB21, ABRERATCAL and ACGTABREMOTIFA2OSEM that are present in 226, 272 and 293 putative gene promoters, respectively. The 665 up-regulated genes include kinases, transcription factors and genes involved in ROS detoxification. The list of up-regulated genes that present at least one occurrence of an ABA-responsive CRE is provided in Additional file 4: Table S4.

The occurrences of CREs in the 1 kb upstream regions were separated into two groups: the putative complex

**Table 4 List of genes common to our experiments and Quinet et al. 2012**

| Gene-ID | $log_2FC$ | Function in RAP-DB | Function in Quinet et al. 2012 |
|---|---|---|---|
| **Carbohydrate metabolism** | | | |
| Os03g0169000 | 2.30 | Similar to predicted protein | Ribulose-5-phosphate-3-epimerase |
| Os04g0275100 | 1.12 | Serine/threonine protein kinase domain containing protein | Wall-associated kinase |
| Os05g0366600 | 1.07 | Similar to Hydroxyisourate hydrolase | Beta-glucosidase |
| Os07g0539900 | 1.40 | Similar to Beta-1,3-glucanase-like protein | O-Glycosyl hydrolase |
| Os07g0538000 | 1.21 | Similar to hydrolase, hydrolyzing O-glycosyl compounds | 1-3,1-4-beta-glucanase |
| Os04g0459500 | 1.01 | Similar to GADPH (383 AA) (Fragment) | Glyceraldehyde-3-phosphate dehydrogenase |
| Os10g0416500 | 1.10 | Similar to Chitinase 1 precursor (EC 3.2.1.14) | Chitinase |
| **Metabolism of secondary products** | | | |
| Os03g0122300 | 1.09 | Similar to Oxidoreductase, 2OG-Fe oxygenase family protein | Flavanone 3-hydroxylase-like protein |
| Os02g0218700 | 1.58 | Similar to Allene oxide synthase (EC 4.2.1.92) | Allene oxide synthase |
| **Metabolism of toxins** | | | |
| Os10g0530900 | 1.85 | Similar to Glutathione S-transferase GST 30 (EC 2.5.1.18) | Glutathione S-transferase (OsGSTU6) |
| **Oxygene and electron transfer** | | | |
| Os04g0600300 | 1.79 | Homodimeric diiron-carboxylate protein, | Alternative oxidase |
| Os08g0105700 | 2.91 | Similar to Bx2-like protein | Cytochrome P450 |
| Os03g0760200 | 1.71 | Cytochrome P450 family protein | Cytochrome P450 |
| Os02g0570700 | 1.21 | Cytochrome P450 family protein | Cytochrome P450 |
| Os07g0418500 | 2.77 | Similar to Cytochrome P450 | Cytochrome P450 |
| Os07g0635500 | 1.89 | Similar to Cytochrome P450 | Cytochrome P450 |
| Os10g0515900 | 1.26 | Cytochrome P450 family protein | Cytochrome P450 |
| Os02g0218700 | 1.58 | Similar to Allene oxide synthase (EC 4.2.1.92) | Allene oxide synthase |
| Os06g0549900 | 1.90 | FAD linked oxidase, N-terminal domain containing protein | Reticulilne oxidase-like protein |
| Os05g0529700 | 1.44 | Heat shock protein DnaJ family protein | DNAJ heat shock protein |
| **Oxidative stress and detoxification** | | | |
| Os07g0531400 | 1.20 | Similar to Peroxidase 27 precursor (EC 1.11.1.7) (PRXR7) | Peroxidase |
| Os11g0112200 | 2.09 | Similar to Cationic peroxidase 1 precursor (EC 1.11.1.7) | Cationic peroxidase |
| Os01g0327100 | 1.81 | Haem peroxidase family protein | Peroxidase |
| Os10g0527400 | 1.54 | Similar to Tau class GST protein 3 | Glutathione S-transferase (OsGSTU3) |
| Os10g0530900 | 1.85 | Similar to Glutathione S-transferase GST 30 (EC 2.5.1.18) | Glutathione S-transferase (OsGSTU6) |
| Os04g0447700 | 1.50 | Similar to Polyketide reductase | NADPH-dependent oxidoreductase |
| Os03g0700700 | 1.73 | Similar to Lipoxygenase | Lipoxygenase |
| **Iron homeostasis** | | | |
| Os09g0396900 | 1.47 | OsVIT1,Protein of unknown function DUF125 | CCC1, iron transporter |
| Os04g0538400 | 1.24 | Similar to Nodulin 21 (N-21) | Vacuolar iron transporter |
| **Hormonal regulation** | | | |
| Os12g0555200 | 1.13 | Similar to Probenazole-inducible protein PBZ1 | Probenazole-inductible protein PBZ1 |
| Os02g0201900 | 1.54 | DNA/RNA helicase, C-terminal domain containing protein | Ethylene response factor (ERF) |
| Os04g0275100 | 1.12 | Serine/threonine protein kinase-related domain containing protein | Wall-associated kinase |
| Os02g0218700 | 1.58 | Similar to Allene oxide synthase (EC 4.2.1.92) | Allene oxide synthase |

regulation group in which CRE occurrence is greater than or equal to the average occurrence in all genes plus two standard deviations (*i.e.* more than 24 different CREs in the 1 kb upstream region); and the putative simple regulation group, when CRE occurrence is smaller than or equal to the average occurrence in all genes minus two standard deviations (*i.e.* less than 4 different CREs in the 1 kb upstream region). Examples of

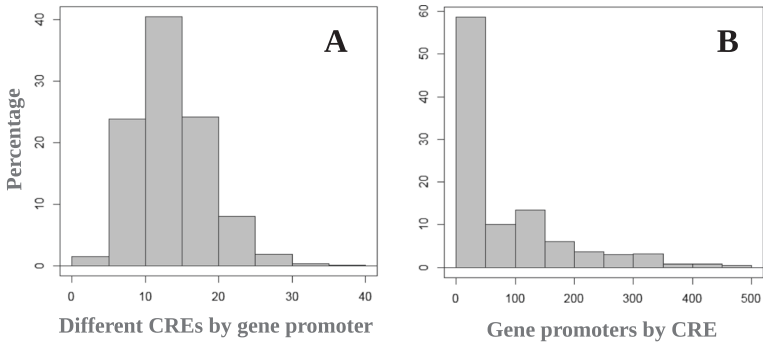



**Figure 3 Histograms of significant (P < 0,05) occurrences of different CREs in each gene promoter and number of gene promoters in which each CRE occurs. A)** Histogram with a percentage of different CREs by gene promoter of up-regulated genes in leaves of 18-day-old rice seedlings (*Oryza sativa* ssp. *japonica* cv. Nipponbare) after four days of iron excess exposure; **B)** Histogram representing the percentage of gene promoters in which each CRE occurs.

simple and complex regulation are the up-regulated gene *Os06g0127800,* a GAI-RGA-SCR (GRAS) family protein involved in a Brassinosteroid signaling, with 35 different CREs with a significant occurrence and thus classified as a gene with putative complex regulation, and the up-regulated gene *Os03g0815200,* that has a function similar to methylenetetrahydrofolate reductase (EC 1.5.1.20), with only one CRE in the 1 kb upstream region, classified as having putative simple regulation.

We found 102 genes in the complex regulation group and 19 in the simple regulation group, and 54 and 270 different CREs in groups of genes with putative simple and complex regulation, respectively. The 54 CREs observed in the putative simple regulation group were also found in the other group. In contrast, 68 CREs were only present in normal genes and 5 only in the putative complex regulation groups (Figure 4). The latter 5 are: C1MOTIFZMBZ2, GBOXSORBCS1, RYREPEAT4, ABRETAEM and ACGTSEED3.

### Rice transposable elements

We completed the gene transcriptomic survey with an analysis of LTR-retrotransposons. We used a microarray that contained oligomers corresponding to previously described elements (Chaparro et al. 2007) to observe the transcriptional activity of all LTR-retrotransposons in the rice genome. More recently, El Baidouri and Panaud (2013) performed a new classification of rice LTR-retrotransposons, defining 369 families and 3623 loci harbouring complete LTR-retrotransposons in the rice genome. We used this latest data for our analysis, retaining only those that had 100% identity and which were unique in the genome among the differentially expressed oligomers.

In the presence of an excess of iron, differential expression was observed for 158 LTR-retrotransposon families among the 369 existing families in rice. When we looked in detail, the stress modified the transcription of 37% (1344 loci/3623 loci present in rice genome) of complete LTR-retrotransposons, among which 95.4% were up-regulated and 4.6% down-regulated. We previously showed in another stress (a rice mutant line derived from an *in vitro* callus culture) that 8 LTR-retrotransposon families were activated at the transpositional level in rice (Sabot et al. 2011). Interestingly, the majority of these families (seven out of eight) were found to be differentially expressed in our study: *BAJIE, osr10, osr37, rn363, rn216, RIRE2* and *RIRE3.*

In addition, we explored if transposable elements can significantly alter the expression of adjacent genes. In the rice genome, LTR-retrotransposons can be present within the region of 3 kb upstream or downstream genes. In our microarray, we do not have the complete flanking region for all genes. However, among the differentially expressed LTR-retrotransposons, 10% are located in the neighborhood of genes. In fact, 57 are located within 3 kb upstream from genes which are differentially expressed under iron stress, while 65 are located in the 3 kb downstream from other differentially expressed genes. We checked whether these elements have an impact on the expression of the adjacent gene. No co-transcription events between genes and LTR-retrotransposons could be detected by RT-PCR (data not shown), suggesting that none of these elements presented direct read-through transcription of adjacent genes. We therefore looked for CREs in the complete LTRs of LTR-retrotransposons differentially expressed in our conditions, because the LTR contains the promoter region of the element. We found 1247 LTRs that contain at least one CRE and, among the 469 analysed CREs, 279 were found with a significant occurrence ($p \leq 0.05$). The most representative was ARFAT, present in more than 470 up-regulated LTR-retrotransposons. Comparing these results with those for genes, we found that,

**Table 5 Number of promoters of up-regulated genes and LTRs of LTR-retrotransposons where ABA responsive CREs are significantly present (P < 0.05)**

| ABA responsive CREs | Number of gene promoters | Number of LTR-retrotranposons |
|---|---|---|
| ABAREG2 | 0 | 0 |
| ABASEED1 | 0 | 0 |
| ABRE2HVA22 | 0 | 3 |
| ABRE3OSRAB16 | 0 | 0 |
| ABRECE3HVA1 | 0 | 0 |
| ABRECE3ZMRAB28 | 0 | 4 |
| ABRE2HVA1 | 1 | 1 |
| ABRE3HVA1 | 1 | 0 |
| ABRE3HVA22 | 1 | 0 |
| ABADESI2 | 2 | 1 |
| ABREDISTBBNNAPA | 2 | 0 |
| ABREMOTIFIOSRAB16B | 2 | 0 |
| ABRETAEM | 2 | 0 |
| ABREBNNAPA | 3 | 0 |
| ABREBZMRAB28 | 3 | 0 |
| ABREMOTIFIIIOSRAB16B | 3 | 3 |
| ABREAZMRAB28 | 6 | 0 |
| ABRECE1HVA22 | 8 | 1 |
| ABADESI1 | 9 | 0 |
| ABREA2HVA1 | 16 | 0 |
| ABREMOTIFAOSOSEM | 29 | 2 |
| ACGTABREMOTIFAOSOSEM | 29 | 2 |
| ABREATRD22 | 38 | 48 |
| ABREZMRAB28 | 67 | 103 |
| ABREATCONSENSUS | 118 | 110 |
| ABRELATERD1 | 154 | 65 |
| ABREOSRAB21 | 226 | 167 |
| ABRERATCAL | 272 | 128 |
| ACGTABREMOTIFA2OSEM | 293 | 152 |

among the 8 most highly-represented CREs in the putative promoters of up-regulated genes, 5 are also found in the LTRs of LTR-up-regulated retrotransposons (TATABOX3, ARFAT, GT1CORE, HEXMOTIF-TAH3H4, LTRE1HVBLT49). Interestingly, when we compared the distribution of TATABOX3 in all LTRs (differentially expressed or not) the overall distribution of this CRE is significant in LTRs of LTR-retrotransposons up-regulated under iron excess (test chi$^2$ $p < 0.001$).

Additionally, among the 29 ABA responsive CREs (Table 5), we identified 15 that have at least one significant occurrence in LTR of up-regulated LTR-retrotransposons in response to iron toxicity. The most frequent CREs were ABREOSRAB21, ABRERATCAL and ACGTABREMOTIFA2OSEM that are present in 167, 128 and 152 LTR of LTR-retrotransposons, respectively. This is exactly the same pattern as the one we found for putative promoters of up-regulated genes. Our analysis shows that genes and LTR-retrotransposons up-regulated under iron toxicity present the same pattern of CREs.

## Discussion and Conclusions

Previous studies on the Nipponbare cultivar led to contradictory results. Wan et al. (2003), using a nutritive solution with 250 mg $L^{-1}$ of $Fe^{2+}$, described it as being tolerant to iron excess, as the leaf bronzing index was not significantly different to Suakoko8 (the iron toxicity tolerant control cultivar) after 28 days exposure to iron excess. However, more recently, Engel et al. (2012) studied different rice genotypes, among which Nipponbare and Kong Pao cultivars, for response after 4, 6 and 8-week-old plants to an Fe pulse of 1,500 mg $L^{-1}$ $Fe^{2+}$ for 6 days. Nipponbare was considered susceptible because they observed that 8-week-old plants of Nipponbare showed a leaf-bronzing score significantly higher than the genotypes considered tolerant. Furthermore, taking into account the above-ground biomass accumulation, no significant differences were found between Nipponbare and tolerant genotypes. The Kong Pao cultivar was classified as tolerant to iron toxicity. The Nipponbare used in this experiment comes from the same stock used for the IRGSP, and has performed in several experiments as medium tolerant to tolerant to iron excess (Costa de Oliveira, pers.commun.).

We compared our observations with those of Quinet et al. (2012), who studied gene expression in 2-week Kong Pao (*indica* rice) plants in response to 125 mg $L^{-1}$ $FeSO_4$ for 3 days (short-term) and 3 weeks (long-term). We find that only a few genes in *japonica* rice are the same as those responsive to iron toxicity in *indica* rice (Table 4). None of the down-regulated genes are the same as in *indica* rice shoots in short-term response. Quinet et al. (2012) find 135 down and 95 up-regulated genes in shoots in the short-term response, whereas we found more than 10 times as many up-regulated genes. This difference could be due to the fact that in our experiment we performed a global analysis with a microarray that contained probes corresponding to all genes of the *japonica* rice genome and all complete LTR-retrotransposons.

Although we did not find exactly the same genes, it is interesting to note that we often found genes belonging to the same family or involved in the same metabolic pathway. In fact, a large number of genes up-regulated in our experiment are involved in carbohydrate and secondary metabolism, oxidative stress and detoxification of ROS, iron homeostasis (YSL, NRAMP, ZIP, metallothionein) and

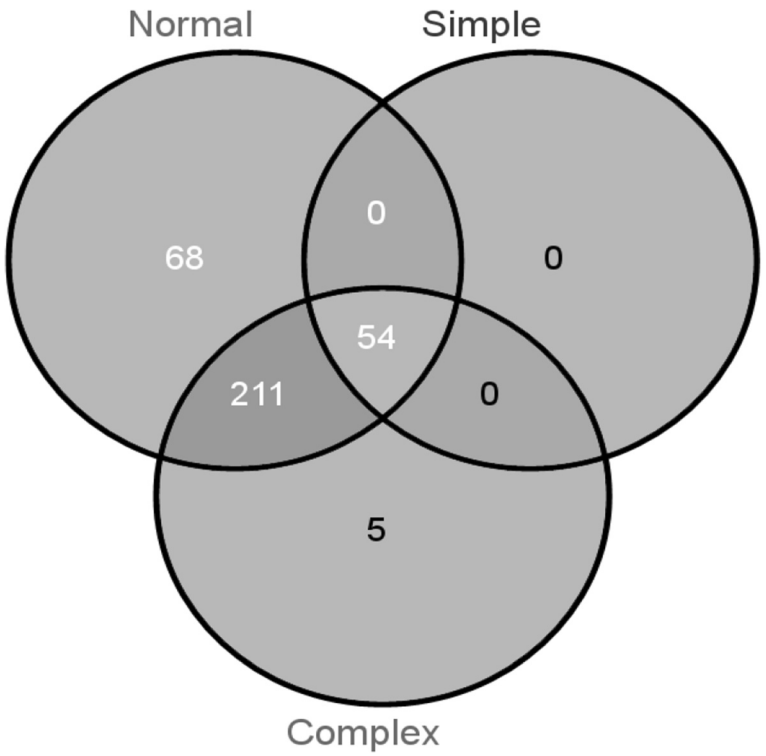



**Figure 4 Venn diagram showing numbers of different CREs present in each group (complex, simple and normal regulation).** Complex regulation was considered when the CRE occurrence was greater than or equal to the average of occurrences in all up-regulated genes plus two standard deviations. The group of simple regulation was considered when the CRE occurrences was smaller than or equal to the average of occurrences in all up- regulated genes minus two standard deviations. CREs that occurs in the gene promoter in the interval between complex and simple regulation are considered to have normal regulation.

hormonal response (ABA signaling pathway, auxin, ethylene), as found by Quinet et al. 2012. In addition, these authors observed no difference in leaf iron accumulation after iron excess exposure, in contrast to our experiment in *japonica* rice where we obtained twice as much iron as in the rice control. The absorbed iron would be stored in ferritins and immobilized by chelation (Briat et al. 1999; Zancani et al. 2004). However, no changes in gene expression were observed for iron storage proteins like ferritins in our study. Similarly, Quinet et al. (2012) did not observe any change in expression of ferritin genes in leaves during long or short-term responses. In their experiments, ferritins were only up-regulated in roots in short-term responses. Studies in which *OsFER1* and *OsFER2* were observed to be over-expressed in rice (Stein et al. 2009b) used higher iron concentrations than in our conditions.

In a recent study, Sperotto et al. (2010) showed that in eight rice cultivars, in the same conditions, the expression pattern of 25 metal-related genes varied among the 8 cultivars and revealed contrasting levels of iron in seeds between cultivars. Subsequently, Banerjee and Chandel (2011) also observed the expression of 43 metal homeostasis related genes in 12 rice cultivars and showed significant genotypic variation in their levels of expression. We find very few genes in common in the transcriptomic profiling with the short-term response to iron excess in *indica* and *japonica* rice. The differences could be inherent to responses of subspecies, or could be associated with the Kong Pao tolerance (as reported by Engel et al. 2012). We speculate that the differences observed between our experiments and Quinet et al. are due to the rice genotypes and the experimental protocols used in the two studies.

In this study, a large number of transporters have been identified and the GO analysis revealed that transport activity is activated in response to iron excess. These proteins may have an important role in transport and sequestration of iron. However, much still remains to be elucidated on intracellular metal transport involving vacuoles, chloroplasts and mitochondria, although some transporters have been identified with specific iron translocation roles for these compartments (Ishimaru et al. 2006; Lee et al. 2010; Ishimaru et al. 2011; Yuan et al. 2011; Zhang et al. 2012). Despite these advances, the coordinated function of different transporters that have a role in iron homeostasis is not fully understood. It will be interesting to characterize the different genes highlighted in our study.

Our study highlights the importance of oxidative stress and detoxification of ROS. In fact, oxidative stress is a harmful process with potentially deleterious effects on plant metabolism that have to be avoided by mobilization of antioxidant defense. However, Kuzniak (2010) pointed out that ROS are key regulators of biological processes in plant biology. It is now generally accepted that the effects of ROS result from direct or indirect responses of sensing systems involving antioxidants, rather than from oxidative damage to bio-molecules. For example, glutathione metabolism is a pathway that is involved in the antioxidative system of plants, reduced glutathione (GSH) playing a central role by scavenging ROS (Ranieri et al. 2005). It is generally considered that GSH content positively correlates with metal stress (Tausz et al. 2004) and participates in plant defense against oxidative stress and toxicity generated from heavy metals (Marrs 1996). The stimulation by iron of superoxide dismutase, ascorbate peroxidase and gluthatione reductase, as observed in our conditions, has already been highlighted in rice (Fang et al. 2001 and Majerus et al. 2007). Another group of up-regulated target genes identified in this work is composed of heavy metal detoxification genes, such as cytochrome P450, metallothioneins, and MATE family of transporters that can help to reduce ROS production in mitochondria. Most up-regulated genes encode proteins that act in mitochondria or plastids, organelles that are involved in iron homeostasis and ROS production.

Very recently, Nakashima et al. (2014) performed promoter analysis of drought-responsive genes in rice plants among which *OsNAC6,* that we found up-regulated under iron excess. Transient assays using promoters indicated that AREB/ABF (ABA-responsive element-binding protein) transcription factors enhanced expression of this gene and GUS assays revealed that the *OsNAC6* promoter was induced by drought, high salinity and ABA treatment (Nakashima et al. 2014). In our study, the *in silico* analysis revealed a large number of up-regulated genes (*ca.* 30% of up-regulated genes) possessing ABA cis-regulatory elements. These results suggest that ABA may have an important role in response to iron toxicity.

The analysis of CREs allowed us to identify common CREs between promoters of genes and the LTRs of LTR-retrotransposons up-regulated in our condition, which are sequences (CREs) potentially involved in driving expression in iron excess. Though these sequence patterns require experimental validation, our current findings may open new avenues for studying the regulation of gene and LTR-retrotransposon expression in iron excess. It has already been demonstrated that some LTRs harbor cis-regulatory signals that confer responsiveness to various external stimuli and play a role in reactivation of transposition (Takeda et al. 1999). In addition, LTRs can serve as alternative promoters or enhancers to regulate genes as far away as 40 kb (Pi et al. 2007; Romanish et al. 2007). Our analysis emphasizes the role of LTR-retrotransposons in the evolutionary and environmental adaptation of plants. Furthermore transposable elements are maintained silent by epigenetic mechanisms in normal conditions (Feschotte et al. 2002). It would be interesting to further analyze whether the silencing mechanisms are alleviated in our iron toxicity conditions.

In this paper, our microarray allowed identification of a large number of up-regulated genes and activated LTR-retrotransposons and highlighted pathways involved in response to iron excess in *japonica* rice. We aimed to identify the CREs associated with this expression data using *in silico* prediction tools. The up-regulated genes present putative cis-regulatory elements in their promoter region indicating that many transcription factors may modulate the expression of these genes, and that these genes are probably multi-stress responsive. In addition, some CREs were common to genes and LTR-retrotransposons. These motifs could be used to design experimental verification of regulatory elements and the identification of transcription factors that regulate gene expression under iron excess. In fact, gene regulation is a crucial step to allow plant adaptation in response of fluctuating environments. Plants induce or repress various genes related to iron homeostasis in response to iron excess. Although our results shed new light on the response to iron excess, much remains to be discovered on the efficient strategies of stress adaptation.

## Methods

### Plant material

Pre-germinated seedlings of rice (cv. Nipponbare) with *ca*. 1 cm of rootlet were transferred to pots containing a complete nutrient solution as described by Camargo and Oliveira (1981): 4 μM $Ca(NO3)_2$; 2 μM $MgSO_4$; 4 μM $KNO_3$; 0.435 μM $(NH4)_2SO_4$; 0.5 μM $KH_2PO_4$; 2 μM $MnSO_4$; 0.3 μM $CuSO_4$; 0.8 μM $ZnSO_4$; 30 μM NaCl; 10 μM Fe-EDTA; 0.10 μM $Na_2MoSO_4$; 10 μM $H_3BO_3$ and were grown in these conditions for 14 days. The nutrient solution was changed every week, the pH was adjusted to 5.5. Subsequently, for iron excess treatment, the plants were transferred to pots containing nutrient solution supplemented with 7 mM of $Fe^{2+}$, at pH 4.5 for four days. Control treatment plants were also changed on the 14th day to a complete nutrient solution at pH 4.5. The experiment consisted of three replicates in a completely random design, each replicate consisting of 100 seedlings. During the experiment, the photoperiod was 16 hours and photon flux density of 25 μmol $m^{-2}$ $s^{-1}$. On the 18th day, the leaves were collected and stored at −80°C until extraction of total RNA.

### RNA extraction and cDNA synthesis

Total RNA was extracted from a 2 g sample of each replicate according to the protocol extraction TRIzol™ Reagent (Invitrogen, Carlsbad, CA, USA) and purified with the RNeasy plant mini kit (Qiagen, Courtaboeuf, France) according to the manufacturer's protocol. The quality of RNA was verified by electrophoresis on an agarose gel (1%) and RNAs assayed by spectrophotometry at a wavelength of 260 nm. cDNA synthesis was performed from mRNA according to the protocol Synthesis of Double-Stranded cDNA (NimbleGen) using the SuperScript™ Double-Stranded cDNA Synthesis Kit (Invitrogen), using oligo-dT(15) primers (Promega). Subsequently, the quantity of cDNA was measured by spectrophotometry, retaining samples with a UV absorbance ratio > 1.7 at A260/A280 and > 1.5 at A260/A230.

### Microarray hybridizations and designs

Six hybridizations were performed, three with samples from control conditions that grew in nutritive solution and three with samples treated with 390 mg $L^{-1}$ of $Fe^{2+}$. We designed an oligomer microarray, similar to one which was described previously (Picault et al. 2009). The oligomer microarray was produced by NimbleGen™ (Madison, WI, USA) and is composed of about 385,000 60mer probes selected for their GC content, Tm and number of cycles needed to synthesize the oligomers. This chip contains 90,000 probes representing 45,000 genes (2 probes per gene) of rice *Oryza sativa* ssp. *japonica* and 290,000 probes representing copies of LTR retrotranposons. Gene probes were designed at the 3′ end of the genes to allow detection by hybridization with potentially partial reverse transcriptase products. In contrast, LTR-retrotransposons are represented throughout their length at the rate of one oligonucleotide for each 500 bp.

When it was possible, oligonucleotides were designed to be unique in the genome (*i.e.* locus specific), to avoid problems of oligonucleotide redundancy on the chip. Hybrids with up to three mismatches were considered to be stable enough to withstand the conditions of washing after the chip hybridization during hybridization between a cDNA and an oligonucleotide. The oligonucleotides are therefore considered to be locus specific when they do not match elsewhere and have less than four mismatches.

### Data verification, spatial effect correction and normalization

Data verification was performed by comparing the average, standard deviation and quantiles of each sample, in order to highlight the samples which present excessively different contributions. This verification was performed with the R statistical software (R Development Core Team 2012) using the boxplot and summary command. This enabled us to identify any sample with inappropriate hybridizations.

The first step consists in reducing the space bias effect for each hybridization. Space biases are due mainly to poor washing after hybridization, or a misallocation of signal, manifested by a gradient of intensity. These biases were corrected with the script SpatialSmooth of NMPP software (NimbleGen Microarray Data Processing Pipeline - NMPP) (Wang et al. 2006) through a global distance-weighted smoothing algorithm. After the spatial smooth, quantile normalization adjusts values in a microarray experiment to improve consistency and reduce technical biases and variations between hybridizations. Normalization between hybridizations was performed with the QuantileNorm script of NMPP software (Wang et al. 2006). Quantile normalizations force the arrays of a set of experiments to have absolutely identical distribution. It is based on the assumption that the RNA populations hybridized to the arrays should be the same. The global normalization adjusts each condition to the same baseline (the median) in order to allow the hybridizations of the different conditions to be comparable. Analysis of variance with a false discovery rate adjustment method was realized (Benjamini and Hochberg 1995).

The results of different treatment comparison were calculated as $Log_2$-fold change. The oligonucleotides selected were those which presented a two-fold increase or decrease in expression, *i.e.*, a log-fold change smaller or equal to −1 for down-regulation, and greater or equal to 1 for up-regulation. Oligonucleotides displaying P-value ≤0.05 for the statistical test were selected.

### Oligonucleotide annotation and gene ontology

A local alignment (BLAST) (Altschul et al. 1990) for all oligonucleotides differentially expressed on the chip by considering a $\log_2$-fold-change ≥ 1 and ≤ −1 was carried out using the database IRGSP 1.0. Gene ontology and metabolic pathway classifications were performed using the software Blast2GO (Conesa, et al. 2005) with Gene Ontology (Ashburner et al. 2000) and KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa et al. 2012) databases, respectively. The Venn diagram was performed with Venny tool (Oliveros 2007).

### Experimental validation of microarray by RT-qPCR

Chip validation was performed by RT-qPCR (reverse transcriptase quantitative PCR) using primer pairs for 16 up-regulated and 3 down-regulated genes. Primer design was performed in Primer Express® software (Applied Biosystems, California, United States). Total RNA was digested with DNase I™. The RNA quality was checked by agarose gel electrophoresis and RNA quantity by measuring the absorbance at 260 nm. cDNA first-strand

synthesis was performed with a SuperScript™ First-Strand Synthesis System for RT-PCR (Invitrogen) from 2 μg of RNA. RT-qPCR was performed in an Applied Biosystems 7500 Fast Real-Time PCR System using a SYBR Green detection system (Applied Biosystems, California, United States). The ΔΔCt relative quantification method (Livak and Schmittgen 2001), in which the expression data of the target genes were normalized with the level of expression of GAPDH (Glyceraldehyde 3-phosphate dehydrogenase) as reference gene, were used, with three technical replicates. The RT-qPCR experiment was performed according to MIQE guidelines (Bustin et al. 2009). Amplification was done with Taq Platinum (Invitrogen) with the following program: 50°C for 30s, 95°C for 10s; 40 cycles of 95°C for 30s, 60°C for 1 min, and 72°C for 1 min and a final elongation at 72°C for 5 min. Pearson's correlation was performed between the data obtained by microarray and RT-qPCR for each gene validated, using the Hmisc Package v. 2.14.0 of the R statistical software (R Development Core Team 2012).

### Quantification of micronutrients in the tissue of rice shoots

The content of micronutrients such as copper (Cu), zinc (Zn), manganese (Mn) and iron (Fe) was determined according to the methodology described by Tedesco et al. (1995) using three replicates for each condition. The data obtained in atomic absorption spectrometry (Thermo Scientific) was calculated based on a standard curve for each element. The non-normality and heterogeneity of the data was checked and an ANOVA F-test was performed using the R statistical software (R Development Core Team 2012).

### Cis-Regulatory Element (CREs) pattern search in putative promoters of up-regulated genes and LTRs of retrotransposons

The putative promoters (1.0 Kbp upstream portions of start site of transcription) of up-regulated genes (2,457 genes considering $\log_2 FC \geq 1$) and the LTRs of LTR-retrotransposons were obtained from RAP-DB database. PLACE – a database of plant *cis*-acting regulatory DNA elements (http://www.dna.affrc.go.jp/PLACE/index.html) (Higo et al. 1999) and PLANTICS (Pegoraro et al. 2013) were employed to search for information (ID, consensus sequences, TFs related) about reported *cis*-acting regulatory elements (CREs). A Z score for the occurrences for each of 469 CREs in the 2,457 putative promoters of up-regulated genes was calculated in order to verify if the probability of the results found was not random. A cutoff of 0.05 (or 5%) was used to eliminate false positives (Rombauts et al. 2003).

## Additional files

**Additional file 1: Table S1.** List of 2525 genes differentially expressed in leaves of 18-day-old rice seedlings (*Oryza sativa* ssp. *japonica* cv. Nipponbare) after four days of iron excess exposure: 2.457 were up-regulated and 68 were down-regulated: locus ID, fold-change (log2FC) and description of genes.

**Additional file 2: Table S2.** Number of up and down-regulated genes in leaves of 18-day-old rice seedlings (*Oryza sativa* ssp. *japonica* cv. Nipponbare) after four days of iron excess exposure, and its respective metabolic pathway generated by Kegg (Kyoto Encyclopedia of Genes and Genomes).

**Additional file 3: Table S3.** Number of up-regulated genes involved in biosynthesis of plant hormones in leaves of 18-day-old rice seedlings (*Oryza sativa* ssp. *japonica* cv. Nipponbare) after four days of iron excess exposure, generated by Kegg (Kyoto Encyclopedia of Genes and Genomes).

**Additional file 4: Table S4.** Number of significant occurrences of ABA responsive *cis*-regulatory elements in promoter (1 kbp upstream) of up-regulated genes in leaves of 18-day-old rice seedlings (*Oryza sativa* ssp. *japonica* cv. Nipponbare) after four days of iron excess exposure: locus ID, fold-change (log2FC) and description of 20-ABADESI1, 21- ABADESI2, 25-ABRE2HVA1, 27-ABRE3HVA1, 28-ABRE3HVA22, 30-ABREA2HVA1, 31-ABREATCONSENSUS, 32-ABREATRD22, 33-ABREAZMRAB28,- 34-BREBNNAPA, 35-ABREBZMRAB28, 36-ABRECE1HVA22, 39-ABREDISTBBNNAPA, 40-ABRELATERD1, 41-ABREMOTIFAOSOSEM, 42-BREMOTIFIIIOSRAB16B, 43-ABREMOTIFIOSRAB16B, 44-ABREOSRAB21, 45-ABRERATCAL, 46-ABRETAEM, 47-ABREZMRAB28, 50-ACGTABREMOTIFA2OSEM, and 51-ACGTABREMOTIFAOSOSEM. NBS – number of binding sites, NDCRE (number of different cis-regulatory elements).

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

OP, ACO, NP and TF designed the study. TF, NP, DRF, CCM, APSB performed the experiments. CC designed the microarray. TF, CC, LGW, LC analyzed the data and performed the bioinformatics' approach. TF and NP wrote the manuscript, which was further edited by ACO and OP. All authors read and approved the final manuscript.

#### Acknowledgements

The authors thank to Prof. Dr. Rogério O. de Sousa and Roberto C.D. Wolter for technical assistance in the laboratory of plant nutrition. This work was supported by CAPES-COFECUB (BEX: 3455/08-4), CNPq and FAPERGS. We thank R. Cooke for his useful comments on the manuscript.

#### Author details

[1]Plant Genomics and Breeding Center, Eliseu Maciel School of Agronomy, Federal University of Pelotas, 96010-610 Pelotas, RS, Brazil. [2]Laboratoire Génome et Développement des Plantes, UMR 5096, Université de Perpignan Via Domitia, F-66860 Perpignan, France. [3]CNRS, Laboratoire Génome et Développement des Plantes, UMR 5096, F-66860 Perpignan, France. [4]Present address: Universidade Tecnológica Federal do Paraná, Campus Pato Branco, 85503-390 Pato Branco, PR, Brazil. [5]Present address: Laboratoire Ecologie et Evolution des Interactions, UMR 5244, F-66860, Université de Perpignan Via Domitia, Perpignan, France.

Received: 21 November 2014 Accepted: 21 January 2015
Published online: 25 February 2015

#### References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25–29

Fofana AA (2009) Rice yiel gap due to iron toxicity in west Africa. J Agron Crop Sci 195:66–76
Banerjee S, Chandel G (2011) Understanding the role of metal homeostasis related candidate genes in Fe/Zn uptake, transport and redistribution in rice using semi-quantitative RT-PCR. J Plant Mol Biol Biotechnol 1:33–46
Becker M, Asch F (2005) Iron toxicity in rice - conditions and management concepts. J Plant Nutr Soil Sci 168:558–573
Belouchi A, Kwan T, Gros P (1997) Cloning characterization of the OsNramp family from *Oryza sativa*, a new family of membrane proteins possibly implicated in the transport of metal ions. Plant Mol Biol 33:1085–1092
Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc B Met 57:289–300
Bennetzen JL (2007) Patterns in grass genome evolution. Curr Opin Plant Biol 10:176–181
Briat JF, Lobréaux S, Grignon N, Vansuyt G (1999) Regulation of plant ferritin synthesis: how and why. Cell Mol Life Sci 56:155–166
Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem 55:611–622
Butelli E, Licciardello C, Zhang Y, Liu J, Mackay S, Bailey P, Reforgiato-Recupero G, Martin C (2012) Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. Plant Cell 3:1242–1255
Camargo CEO, Oliveira OF (1981) Tolerance of wheat cultivars to different aluminum levels in nutrient solution and soil (in Portuguese with English abstract). Bragantia 49:21–23
Chaparro C, Guyot R, Zuccolo A, Piégu B, Panaud O (2007) RetrOryza: a database of the rice LTR-retrotransposons. Nucleic Acids Res 35:D66–D70
Chen WW, Yang JL, Qin C, Jin CW, Mo JH, Ye T, Zheng SJ (2010) Nitric oxide acts downstream of auxin to trigger root ferric-chelate reductase activity in response to iron deficiency in *Arabidopsis*. Plant Physiol 154:810–819
Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676
Connolly EL, Guerinot M (2002) Iron stress in plants. Genome Biol 3:REVIEWS1024
Curie C, Cassin G, Couch D, Divol F, Higuchi K, Le Jean M, Misson J, Schikora A, Czernic P, Mari S (2009) Metal movement within the plant: contribution of nicotianamine and yellow stripe 1-like transporters. Ann Bot 103:1–11
Dabarni B, Briat JF, Holm PB, Husted S, Noeparvar S, Borg S (2013) Dissecting plant iron homeostasis under short and long-term iron fluctuations. Biotechnol Adv 31:1292–1307
Duc C, Celler F, Lobreaux S, Briat JF, Gaymard F (2009) Regulation of iron homeostasis in *Arabidopsis thaliana* by the clock regulator time for coffee. J Biol Chem 284:36271–36281
El Baidouri M, Panaud O (2013) Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. Genome Biol Evol 5:954–965
Engel K, Asch F, Becker M (2012) Classification of rice genotypes based on their mechanisms of adaptation to iron toxicity. J Plant Nutr Soil Sci 175:871–881
Fang WC, Wang JW, Lin CC, Kao CH (2001) Iron induction of lipid peroxidation and effects on antioxidative enzyme activities in rice leaves. Plant Growth Regul 35:75–80
Feng H, Guan D, Sun K, Wang Y, Zhang T, Wang R (2013) Expression and signal regulation of the alternative oxidase genes under abiotic stresses. Acta Biochim Biophys Sin 45:985–994
Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meet genomics. Genetics 3:329–341
Foster R, Izawa T, Chua NH (1994) Plant bZIP proteins gather at ACGT elements. FASEB J 8:192–200
Fu Y, Zhao W, Peng Y (2007) Induced expression of oryzain alpha gene encoding a cysteine proteinase under stress conditions. J Plant Res 120:465–469
Grandbastien MA (1998) Activation of plant retrotransposons under stress conditions. Trends Plant Sci 3:181–187
Green PJ, Yong MH, Cuozzo M, Kano-Murakami Y, Silverstein P, Chua NH (1988) Binding site requirements for pea nuclear protein factor GT-1 correlate with sequences required for light-dependent transcriptional activation of the rbcS-3A gene. EMBO J 7:4035–4044
Grennan AK (2006) Abiotic stress in rice: an "omic" approach. Plant Physiol 140:1139–1141
Gross J, Stein RJ, Fett-Neto AG, Fett JP (2003) Iron homeostasis related genes in rice. Genet Mol Biol 26:477–497
Guerinot ML (2000) The ZIP family of metal transporters. Biochim Biophys Acta 1465:190–198
Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database. Nucleic Acids Res 27:297–300
Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, Kanda M (1996) Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci U S A 93:7783–7788
Huang N, Sutliff TD, Litts JC, Rodriguez RL (1990) Classification and characterization of the rice alpha-amylase multigene family. Plant Mol Biol 14:655–668
Inoue H, Takahashi M, Kobayashi T, Suzuki M, Nakanishi H, Mori S, Nishizawa NK (2008) Identification and localisation of the rice nicotianamine aminotransferase gene *OsNAAT1* expression suggests the site of phytosiderophore synthesis in rice. Plant Mol Biol 66:193–203
International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800
Ishimaru Y, Suzuki M, Kobayashi T, Takahashi M, Nakanishi H, Mori S, Nishizawa NK (2005) OsZIP4, a novel zinc-regulated zinc transporter in rice. J Exp Bot 56:3207–3214
Ishimaru Y, Suzuki M, Tsukamoto T, Suzuki K, Nakazono M, Kobayashi T, Wada Y, Watanabe S, Matsuhashi S, Takahashi M, Nakanishi H, Mori S, Nishizawa NK (2006) Rice plants take up iron as an $Fe^{3+}$- phytosiderophore and as $Fe^{2+}$. Plant J 45:335–346
Ishimaru Y, Masuda H, Bashir K, Inoue H, Tsukamoto T, Takahashi M, Nakanishi H, Aoki N, Hirose T, Ohsugi R, Nishizawa NK (2010) Rice metal-nicotianamine transporter, OsYSL2, is required for the long-distance transport of iron and manganese. Plant J 62:379–390
Ishimaru Y, Kakei SH, Bashir K, Sato Y, Sato Y, Uozumi N, Nakanishi H, Nishizawa NK (2011) A rice phenolic efflux transporter is essential for solubilizing precipitated apoplasmic iron in the plant stele. J Biol Chem 286:24649–24655
Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40:D109–D114
Kim SA, Guerinot ML (2007) Mining iron: iron uptake and transport in plants. FEBS Lett 581:2273–2280
Kobayashi T, Nishizawa NK (2012) Iron uptake, translocation, and regulation in higher plants. Annu Rev Plant Biol 63:131–152
Koike S, Inoue H, Mizuno D, Takahashi M, Nakanishi H, Mori S, Nishizawa NK (2004) OsYSL2 is a rice metal-nicotianamine transporter that is regulated by iron and expressed in the phloem. Plant J 39:415–424
Kumar A, Bennetzen JL (1999) Plant retrotransposons. Annu Rev Genet 33:479–532
Kuzniak E (2010) The ascorbate-glutathione cycle and related redox signals in plant-pathogen interactions. In Anjum, NA, Umar, S, Chan, M-T (eds) Ascorbate-Glutathione Pathway and Stress Tolerance in Plants. Springer; New York p 115–136.
Lee S, An G (2009) Over-expression of OsIRT1 leads to increased iron and zinc accumulations in rice. Plant Cell Environ 32:408–416
Lee S, Kim SA, Lee J, Guerinot ML, An G (2010) Zinc deficiency-inductible OsZIP8 encodes a plasma membrane-localized zinc transporter in rice. Mol Cells 29:551–558
Lelievre JM, Oliveira LO, Nielsen NC (1992) 5′CATGCAT-3′ elements modulate the expression of glycinin genes. Plant Physiol 98:387–391
Lingam P, Mohrbacher J, Brumbarova T, Potuschak T, Fink-Straube C, Blondet E, Genschik P, Bauer P (2011) Interaction between the bHLH transcription factor FIT and ETHYLENE INSENSITIVE3/ETHYLENE INTENSIVE3-LIKE1 reveals molecular linkage between the regulation of iron acquisition and ethylene signaling in *Arabidopsis*. Plant Cell 23:1815–1829
Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) Method. Methods 25:402–408
Ma J, Devos KM, Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res 14:860–869
Majerus V, Bertin P, Swenden V, Fortemps A, Lobreaux S, Lutts S (2007) Organ-dependent responses of the African rice to short-term iron toxicity: ferritin regulation and antioxidative responses. Biol Plant 51:303–331
Mangeon A, Magioli C, Tarré E, Cardeal V, Araujo C, Falkenbach E, Rocha CAB, Rangel-Lima C, Sachetto-Martins G (2010) The tissue expression pattern of the AtGRP5 regulatory region is controlled by a combination of positive and negative elements. Plant Cell Rep 29:461–471
Marrs KA (1996) The functions and regulation of glutathione S-transferase in plants. Annu Rev Plant Physiol Plant Mol Biol 47:127–158

Marschner H (1995) Mineral nutrition of higher plants, 2nd edn. Academic, London
Mitsuda N, Ohme-Takagi M (2009) Functional analysis of transcription factors in *Arabidopsis*. Plant Cell Physiol 50:1232–1248
Nakashima K, Tran LS, Van Nguyen D, Fujita M, Maruyama K, Todaka D, Ito Y, Hayashi N, Shinozaki K, Yamaguchi-Shinozaki K (2007) Functional analysis of NAC-type transcription factor OsNAC6 involved in abiotic and biotic stress-responsive gene expression in rice. Plant J 51:617–630
Nakashima K, Jan A, Todaka D, Maruyama K, Goto S, Shinozaki K, Yamaguchi-Shinozaki K (2014) Comparative functional analysis of six drought-responsive promoters in transgenic rice. Planta 239:47–60
Narayanan NN, Vasconcelos MW, Gruzak MA (2007) Expression profiling of *Oryza sativa* metal homeostasis genes in different rice cultivars using a cDNA macroarray. Plant Physiol Biochem 45:277–286
Nath S, Panda P, Mishra S, Dey M, Choudhury S, Sahoo L, Panda SK (2014) Arsenic stress in rice: redox consequences and regulation by iron. Plant Physiol Biochem 80:203–210
Nielsen KA, Moller BL (2005) Cytochrome P450s in plants. In: Ortiz de Montellano PR (ed) Cytochrome P450: structure, mechanism, and biochemistry, 3rd edn. Kluwer Academic/Plenum Publishers, New York, pp 553–583
Oliveros JC (2007) VENNY: an interactive tool for comparing lists with Venn Diagrams., http://bioinfogp.cnb.csic.es/tools/venny/index.html
Pan Y, Michael TP, Hudson ME, Kay EA, Chory J, Schuler MA (2009) Cytochrome P450 monooxygenases as reporters for circadian-regulated pathways. Plant Physiol 150:858–878
Park SY, Yu JW, Park JS, Li J, Yoo SC, Lee NY, Lee SK, Jeong SW, Seo HS, Koh HJ, Jeon JS, Park YI, Paek NC (2007) The senescence-induced staygreen protein regulates chlorophyll degradation. Plant Cell 19:1649–1664
Pegoraro C, Farias DR, Mertz LM, Santos RS, Maia LC, Rombaldi CV, Costa de Oliveira A (2013) Ethylene response factors gene regulation and expression profiles under different stresses in rice. Theor Exp Plant Physiol 25(4):261–274
Pi W, Zhu X, Wu M, Wang Y, Fulzele S, Eroglu A, Ling J, Tuan D (2007) Long-range function of an intergenic retrotransposon. Proc Natl Acad Sci U S A 107:12992–12997
Picault N, Chaparro C, Piegu B, Stenger W, Formey D, Llauro C, Descombin J, Sabot F, Lasserre E, Meynard D, Guiderdoni E, Panaud O (2009) Identification of an active LTR retrotransposon in rice. Plant J 58:754–765
Polle A, Schützendübel A (2003) Heavy metal signalling in plants: linking cellular and organismic responses. In: Hirt H, Shinozaki K (eds) Topics in current genetics, plant responses to abiotic stress, vol 4. Springer, Berlin Heidelberg
Quinet M, Vromman D, Clippe A, Bertin P, Lequeux H, Dufey I, Lutts S, Lefèvre I (2012) Combined transcriptomic and physiological approaches reveal strong differences between short- and long-term response of rice (*Oryza sativa*) to iron toxicity. Plant Cell Environ 35:1837–1859
R Development Core Team, R Foundation for Statistical Computing, (2012), Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org.
Ranieri A, Castagna A, Scebba F, Careri M, Zagnoni I, Predieri G, Pagliari M, di Toppi LS (2005) Oxidative stress and phytochelatin characterisation in bread wheat exposed to cadmium excess. Plant Physiol Biochem 43:45–54
Ricachenevsky FK, Sperotto RA, Menguer PK, Fett JP (2010) Identification of Fe-excess-induced genes in rice shoots reveals a WRKY transcription factor responsive to Fe, drought and senescence. Mol Biol Rep 37:3735–3745
Ricachenevsky FK, Sperotto RA, Menguer PK, Sperb AR, Lopes KL, Fett JP (2011) ZINC-INDUCED FACILIATTOR-LIKE family in plants: lineage-specific expansion in monocotyledons and conserved genomic and expression features among rice (*Oryza sativa*) paralogs. BMC Plan Biol 11:20
Romanish MT, Lock WM, van de Lagemaat LN, Dunn CA, Mager DL (2007) Repeated recruitment of LTR retrotransposons as promoters by the anti-apoptotic locus NAIP during mammalian evolution. PLoS Genet 3:e10
Rombauts S, Florquin K, Lescot M, Marchal K, Rouze P, Van De Peer Y (2003) Computational approaches to identify promoters and cis-regulatory elements in plant genomes. Plant Physiol 132:1162–1176
Sabot F, Picault N, El-Baidouri M, Llauro C, Chaparro C, Piegu B, Roulin A, Guiderdoni E, Delabastide M, McCombie R, Panaud O (2011) Transpositional landscape of the rice genome revealed by paired-end mapping of high-throughput re-sequencing data. Plant J 66:241–246
Saika H, Ohtsu K, Hamanaka S, Nakazono M, Tsutsumi N, Hirai A (2002) AOX1c, a novel rice gene for alternative oxidase; comparison with rice AOX1a and AOX1b. Genes Genet Syst 77:31–38
Schuler MA (1996) The role of cytochrome P450 monooxygenases in plant-insect interactions. Plant Physiol 112:1411–1419
da Silveira VC, Fadanelli C, Sperotto RA, Stein RJ, Basso LA, Santos DS, Vaz Junior IS, Dias JF, Fett JP (2009) Role of ferritin in the rice tolerance to iron overload. Sci Agric 66:549–555
Sperotto RA, Boff T, Duarte GL, Santos LS, Grusak MA, Fett JP (2010) Identification of putative target genes to manipulate Fe and Zn concentrations in rice grains. J Plant Physiol 167:1500–1506
Stein RJ, Duarte GL, Spohr MG, Lopes SIG, Fett JP (2009a) Distinct physiological responses of two rice cultivars subjected to iron toxicity under field conditions. Ann Appl Biol 154:269–277
Stein RJ, Fett JP, Ricachenevsky FK (2009b) Differential regulation of the two rice ferritin genes (*OsFER1* and *OsFER2*). Plant Sci 177:563–569
Takagi S (1976) Naturally occurring iron-chelating compounds in oat- and rice-root washing. I: activity measurement and preliminary characterization. Soil Sci Plant Nutr 22:423–433
Takagi S, Nomoto K, Takemoto T (1984) Physiological aspect of mugineic acid, a possible phytosiderophore of graminaceous plants. J Plant Nutri 7:1–5
Takahashi R, Ishimaru Y, Senoura T, Shimo H, Ishikawa S, Arao T, Nakanishi H, Nishizawa NK (2011) The OsNRAMP1 iron transporter is involved in Cd accumulation in rice. J Exp Bot 62:4843–4850
Takeda S, Sugimoto K, Otsuki H, Hirochika H (1999) A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon Tto1 is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. Plant J 18:383–393
Tausz M, Sircelj H, Grill D (2004) The glutathione system as a stress marker in plant ecophysiology: is a stress-response concept valid? J Exp Bot 55:1955–1962
Tedesco MJ, Gianello C, Bissani CA, Bohnen H, Volkweiss SJ (1995) Análise de solo, plantas e outros materiais, 2nd edn. Universidade Federal do Rio Grande do Sul, Porto Alegre, p 147, Boletim Técnico, 5
Thongbai P, Goodman BA (2000) Free radical generation and post anoxic injury in an iron toxic soil. J Plant Physiol 23:1887–1990
Todaka D, Nakashima K, Shinozaki K, Yamaguchi-Shinozaki K (2012) Toward understanding transcriptional regulatory networks in abiotic stress responses and tolerance in rice. Rice 5:6
Tsutsui T, Yamaji N, Feng Ma J (2011) Identification of a cis-acting element of ART1, a C2H2-type zinc-finger transcription factor for aluminum tolerance in rice. Plant Physiol 156:925–931
Ulmasov T, Hagen G, Guilfoyle TJ (1999) Dimerization and DNA binding of auxin response factors. Plant J 19:309–319
De Victoria FC, Bervald CMP, Da Maia LC, De Sousa RO, Panaud O, De Oliveira AC (2012) Phylogenetic relationships and selective pressure on gene families related to iron homeostasis in land plants. Genome 55:883–900
Wan JL, Zhai HQ, Wan JM, Ikehashi H (2003) Detection and analysis of QTLs for ferrous iron toxicity tolerance in rice, *Oryza sativa* L. Euphytica 131:201–206
Wang H, Chai Y, Chu X, Zhao Y, Wu Y, Zhao J, Ngezahayo F, Xu C, Liu B (2009) Molecular characterization of a rice mutator-phenotype derived from an incompatible cross-pollination reveals transgenerational mobilization of multiple transposable elements and extensive epigenetic instability. BMC Plant Biol 9:63
Wang X, He H, Li L, Chen R, Deng XW, Li S (2006) NMPP: a user-customized nimblegen microarray data processing pipeline. Bioinformatics 22:2955–2957
Wang M, Gruissem W, Bhullar NK (2013) Nicotianamine synthase overexpression positively modulates iron homeostasis-related genes in high iron rice. Front Plant Sci 4:156
Werck-Reichhart D, Bak S, Paquette S (2002) Cytochromes p450. Arabidopsis Book 1:e0028
Wu J, Wang C, Zheng L, Wang L, Chen Y, Whelan J, Shou H (2011) Ethylene is involved in the regulation of iron-acquisition-related genes in *Oryza sativa*. J Exp Bot 62:667–674
Yang A, Li Y, Xu Y, Zhang WH (2013) A receptor-like protein RMC is involved in regulation of iron acquisition in rice. J Exp Bot 16:5009–5020
Yokosho K, Yamaji N, Ueno D, Mitani N, Ma JF (2009) OsFRDL1 is a citrate transporter required for efficient translocation of iron in rice. Plant Physiol 149:297–305
Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *Indica*). Science 296:79–92
Yuan M, Li X, Xiao J, Wang S (2011) Molecular and functional analyses of COPT/Ctr-type copper transporter-like gene family in rice. BMC Plant Biol 11:69
Zancani M, Peresson C, Biroccio A, Federici G, Urbani A, Murgia I, Soave C, Micali F, Vianello A, Macrì F (2004) Evidence for the presence of ferritin in plant mitochondria. Eur J Biochem 271:3657–3664

Zhang Y, Xu YH, Yi HY, Gong JM (2012) Vacuolar membrane transporters OsVIT1 and OsVIT2 modulate iron translocation between flag leaves and seeds in rice. Plant J 72:400–410
Zheng L, Huang F, Narsai R, Wu J, Giraud E, He F, Cheng L, Wang F, Wu P, Whelan J, Shou H (2009) Physiological and transcriptome analysis of iron and phosphorus interaction in rice seedlings. Plant Physiol 151:262–274
Zhou G, Xu Y, Li J, Yang L, Liu JY (2006) Molecular analyses of the metallothionein gene family in rice (*Oryza sativa* L.). J Biochem Mol Biol 39:595–606
Zhou X, Yang Y (2004) Differential expression of rice Nramp genes in response to pathogen infection, defense signal molecules and metal ions. Physiol Mol Plant P 65:235–243

Lors d'un stress ferrique de 4 jours, nous avons mis en évidence que la majorité des gènes différentiellement exprimés est impliquée dans les voies métaboliques des lipides, le métabolisme des carbohydrates, la biosynthèse des métabolites secondaires et des hormones. Nous avons aussi identifié des gènes impliqués dans l'acquisition et la mobilisation du fer, le transport des cations et la détoxification des ROS.

La variété que nous avons utilisée (Nipponbare) est une variété de *japonica* classée comme sensible à l'excès de fer (Engel et al., 2012). Nous avons donc pu comparer nos résultats avec ceux d'une variété d'*indica* tolérante à l'excès de fer (Quinet et al., 2012) et nous avons ainsi pu mettre en évidence les gènes communs et ceux différentiellement exprimés entre ces deux cultivars en fonction de la sensibilité ou de la tolérance au fer.

De plus, pour la première fois, nous avons identifié que le stress ferrique entraîne la sur-expression d'un grand nombre de rétrotransposons à LTR (37%). Grâce à notre base de données, nous avons pu réaliser un inventaire exhaustif des éléments de manière locus spécifique. Nous avons essayé de valider expérimentalement les candidats obtenus pour les co-transcrits gène-ET locus par locus, afin de confirmer chaque corrélation, mais aucun candidat n'a pu être confirmé.

Afin de comprendre les mécanismes contrôlant l'expression des gènes en réponse à un excès de fer, il était important de savoir si des éléments régulateurs (cis-acting regulatory elements (CREs)) spécifiques étaient présents dans les promoteurs des gènes différentiellement exprimés. Nous avons étudié les CREs présents dans les promoteurs des gènes mais aussi dans les LTR des rétrotransposons à LTR. Nous avons observé qu'au moins 27 % des gènes sur-exprimés possèdent au moins un CRE de réponse à l'acide abscissique (ABA) dans leur promoteur, cette hormone pourrait donc avoir un rôle important dans la réponse à la toxicité du fer. De plus, parmi les gènes et les éléments sur-exprimés, nous avons identifié des CREs communs entre les LTR des rétrotransposons à LTR et les promoteurs des gènes. Il serait intéressant de tester si ces CREs sont potentiellement impliqués dans la régulation de l'expression lors d'un excès de fer.

Nos travaux renforcent l'idée que les rétrotransposons à LTR participent à la réponse transcriptionnelle lors d'un stress et peuvent conférer un avantage adaptatif à la plante. Bien que nos résultats donnent un nouvel éclairage sur la compréhension des mécanismes de réponse à un excès de fer, beaucoup reste à découvrir sur les stratégies d'adaptation des plantes.

### *2.2.2. Coopération riz-bactérie*

De 2009 à 2012, j'ai été responsable du projet ANR AZORIZ (coord. F. Wisniewski-Dye, Université Lyon1) : « Specificity of the phytostimulatory cooperation between *Azospirullum lipoferum* and rice » (2009-2011).

Divers microorganismes naturellement présents dans le sol peuvent s'associer avec les racines des plantes et stimuler leur croissance. Certaines bactéries représentent une alternative intéressante dans un contexte d'agriculture durable et sont utilisées comme biofertilisant ou comme agent de lutte biologique. *Azospirillum* est la principale bactérie phytostimulatrice des céréales et est utilisée comme biofertilisant dans divers pays (notamment en Amérique latine et centrale) pour subvenir aux besoins de l'alimentation humaine et animale. Contrairement à la symbiose entre les légumineuses et *Rhizobium*, l'association entre les céréales et *Azospirillum* ne conduit pas à la formation d'organes nouveaux et les mécanismes moléculaires mis en jeu lors de cette interaction sont peu documentés. De plus, l'association préférentielle entre une variété de plante à un type d'*Azospirillum* a été fréquemment observée, mais jamais expliquée au niveau moléculaire. Le projet AZORIZ visait à élucider les mécanismes de l'association *Azospirillum*-plante, avec pour objectif ultime une utilisation efficace et rationnelle de cette bactérie en agronomie.

Deux variétés de riz (Nipponbare et Cigalon) ont été inoculées par deux *Azospirillum* (B510 et 4B) dans un système simplifié (en l'absence d'autres microorganismes). L'association a été analysée au niveau de chaque partenaire 7 jours après inoculation par : (i) une analyse par microscopie permettant de visualiser la colonisation des racines du riz par *Azospirillum*, (ii) une mesure des paramètres de croissance des plantes, (iii) une analyse du contenu des métabolites secondaires de la plante (les métabolites secondaires ne sont pas impliqués dans le développement de la plante mais permettent notamment à la plante de se défendre et d'attirer des pollinisateurs), (iv) une analyse de l'expression des gènes de la plante, (v) une analyse de l'expression des gènes de la bactérie.

Le volet me concernant correspondait plus particulièrement à la partie (iv) du projet. Nous avons utilisé la puce « transposome », décrite précédemment, pour étudier les interactions entre le riz et les bactéries rhizosphériques phytostymulatrices (PGRP). Les analyses globales d'expression des gènes et des rétrotransposons à LTR chez les 2 cultivars ont été réalisées sur les racines 7 jours après inoculation avec *Azospirillum lipoferum* B510 (isolé de Nipponbare) ou *Azospirillum sp*. 4B (isolé de Cigalon). Chaque cultivar a été inoculé avec la souche d' *Azospirillum* d'origine

(Nip_B510, Cig_4B). Pour tester la spécificité de la réponse, nous avons aussi réalisé les inoculations inverses (Nip_4B, Cig_B510). Des plantes non inoculées ont servi de témoins. Je me suis rendue à Lyon pour former l'étudiante en thèse (Amel Chamam) pour la réalisation des expériences liées à l'approche transcriptomique. J'ai ensuite aidé le doctorant Benoît Drogue, pour l'analyse et l'interprétation des résultats. Dans le cadre de ce projet, nous avons recruté un CDD (Michaël Mozar) d'un an que j'ai supervisé pour l'analyse des données de bio-informatique. J'ai aussi réalisé toutes les validations par Q-PCR et ce travail a donné lieu à une publication (Drogue et al., 2014).

frontiers in
PLANT SCIENCE

ORIGINAL RESEARCH ARTICLE
published: 06 November 2014
doi: 10.3389/fpls.2014.00607

# Plant root transcriptome profiling reveals a strain-dependent response during *Azospirillum*-rice cooperation

**Benoît Drogue[1], Hervé Sanguin[1†], Amel Chamam[1], Michael Mozar[2†], Christel Llauro[2], Olivier Panaud[2], Claire Prigent-Combaret[1]*, Nathalie Picault[2] and Florence Wisniewski-Dyé[1]**

[1] Ecologie Microbienne, UMR CNRS 5557/USC INRA 1364, Université Lyon 1, Villeurbanne, France
[2] Laboratoire de Génome et Développement des Plantes, UMR 5096 CNRS/IRD/Université de Perpignan Via Domitia, Perpignan, France

**Edited by:**
*Essaid Ait Barka, Reims University, France*

**Reviewed by:**
*Raffaella Balestrini, Consiglio Nazionale delle Ricerche, Italy*
*Fanhong Meng, Texas A&M University, USA*

***Correspondence:**
*Claire Prigent-Combaret, Ecologie Microbienne, UMR CNRS 5557/USC INRA 1364, Université Lyon 1, 43 Boulevard du 11 Novembre 1918, 69622 Villeurbanne Cedex, France*
*e-mail: claire.prigent-combaret@univ-lyon1.fr*

**†Present address:**
*Hervé Sanguin, CIRAD, UMR LSTM, Montpellier, France;*
*Michael Mozar, Laboratoire Information Génomique et Structurale, CNRS - UMR7256 Marseille, France*

Cooperation involving Plant Growth-Promoting Rhizobacteria results in improvements of plant growth and health. While pathogenic and symbiotic interactions are known to induce transcriptional changes for genes related to plant defense and development, little is known about the impact of phytostimulating rhizobacteria on plant gene expression. This study aims at identifying genes significantly regulated in rice roots upon *Azospirillum* inoculation, considering possible favored interaction between a strain and its original host cultivar. Genome-wide analyzes of *Oryza sativa japonica* cultivars Cigalon and Nipponbare were performed, by using microarrays, seven days post-inoculation with *Azospirillum lipoferum* 4B (isolated from Cigalon) or *Azospirillum sp.* B510 (isolated from Nipponbare) and compared to the respective non-inoculated condition. A total of 7384 genes were significantly regulated, which represent about 16% of total rice genes. A set of 34 genes is regulated by both *Azospirillum* strains in both cultivars, including a gene orthologous to PR10 of *Brachypodium*, and these could represent plant markers of *Azospirillum*-rice interactions. The results highlight a strain-dependent response of rice, with 83% of the differentially expressed genes being classified as combination-specific. Whatever the combination, most of the differentially expressed genes are involved in primary metabolism, transport, regulation of transcription and protein fate. When considering genes involved in response to stress and plant defense, it appears that strain B510, a strain displaying endophytic properties, leads to the repression of a wider set of genes than strain 4B. Individual genotypic variations could be the most important driving force of rice roots gene expression upon *Azospirillum* inoculation. Strain-dependent transcriptional changes observed for genes related to auxin and ethylene signaling highlight the complexity of hormone signaling networks in the *Azospirillum*-rice cooperation.

**Keywords: *Azospirillum*, hormone signaling, plant defense, plant growth-promoting rhizobacteria, rice, transcriptome**

## INTRODUCTION

Rhizodeposition supports growth of a wide range of microorganisms able to establish intimate interactions with plant roots. In the case of parasitism, nutritional requirements of the microbe partner are supported at the expense of plant development and reproduction (O'Brien et al., 2011; Schumacher and Tudzynski, 2012). In the case of mutualism, the interaction leads to a nutritional exchange so that costs and benefits are reciprocally shared by both microbial and plant partners (Smith and Read, 2008). Whether engaged in a parasitic or mutualistic interaction, the microbial partner is perceived as an intruder and the success of the adaptation strategy partly depends on the microbe's ability to bypass defense mechanisms and invade plant tissues (Soto et al., 2009). Then, plant immune response involves gene expression changes that mediate trade-off between defense and development to ensure plant survival through an efficient allocation of resources (Buscaill and Rivas, 2014). Cooperation involving Plant Growth-Promoting Rhizobacteria (PGPR) results in improvements of plant growth and health; however, the invasion of root tissues is not a critical step in successful interaction as several efficient strains are described as root-surface colonizers (Lugtenberg and Kamilova, 2009; Chamam et al., 2013). If mechanisms directly implicated in plant growth-promotion have been extensively studied, most of these works have assessed the impact of PGPR on plant morphological traits and little is known about changes induced at the molecular level (Bashan and de-Bashan, 2010; Galland et al., 2012; van de Mortel et al., 2012; Wisniewski-Dyé et al., 2013).

For more than 50 years, PGPR of a wide range of genera including *Acetobacter*, *Azospirillum*, *Bacillus*, *Burkholderia*, *Herbaspirillum*, *Phyllobacterium* or *Pseudomonas* have been

known for stimulating the growth of numerous host plants (Desbrosses et al., 2009; Lugtenberg and Kamilova, 2009; Richardson et al., 2009; Kumar et al., 2011; Saharan and Nehra, 2011). In particular, the genus *Azospirillum* constitutes an important phytostimulator and an increasing number of field trials are undertaken, principally in India and Latin America where several *Azospirillum* inoculants are commercialized (Steenhoudt and Vanderleyden, 2000; Bashan et al., 2004; Fuentes-Ramirez and Caballero-Mellado, 2012). In most cases, successful inoculation results in root and shoot morphological changes, plant nutrition improvements, and yield enhancements (Richardson et al., 2009; Vacheron et al., 2013; Wisniewski-Dyé et al., 2013). If the phytostimulating effect of *Azospirillum* was originally attributed to its ability to fix atmospheric nitrogen, it is now admitted that the modulation of the phytohormonal balance is the most important mechanism resulting in the modification of root architecture and higher nutrient uptake by the plant (Steenhoudt and Vanderleyden, 2000; Somers et al., 2004; Prigent-Combaret et al., 2008). Besides morphological changes, *Azospirillum* also increases root exudation and modifies the chemical structure of root cell wall (Heulin et al., 1987; El Zemrany et al., 2007). In addition, *Azospirillum* was found capable of increasing the resistance of the host plant against pathogen through mechanisms independent of salicylic acid signaling (Yasuda et al., 2009). Investigation on maize secondary metabolism revealed that major qualitative and quantitative modifications occur following *Azospirillum* inoculation (Walker et al., 2011). Moreover, these modifications depend on bacterial strain/maize cultivar combinations suggesting that a genotype specific perception of *Azospirillum* occurs during the cooperation with maize. These observations were recently strengthened by a study made on two rice cultivars, Cigalon and Nipponbare, after the inoculation of two *Azospirillum* strains isolated from each cultivar (Chamam et al., 2013): *Azospirillum lipoferum* 4B isolated from Cigalon roots (Thomas-Bauzon et al., 1982) and *Azospirillum* sp. B510 isolated from Nipponbare (Elbeltagy et al., 2001). Profiling of secondary metabolites and morphological measurements evidenced that the impact of *Azospirillum* differs according to strain/cultivar combinations and that a specific interaction leading to a stronger phytostimulation occurs between a strain and its original host cultivar. In addition, the endophyte strain B510 was shown to trigger a systemic response, as metabolic changes were observed in both roots and shoots. However, whether or not perception of *Azospirillum* involved plant immune response remains an unanswered question and regulatory mechanisms underlying host-specific metabolic changes have to be unraveled.

In this context, our study aims at characterizing genetic determinants regulated in rice roots at an early stage of the interaction with *Azospirillum*, considering possible favored interaction between a strain and its original host cultivar. Thus, genome-wide analyzes of root gene expression of *Oryza sativa japonica* cultivars Cigalon and Nipponbare were performed 7 days post-inoculation with *A. lipoferum* 4B or *Azospirillum* sp. B510 and compared to the respective non-inoculated condition. A focus was made on genes potentially involved in plant defense and hormone signaling.

## MATERIALS AND METHODS

### BIOLOGICAL MATERIAL

In this study, two rice (*Oryza sativa* L.) cultivars belonging to the *japonica* group, cv. Cigalon (Center Français du Riz, France) and cv. Nipponbare (J. B. Morel, BGPI, Montpellier, France) were inoculated with two diazotrophic strains of the genus *Azospirillum*: *A. lipoferum* 4B initially isolated from rice roots of the cv. Cigalon in France (Thomas-Bauzon et al., 1982) and *Azospirillum* sp. B510 initially isolated from surface sterilized rice stems of the cv. Nipponbare (Elbeltagy et al., 2001).

### RNA SAMPLES AND cDNA SYNTHESIS

Six independent experiments were performed per condition (three for microarray hybridization and three for qRT-PCR validation). Seed sterilization, plant inoculation and plant growth were performed as previously described (Chamam et al., 2013; Drogue et al., 2014). Rice seeds were surface sterilized by washing for 40 min in a sodium hypochlorite solution, rinsed 5 times in demineralized sterile water, and then chlorine traces were removed by washing 3 times in sterile-filtered 2% (w/v) sodium thiosulfate before rinsing 5 times in demineralized sterile water. Surface sterilized seeds were germinated on sterile plant agar ($8\,g{\cdot}L^{-1}$) (Sigma Chemical Co, Saint Louis, USA) for 2 days in the dark at 28°C. Bacterial cells in late-exponential phase were mixed with 50 mL of plant agar ($8\,g{\cdot}L^{-1}$) (to a final concentration of $2.10^7$ cells$\cdot mL^{-1}$) and introduced into 120 × 120 × 17 mm square plates. For both rice cultivars, five disinfected germinated seeds were laid onto the plates and plates were incubated vertically, for 7 days in a growth chamber (MLR350, SANYO, UK) with a photoperiod of 16 h at 28°C (light 150 $\mu E\ m^{-2}\ s^{-1}$), and 8 h at 22°C in the dark. For each experiment, 30 plant root systems were pooled and frozen using liquid nitrogen. Root cell lysis was performed by grinding root systems with a mortar and pestle under liquid nitrogen. Total RNA was isolated using the TRIzol method (Invitrogen, Carlsbad, CA, USA). RNA samples were purified using RNeasy plant mini kit (Qiagen, Courtaboeuf, France) according to the manufacturer's protocol. RNA integrity was assessed using Agilent RNA 6000 Pico Kit (Agilent Technologies, Waldbronn, Germany) and the Agilent 2100 Bioanalyzer (Agilent Technologies) device.

In order to increase mRNA representation in RNA samples, total RNA were digested with mRNA ONLYTM Procaryotic mRNA isolation kit (Epicenter Biotechnologies, Madison, WI, USA) according to the provided protocol.

The microarray cDNA (three independent samples per condition) was synthesized with the Superscript® Double-Stranded cDNA Synthesis Kit (Invitrogen), using a mix (1:1) of random primers (Promega Corporation, Madison, WI, USA) and Oligo-dT (15) primers (Promega).

### MICROARRAY HYBRIDIZATION AND DATA ANALYSIS

We designed an oligo microarray, which was produced by NimbleGenTM (Madison, WI, USA) derived of one which was described previously (Picault et al., 2009). This microarray is composed of about 385,000 60 mer probes selected for their GC content, Tm, and number of cycles needed to synthesize the oligo. This chip contains 90,000 probes representing 45,000 genes

(two probes per gene) of rice *Oryza sativa* ssp. *japonica*, based on the TIGR rice genome annotation version 3.1 genes (Yuan et al., 2005) and 201,691 oligomers corresponding to previously described copies of LTR retrotranposons available on the retrOryza database (www.retroryza.fr; Chaparro et al., 2007). Probes represent 1000 bp of the LTR-retrotransposon flanking regions at the 3′ and 5′ side. The oligonucleotides have been designed at the 3′ end of the genes to detect the readings of reverse transcriptase. On the other hand, the LTR-retrotransposons are represented throughout their length at the rate of a probe every 500 bp. The analysis was performed using the classification proposed by El Baidouri and Panaud (2013) that consists in 369 families and 3623 loci harboring complete elements. Among the differentially expressed oligomers, only those displaying 100% identity and which are unique in the genome were analyzed.

When it was possible, probes have been designed to be unique in the genome (i.e., locus specific) to overcome the problems of oligonucleotides redundancy on the chip. When there were three mismatches during hybridization between a cDNA and an oligonucleotide, hybridization was considered stable enough to withstand the conditions of washing after the chip hybridization. The oligonucleotides are therefore regarded as locus specific when they are not matching elsewhere, but having at most three mismatches, which represents 5% of all oligonucleotides.

For each condition, three independent cDNA samples were labeled and hybridized by Roche Nimblegen according to their standard protocol. Data analysis was performed using Bioconductor microarray packages for R software (http://www.bioconductor.org/). The robust multi-array average (RMA) method associated with quantile normalization was applied (Bolstad et al., 2003; Irizarry et al., 2003). Analysis of variance with a false discovery rate adjustment method was realized (Benjamini and Hochberg, 1995). The results of different treatment comparison were obtained in Log2-fold change. The oligonucleotides selected were those which present a two fold increase or decrease in expression, i.e., a log-fold change smaller or equal to −1 for down-regulation, and greater or equal to 1 for up-regulation. Oligonucleotides displaying $P \leq 0.05$ for the statistical test were selected. For each cultivar, the respective uninoculated condition was used as control. All oligonucleotides differentially expressed were remapped to Os-Nipponbare-Reference-IRGSP version 1.0 (Rice Annotation Project et al., 2008) using BLAST.

### RT-qPCR

For each condition, three independent RNA samples were used to validate gene expression level by performing reverse transcription quantitative real-time PCR (RT-qPCR). Validation was made on a group of 14 representative genes (**Table 1**) using LightCycler® 480 SYBR Green I Master kit (Roche Diagnostics GmbH, Mannheim, Germany) on a LightCycler® 480 Real-Time PCR System (Roche). The actin gene (Os03g0718100) showing an invariant expression was used as reference to normalize RT-qPCR values. After DNase I treatment, total RNA (800 ng) was used for cDNA synthesis using GoScriptTM Reverse transcription system (Promega) with oligo-dT (15) primer in accordance with the manufacturer's protocol.

DNA contamination was checked with reactions that lacked reverse transcriptase as negative controls. Specific primers were designed using Light Cycler Probe design Software 2.0 (Roche) with the following criteria: product size ranges 100–400 pb, primer size comprised between 17 and 22 bases, optimal primer Tm 60°C (Additional file 1: Table S1). Real time PCR conditions were: a denaturation stage of 10 min at 95°C; an amplification stage of 45 cycles of 15 s at 94°C, 10 s at 60°C and 20 s at 72°C; and a melting curve stage of 5 s at 95°C and 1 min at 65°C increased to 97°C with a ramp rate of 0.11°C $s^{-1}$. All reactions were performed in three technical replicates and carried out in LightCycler 480 Multiwell plate 96 (Roche) with adhesive sealings foils (Roche) in a final volume of 10 µl containing 1 µl of each primer (5 µM), 5 µl of master mix and 3 µl of cDNA diluted 50 times. For each cultivar, the respective uninoculated condition was used as the calibrator condition and relative gene expression was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

Data were statistically validated by a correlation test using the Pearson's method.

### AVAILABILITY OF MICROARRAY DATA

Microarray data are available on GEO database through the following accession number GSE59137.

## RESULTS

Transcriptomic profiles were obtained using microarrays and the four following combinations were analyzed (three independent replicates per combination): Cigalon/*A. lipoferum* 4B (Cig_4B) and Cigalon/*Azospirillum* sp. B510 (Cig_B510) compared to non-inoculated Cigalon; Nipponbare/*A. lipoferum* 4B (Nip_4B) and Nipponbare/*Azospirillum* sp. B510 (Nip_B510) compared to non-inoculated Nipponbare. Genes significantly regulated were selected using a $P_{\text{adjusted}}$-value ($P_{\text{adj}}$) threshold of 0.05 and a fold change cutoff of 2 ($|\text{Log}_2(\text{FC})| \geq 1$). According to the cultivar of which each strain was originally isolated, Cig_4B and Nip_B510 combinations constitute interactions between a strain and its original host cultivar, which will be hereafter designed as host combinations, while Cig_B510 and Nip_4B combinations constitute interactions with non-host cultivars, which will be hereafter designed as non-host combinations.

### GENERAL FEATURES OF RICE-ROOT TRANSCRIPTOME PROFILING IN RESPONSE TO *AZOSPIRILLUM* INOCULATION

Microarray design was based on the genome sequence of cultivar Nipponbare. To ensure that it could be used to hybridize cDNA obtained from cultivar Cigalon, the genetic proximity between both cultivars was analyzed by sequencing eight genes (including the gene encoding actin) after PCR amplification from Cigalon DNA (at least 500 pb per gene) (Additional file 2: Table S2). For all the sequenced genes, an identity of 100% was observed with the corresponding genes of Nipponbare. In addition, when comparing the non-inoculated Cigalon transcriptome profile to the non-inoculated Nipponbare profile, only 193 genes are differentially transcribed between the two cultivars (87 up-regulated and 106 down-regulated). This represents only 0.43% of the targeted genes (i.e., 45,000 genes) and may be explained by physiological differences. Thus, the microarray was considered suitable

**Table 1 | Validation of microarray data.**

| Gene | | FC Cig_4B[a] | | FC Cig_B510 | | FC Nip_4B | | FC Nip_B510[b] | |
|---|---|---|---|---|---|---|---|---|---|
| | | $Log_2$ (qPCR) | $Log_2$ (array) | $Log_2$ (qPCR) | $Log_2$ (array) | $Log_2$ (qPCR) | $Log_2$ (array) | $Log_2$ (qPCR) | $Log_2$ (array) |
| **Differentially expressed in all conditions** | | | | | | | | | |
| Os09g0358000 | Tyrosine kinase domain | 2.27 | 1.77 | 2.53 | 2.46 | 1.87 | 3.01 | 2.06 | 1.32 |
| **Differentially expressed in Cig_4B and Nip_4B only** | | | | | | | | | |
| Os09g0417800 | WRKY transcription factor 62 | 1.49 | 1.75 | 0.66 | _ | 1.00 | 1.88 | −0.01 | _ |
| Os11g0686900 | Similar to NB-ARC domain | 0.83 | 1.16 | −0.23 | _ | 1.74 | 2.40 | nd | _ |
| **Differentially expressed in Cig_B510 and Nip_B510 only** | | | | | | | | | |
| Os06g0115600 | Common symbiosis signaling (SYM) pathway | −0.09 | _ | −2.00 | −1.84 | 0.01 | _ | −1.18 | −1.47 |
| Os05g0583000 | Similar to WRKY8 | 0.32 | _ | 2.92 | 1.71 | 0.16 | _ | 2.24 | 1.39 |
| Os12g0139400 | A-type response regulator, cytokinin signaling | −0.15 | _ | −2.84 | −1.78 | −0.09 | _ | −2.94 | −1.10 |
| Os11g0143300 | A-type response regulator, cytokinin signaling | 0.15 | _ | −1.09 | −1.95 | 0.04 | _ | −2.25 | −1.62 |
| Os02g0805100 | Similar to auxin-responsive protein IAA12 | −0.01 | _ | −1.15 | −1.34 | −0.23 | _ | −2.47 | −1.16 |
| **Differentially expressed in Cig_4B and Cig_B510 only** | | | | | | | | | |
| Os05g0196600 | Similar to ACC synthase | 0.38 | 1.33 | 0.95 | 1.63 | −0.15 | _ | 0.16 | _ |
| **Differentially expressed in Cig_4B only** | | | | | | | | | |
| Os08g0136100 | Homeobox-leucine zipper domain | 0.87 | 1.11 | 0.25 | _ | 0.28 | _ | −0.14 | _ |
| **Significantly regulated in Cig_B510 only** | | | | | | | | | |
| Os06g0179200 | Similar to Nodulin-like protein | 0.16 | _ | −1.60 | −1.78 | 0.18 | _ | 0.12 | _ |
| Os08g0499300 | WRKY transcription factor 30 | 0.10 | _ | 0.52 | 1.34 | 0.12 | _ | −0.06 | _ |
| **Significantly regulated in Nip_B510 only** | | | | | | | | | |
| Os01g0904700 | B-type response regulator, cytokinin signaling | −0.12 | _ | −0.07 | _ | −0.04 | _ | −1.09 | −1.21 |
| Os05g0515400 | Similar to auxin response factor 14 | −0.09 | _ | 0.10 | _ | −0.09 | _ | −1.15 | −1.42 |
| **Reference gene** | | | | | | | | | |
| Os03g0718100 | Actin | | | | | | | | |

[a] *Dashes replace non-significant fold change values obtained from microarray data ($|Log_2(FC)| \leq 1$ and $P_{adj} > 0.05$).*

[b] *nd, not determined.*

for analysis and comparison of both Nipponbare and Cigalon transcriptomes.

When considering the four combinations, a total of 7384 genes are differentially expressed in rice roots, which represent about 16% of the entire set of rice genes. Each strain/cultivar combination displays specific expression profiles highlighting a strain-specific response of the host plant. The most important changes are observed when strain B510, isolated from Nipponbare, is inoculated on Cigalon roots (Cig_B510), with 3865 regulated genes, equally induced and repressed (1993 up-regulated; 1872 down-regulated) (**Figure 1**). Conversely, the inoculation of strain 4B on its original cultivar Cigalon (Cig_4B) is accompanied by the differential expression of only 1243 genes, mostly induced (1196 up-regulated; 47 down-regulated). When considering Nipponbare roots, the number of regulated genes is similar for Nip_4B and Nip_B510 combinations with respectively, 2141 and 2539 regulated genes. However, these genes are mostly induced in Nip_4B (1965 up-regulated; 176 down-regulated) while they are repressed in Nip_B510 combination (203 up-regulated; 2336 down-regulated).

A Venn diagram analysis conducted on expression profiles obtained for the four combinations, unveils 15 sets of genes (**Figure 1**). Four sets, named combination-specific genes (genes induced or repressed only in one combination), represent 83% of all differentially expressed genes, with 358 genes for Cig_4B (347 up-regulated, 11 down-regulated), 2317 genes for Cig_B510 (1076 up-regulated, 1241 down-regulated), 1786 genes for Nip_4B (1677 up-regulated, 109 down-regulated) and 1697 genes for Nip_B510 (26 up-regulated, 1671 down-regulated) (Additional file 3: Table S3). Two other sets regroup genes that are common only to Cig_4B and Cig_B510 (700 up-regulated, 24 down-regulated) or only to Nip_4B and Nip_B510 (49 up-regulated, 59 down-regulated), representing *Azospirillum*-regulated cultivar-specific genes (Additional File 4: Table S4). Inversely, two sets comprise strain-specific genes that display the same regulation only in both Cig_4B and Nip_4B (59

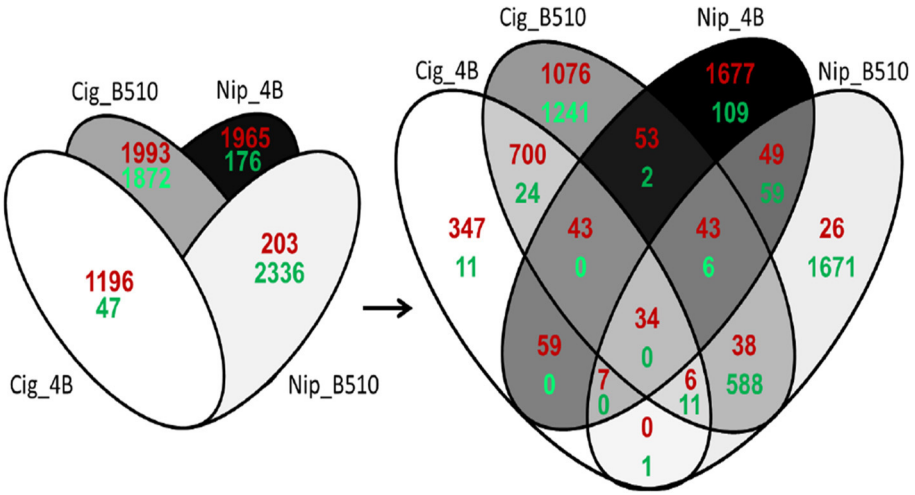



**FIGURE 1 | Venn diagram of genes differentially expressed in rice roots ($|Log_2(FC)| \geq 1$ and $P_{adj} \leq 0.05$).** *Oryza sativa* L. *japonica* cultivar Cigalon and cultivar Nipponbare were inoculated with *Azospirillum* strains 4B and B510. For each cultivar, the respective non-inoculated condition was used as reference to evidence genes up-regulated (red) and down-regulated (green) after *Azospirillum* inoculation ($|Log_2(FC)| \geq 1$ and $P_{adj} < 0.05$). Cig_4B and Nip_B510 combinations constitute the interaction between a strain and its original host cultivar (host combinations). Cig_B510 and Nip_4B combinations constitute interactions with non-host cultivars (non-host combinations).

up-regulated, zero down-regulated) or only in both Cig_B510 and Nip_B510 (38 up-regulated, 588 down-regulated) (Additional File 5: Table S5). Two additional sets include genes displaying the same regulation in both host combinations, or in both non-host combinations (Additional File 6: Table S6). One set contains 34 up-regulated genes common to the four combinations and no down-regulated genes. The four remaining sets comprise genes that are common to three of the four combinations and will no longer be discussed in the current analysis. Microarray data were confirmed by analyzing expression levels of 14 representative genes using reverse transcription quantitative polymerase chain reaction (RT-qPCR) (**Table 1**). This includes seven up-regulated genes and seven down-regulated genes belonging to seven of the 15 categories described in **Figure 1**. These results show that the array data are in accordance with the RT-qPCR data ($R^2$= 0.83; $P_{value} = 2.7.10^{-10}$).

The transcriptome survey was completed with an analysis of the LTR-retrotransposons. Indeed, LTR-retrotransposons, a particular type of transposable elements, represent 25% of the total genomic sequence of rice (Rice Annotation Project et al., 2008). The transcriptional activation of LTR-retrotransposons can lead to the activation of a flanking gene, either through the action of the enhancer regions of the element or by co-transcription (Michaud et al., 1994); conversely, LTR-retrotransposons can also act as suppressors of gene expression when they are inserted in antisense in the 3′ region of a gene. Interestingly, *Azospirillum* inoculation leads to a differential expression of a large number of LTR-retrotransposon families for both rice cultivars. To avoid bias related to cultivar polymorphism, the complete analysis was performed only for the cultivar Nipponbare. For this latter, differential expression is observed for 115 and 148 LTR-retrotransposon families for strains B510 and 4B, respectively. In addition, strain 4B modifies the transcription of 21.4% of all LTR-retrotransposons, among which 51% are up-regulated and 49% down-regulated. For strain B510, 17.8% of all LTR-retrotransposons are differentially expressed (37% up-regulated and 62% down-regulated). Among the differentially expressed elements, 22% are common to Nip_4B and Nip_B510 combinations. In addition, we explored if transposable elements can significantly alter the expression of adjacent genes. In this analysis, two and six LTR-retrotransposons are located in the 3 kb upstream regions of regulated genes while four and five ones are located in the 3 kb downstream regions, for Nip_B510 and Nip_4B, respectively (Additional File 7: Table S7). However, no co-transcription events between gene and LTR-retrotransposon was evidenced by using RT-PCR method (data not shown), suggesting that none of these elements have a direct impact on the expression of adjacent genes.

## IMPACT OF STRAIN LIFESTYLE ON RICE ROOTS GENE EXPRESSION

As mentioned above, most of the differentially expressed genes are induced in combinations involving strain 4B while they are mostly repressed for combinations involving strain B510 (**Figure 1**). In addition, when considering functional classification available for only 20% of all differentially expressed genes, several differences are evidenced at the strain level (**Figure 2**). Whatever the combination, the most important numbers of differentially expressed genes are observed for (i) primary metabolism, (ii) transport, (iii) regulation of transcription, and (iv) protein fate, four categories in which genes seem quasi exclusively repressed in Nip_B510 while they are quasi exclusively induced in Nip_4B. Similarly, when considering genes involved in response to stress and plant defense, two substantial categories in plant-microbe interactions, it appears that strain B510 leads to the repression of a more important number of genes than strain 4B. These results are of particular interest as they may

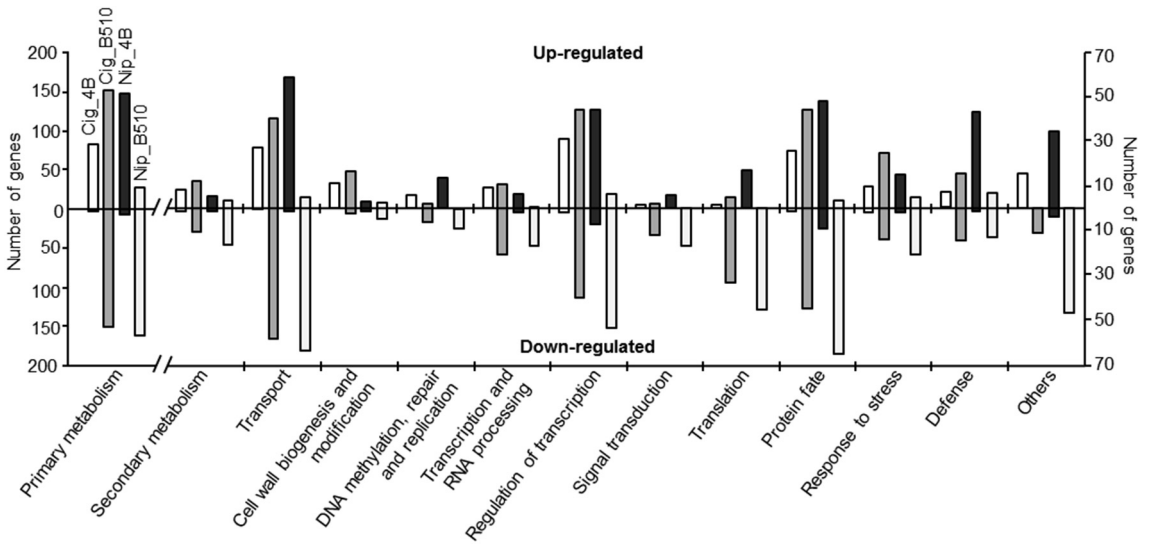



**FIGURE 2 | Functional classification of the differentially expressed genes in *Oryza sativa* L. *japonica* in response to *Azospirillum* inoculation.** Numbers of genes differentially expressed are shown per functional categories for each *Azospirillum*-rice combination: Cig_4B (white), Cig_B510 (dark gray), Nip_4B (black), and Nip_B510 (light gray). Genes were classified according to Biological Process assignation taken from the Rap-DB database. Genes with no Biological Process assignation represent about 80% of differentially expressed genes for each condition. Category "Others" includes principally genes implicated in photosynthesis, exocytosis, sexual reproduction, cytoskeleton organization, cell cycle, and cell death.

be related to lifestyle differences between *Azospirillum* strains; indeed, strain 4B was shown to colonize only rice-root surface while strain B510 is able to colonize the outer layers of rice-root tissues (Elbeltagy et al., 2001; Chamam et al., 2013). To consider the impact of strain lifestyle on plant defense response, the analysis was focused on genes potentially involved in biotic stress. Identification of these genes was improved using Mapman annotation software as recently described for the *Azospirillum-Arabidopsis* interaction (Spaepen et al., 2014). This additional analysis, illustrated in **Figure 3**, confirms observations made above. Then, strain 4B leads quasi exclusively to the induction of genes related to biotic stress in Nipponbare and to a lower extent in Cigalon. On the contrary, genes involved in biotic stress response are quasi exclusively repressed in the Nip_B510 host combination, while the number of up-regulated genes and down-regulated genes is similar in the Cig_B510 non-host combination. In addition, Mapman visualization analysis highlights relevant categories of genes being differentially expressed such as genes encoding peroxidases, transcription factors of the MYB, WRKY, and ERF families as well as Pathogenesis-Related (PR) genes.

Another way to analyze the impact of strain lifestyle on rice roots gene expression was to consider the 4B_specific genes (59) and B510-specific genes (626) highlighted in the Venn diagram analysis (**Figure 1**; Additional File 5: Table S5). Among the 59 genes specifically induced by the surface-colonizing strain 4B, we identified 14 genes that could be involved in biotic stress response among which two genes encoding peroxidases, two WRKY transcription factors and one PR gene. While all the 4B-specific genes are up-regulated, 38 of the B510-specific genes are up-regulated and 588 are down-regulated including at least 14 and 102 genes associated to biotic stress response, respectively. The endophyte strain B510 triggers the induction of four genes potentially involved in hormone signaling, as well as genes encoding a peroxidase, a WRKY transcription factor and a PR protein. Moreover, a larger number of genes are down-regulated, including five PR genes, three genes encoding peroxidases but no WRKY transcription factor. Finally, the repression of a putative indole synthase (Os03g0797000), three putative auxin responsive proteins (Os02g0805100, Os08g0335600, Os11g0523800), two response regulators involved in cytokinin signaling (Os11g0143300, Os12g0139400), an isochorismate synthase (Os09g0361500) required for salicylic acid synthesis, as well as four genes potentially involved in ethylene signaling (Os03g0439500, Os04g0667400, Os09g0115500, 0s09g0309700) suggest that a B510-specific hormone signaling occurs in rice roots.

## PLANT MARKERS OF *AZOSPIRILLUM*-RICE COOPERATION AND SIMILARITIES WITH PATHOGEN INFECTION

As revealed in the Venn diagram analysis (**Figure 1**), a total of 34 genes (12 genes of unknown function) are induced in the four combinations and could be defined as plant markers of *Azospirillum*-rice interactions (**Table 2**). Interestingly, a PR gene orthologous to PR10 of *Brachypodium* is induced in the four conditions (Os12g0555200). In addition, two genes encoding Cys-Rich domain containing protein (Os02g0579800, Os02g0580000) and an orthologous gene to AT1G59950 of *Arabidopsis thaliana* (Os03g0237100), which are potentially implicated in stress response, as well as a terpene synthase involved in gibberellin synthesis (Os04g0178300), and a gene encoding a putative precursor of phytoalexin (Os11g0474800, encoding a stemar-13-ene synthase) are induced in the four combinations. Finally, the set of potential plant markers of *Azospirillum*-rice cooperation also includes a transcription factor (Os01g0952800) and four

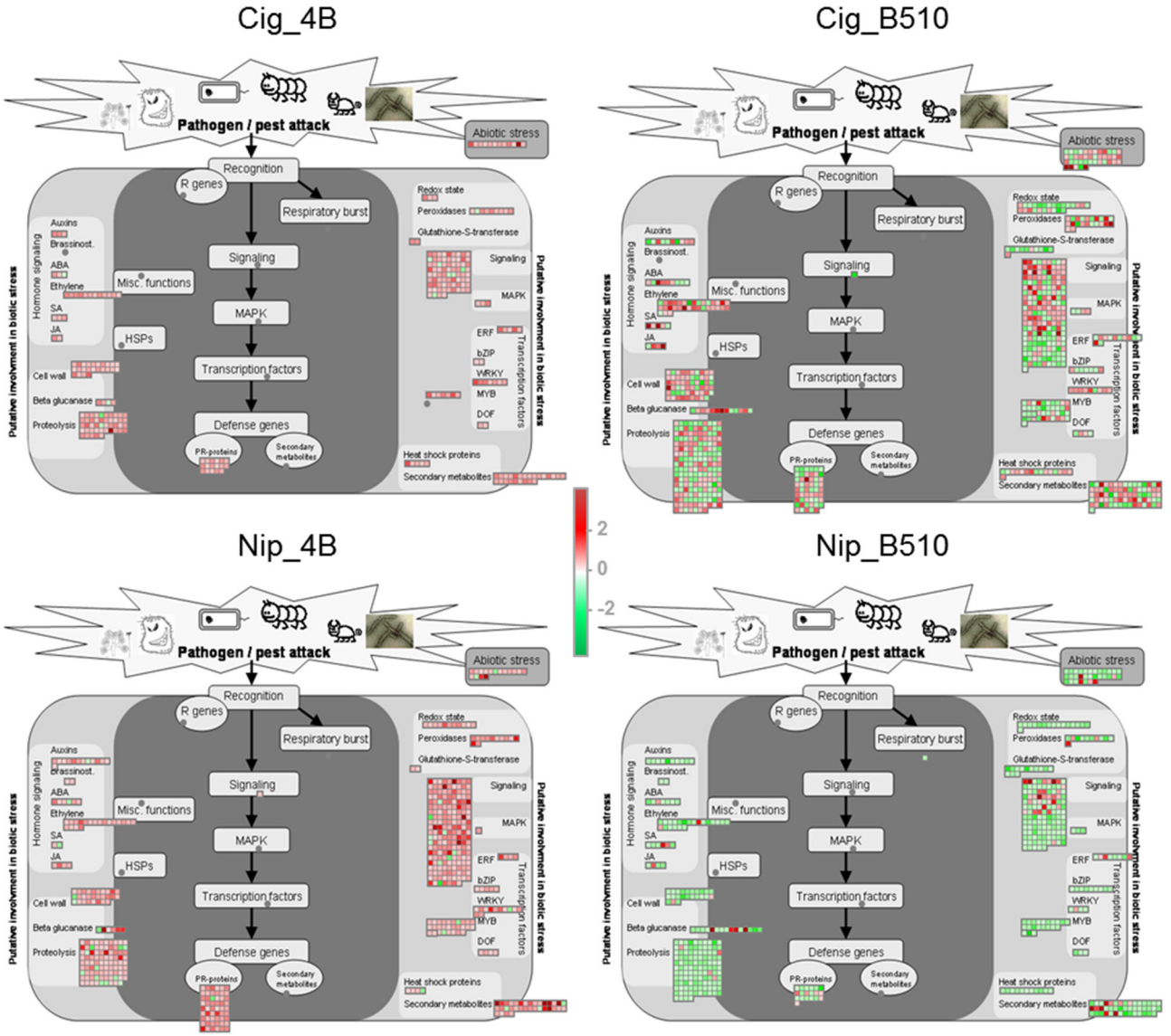



**FIGURE 3 | Mapman software visualization of genes related to biological stress.** Genes differentially expressed ($|\text{Log}_2(\text{FC})| \geq 1$ and $P_{adj} \leq 0.05$) after *Azospirillum* inoculation on rice roots are represented according to their respective $\text{Log}_2$ rescaled fold change. Red color represents up-regulated genes and green color represents down-regulated genes with the most intense color on the scale representing a $\text{Log}_2(\text{FC}) = 4.8$. The dark gray rectangle includes genes directly involved in biotic stress responses and the light gray rectangle includes genes potentially involved in biotic stress response. ABA, abscisic acid; bZIP, basic region leucine zipper; DOF, DNA-binding with one finger; ERF, ethylene responsive factor; HSP, heat shock protein; JA, jasmonic acid; MAPK, mitogen activated protein kinase; MYB, myeloblast; PR, pathogenesis-related; R, resistance; SA, salicylic acid.

genes potentially involved in signal transduction (Os07g0537900, Os08g0501500, Os08g0203400, Os09g0358000, this latter being validated by RT-qPCR, **Table 1**).

Making use of the Rice Oligonucleotide Array Database (Cao et al., 2012), we identified experiments analyzing gene expression in rice during plant-microbe interactions. Prior to this analysis, RT-qPCR expression profiles observed in rice roots for the 14 genes used to validate our microarray data (**Table 1**) were compared to RT-qPCR expression profiles obtained in rice leaves and none displays differential expression when considering shoot compartment (data not shown). Then, the array database analysis was focused on rice roots and only one experiment analyzing gene expression in the root compartment was evidenced. This work considered the impact of the fungus phytopathogen *Magnaporthe oryzae* on rice transcriptome 6 days after infection (Marcel et al., 2010). The list of 34 potential markers of *Azospirillum*-rice cooperation, up-regulated in the four combinations, was compared to the list of genes differentially expressed after *M. oryzae* inoculation**,** evidencing that 10 of the potential markers (3 genes of unknown function) are also up-regulated in response to *Magnaporthe* infection while none of these markers are down-regulated after *Magnaporthe* infection (**Table 2**). Thus, the PR

**Table 2 | List of 34 genes induced in the four combinations.**

| Gene ID | Log$_2$(FC) | | | | Gene description | MSU Gene ID | *M. oryzae*[a] |
|---|---|---|---|---|---|---|---|
| | Cig_4B | Nip_4B | Cig_B510 | Nip_B510 | | | |
| Os01g0191200 | 1.34 | 3.28 | 2.18 | 3.09 | Putative HAD phosphatase | LOC_Os01g09540 | |
| Os01g0495701 | 1.14 | 2.60 | 2.02 | 2.12 | Conserved hypothetical protein | None | |
| Os01g0608101 | 1.33 | 2.73 | 2.42 | 2.39 | Conserved hypothetical protein | None | |
| Os01g0647200 | 1.07 | 2.97 | 1.91 | 1.71 | Conserved hypothetical protein | LOC_Os01g45914 | |
| Os01g0952800 | 1.70 | 3.32 | 2.75 | 2.65 | Transcription factor | LOC_Os01g72370 | |
| Os02g0569400 | 1.54 | 2.42 | 2.23 | 1.78 | Cytochrome P450 family protein | LOC_Os02g36070 | |
| Os02g0579800 | 1.49 | 3.08 | 1.75 | 1.91 | Cys-rich domain containing protein | LOC_Os02g36940 | + |
| Os02g0580000 | 1.51 | 2.80 | 1.78 | 1.32 | Cys-rich domain containing protein | LOC_Os02g36950 | |
| Os02g0582900 | 2.20 | 4.95 | 3.77 | 4.98 | Conserved hypothetical protein | LOC_Os02g37190 | + |
| Os02g0594232 | 1.18 | 2.28 | 2.09 | 1.8 | Conserved hypothetical protein | None | |
| Os02g0791300 | 1.41 | 3.00 | 2.91 | 2.58 | Conserved hypothetical protein | LOC_Os02g54870 | |
| Os03g0129800 | 1.01 | 1.79 | 1.01 | 1.58 | Uncharacterized protein | LOC_Os03g03730 | + |
| Os03g0237100 | 1.42 | 3.43 | 2.40 | 2.94 | Putative NADPH-dependent codeinone reductase | LOC_Os03g13390 | |
| Os03g0307300 | 1.34 | 3.33 | 2.29 | 2.54 | Nicotianamine synthase 1 | LOC_Os03g19427 | |
| Os04g0178300 | 1.25 | 3.22 | 2.39 | 2.91 | Putative syn-copalyl diphosphate synthase | LOC_Os04g09900 | |
| Os04g0178400 | 1.92 | 3.56 | 2.60 | 2.82 | Putative cytochrome P450 | LOC_Os04g09920 | + |
| Os06g0293500 | 2.40 | 3.26 | 2.62 | 1.73 | Conserved hypothetical protein | LOC_Os06g18960 | + |
| Os06g0486800 | 1.10 | 2.31 | 1.84 | 1.42 | Putative mitochondrial formate dehydrogenase | LOC_Os06g29180 | |
| Os06g0718400 | 1.39 | 1.68 | 1.66 | 1.53 | Plastocyanin-like domain containing protein | LOC_Os06g50420 | |
| Os07g0190000 | 1.69 | 3.02 | 2.77 | 2.85 | Putative 1-deoxy-D-xylulose 5-phosphate synthase | LOC_Os07g09190 | + |
| Os07g0258400 | 1.61 | 2.25 | 2.60 | 1.70 | Putative metal transporter Nramp6 | LOC_Os07g15460 | |
| Os07g0416900 | 2.76 | 3.78 | 2.35 | 2.43 | Omega-6 fatty acid desaturase | LOC_Os07g23410 | + |
| Os07g0537900 | 1.12 | 1.24 | 1.83 | 1.62 | Ser/Thr kinase receptor domain | LOC_Os07g35340 | |
| Os07g0664000 | 1.52 | 2.24 | 2.59 | 1.85 | Putative short chain dehydrogenase/reductase | LOC_Os07g46870 | + |
| Os08g0203400 | 1.66 | 4.69 | 2.54 | 3.59 | Protein kinase/core domain containing protein | LOC_Os08g10310 | |
| Os08g0501500 | 1.03 | 1.27 | 2.26 | 1.70 | OsWAK receptor-like protein kinase | LOC_Os08g39210 | + |
| Os09g0358000 | 1.77 | 3.07 | 2.71 | 2.51 | Tyrosin kinase domain containing protein | LOC_Os09g19350 | |
| Os09g0358100 | 1.27 | 2.05 | 1.79 | 1.05 | Senescence-induced serine/threonine kinase | LOC_Os09g19360 | |
| Os10g0195250 | 1.24 | 2.08 | 2.01 | 1.68 | Conserved hypothetical protein | LOC_Os10g11889 | |
| Os11g0262600 | 1.31 | 2.45 | 2.29 | 2.29 | Conserved hypothetical protein | LOC_Os11g15624 | |
| Os11g0474800 | 2.83 | 2.81 | 3.20 | 2.57 | Putative stemar-13-ene synthase | LOC_Os11g28530 | |
| Os12g0171801 | 1.12 | 2.77 | 2.75 | 1.64 | Hypothetical protein | None | |
| Os12g0236100 | 1.17 | 2.03 | 1.70 | 1.60 | Conserved hypothetical protein | LOC_Os12g13340 | |
| Os12g0555200 | 2.22 | 3.97 | 2.77 | 3.31 | Putative pathogenesis-related Bet v family protein | LOC_Os12g36850 | + |

[a] *Genes previously shown to be induced in rice root 6 days post-inoculation with M. oryzae (Marcel et al., 2010) are indicated with the sign +.*

gene orthologous to PR10 of *Brachypodium* (Os12g0555200) discussed above, as well as a gene related to signal transduction (Os08g0501500) and a gene encoding a Cys-Rich domain containing protein (Os02g057980) are up-regulated in response to both *Azospirillum* and *Magnaporthe*.

To understand whether rice biotic stress responses could be specific to either host or non-host interactions, the analysis was focused on the comparison of host combinations and non-host combinations. According to **Figure 1**, the number of up-regulated genes appears to be lower when a strain is inoculated on its original host cultivar (Cig_4B and Nip_B510). In addition, fewer genes implicated in response to stress and plant defense are differentially expressed in host combinations (**Figure 2**), the most striking example being PR genes as highlighted in **Figure 3**. However, among all differentially expressed genes, only one down-regulated gene is common to both host combinations (Os07g0638600 encoding a peroxidase) and 55 genes (53 up-regulated, 2 down-regulated) are common to non-host combinations (**Figure 1**). Among the latter, two genes implicated in successive steps of ethylene synthesis (Os06g0524900, Os09g0451400), a NB-ARC domain containing gene (Os06g0524900) and a gene encoding an expansin (Os05g0276500) were identified. Several genes implicated in oxido-reductive processes (Os01g0327000, Os07g0164900, Os12g0260500) as well as a gene encoding a glycoside hydrolase (Os09g0395600) and a gene encoding an UDP-glucuronosyl transferase (Os01g0597800) are also induced. When comparing to the list of genes differentially expressed after *Magnaporthe* infection, 6 of the 53 up-regulated genes common to both *Azospirillum*-rice non-host combinations (Cig_B510 and Nip_4B) are also induced after

*Magnaporthe* inoculation and none of these genes are repressed (Additional File 6: Table S6). However, these genes do not seem to be directly involved in biotic stress response of rice roots.

### COMBINATION-SPECIFIC EXPRESSION PROFILES

Many differences are observed between the four cultivar/strain combinations when comparing expression profiles, particularly for genes involved in biotic stress responses (**Figure 3**). This includes genes potentially involved in hormone signaling and both ethylene and auxin signaling occur to be finely regulated during *Azospirillum*-rice cooperation.

Indeed, if only one of the genes related to ethylene signaling was classified as Cig_4B specific, 15 of these genes were classified as Cig_B510-specific, 11 were classified as Nip_4B-specific and 13 were classified as Nip_B510-specific (**Table 3**). All the Cig_4B- and Nip_4B-specific genes related to ethylene signaling are up-regulated, while two of the Cig_B510-specific genes and

**Table 3 | List of combination-specific genes related to ethylene signaling.**

| Gene ID | Log2(FC) | | | | Genome annotation |
|---|---|---|---|---|---|
| | Cig_4B | Cig_B510 | Nip_4B | Nip_B510 | |
| Os02g0594300 | 1.29 | | | | Similar to enhancer of shoot regeneration ESR1 |
| Os01g0797600 | | −1.77 | | | AP2 domain-containing ethylene responsive protein |
| Os01g0536400 | | −1.66 | | | Similar to 1-aminocyclopropane-1-carboxylate oxidase |
| Os04g0257500 | | 1.10 | | | Similar to ethylene-responsive transcription factor TSFR1 |
| Os09g0451000 | | 1.17 | | | Similar to 1-aminocyclopropane-1-carboxylate oxidase |
| Os10g0523900 | | 1.19 | | | AP2 domain containing ethylene responsive protein |
| Os05g0437100 | | 1.26 | | | Ethylene-responsive transcription factor |
| Os01g0757200 | | 1.30 | | | Similar to gibberellin 2-beta-dioxygenase |
| Os03g0860600 | | 1.35 | | | Similar to 1-aminocyclopropane-1-carboxylate oxidase |
| Os02g0767300 | | 1.36 | | | Similar to flavonol synthase |
| Os09g0570800 | | 1.41 | | | Similar to 1-aminocyclopropane-1-carboxylate oxidase |
| Os01g0832600 | | 1.66 | | | Similar to leucoanthocyanidin dioxygenase |
| Os09g0248900 | | 1.70 | | | Similar to ethylene-responsive protein |
| Os02g0797100 | | 1.89 | | | AP2 domain containing ethylene responsive protein |
| Os05g0127500 | | 1.95 | | | Similar to leucoanthocyanidin dioxygenase |
| Os04g0182200 | | 2.60 | | | 2OG-Fe(II) oxygenase domain containing protein |
| Os01g0230200 | | | 1.06 | | Ethylene-responsive protein |
| Os07g0169600 | | | 1.12 | | 2OG-Fe(II) oxygenase domain containing protein |
| Os03g0100900 | | | 1.15 | | Ethylene-responsive element-binding protein |
| Os10g0536400 | | | 1.18 | | 2OG-Fe(II) oxygenase domain containing protein |
| Os04g0407800 | | | 1.24 | | 2OG-Fe(II) oxygenase domain containing protein |
| Os06g0162500 | | | 1.29 | | Similar to Naringenin 3-dioxygenase like protein |
| Os02g0202000 | | | 1.32 | | Similar to ethylene responsive protein |
| Os04g0548000 | | | 1.32 | | Ethylene-responsive element-binding protein |
| Os02g0654700 | | | 1.50 | | Ethylene-responsive transcription factor |
| Os03g0122300 | | | 1.53 | | Similar to Flavanone 3-hydroxylase-like protein |
| Os08g0366100 | | | 1.75 | | Endothelial differentiation-related factor |
| Os06g0177600 | | | | −2.07 | 2OG-Fe(II) oxygenase domain containing protein |
| Os11g0186900 | | | | −1.82 | Similar to 1-aminocyclopropane-1-carboxylate oxidase |
| Os04g0522500 | | | | −1.58 | 2OG-Fe(II) oxygenase domain containing protein |
| Os05g0155200 | | | | −1.41 | Similar to ethylene receptor |
| Os01g0935400 | | | | −1.41 | 2OG-Fe(II) oxygenase domain containing protein |
| Os02g0520000 | | | | −1.41 | Ethylene responsive protein |
| Os04g0643500 | | | | −1.30 | 2OG-Fe(II) oxygenase domain containing protein |
| Os02g0276900 | | | | −1.28 | Ethylene-responsive protein |
| Os04g0565900 | | | | −1.22 | Ethylene-responsive protein |
| Os04g0493100 | | | | −1.21 | Ethylene-responsive protein |
| Os04g0578000 | | | | −1.18 | Similar to 1-aminocyclopropane-1-carboxylate synthase |
| Os03g0690500 | | | | −1.16 | Similar to 1-aminocyclopropane-1-carboxylate oxidase |
| Os06g0573900 | | | | −1.10 | Similar to 1-aminocyclopropane-1-carboxylate oxidase |

all the Nip_B510-specific genes related to ethylene signaling are down-regulated. When considering genes related to auxin signaling, one was classified as Cig_4B specific, seven were classified as Cig_B510-specific, 12 were classified as Nip_4B-specific and eight were classified as Nip_B510-specific. In addition, nine genes involved in abscisic acid signaling and seven genes involved in salicylic acid signaling were shown to be differentially expressed in only one of the four combinations, suggesting that a wide modification of hormone signals occurs in rice roots after *Azospirillum* inoculation. Auxin, ethylene, abscissic acid, and salicylic acid are known to be involved in plant immunity, and many genes related to biotic stress response display a combination-specific profile in the current study. For example, 76 genes similar to PR-genes were identified to be combination-specific. As previously mentioned, fewer genes are differentially expressed in host combinations than in non-host combinations. Indeed, three and 13 of these genes were identified for Cig_4B and Nip_B510, respectively, while 25 and 35 of these genes were identified for Nip_4B and Cig_B510, respectively. Taken all together, these results suggest that *Azospirillum* inoculation leads to important changes in rice root hormone signaling and plant defense, depending on the strain/cultivar combination.

Besides genes related to hormone signaling, the combination-specific response also includes nine genes related to hydroxycinnamic acids metabolism (Additional File 3: Table S3). This includes seven genes encoding cinnamyl alcohol dehydrogenase (Os01g0528800, Os03g0223200, Os09g0400200, Os11g0622800 which are down-regulated; Os04g0612700, Os09g0399800, Os09g0400400 which are up-regulated), one gene encoding a cinnamoyl-CoA reductase (Os01g0828100, down-regulated) and one gene encoding a coumarate-CoA synthase (Os08g0245200, downregulated). In addition, four genes related to chalcones metabolism and five genes related to anthocyanin metabolism display a combination-specific profile. Such a result is of particular interest as some secondary metabolites stemming from these pathways were previously shown to be discriminant in the analysis of root methanolic extract composition, following *Azospirillum* and *Frankia* inoculations (Popovici et al., 2011; Chamam et al., 2013).

## DISCUSSION

Unraveling the molecular basis of host-specific adaptations in PGPR-plant cooperation helps understanding frontiers between the perception of symbiotic, cooperative, and pathogenic microbes hosted by plant roots. Based on previous studies detecting metabolic and morphological changes of PGPR-inoculated plants at early stages (Cassán et al., 2009; Walker et al., 2011, 2012; Chamam et al., 2013), we analyzed the transcriptomic response of rice roots 7 days after inoculation. To our knowledge, this study constitutes the first investigation of wide transcriptomic response of rice roots to PGPR-inoculation, aiming at deciphering interaction specificity in plant-microbe cooperation.

Comparison of expression profiles obtained for Cigalon/*A. lipoferum* 4B (Cig_4B), Cigalon/*Azospirillum* sp. B510 (Cig_B510), Nipponbare/*A. lipoferum* 4B (Nip_4B) and Nipponbare/*Azospirillum* sp. B510 (Nip_B510) combinations evidences a fine-tuned transcriptomic response depending on both *Azospirillum* and rice genotypes. As revealed by the high percentage (83%) of genes up-regulated or down-regulated only in one of the four conditions, individual genotypic variations are the most important driving force of rice roots gene expression, in the tested conditions. Indeed, only 34 markers of the *Azospirillum*-rice cooperation, induced in the four combinations, were identified and further investigations should be undertaken to identify the impact of a larger range of *Azospirillum* strains on the regulation of these markers. Besides combination-specific traits, expression profiles showed strain-specific and cultivar-specific characteristics, highlighting potential differences in the strategies of interaction. Indeed, strain 4B causes few repressions while at least half of the genes regulated in response to strain B510 are repressed, regardless of the cultivar. In addition, only 59 genes display similar regulation in both combinations involving strain 4B whereas it represents 626 genes for those involving strain B510. Accordingly, strain-specific responses were observed when considering the respective impact of 4B and B510 on development and secondary metabolism of rice cultivars Cigalon and Nipponbare (Chamam et al., 2013). While 4B promotes shoot and root growth of both cultivars, B510 promotes development of cultivar Nipponbare exclusively. However, B510 was shown to be the only strain inducing a systemic response, as revealed by variation of secondary metabolite profiles of both shoots and roots. These strain-specific responses could be due to differences in strain lifestyle as 4B colonizes only the surface of rice roots while B510 has the ability to colonize the cortex layers in rice (Thomas-Bauzon et al., 1982; Elbeltagy et al., 2001; Chamam et al., 2013). In addition to their impact on plant growth, endophytic PGPR induce stress and defense responses and the inoculation of B510 was shown to enhance resistance against rice blast disease and rice blight disease (Miché et al., 2006; Rosenblueth and Martínez-Romero, 2006; Yasuda et al., 2009). However, whether these changes were the result of major gene induction or repression was not addressed. Recently, a study on differential gene expression of rice roots inoculated with the endophyte *Herbaspirillum seropedicae* evidenced a decrease in expression of defense related protein PBZ1 and thionins, suggesting that bacteria modulate plant defense to allow the establishment of an efficient cooperation (Brusamarello-Santos et al., 2011). Indeed, colonization of root tissues by bacteria depends on the balance between the plant's ability to induce efficient defenses in response to the intrusive microbe and the microorganisms' ability to bypass plant immunity (Pieterse et al., 2009).

Several genes associated to plant defense mechanisms are still regulated 7 days post-inoculation. Particularly, a gene orthologous to the PR10 gene of *Brachypodium* is induced in all combinations. Interestingly, this gene was previously shown to be up-regulated in rice roots 6 days after *Magnaporthe* infection (Marcel et al., 2010). PR genes are known to be induced during PAMPs-triggered immunity (PTI), the first step of plant defense that involves Pattern Recognition Receptors (PRRs) (Chisholm et al., 2006; Pieterse et al., 2009). PRRs recognize universal microbial determinants such as flagellin, chitin, glycoproteins, and lipoproteins designed as PAMPs/MAMPs for Pathogens/Microbes–Associated Molecular

Patterns (Schwessinger and Zipfel, 2008). Even if symbiotic and pathogenic interactions exhibit a number of unique characteristics, it was suggested that similar response mechanisms were adapted to cope with different biotic and abiotic stresses (Baron and Zambryski, 1995). Surprisingly, down-regulated mechanisms seem to be more conserved than up-regulated mechanisms between pathogenic and symbiotic interactions, genes involved in plant defense and stress response being a major part of these repressions (Damiani et al., 2012). In this context, endophytic colonization being more intrusive than surface colonization, plant response to endophyte PGPR could be more similar to pathogenic response than plant response to surface colonizers. This hypothesis is supported by the fact that 30 genes common to Cig_B510 and Nip_B510 are also down-regulated during *Magnaporthe* infection while none of the genes common to Cig_4B and Nip_4B are down-regulated (data not shown).

Most of the genes containing a NB-ARC domain are differentially expressed on a strain/cultivar dependent manner, the most striking effect being observed for the non-host combinations. Most of NB-ARC genes regulated during interaction with strain B510 are down-regulated and repression occurs for a higher number of NB-ARC genes in the Nip_B510 host interaction. NB-ARC domain is generally associated to R genes that are involved in Effector-Triggered Immunity (ETI) (Chisholm et al., 2006; Pieterse et al., 2009). Differences observed between host and non-host combinations suggests that the way a strain is perceived by rice roots could have been subjected to long-lasting co-adaptation events between a strain and its original host cultivar, a hypothesis that has already been proposed based on secondary metabolites profiling of rice inoculated with strain 4B and B510, and on transcriptomic response of strain 4B colonizing rice (Chamam et al., 2013; Drogue et al., 2014). The involvement of plant defense systems in PGPR-plant cooperations was mainly considered in the context of biocontrol agents, their perception leading to the induction of long-lasting and broad-spectrum systemic resistance (Van Loon et al., 1998; Van Wees et al., 2008; Pieterse et al., 2009). Induced systemic resistance (ISR) is associated with priming effect for enhanced defense inducing a few reprogramming of plant transcriptome (Verhagen et al., 2004; Wang et al., 2005; Van Wees et al., 2008). In the case of phytostimulating PGPR, it was reported that members of genus *Azospirillum* and *Burkholderia* induce defense response at a lower extent than pathogens (Bashan, 1998; Bordiec et al., 2011). Moreover, induction of plant defense mechanisms was shown to control the establishment of compatible and incompatible interactions between plants and endophytic PGPR (Miché et al., 2006; Rosenblueth and Martínez-Romero, 2006; Reinhold-Hurek and Hurek, 2011). Thus, discrepancies between endophyte and surface colonizer effects on plant defense system should be further investigated by studying the impact of other *Azospirillum* strains that display endophytic properties, such as *A. brasilense* Sp245. A recent study analyzed the impact of *A. brasilense* Sp245 inoculation on *A. thaliana* gene expression and evidenced that root transcriptome undergoes significant changes on genes related to hormone signaling and plant defense (Spaepen et al., 2014) Taking into account that plant immunity involves dynamic mechanisms that lead to temporal changes in the expression of defense related genes remains an important issue to measure the sustainability of PGPR-plant cooperations.

Modulating plant hormone balance is an important trait of phytostimulating PGPR (Richardson et al., 2009; Bashan and de-Bashan, 2010; Vacheron et al., 2013). Especially, several members of the genus *Azospirillum* are able to produce auxin, cytokinin, and gibberellin (Richardson et al., 2009; Bashan and de-Bashan, 2010). In the case of strains 4B and B510, genes related to indole-3-acetic acid (IAA) biosynthesis pathway, *ipdC*/*ppdC*, are absent from their genomes and 1-aminocyclopropane-1 carboxylate (ACC) deamination could be a relevant mechanism for hormone modulation and plant-growth promotion (Blaha et al., 2006; Prigent-Combaret et al., 2008; Kaneko et al., 2010; Wisniewski-Dyé et al., 2011). This property is encoded by *acdS* found in both pathogenic and non-pathogenic bacteria (Blaha et al., 2006). ACC is a precursor of ethylene, a gaseous hormone that represses root-growth and induces systemic resistance against pathogens (Bleecker and Kende, 2000; Pieterse et al., 2009; Galland et al., 2012). It was proposed that bacterial deamination of ACC could lead to a decrease of ethylene levels in plant roots and consequently an increase in root development (Glick, 2005; Galland et al., 2012). Interestingly, the impacts of strain 4B and strain B510 on rice root morphological traits differ. Indeed, strain 4B improves total root length, principal root length and the number of roots per plant while strain B510 seems to improve mostly principal root length (Chamam et al., 2013). Then, differential impact of strain/cultivar combinations on rice root architecture could be linked to the combination-specific regulation of many genes related to ethylene biosynthesis.

Many plant hormones are involved in both plant growth and plant immunity, two physiological traits regulated by a network of interconnected signaling pathways (Pieterse et al., 2009). As such, auxin contributes to both plant development and disease resistance in a pathway interconnected with salicylic acid signaling (Wang et al., 2007). Cross-communication between plant immunity and plant development may contribute to quick adaptation in a cost-efficient manner according to numerous trade-offs reported between growth rate and disease resistance (Walters and Heil, 2007; Pieterse et al., 2009). Thus, the strain-specific effect of *Azospirillum* on the regulation of hormone-related genes must be taken into account to appraise the cost-benefit balance of each strain/cultivar cooperation. While these two strains display similar characteristics for hormonal production (Wisniewski-Dyé et al., 2012), regulation of rice genes related to hormone signaling, notably auxin signaling, is strikingly different when considering each strain (see above). These results highlight the complexity of hormone signaling networks involved in *Azospirillum*-rice cooperation.

This study aimed at identifying genetic determinants regulated in rice roots upon *Azospirillum* inoculation, considering possible favored interaction between a strain and its original host cultivar. Thus, two rice cultivars were inoculated with two *Azospirillum* strains, resulting in four strain/cultivar combinations. The wide set of genes differentially expressed only in one of the four combinations suggest that individual genotypic variations could be the most important driving force of rice root

gene expression upon *Azospirillum* inoculation. Strain-dependent transcriptional changes observed for genes related to auxin and ethylene signaling highlight the complexity of hormone signaling networks in the *Azospirillum*-rice cooperation. In this context, unraveling cross-connected hormone networks involved in both growth promotion and plant defense response appears to be an important issue to understand mechanisms involved in beneficial interactions between PGPR and plants.

## AUTHOR CONTRIBUTIONS

Florence Wisniewski-Dyé and Claire Prigent-Combaret conceived and designed the experiments with the help of Nathalie Picault. Amel Chamam and Hervé Sanguin performed the experiments and Benoît Drogue contributed to the preparation of rice RNA samples. Benoît Drogue and Hervé Sanguin performed the analysis of microarray data, with the help of Michael Mozar for the analysis of raw data. Christel Llauro, Nathalie Picault, and Olivier Panaud performed the qRT-PCR and the part on LTR-retrotransposons. Benoît Drogue, Hervé Sanguin, Claire Prigent-Combaret, and Florence Wisniewski-Dyé wrote the manuscript; all authors contributed to the discussion and approved the final manuscript.

## ACKNOWLEDGMENTS

We are grateful to C. Louvel (Center français du riz, Arles, France) and J. B. Morel (BGPI, Montpellier, France) for gift of Cigalon and Nipponbare seeds, respectively. The platforms "Serre" and DTAMB of FR41 (Université Lyon 1) were used to carry out this work. Benoît Drogue and Amel Chamam receive fellowships from, respectively ≪Région Rhône-Alpes≫ and ≪Centre National de la Recherche Scientifique≫. This work was funded by the ANR project AZORIZ (ANR-08-BLAN-0098).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpls.2014.00607/abstract

## REFERENCES

Baron, C., and Zambryski, P. C. (1995). The plant response in pathogenesis, symbiosis, and wounding: variations on a common theme? *Annu. Rev. Genet.* 29, 107–129. doi: 10.1146/annurev.ge.29.120195.000543

Bashan, Y. (1998). Inoculants of plant growth-promoting bacteria for use in agriculture. *Biotechnol. Adv.* 16, 729–770. doi: 10.1016/S0734-9750(98)00003-2

Bashan, Y., and de-Bashan, L. E. (2010). How the plant growth-promoting bacterium *Azospirillum* promotes plant growth—a critical assessment. *Adv. Agron.* 108, 77–136. doi: 10.1016/S0065-2113(10)08002-8

Bashan, Y., Holguin, G., and de-Bashan, L. E. (2004). *Azospirillum*-plant relationships: physiological, molecular, agricultural, and environmental advances (1997–2003). *Can. J. Microbiol.* 50, 521–577. doi: 10.1139/w04-035

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* 57, 289–300.

Blaha, D., Prigent-Combaret, C., Mirza, M. S., and Moënne-Loccoz, Y. (2006). Phylogeny of the 1-aminocyclopropane-1-carboxylic acid deaminase-encoding gene *acdS* in phytobeneficial and pathogenic Proteobacteria and relation with strain biogeography. *FEMS Microbiol. Ecol.* 56, 455–470. doi: 10.1111/j.1574-6941.2006.00082.x

Bleecker, A. B., and Kende, H. (2000). Ethylene: a gaseous signal molecule in plants. *Annu. Rev. Cell Dev. Biol.* 16, 1–18. doi: 10.1146/annurev.cellbio.16.1.1

Bolstad, B. M., Irizarry, R. A., Astrand, M., and Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185–193. doi: 10.1093/bioinformatics/19.2.185

Bordiec, S., Paquis, S., Lacroix, H., Dhondt, S., Ait Barka, E., Kauffmann, S., et al. (2011). Comparative analysis of defence responses induced by the endophytic plant growth-promoting rhizobacterium *Burkholderia phytofirmans* strain PsJN and the non-host bacterium *Pseudomonas syringae* pv. *pisi* in grapevine cell suspensions. *J. Exp. Bot.* 62, 595–603. doi: 10.1093/jxb/erq291

Brusamarello-Santos, L. C. C., Pacheco, F., Aljanabi, S. M. M., Monteiro, R. A., Cruz, L. M., Baura, V. A., et al. (2011). Differential gene expression of rice roots inoculated with the diazotroph *Herbaspirillum seropedicae*. *Plant Soil* 356, 113–125. doi: 10.1007/s11104-011-1044-z

Buscaill, P., and Rivas, S. (2014). Transcriptional control of plant defence responses. *Curr. Opin. Plant Biol.* 20, 35–46. doi: 10.1016/j.pbi.2014.04.004

Cao, P., Jung, K.-H., Choi, D., Hwang, D., Zhu, J., and Ronald, P. C. (2012). The rice oligonucleotide array database: an atlas of rice gene expression. *Rice* 5:17. doi: 10.1186/1939-8433-5-17

Cassán, F., Perrig, D., Sgroy, V., Masciarelli, O., Penna, C., and Luna, V. (2009). *Azospirillum brasilense* Az39 and *Bradyrhizobium japonicum* E109, inoculated singly or in combination, promote seed germination and early seedling growth in corn (*Zea mays* L.) and soybean (*Glycine max* L.). *Eur. J. Soil Biol.* 45, 28–35. doi: 10.1016/j.ejsobi.2008.08.005

Chamam, A., Sanguin, H., Bellvert, F., Meiffren, G., Comte, G., Wisniewski-Dyé, F., et al. (2013). Plant secondary metabolite profiling evidences strain-dependent effect in the *Azospirillum–Oryza sativa* association. *Phytochemistry* 87, 65–77. doi: 10.1016/j.phytochem.2012.11.009

Chaparro, C., Guyot, R., Zuccolo, A., Piegu, B., and Panaud, O. (2007). RetrOryza: a database of the rice LTR-retrotransposons. *Nucleic Acids Res.* 35, D66–D70. doi: 10.1093/nar/gkl780

Chisholm, S. T., Coaker, G., Day, B., and Staskawicz, B. J. (2006). Host-microbe interactions: shaping the evolution of the plant immune response. *Cell* 124, 803–814. doi: 10.1016/j.cell.2006.02.008

Damiani, I., Baldacci-Cresp, F., Hopkins, J., Andrio, E., Balzergue, S., Lecomte, P., et al. (2012). Plant genes involved in harbouring symbiotic rhizobia or pathogenic nematodes. *New Phytol.* 194, 511–522. doi: 10.1111/j.1469-8137.2011.04046.x

Desbrosses, G., Contesto, C., Varoquaux, F., Galland, M., and Touraine, B. (2009). PGPR-Arabidopsis interactions is a useful system to study signaling pathways involved in plant developmental control. *Plant Signal. Behav.* 4, 321–323. doi: 10.4161/psb.4.4.8106

Drogue, B., Sanguin, H., Borland, S., Prigent-Combaret, C., and Wisniewski-Dyé, F. (2014). Genome wide profiling of *Azospirillum lipoferum* 4B gene expression during interaction with rice roots. *FEMS Microbiol. Ecol.* 87, 543–555. doi: 10.1111/1574-6941.12244

El Baidouri, M., and Panaud, O. (2013). Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. *Genome Biol. Evol.* 5, 954–965. doi: 10.1093/gbe/evt025

Elbeltagy, A., Nishioka, K., Sato, T., Suzuki, H., Ye, B., Hamada, T., et al. (2001). Endophytic colonization and in planta nitrogen fixation by a *Herbaspirillum* sp. isolated from wild rice species. *Appl. Environ. Microbiol.* 67, 5285–5293. doi: 10.1128/AEM.67.11.5285-5293.2001

El Zemrany, H., Czarnes, S., Hallett, P. D., Alamercery, S., Bally, R., and Jocteur Monrozier, L. (2007). Early changes in root characteristics of maize (*Zea mays*) following seed inoculation with the PGPR *Azospirillum lipoferum* CRT1. *Plant Soil* 291, 109–118. doi: 10.1007/s11104-006-9178-0

Fuentes-Ramirez, L. E., and Caballero-Mellado, J. (2012). "Bacterial biofertilizers," in *PGPR: Biocontrol and Biofertilization*, ed Z. A. Siddiqui (Berlin; Heidelberg: Springer-Verlag), 143–172.

Galland, M., Gamet, L., Varoquaux, F., Touraine, B., and Desbrosses, G. (2012). The ethylene pathway contributes to root hair elongation induced by the beneficial bacteria *Phyllobacterium brassicacearum* STM196. *Plant Sci.* 190, 74–81. doi: 10.1016/j.plantsci.2012.03.008

Glick, B. R. (2005). Modulation of plant ethylene levels by the bacterial enzyme ACC deaminase. *FEMS Microbiol. Lett.* 251, 1–7. doi: 10.1016/j.femsle.2005.07.030

Heulin, T., Guckert, A., and Balandreau, J. (1987). Stimulation of root exudation of rice seedlings by *Azospirillum* strains: carbon budget under gnotobiotic conditions. *Biol. Fertil. Soils* 4, 9–14. doi: 10.1007/BF00280344

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., et al. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4, 249–264. doi: 10.1093/biostatistics/4.2.249

Kaneko, T., Minamisawa, K., Isawa, T., Nakatsukasa, H., Mitsui, H., Kawaharada, Y., et al. (2010). Complete genomic structure of the cultivated rice endophyte *Azospirillum sp.* B510. *DNA Res.* 17, 37–50. doi: 10.1093/dnares/dsp026

Kumar, A., Prakash, A., and Johri, B. N. (2011). "*Bacillus* as PGPR in crop ecosystem," in *Bacteria in Agrobiology: Crop Ecosystems*, ed D. K. Maheshwari (Berlin; Heidelberg: Springer), 37–59.

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Lugtenberg, B., and Kamilova, F. (2009). Plant-growth-promoting rhizobacteria. *Annu. Rev. Microbiol.* 63, 541–556. doi: 10.1146/annurev.micro.62.081307.162918

Marcel, S., Sawers, R., Oakeley, E., Angliker, H., and Paszkowski, U. (2010). Tissue-adapted invasion strategies of the rice blast fungus *Magnaporthe oryzae*. *Plant Cell* 22, 3177–3187. doi: 10.1105/tpc.110.078048

Miché, L., Battistoni, F., Gemmer, S., Belghazi, M., and Reinhold-Hurek, B. (2006). Upregulation of jasmonate-inducible defense proteins and differential colonization of roots of *Oryza sativa* cultivars with the endophyte *Azoarcus sp. Mol. Plant. Microbe Interact.* 19, 502–511. doi: 10.1094/MPMI-19-0502

Michaud, E. J., Vugt, M. J., van Bultman, S. J., Sweet, H. O., Davisson, M. T., and Woychik, R. P. (1994). Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage. *Genes Dev.* 8, 1463–1472. doi: 10.1101/gad.8.12.1463

van de Mortel, J. E., de Vos, R. C. H., Dekkers, E., Pineda, A., Guillod, L., Bouwmeester, K., et al. (2012). Metabolic and transcriptomic changes induced in *Arabidopsis* by the rhizobacterium *Pseudomonas fluorescens* SS101. *Plant Physiol.* 160, 2173–2188. doi: 10.1104/pp.112.207324

O'Brien, H. E., Thakur, S., and Guttman, D. S. (2011). Evolution of plant pathogenesis in *Pseudomonas syringae*: a genomics perspective. *Annu. Rev. Phytopathol.* 49, 269–289. doi: 10.1146/annurev-phyto-072910-095242

Picault, N., Chaparro, C., Piegu, B., Stenger, W., Formey, D., Llauro, C., et al. (2009). Identification of an active LTR retrotransposon in rice. *Plant J.* 58, 754–765. doi: 10.1111/j.1365-313X.2009.03813.x

Pieterse, C. M. J., Leon-Reyes, A., Van der Ent, S., and Van Wees, S. C. M. (2009). Networking by small-molecule hormones in plant immunity. *Nat. Chem. Biol.* 5, 308–316. doi: 10.1038/nchembio.164

Popovici, J., Walker, V., Bertrand, C., Bellvert, F., Fernandez, M. P., and Comte, G. (2011). Strain specificity in the *Myricaceae–Frankia* symbiosis is correlated to plant root phenolics. *Funct. Plant Biol.* 38, 682–689. doi: 10.1071/FP11144

Prigent-Combaret, C., Blaha, D., Pothier, J. F., Vial, L., Poirier, M.-A., Wisniewski-Dyé, F., et al. (2008). Physical organization and phylogenetic analysis of *acdR* as leucine-responsive regulator of the 1-aminocyclopropane-1-carboxylate deaminase gene *acdS* in phytobeneficial *Azospirillum lipoferum* 4B and other Proteobacteria. *FEMS Microbiol. Ecol.* 65, 202–219. doi: 10.1111/j.1574-6941.2008.00474.x

Reinhold-Hurek, B., and Hurek, T. (2011). Living inside plants: bacterial endophytes. *Curr. Opin. Plant Biol.* 14, 435–443. doi: 10.1016/j.pbi.2011.04.004

Rice Annotation Project, Tanaka, T., Antonio, B. A., Kikuchi, S., Matsumoto, T., Nagamura, Y., et al. (2008). The Rice Annotation Project Database (RAP-DB): 2008 update. *Nucleic Acid Res.* 36, D1028–D1033. doi: 10.1093/nar/gkm978

Richardson, A. E., Barea, J.-M., McNeill, A. M., and Prigent-Combaret, C. (2009). Acquisition of phosphorus and nitrogen in the rhizosphere and plant growth promotion by microorganisms. *Plant Soil* 321, 305–339. doi: 10.1007/s11104-009-9895-2

Rosenblueth, M., and Martínez-Romero, E. (2006). Bacterial endophytes and their interactions with hosts. *Mol. Plant. Microbe Interact.* 19, 827–837. doi: 10.1094/MPMI-19-0827

Saharan, B., and Nehra, V. (2011). Plant growth promoting rhizobacteria: a critical review. *Life Sci. Med. Res.* 2011, 1–30.

Schumacher, J., and Tudzynski, P. (2012). "Morphogenesis and infection in *Botrytis cinerea*," in *Morphogenesis and Pathogenicity in Fungi*, eds J. Pérez-Martín and A. Di Pietro (Berlin; Heidelberg: Springer), 225–241.

Schwessinger, B., and Zipfel, C. (2008). News from the frontline: recent insights into PAMP-triggered immunity in plants. *Curr. Opin. Plant Biol.* 11, 389–395. doi: 10.1016/j.pbi.2008.06.001

Smith, S. E., and Read, D. J. (2008). *Mycorrhizal Symbiosis*. London: Academic Press.

Somers, E., Vanderleyden, J., and Srinivasan, M. (2004). Rhizosphere bacterial signalling: a love parade beneath our feet. *Crit. Rev. Microbiol.* 30, 205–240. doi: 10.1080/10408410490468786

Soto, M. J., Domínguez-Ferreras, A., Pérez-Mendoza, D., Sanjuán, J., and Olivares, J. (2009). Mutualism versus pathogenesis: the give-and-take in plant-bacteria interactions. *Cell. Microbiol.* 11, 381–388. doi: 10.1111/j.1462-5822.2009.01282.x

Spaepen, S., Bossuyt, S., Engelen, K., Marchal, K., and Vanderleyden, J. (2014). Phenotypical and molecular responses of *Arabidopsis thaliana* roots as a result of inoculation with the auxin-producing bacterium *Azospirillum brasilense*. *New Phytol.* 201, 850–861. doi: 10.1111/nph.12590

Steenhoudt, O., and Vanderleyden, J. (2000). *Azospirillum*, a free-living nitrogen-fixing bacterium closely associated with grasses: genetic, biochemical and ecological aspects. *FEMS Microbiol. Rev.* 24, 487–506. doi: 10.1111/j.1574-6976.2000.tb00552.x

Thomas-Bauzon, D., Weinhard, P., Villecourt, P., and Balandreau, J. (1982). The spermosphere model. I. Its use in growing, counting, and isolating $N_2$-fixing bacteria from the rhizosphere of rice. *Can. J. Microbiol.* 28, 922–928. doi: 10.1139/m82-139

Vacheron, J., Desbrosses, G., Bouffaud, M.-L., Touraine, B., Moënne-Loccoz, Y., Muller, D., et al. (2013). Plant growth-promoting rhizobacteria and root system functioning. *Front. Plant Sci.* 4:356. doi: 10.3389/fpls.2013.00356

Van Loon, L. C., Bakker, P. A., and Pieterse, C. M. (1998). Systemic resistance induced by rhizosphere bacteria. *Annu. Rev. Phytopathol.* 36, 453–483. doi: 10.1146/annurev.phyto.36.1.453

Van Wees, S. C., Van der Ent, S., and Pieterse, C. M. (2008). Plant immune responses triggered by beneficial microbes. *Curr. Opin. Plant Biol.* 11, 443–448. doi: 10.1016/j.pbi.2008.05.005

Verhagen, B. W. M., Glazebrook, J., Zhu, T., Chang, H.-S., van Loon, L. C., and Pieterse, C. M. J. (2004). The transcriptome of rhizobacteria-induced systemic resistance in *Arabidopsis*. *Mol. Plant. Microbe Interact.* 17, 895–908. doi: 10.1094/MPMI.2004.17.8.895

Walker, V., Bertrand, C., Bellvert, F., Moënne-Loccoz, Y., Bally, R., and Comte, G. (2011). Host plant secondary metabolite profiling shows a complex, strain-dependent response of maize to plant growth-promoting rhizobacteria of the genus *Azospirillum*. *New Phytol.* 189, 494–506. doi: 10.1111/j.1469-8137.2010.03484.x

Walker, V., Couillerot, O., Felten, A. V., Bellvert, F., Jansa, J., Maurhofer, M., et al. (2012). Variation of secondary metabolite levels in maize seedling roots induced by inoculation with *Azospirillum*, *Pseudomonas* and *Glomus* consortium under field conditions. *Plant Soil* 356, 151–163. doi: 10.1007/s11104-011-0960-2

Walters, D., and Heil, M. (2007). Costs and trade-offs associated with induced resistance. *Physiol. Mol. Plant Pathol.* 71, 3–17. doi: 10.1016/j.pmpp.2007.09.008

Wang, D., Pajerowska-Mukhtar, K., Culler, A. H., and Dong, X. (2007). Salicylic acid inhibits pathogen growth in plants through repression of the auxin signaling pathway. *Curr. Biol.* 17, 1784–1790. doi: 10.1016/j.cub.2007.09.025

Wang, Y., Ohara, Y., Nakayashiki, H., Tosa, Y., and Mayama, S. (2005). Microarray analysis of the gene expression profile induced by the endophytic plant growth-promoting rhizobacteria, *Pseudomonas fluorescens* FPT9601-T5 in *Arabidopsis*. *Mol. Plant Microbe Interact.* 18, 385–396. doi: 10.1094/MPMI-18-0385

Wisniewski-Dyé, F., Borziak, K., Khalsa-Moyers, G., Alexandre, G., Sukharnikov, L. O., Wuichet, K., et al. (2011). *Azospirillum* genomes reveal transition of bacteria from aquatic to terrestrial environments. *PLoS Genet.* 7:e1002430. doi: 10.1371/journal.pgen.1002430

Wisniewski-Dyé, F., Drogue, B., Borland, S., and Prigent-Combaret, C. (2013). "*Azospirillum*-plant interaction: from root colonization to plant growth promotion," in *Beneficial Plant-Microbial Interactions: Ecology and Applications,* eds M. Belén Rodelas González and J. Gonzalez-López (Boca Raton, FL: CRC Press), 237–269.

Wisniewski-Dyé, F., Lozano, L., Acosta-Cruz, E., Borland, S., Drogue, B., Prigent-Combaret, C., et al. (2012). Genome sequence of *Azospirillum*

*brasilense* CBG497 and comparative analyses of *Azospirillum* core and accessory genomes provide insight into niche adaptation. *Genes* 3, 576–602. doi: 10.3390/genes3040576

Yasuda, M., Isawa, T., Shinozaki, S., Minamisawa, K., and Nakashita, H. (2009). Effects of colonization of a bacterial endophyte, *Azospirillum* sp. B510, on disease resistance in rice. *Biosci. Biotechnol. Biochem.* 73, 2595–2599. doi: 10.1271/bbb.90402

Yuan, Q., Ouyang, S., Wang, A., Zhu, W., Maiti, R., Lin, H., et al. (2005). The institute for genomic research Osa1 rice genome annotation database. *Plant Physiol.* 138, 18–26. doi: 10.1104/pp.104.059063

**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.

*Received: 24 September 2014; accepted: 18 October 2014; published online: 06 November 2014.*

*Citation: Drogue B, Sanguin H, Chamam A, Mozar M, Llauro C, Panaud O, Prigent-Combaret C, Picault N and Wisniewski-Dyé F (2014) Plant root transcriptome profiling reveals a strain-dependent response during Azospirillum-rice cooperation. Front. Plant Sci. **5**:607. doi: 10.3389/fpls.2014.00607*

*This article was submitted to Plant-Microbe Interaction, a section of the journal Frontiers in Plant Science.*

Les résultats montrent que les deux variétés de riz répondent différemment aux deux *Azospirillum* en termes de stimulation de croissance. Une réponse différentielle est observée au niveau du métabolisme secondaire du riz, et dépend de la combinaison variété de riz – *Azopirillum* considérée. L'association induit également des modifications de l'expression de nombreux gènes chez les deux partenaires (bactérie et plante) et ces modifications dépendent aussi de la combinaison plante-bactérie. Cette étude a conduit à l'identification de gènes clés impliqués dans cette association.

En effet, un jeu de 34 gènes est différentiellement exprimé et commun dans les 2 cultivars avec les 2 souches d'*Azospirillum*. Parmi ces gènes, on retrouve un gène orthologue de PR10 (Pathogen Related) de *brachypodium* (connu pour son rôle dans les réactions de défense des plantes), ainsi que d'autres gènes contenant un domaine NB-ARC (caractéristique des fonctions de résistance), également induits au cours de l'inoculation. Ces 34 gènes sont utilisable comme des marqueurs des interactions *Azospirillum*-riz.

Les résultats ont aussi mis en évidence une réponse du riz dépendante de la souche, avec 83 % des gènes différentiellement exprimés, classés comme spécifiques de la combinaison. Le grand jeu de gènes identifiés dans seulement une des 4 combinaisons suggère que les variations génotypiques individuelles pourraient être la force motrice la plus importante de l'expression des gènes dans les racines du riz au cours de l'inoculation avec *Azospirillum.*

Quelque soit la combinaison, la plupart des gènes différentiellement exprimés sont impliqués dans le métabolisme primaire, le transport et la régulation de la transcription.

Lorsque l'on regarde les gènes impliqués dans la réponse au stress et dans les réactions de défense de la plante, il apparaît que la souche B510, une souche présentant des propriétés endophytes, entraîne la répression d'un grand jeu de gènes par rapport à la souche 4B.

Les changements transcriptionnels observés pour les gènes liés à l'auxine et à l'éthylène en fonction de la souche observée démontrent une grande complexité des voies de signalisation hormonales dans la coopération *Azospirillum*-riz. Dans ce contexte, la compréhension des réseaux de régulation impliquant de nombreuses hormones et ayant un impact sur la stimulation de la croissance et les réponses de défense de la plante paraît importante. Ces interconnections permettent notamment aux végétaux de s'adapter rapidement aux variations environnementales et d'ajuster leur réponse de manière efficace. Cette voie semble donc être prometteuse pour comprendre les mécanismes impliqués dans les interactions bénéfiques entre les plantes et les PGPR.

L'ensemble de ces résultats tend à montrer que l'expression des gènes de chaque partenaire de la coopération *Azospirillum*-riz dépend à la fois de la souche bactérienne et du génotype de la plante. Ces réponses transcriptomiques spécifiques s'observent à l'échelle d'une souche, d'un cultivar ou d'une combinaison souche/cultivar et soulignent la complexité des mécanismes cellulaires mis en place au cours de l'interaction entre *Azospirillum* et les racines de riz. Par ailleurs, ces résultats semblent en faveur de l'hypothèse selon laquelle la spécificité de l'hôte dans les interactions PGPR-plantes serait régie par des réponses adaptatives réciproques qui dépendent des caractéristiques génotypiques de chaque partenaire de la coopération.

En ce qui concerne les rétrotransposons à LTR, ne connaissant pas leurs polymorphismes d'insertion entre Cigalon et Nipponbare, nous avons seulement interprété les données obtenues pour Nipponbare. L'inoculation en présence d'*Azospirillum* entraîne la modification de l'expression d'une centaine de familles de rétrotransposons à LTR pour les 2 souches. Parmi ces éléments, 22 % sont communs entre les 2 combinaisons d'inoculation. Lorsque nous comparons les familles différentiellement exprimées par rapport à celles connues comme actives dans les mutants de riz re-séquencés (Sabot et al., 2011), nous retrouvons les familles *BAJIE*, *Osr10*, *Osr37*, *RIRE2* et *RIRE3*.

De plus, nous avons recherché si ces éléments peuvent modifier l'expression de gènes adjacents. Nous avons identifié 8 rétrotransposons à LTR différentiellement exprimés dans la région 5' de gènes eux-mêmes différentiellement exprimés et 9 dans la région 3' d'autres gènes. Cependant, aucun co-transcrit gène-ET et aucun impact direct de l'élément sur l'expression du gène adjacent n'ont pu être mis en évidence.

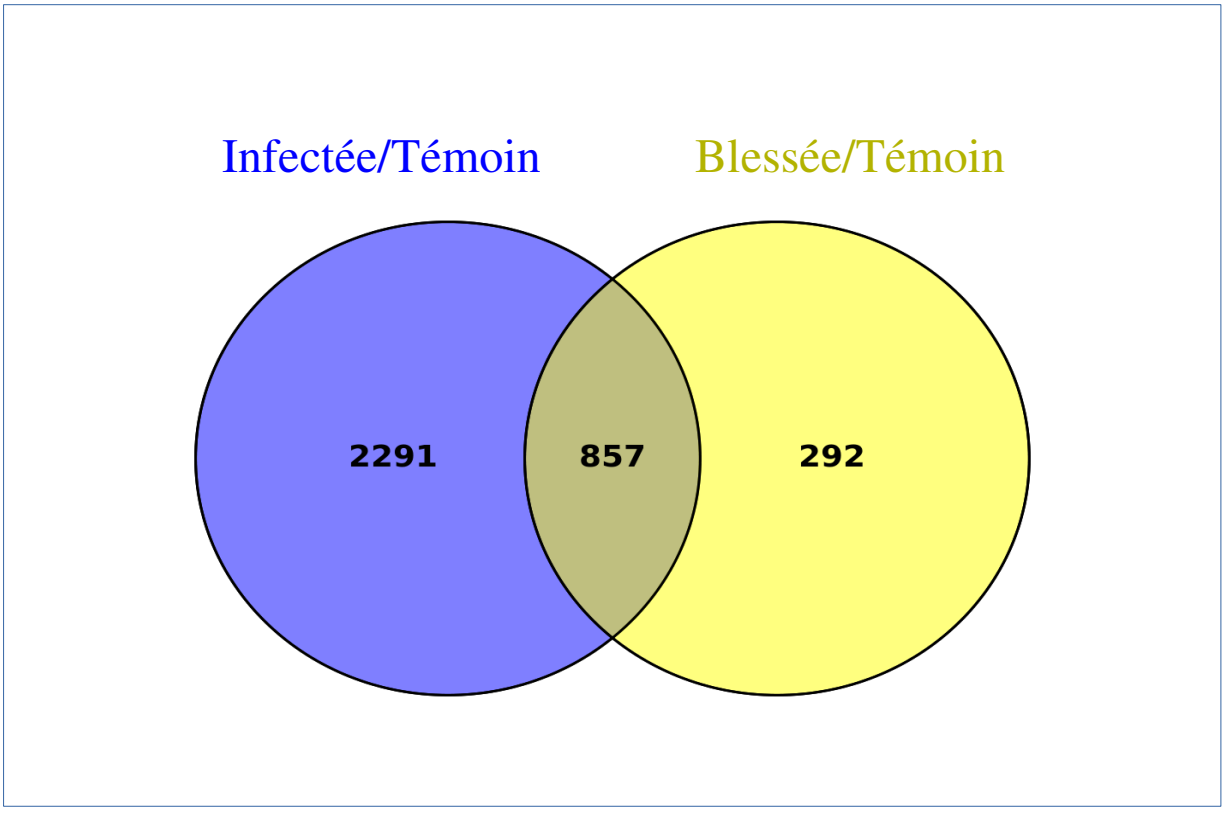



**Figure 13 : Diagramme de Venn représentant les gènes sur-exprimés au cours de l'infection par RYMV.** Dans ce diagramme sont indiqués le nombre de gènes sur-exprimés (1) entre les plantes infectées et les plantes témoins (2) entre les plantes blessées et les témoins. Pour chaque analyse, seuls les gènes présentant les caractéristiques suivantes sont retenus : $Log_2(FC)>1$ et $P_{adj}<0,05$.

### *2.2.3. RYMV*

Nous avons aussi utilisé l'approche transcriptomique pour identifier les gènes et les ETs différentiellement exprimés au cours de l'infection des feuilles de riz par le virus de la panachure jaune du riz (RYMV) en collaboration avec C. Brugidou de l'IRD Montpellier. Le virus RYMV est un virus à ARN (+) qui appartient au genre *Sobemovirus*. Il est endémique en Afrique et est le principal pathogène du riz irrigué (Yassi et al., 1994). Il entraîne un jaunissement des feuilles et une perte de rendement pouvant atteindre 20 à 50 %. Les différentes espèces du riz présentent des niveaux de résistance variable et *japonica* a été caractérisé comme partiellement résistant à ce virus. Notre objectif était d'identifier les gènes et ETs impliqués dans les voies biochimiques permettant à la plante de résister à cette infection.

Les plantes ont été cultivées pendant 2 semaines dans des conditions standards, puis infectées par le virus RYMV. Au laboratoire, pour réaliser l'infection par le virus, nous sommes obligés de faire une blessure pour permettre au virus de pénétrer dans la plante. Nous avons réalisé un lot témoin où les plantes ont seulement subit une blessure avec un scalpel et un autre lot témoin où les plantes n'ont subit ni blessure, ni infection.

Sur la puce « transposome » nous avons mis les oligonucléotides correspondants au virus RYMV et avons pu nous assurer que les plantes étaient bien infectées par le présence de messagers du virus.

Je ne détaillerai pas dans cette partie toutes les démarches réalisées pour l'analyse des données des hybridations car nous avons effectué exactement les mêmes analyses que pour les deux projets de recherche précédents.

En ce qui concerne l'analyse de l'expression des gènes, elle a tout d'abord révélé dans les plantes blessées 1149 gènes différentiellement exprimés (853 sur-exprimés et 296 sous-exprimés) : certains connus pour avoir un rôle lors de blessures. Pour les plantes infectées par le virus, 3148 gènes présentent une modification de leur expression (2450 sur-exprimés et 698 sous-exprimés). Le diagramme de Venn (Figure 13) regroupant les différents gènes montre que 2291 gènes sont spécifiques de l'infection. Une vingtaine de candidats a été validée par PCR quantitative. Parmi ces gènes, nous en retrouvons un grand nombre déjà connus pour leur implication dans les réponses aux stress, tel que des protéines de réponse à des éliciteurs, des facteurs de transcription en réponse à l'éthylène, des gènes jouant un rôle dans la réparation des parois ou dans la résistance aux pathogènes. L'analyse est encore en cours mais l'étude de la 'Gene Ontology' (Figure 14) indique que la plupart des gènes différentiellement exprimés au cours de l'infection sont impliqués dans la

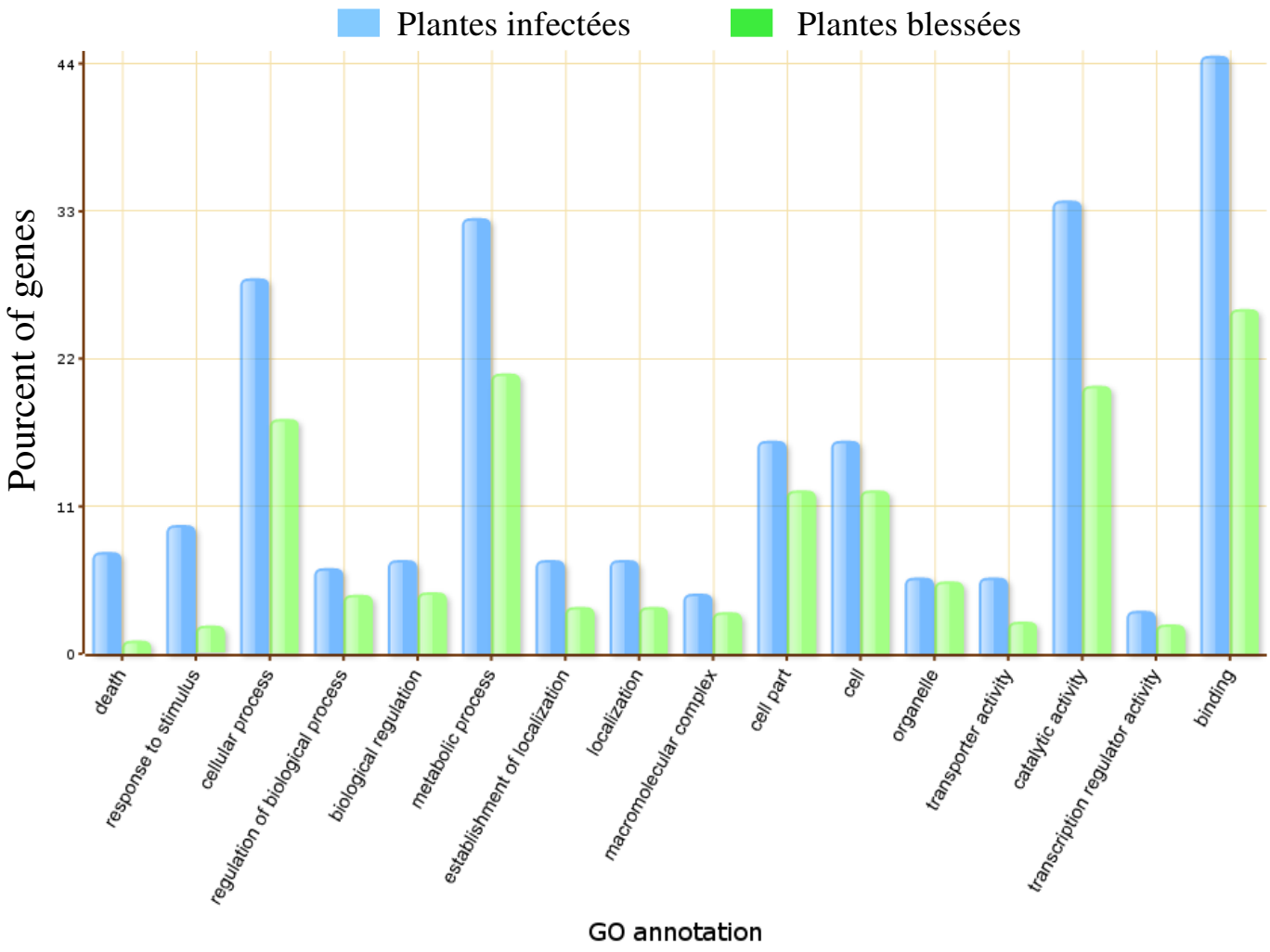



**Figure 14 : Pourcentage de gènes sur-exprimés dans les feuilles de riz suite à l'infection par RYMV.** La Gene Ontology a été générée par GOSlim. En abscisse sont représentées les différentes classes de fonctions moléculaires et de processus biologiques. En ordonnée est indiqué le pourcentage de gènes différentiellement exprimés entre les plantes infectées (en bleu) et les plantes blessées (en vert).

mort cellulaire, la réponse aux stimuli et aux stress, les processus cellulaires et métaboliques ainsi que dans les activités catalytiques et de fixation (Figure 14).

Cette étude transcriptomique a mis en évidence une forte activation de la transcription des rétrotransposons à LTR dans les plantes blessées ainsi que dans les plantes blessées et infectées (plus de 200). Nous retrouvons les familles connues comme actives transpositionnellement :*BAJIE*, *Osr10*, *Osr37*, *RIRE2* et *RIRE3*.

### *2.2.4. Bilan global des analyses transcriptomiques*

Depuis 10 ans, j'ai développé des stratégies qui ont permis le développement d'outils novateurs : les puces « Transposome ».

La première puce « transposome » que nous avons élaboré a permis de détecter efficacement l'expression des rétrotransposons à LTR et de valider cette approche, jamais utilisée auparavant pour l'analyse des ETs chez les plantes.

Nous avons ensuite développé une nouvelle puce dans l'optique de mettre en place une approche génomique globale et détecter l'expression des ETs, de tous les gènes du riz, et d'éventuels co-transcrits gène-ET. Cette approche a permis : (i) d'identifier les gènes et les différentes voies métaboliques impliqués lors de stress variés (ferrique, coopération bactérienne et RYMV), (ii) de montrer la présence importante des ETs dans le transcriptome végétal et de confirmer leur activation transcriptionnelle en situations de stress et (iii) un recensement des conditions d'expression de nombreux ETs. Dans les différentes conditions testées, généralement, nous ne retrouvons pas les mêmes éléments actifs, mais par contre souvent les mêmes familles. Par exemple, dans nos 3 conditions, nous retrouvons des familles connues comme étant actives transpositionnellement : *BAJIE*, *Osr10*, *Osr37*, *RIRE2* et *RIRE3*. En revanche, parmi les autres familles décrites comme active transpositionnellement chez le riz, *RN216* et *RN363* sont seulement sur-exprimées lors du stress ferrique.

Par contre, la partie concernant les co-transcrits gènes-ET fut infructueuse. Nous avons eu des difficultés à valider les stratégies expérimentales de détection globale de co-transcrits ETs/gènes. Dans les trois expériences, des essais ont été réalisés lorsque l'on avait un rétrotransposon à LTR différentiellement exprimé dans une zone de 3 kb en amont ou en aval d'un gène différentiellement exprimé. En raison de l'accumulation importante de données et de la nécessité de valider chaque association locus par locus, nous nous sommes attachés, dans cette partie, à l'étude d'un nombre restreint de cas qui nous parurent intéressants, soit en raison de certaines particularités de l'association gène-ET, soit en raison de l'intérêt du gène impliqué, mais sans succès. Nous avons ainsi obtenu une vision globale de leurs conditions d'expression mais malheureusement nous n'avons pas pu mettre en évidence d'exemple dans lequel des séquences géniques étaient placées sous contrôle direct de l'expression d'un ET. Il serait intéressant d'utiliser les nouvelles technologies comme le RNAseq qui nous permettrait de détecter des co-transcrits gènes-ET dans différentes conditions de stress.

## 3. Impact des rétrotransposons à LTR sur l'évolution des génomes

Au sein de l'équipe, nous avons ensuite décidé de nous concentrer sur les origines de l'augmentation très rapide et brutale de la transposition des éléments transposables. Pour essayer de répondre à cette question, plusieurs approches sont possibles mais des résultats récemment obtenus au sein de l'équipe ont mis en évidence que des transferts horizontaux d'éléments transposables ont lieu dans les plantes. De plus, ces éléments transférés sont actifs dans le nouveau génome. Il a été montré que dans certains cas ils ont subit une explosion de la transposition, entraînant l'amplification d'une centaine de copie.

D'après ces données, nous proposons un nouveau modèle, où les transferts horizontaux des éléments transposables pourraient être un processus de l'évolution permettant leur survie dans les génomes eucaryotes. L'hypothèse principale à la base de ce modèle est que les transferts horizontaux permettraient l'introduction des éléments transposables dans un génome où il pourrait échapper aux différentes voies d'inactivation ciblant les éléments endogènes. Cependant, bien que les transferts horizontaux soient très courant chez les bactéries, l'évidence des transferts horizontaux chez les eucaryotes reste encore assez rare.

Pour tester ce modèle, nous avons mis au point une nouvelle méthode de détection de transferts horizontaux dans les génomes séquencés. Des programmes bio-informatiques permettant de rechercher chez tous les organismes les transferts horizontaux ont été développés. Cette recherche est basée sur les « incongruences » phylogénétiques. Cette procédure a été utilisée à petite échelle et avec succès dans l'équipe. Il était donc pertinent d'employer cette méthode à grande échelle, chez tous les organismes séquencés. Nous avons ainsi réalisé des comparaisons de séquences et détecté les séquences présentant une identité entre deux espèces supérieure à celle attendue par rapport à leur temps de radiation.

Cette étude a ensuite été complétée par des analyses portant sur l'activité récente des éléments transposables impliqués dans les transferts horizontaux dans les génomes concernés. Ce test a été réalisé à travers une analyse phénétique à partir des données *in silico* des génomes des différentes espèces impliquées dans ces transferts. J'ai participé aux validations *in vivo* qui ont été réalisées grâce au clonage et séquençage des différents candidats. Ce travail a été publié dans Genome Research en 2014 (El Baidouri et al., 2014).

# Widespread and frequent horizontal transfers of transposable elements in plants

Moaine El Baidouri,[1,2] Marie-Christine Carpentier,[1] Richard Cooke,[1] Dongying Gao,[2] Eric Lasserre,[1] Christel Llauro,[1] Marie Mirouze,[3] Nathalie Picault,[1] Scott A. Jackson,[2] and Olivier Panaud[1,4]

[1]*Université de Perpignan Via Domitia, Laboratoire Génome et Développement des Plantes, UMR5096 CNRS/UPVD, 66860 Perpignan Cedex, France;* [2]*Center for Applied Genetic Technologies, University of Georgia, Athens, Georgia 30602, USA;* [3]*Institut de Recherche pour le Développement, UMR232, 34394 Montpellier, France*

**Vertical, transgenerational transmission of genetic material occurs through reproduction of living organisms. In addition to vertical inheritance, horizontal gene transfer between reproductively isolated species has recently been shown to be an important, if not dominant, mechanism in the evolution of prokaryotic genomes. In contrast, only a few horizontal transfer (HT) events have been characterized so far in eukaryotes and mainly concern transposable elements (TEs). Whether these are frequent and have a significant impact on genome evolution remains largely unknown. We performed a computational search for highly conserved LTR retrotransposons among 40 sequenced eukaryotic genomes representing the major plant families. We found that 26 genomes (65%) harbor at least one case of horizontal TE transfer (HTT). These transfers concern species as distantly related as palm and grapevine, tomato and bean, or poplar and peach. In total, we identified 32 cases of HTTs, which could translate into more than 2 million among the 13,551 monocot and dicot genera. Moreover, we show that these TEs have remained functional after their transfer, occasionally causing a transpositional burst. This suggests that plants can frequently exchange genetic material through horizontal transfers and that this mechanism may be important in TE-driven genome evolution.**

**[Supplemental material is available for this article.]**

Transposable elements (TEs) are mobile genomic DNA sequences that are found in almost all living organisms (Finnegan 1985). They so densely populate the genomes of many eukaryotic species that they are often the major components, as in human (>50%) (Prak and Kazazian 2000) or bread wheat (>95%) (Bennetzen 2000). In this regard, TEs have been shown to have a major impact on both structural and functional modifications of genomes (Bennetzen 2000; Feschotte 2008). They are usually classified into two distinct types. Class I elements (retrotransposons) transpose via an RNA intermediate through a copy and paste mechanism, whereas class II elements (transposons) transpose through a cut-and-paste mechanism (Wicker et al. 2007). Both classes comprise various types (orders and superfamilies in Wicker's classification). Although most of these types can be found in all the plant genomes sequenced so far, LTR retrotransposons represent by far the major genomic constituents in the kingdom. In this regard, LTR retrotransposons have been shown to strongly impact genome structure (Piegu et al. 2006), whereas several reports have demonstrated their putative functional impact as epigenetic mediators (Kobayashi et al. 2004).

TEs achieve their transposition cycle within their host and are thus considered to be lineage specific because, like genes, they are inherited vertically from one generation to another. However, unlike genes, they do not encode any information essential for their host, and their insertion into genes can in some cases have a negative effect on fitness. This "selfish" and potentially deleterious nature has raised the question of their persistence in eukaryotic lineages, especially after it was shown that TEs are strictly controlled by several silencing pathways (Slotkin and Martienssen 2007; Rigal and Mathieu 2011) and efficiently eliminated from their host genomes through deletions (Vitte and Panaud 2005). Horizontal transfers could allow TEs to escape this process by transposing into a new "naive" host genome, therefore ensuring their long-term survival. However, although horizontal gene transfers are very common in Bacteria (Rocha 2013), evidence of HTTs in eukaryotes remains scarce, although recent reports suggest their potential impact in genome evolution (Schaack et al. 2010). Three criteria have been defined for the detection of HTTs: patchy distributions of TEs in phylogenies; identification of TEs exhibiting high sequence similarity between distantly related taxa; and phylogenetic incongruence between the host and TEs (Gilbert et al. 2010; Kuraku et al. 2012; Wallau et al. 2012; Walsh et al. 2013). An exhaustive search for HTTs that meet these three criteria in a wide taxonomic range thus requires a comprehensive set of genomic resources. Next-generation sequencing (NGS) has made available full genome sequences for many organisms, enabling genome-wide comparative surveys for a large panel of evolutionary lineages. Using such resources, we surveyed HTTs across the plant kingdom, and we show that they are very frequent and widespread among monocots and dicots.

## Results

We conducted an ab initio search for HTTs among 40 angiosperm species belonging to 36 monocot and eudicot genera for which

**[4]Corresponding author**
**E-mail panaud@univ-perp.fr**
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.164400.113. Freely available online through the *Genome Research* Open Access option.

high-quality genome sequences are available (Fig. 1; Supplemental Table 1). We focused on LTR retrotransposons (LTR-RTs), class I TEs constituting the largest portion of the TE repertoire of plant genomes (Wicker et al. 2007). First, we retrieved full-size elements from genome sequences (see Methods). Around 300,000 elements were clustered into distinct families following our recently published method (El Baidouri and Panaud 2013). HTT candidates were detected by applying a 90% identity threshold within either the monocot or dicot classes and an 85% identity threshold between these classes (to take into account their greater divergence time). The former value corresponds to a date of ~3 My (using an average LTR-RT divergence rate of $1.6 \times 10^{-8}$ substitutions/site/year estimated in plants) (Ma and Bennetzen 2004). It is lower than the divergence times between the genera in our data set (5 My for the two *Arabidopsis* species, >150 My for the monocot-dicot split), ensuring the detection of horizontally, as opposed to vertically, inherited TE sequences. Thirty-two families containing elements from at least two distinct genera were identified (Supplemental Data 1), suggesting at least 32 horizontal transfers, based on their patchy distribution in the plant phylogenetic tree (Table 1; Fig. 1). Among these, BG12 was previously described by our group using a different approach (Roulin et al. 2009).

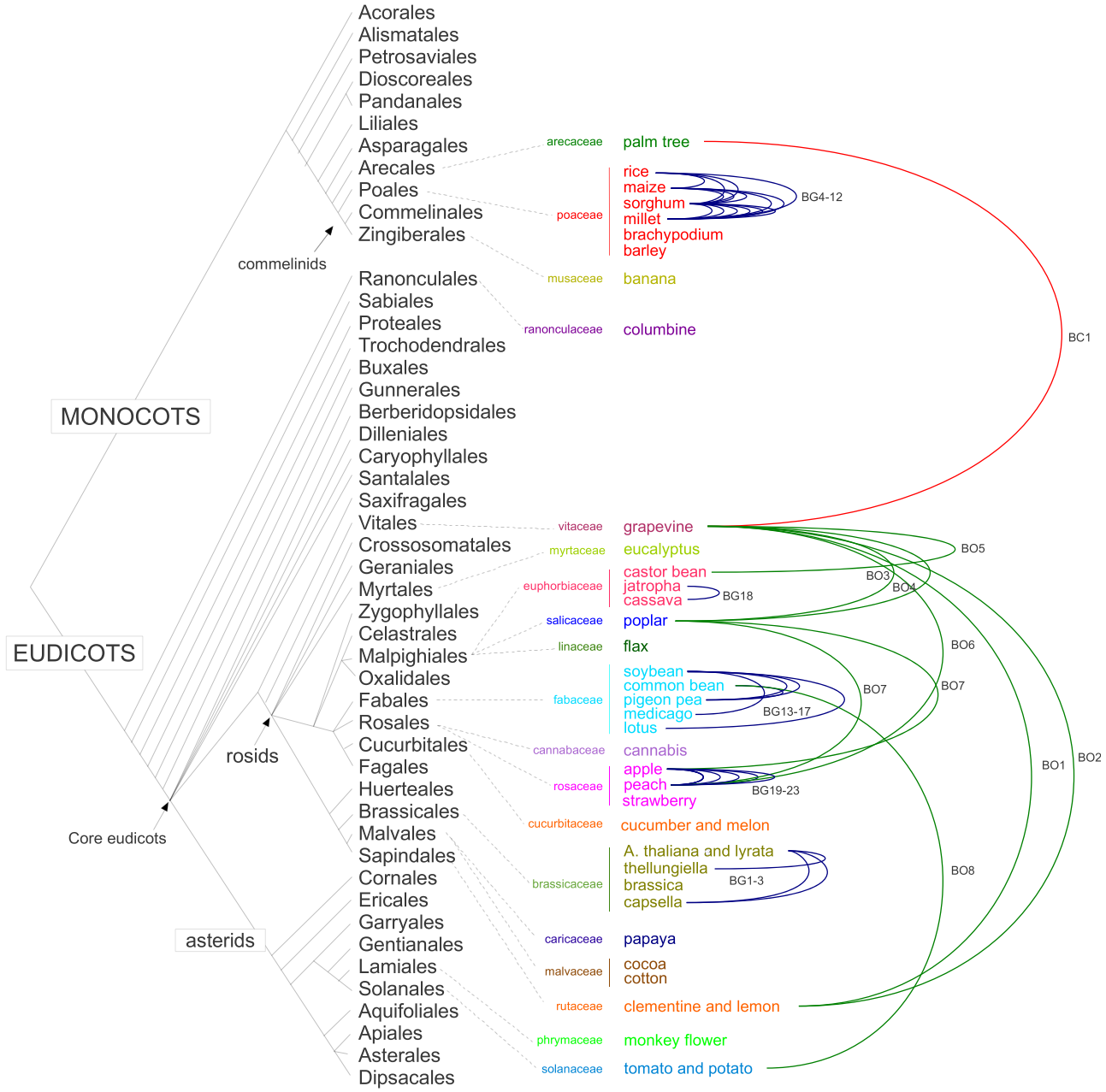



**Figure 1.** Horizontal transposon transfers (HTTs) identified in our survey of 40 fully sequenced plant genomes. The 40 species used in this study together with the color-coded families to which they belong are positioned in the monocot/dicot phylogenetic tree obtained from APG3 (http://www.mobot.org/MOBOT/research/APweb/) (see Supplemental Table 1 for details). Each HTT is represented by a line connecting the species involved (red line, transfer between classes [BC]; green line, transfer between orders [BO]; blue line, transfer between genera [BG]).

**Table 1. Description of HTTs identified in the survey of 40 plant genomes**

**HTTs between classes or between orders[a]**

| HTT | Species | LTR-RT identity (%) | Gene's identity based on CDS (%) | Incongruence | Validation |
|---|---|---|---|---|---|
| BC1 | Grapevine/palm | 86 | 80 (±3) | + | + |
| BO1 | Clementine/grapevine | 95 | 82 (±3) | + | + |
| BO2 | Clementine/grapevine | 94 | 82 (±3) | + | + |
| BO3 | Grapevine/poplar | 93 | 82 (±3) | + | + |
| BO4 | Grapevine/poplar | 90 | 82 (±3) | + | + |
| BO5 | Castor bean/grapevine | 92 | 82 (±3) | + | n.a. |
| BO6 | Grapevine/peach | 92 | 82 (±3) | + | + |
| BO7 | Peach/poplar | 92 | 81 (±3) | + | + |
| BO8 | Common bean/tomato | 94 | 79 (±4) | + | + |

**HTTs between genera[b]**

| HTT | Species | LTR-RT identity (%) | Gene's identity based on Ks (%) | Incongruence | Validation |
|---|---|---|---|---|---|
| BG1 | *Arabidopsis lyrata*/*Thellungiella* | 92 | 70 (±8) | n.a. | n.a. |
| BG2 | *Arabidopsis lyrata*/*Capsella* | 91 | 75 (±7) | n.a. | n.a. |
| BG3 | *Arabidopsis lyrata*/*Capsella* | 91 | 75 (±7) | n.a. | n.a. |
| BG4 | Millet/sorghum | 91.5 | 69 (±11) | + | + |
| BG5 | Millet/sorghum | 90 | 69 (±11) | n.a. | + |
| BG6 | Millet/sorghum | 91 | 69 (±11) | n.a. | + |
| BG7 | Millet/sorghum | 93 | 69 (±11) | n.a. | + |
| BG8 | Millet/rice | 91.5 | 39 (±12) | + | + |
| BG8 | Millet/sorghum | 90 | 69 (±11) | + | + |
| BG8 | Rice/sorghum | 95 | 37 (±11) | + | + |
| BG9 | Maize/millet | 93 | 62 (±15) | n.a. | + |
| BG9 | Maize/sorghum | 97 | 85 (±7) | n.a. | + |
| BG9 | Millet/sorghum | 93 | 69 (±11) | n.a. | + |
| BG10 | Maize/sorghum | 91.5 | 85 (±7) | n.a. | + |
| BG11 | Maize/millet | 91 | 62 (±15) | + | + |
| BG12 | Rice/sorghum | 95 | 37 (±11) | + | n.a. |
| BG13 | Common bean/pigeon pea | 91 | 68 (±9) | n.a. | n.a. |
| BG14 | *Medicago*/soybean | 90 | 45 (±11) | n.a. | + |
| BG15 | Pigeon pea/soybean | 93 | 70 (±13) | n.a. | n.a. |
| BG16 | Pigeon pea/soybean | 90 | 70 (±13) | n.a. | n.a. |
| BG17 | Lotus/pigeon pea | 92 | 56 (±10) | n.a. | n.a. |
| BG18 | Cassava/jatropha | 90.5 | 46 (±11) | n.a. | n.a. |
| BG19 | Apple/peach | 90 | 66 (±9) | + | + |
| BG20 | Apple/peach | 91 | 66 (±9) | n.a. | + |
| BG21 | Apple/peach | 95 | 66 (±9) | n.a. | + |
| BG22 | Apple/peach | 90 | 66 (±9) | n.a. | + |
| BG23 | Apple/peach | 95 | 66 (±9) | n.a. | + |

(n.a.) Not available.
[a]Average identity between best pairs of all CDS is given with standard error.
[b]Average identity between homologs of 20 genes is given together with standard error.

Transfers were confirmed by comparing sequence identity between elements with the average identity of gene sequences between the species. For intra-family transfers, we used the $K_s$ value of a set of 20 highly conserved genes previously used for phylogenetic studies in monocots and dicots (Zhang et al. 2012). TE sequence identity was always higher than the $K_s$ value (Table 1; Supplemental Table 2), thus meeting the high similarity criterion for HTTs. For more distant transfers (inter-order or inter-monocot/dicot), potential saturation of synonymous sites made the use of $K_s$ values inappropriate. Instead, we calculated genome-wide sequence identity between all annotated genes for each group of species involved (see Methods) and compared these values to those of the transferred LTR-RTs (Table 1; Fig. 2). Distributions of pairwise sequence identities between the closest gene homologs were always unimodal (Fig. 2). We concluded that sequence identity at the peak of the distribution should be a good indicator of overall genomic divergence. Peak values were always lower than 90% (our threshold for the detection of LTR-RT HTTs).

Finally, we checked phylogenetic incongruence for the 32 HTTs. Sequences homologous to horizontally transferred elements were identified in the NCBI nucleotide databases, excluding the species involved in the transfer, and a tree built including these sequences. Fifteen trees showed incongruence (Supplemental Fig. 1). For the remaining 17, BLAST searches showed no homologous sequences from taxonomic groups more closely related than the two species involved in the transfer. The presence in the plant genomes of the elements detected in silico was tested by PCR amplification and sequencing for 22 HTTs. Transfers were confirmed in all cases (Supplemental Fig. 2). Overall, we identified one HTT between a monocot and a dicot species (between palm tree and grapevine, labeled BC), eight between distinct orders within either the dicots or monocots (BO), and 23 between

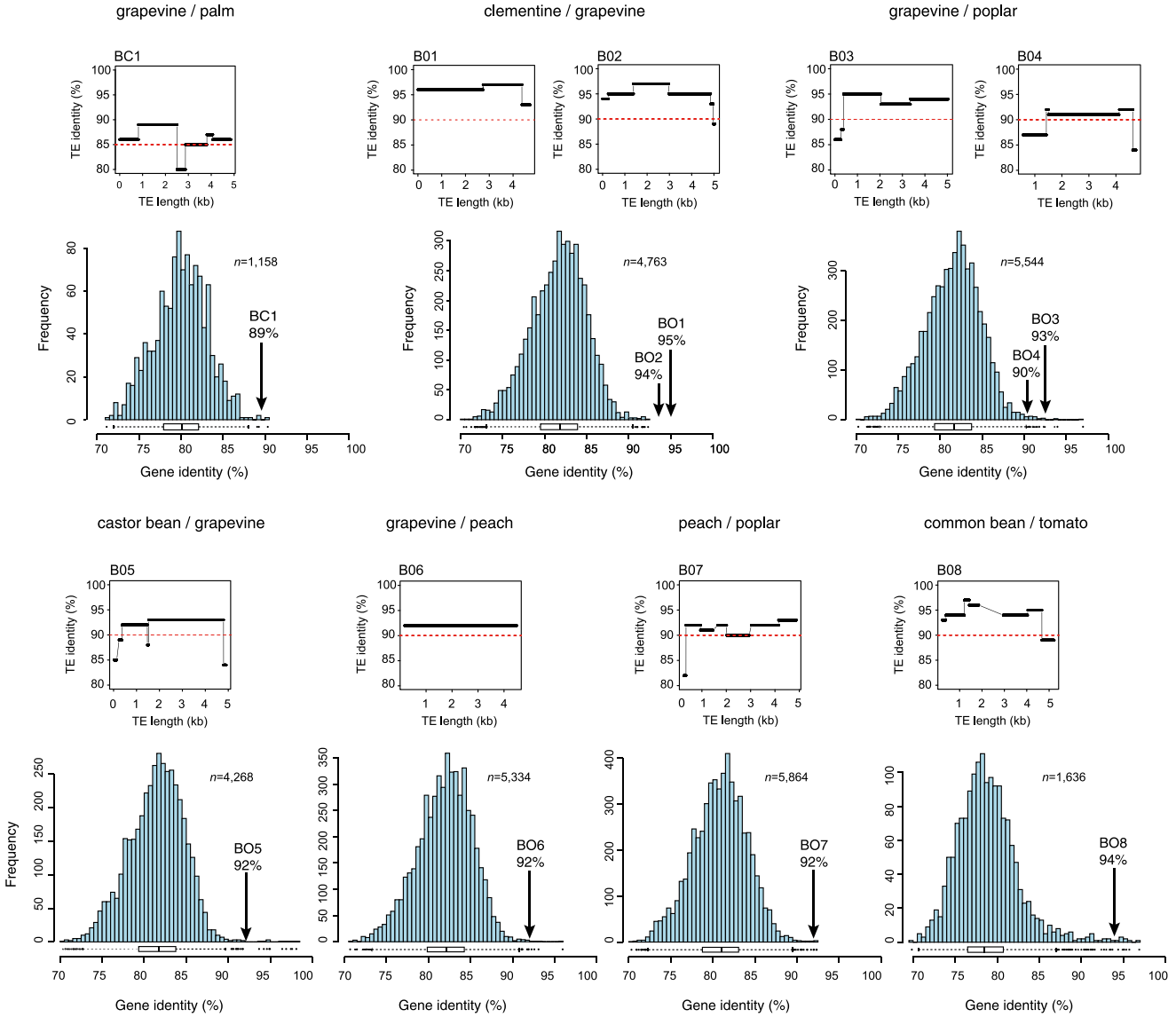



**Figure 2.** Comparison between the sequence identity of LTR-RTs and the genomic distance between the species involved in BC and BO transfers. In each panel, the *top* graph represents the sequence identity along the complete length of the LTR-RTs involved in the transfer in both species as indicated, with the red line representing the detection threshold (85% and 90% identity for BC and BO HTTs, respectively). The histogram (in blue) represents the distribution of pairwise gene identity based on CDS comparisons (see Methods). Numbers of CDS pairs analyzed are as indicated (*n*). Arrows correspond to average sequence identity between the transferred LTR-RTs.

genera of the same family (in both monocots and dicots, BG) (see Fig. 1).

Our approach only allowed detection of HTTs for LTR-RTs. We focused our study on this particular type of TEs because of their prevalence in plant genomes and because they are the most easily characterized and precisely annotated. However, we could not exclude that HTTs may concern other TEs such as transposons. However, the identification of such events at the whole genome scale would require initial genome-wide pairwise comparisons to identify highly identical sequences and characterization of the horizontally transferred DNA identified. For our 40 genomes, this would necessitate 780 comparisons (40 × 39/2), which we estimated would require nearly 2 yr of computation time on a 400-core cluster. We nevertheless performed comparisons a posteriori among species for which we detected HTTs of LTR-RTs (see Methods; Supplemental Table 3). For the most distant pairwise comparisons (the inter-monocot/dicot and the eight inter-order transfers), searches yielded the LTR-RT sequences identified using the initial detection procedure, together with ribosomal and mitochondrial sequences, confirming the first results. The BG transfers involve species from the same family, which are more closely related than those involved in the BC and BO transfers, and full genome comparisons yielded too many hits to be analyzed in extenso. We nevertheless completed such analysis for the BG8 transfer that involves rice and millet. BG8 was chosen because previous studies had identified the transfer of a class II element between these two species (Diao et al. 2006). The whole-genome comparison yielded similar results as the BO and BC transfers, and

we could in addition detect the sequence of the previously identified transposon among the many other highly conserved sequences. Moreover, we identified a second transposon that may have been horizontally transferred between the two species (Supplemental Table 3).

The transpositional activity of horizontally transferred LTR-RTs was surveyed by determining their copy numbers in their host genome and tentatively dating their insertion by comparing their LTR sequences (Fig. 3; Supplemental Table 6). Thirty families were repeated with 1–400 complete copies, the remaining two were single copy in both species. All 32 families harbor the LTR-RT gag-pol domains, suggesting that they may be functional (Supplemental Fig. 3). Furthermore, sequence divergence between the two LTRs of each copy was always lower than the average sequence divergence between the elements from the two species (Fig. 3), which strongly suggests that horizontally transferred LTR-RTs have remained transpositionally active, with two cases showing a significant increase in copy number (BG10 and BG11) (Fig. 3).

Based on our analysis of 40 plant species, we estimated the total number of LTR-RT horizontal transfers that may have occurred among dicots and monocots within the last 3 million years. The species belong to 36 genera, 18 families, 14 orders, and two classes (monocots and eudicots). Our study therefore consisted of 776 pairwise comparisons: 256 monocot/dicot, 467 inter-order, 12 inter-family, and 41 intra-family (the four intra-genus comparisons in *Arabidopsis*, *Cucumis*, *Citrus*, and *Solanum* were not taken into account). To test whether this data set was representative, we estimated the total number of these four types of taxonomic comparisons by analyzing 1000 random draws of 36 genera among the 13,551 in monocots and dicots (see Methods; Fig. 4). Only the

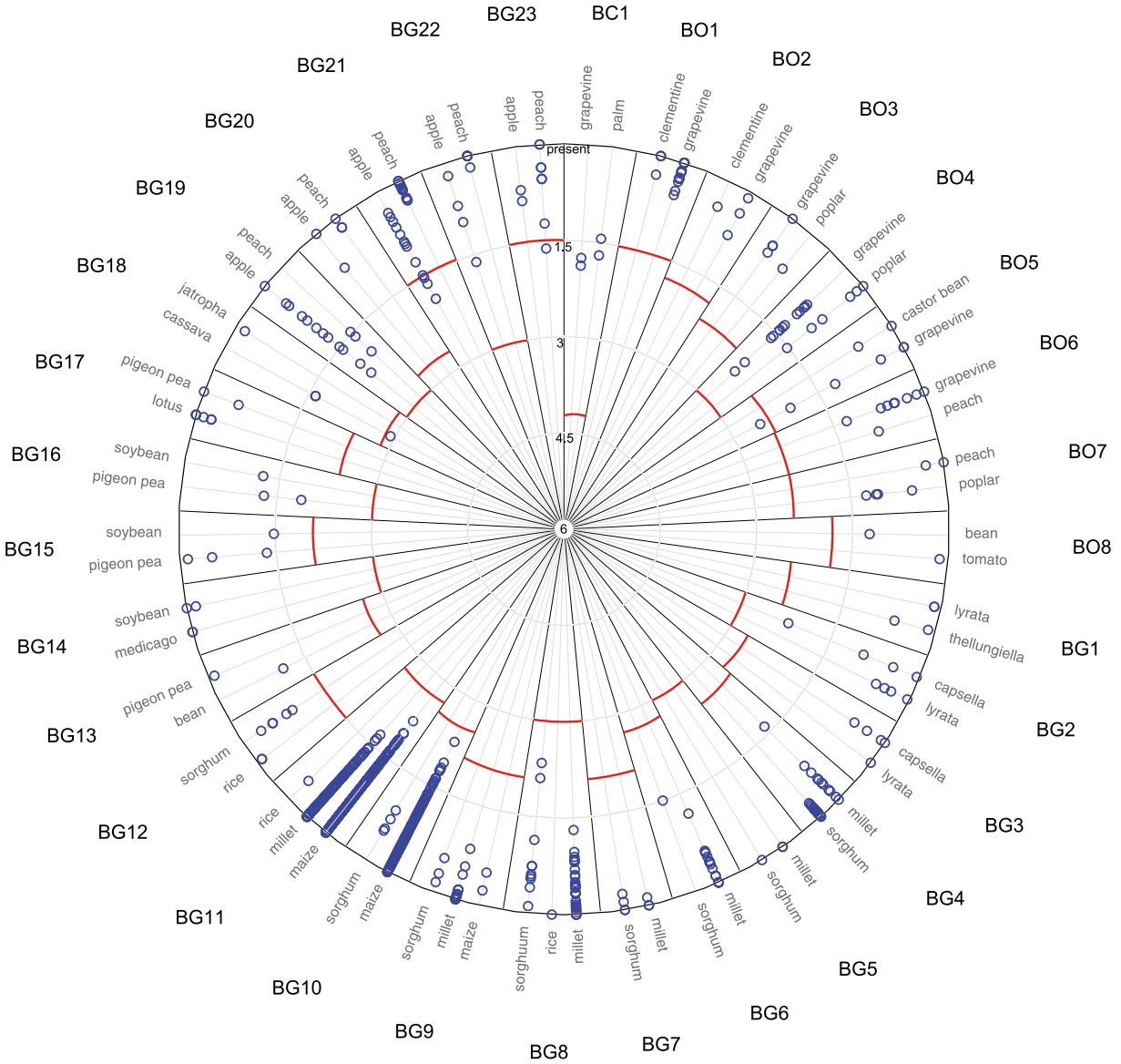



**Figure 3.** Transpositional activity of horizontally transferred LTR-RTs. Concentric circles represent the time scale for insertion dates: from 6 My (*center*) to present (*outer* circle). For each HTT, the red line illustrates the estimated date of the transfer (based on percent identity between the LTR-RTs involved in the transfer). For each species, the insertion date of each element (illustrated by the percent identity between both its LTR sequences) is represented as blue circles. Graphs are plotted with R (library plotrix) and edited with Illustrator.

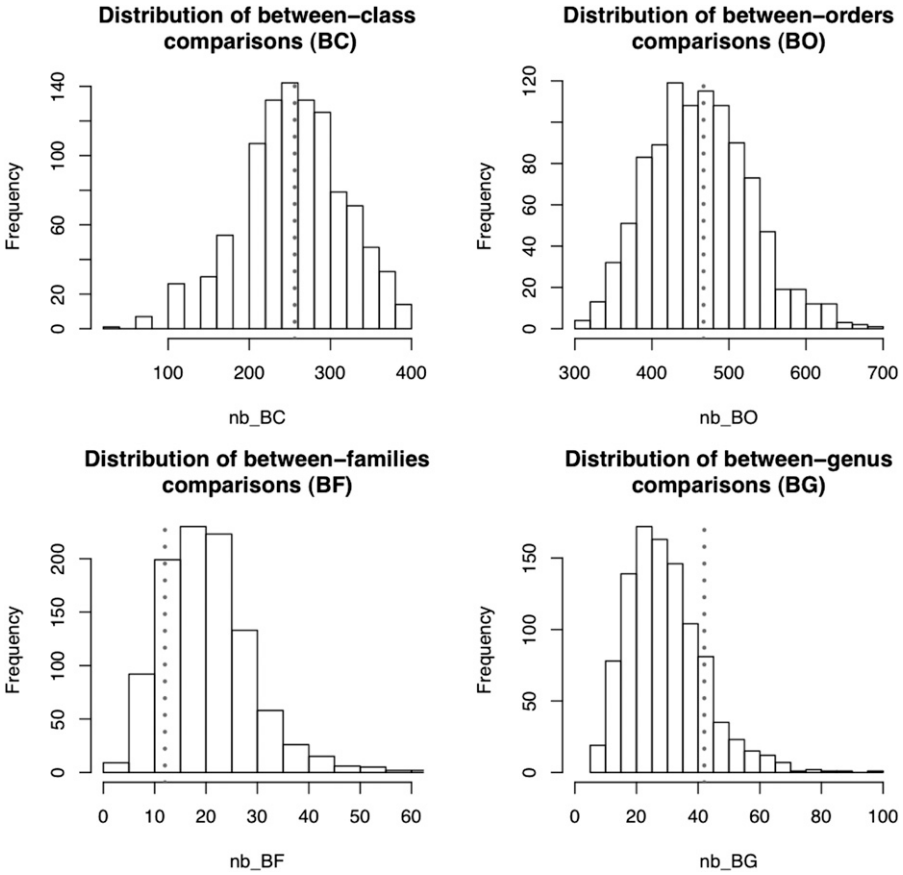



**Figure 4.** Distributions of the four types of comparisons based on the simulation of 1000 random draws of 36 genera. The dotted lines represent the number of each of the comparison types in our sample of 36 genera from which sequenced genomes were analyzed.

between-genera comparisons were overrepresented, and we corrected the estimation accordingly. We thus estimate that more than 2 million HTTs may have occurred in monocots and dicots within the last 3 million years (see Methods for a detailed description of the computation).

## Discussion

Our discovery leads to a reconsideration of LTR-RT biology: these TEs have long been considered as genomic components inherited exclusively vertically. They are often compared to retroviruses, which are presumed to have derived from LTR-RTs and acquired an envelope gene, allowing horizontal transmission within or between populations through contamination (and to a lesser extent vertically through sexual reproduction of their host) (Eickbush and Malik 2002). The occurrence of millions of HTTs in flowering plants in a recent past suggests that LTR-RTs may also have a strong propensity to be transmitted between distinct species. The mechanisms of these HTTs are not yet fully understood. We found several cases of multiple HTTs between the same species (e.g., five between sorghum and millet and five between apple and peach) (Table 1; Fig. 1). This suggests either that this mechanism should enable the transfer of several TE families at once, or that these multiple transfers result from prolonged sympatric distributions. The proximity of species, as in parasitism, has been proposed to favor HTTs in both plants and animals (Mower et al. 2004; Gilbert et al. 2010). However, none of the species we analyzed are parasitic, and we can exclude this as the only cause of HTTs in plants. Other studies suggest that some pathogens may act as horizontal transfer vectors (Sun et al. 2013). We show that HTTs are more frequent in closely related taxonomic groups (one inter-monocot/dicot transfer versus 29 intra-family transfers). If the latter hypothesis is true, HTTs may be favored by pathogens with narrower infectious spectra. In any case, because they are widespread in plants, there is a need to understand the vectors of these transfers and the mechanisms involved.

The estimation of the number of HTTs that may have occurred in all monocots and dicots must be taken with caution because it assumes that all plant lineages may be equally subjected to HTT, which cannot be tested without any prior knowledge on the putative mechanisms of the transfers (see above). In addition, one could anticipate that the geographical distributions of monocot and dicot lineages over the last 3 million years, and especially during the last glaciation/deglaciation periods, should have had an impact on HTTs. However, these distributions are not known, and consequently, the sympatric relationships between the species that belong to the 13,551 monocots and dicots genera cannot be established. Nevertheless, our estimation exceeds by several orders of magnitude the number of HTTs documented until now in both plants and animals (Wallau et al. 2012).

The last 20 yr of genomic studies in plants have demonstrated the impact of TEs on the structure, evolution, and function of eukaryotic genomes (Feschotte 2008; Rebollo et al. 2012). The question of their survival and evolutionary success has often been raised, as their putative mutagenic and therefore deleterious nature should theoretically result in elimination. Recent studies have shown that TEs are in fact efficiently silenced by several epigenetic pathways and subsequently quickly eliminated through deletion from their host genomes, providing an explanation for their limited negative biological impact, but certainly not for their long-term survival in most lineages. Our results provide a possible answer to this paradox because we show that transferred elements remain transpositionally active in both species. We propose that HTTs provide an escape route from silencing and elimination and are thus essential for their survival in plants.

## Methods

### Identification and characterization of HTTs

Genome sequences from 40 angiosperm species were downloaded mainly from Phytozome v9.0 (http://www.phytozome.net/) (the complete list with all sources is given in Supplemental Table 1). Full-size LTR-RT elements were identified in these genomes (by in silico analysis using *LTRharvest* prediction software, http://www.zbh.uni-hamburg.de/?id=206) (Ellinghaus et al. 2008). Default parameters

were used except for the following: -xdrop 37 -motif tgca -motifmis 1 -minlenltr 100 -maxlenltr 3000 -mintsd 2. A total of approximately 300,000 copies were obtained and merged into one multi-FASTA database. To detect HTT candidates, we used a clustering strategy that we described previously for LTR-RT family classification (El Baidouri and Panaud 2013). This method, based on an all-against-all comparison of LTR-RT sequences, was used to retrieve elements sharing high sequence identity (>90% within monocots or dicots and >85% between monocots and dicots because of the greater divergence time between these classes) between different species. As a first step, a nucleotide BLAST (all against all) of these elements was performed using the following parameters: -r 2 (reward for nucleotide match), -e 1e-20 (*E*-value), -F F (Filter = false), and -m 8 (for tabular output). The second step consisted of clustering sequences (based on the BLAST results) using SiLiX software (http://lbbe.univ-lyon1.fr/SiLiX) (Miele et al. 2001) in order to define highly similar LTR-RT copies. About 124,000 distinct clusters were obtained. The vast majority of these clusters contain several copies of the same LTR-RT family from a single species. This is what is expected in the case of a vertical TE transmission. However, 32 clusters contain LTR-RT elements that share high sequence similarity and belong to at least two different species. All potential candidates were validated by checking that the LTR-RT sequences were located on large contigs and not on isolated, short sequences in genome assemblies, and that high sequence identity was limited to the elements themselves—to eliminate possible contamination or annotation errors. These elements are our HTT candidates. Sequence identity between the elements involved in HTTs was computed after alignment using the SeaView software (Galtier et al. 1996). Phylogenies of the elements were built using the maximum likelihood method.

### Estimation of genomic distances

Gene identities were computed following two methods. (1) For BG HTTs, a set of 20 genes (Zhang et al. 2012) was used to determine an average $K_s$ value for each species pair involved in the transfer. $K_a/K_s$ were computed using $K_a/K_s$ calculator software (Table 1; Supplemental Table 2; Zhang et al. 2006). (2) For the BO and BC HTTs between more distantly related species, in which $K_s$ values are subject to caution due to potential saturation of sites, full genome comparisons of the complete gene sets were performed using a BLASTN homology search procedure (Fig. 2). For each pair of species involved in HTT, a nucleotide BLAST was performed (BLAST 2.2.26+) using all coding DNA sequence (CDS) of one species as a query against all CDS of the other species, with the default parameters. Multi-FASTA files of CDS primary transcripts were retrieved from Phytozome v9.0 (http://www.phytozome.net). The best hit for each BLAST was selected and a filter was applied on the percentage of query coverage (>60%). The distribution of the sequence identity of best hits was plotted and genomic distance considered to correspond to the percentage identity at the mode of the distribution.

### Phylogenetic incongruences between horizontally transferred LTR-RTs and species trees

For each transferred element, the reverse transcriptase domain was used as a query for a homology search against the NCBI nr nucleotide database using the BLASTN algorithm and excluding the species involved in the transfer. All sequences, including queries, were aligned using Muscle, and a phylogenetic tree was built using a maximum likelihood method and 100 bootstrap replicates. Alignments and phylogenetic analyses were performed using SeaView software on a LINUX platform.

### Pairwise full genome comparisons

Validation of selected comparisons was carried out a posteriori by whole-genome comparisons. The genomic sequences of the species implicated in inter-order and inter-class HTTs were split into small fragments of 1 kbp using splitter software from the EMBOSS package (http://emboss.bioinformatics.nl/cgi-bin/emboss/help/splitter). A BLAST2seq search was performed using the sequence fragments of the two species implicated in an HTT event as query and subject. All genomic regions that produced significant BLAST hits (sequence identity >90% and HSP length >200 bp) were retained for further analysis. Nucleotide BLASTN and protein BLASTX searches for highly similar regions were performed against the NCBI nr databases (http://www.ncbi.nlm.nih.gov), and hits are reported in Supplemental Table 3.

### Wet-lab validation of HTTs

For 22 HTTs, pairs of PCR primers were defined that should amplify the LTR-RT sequences in both species (Supplemental Tables 4 and 5). PCR products were purified according to the manufacturer's instructions (MP Biomedicals) and sequenced on an Applied Biosystems (Life Technologies) 3130×l sequencer. These sequences were aligned with the original genomic sequences using ClustalW in SeaView (Supplemental Fig. 2; Galtier et al. 1996).

### Functional annotation of LTR-RT families involved in HTTs

For each transfer, one element per species was analyzed for the presence of functional domains. Functional domains were defined using the CDD tool for conserved domain annotation at NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

### Estimation of the number of HTTs among monocots and dicots

The complete list of the 13,551 genera of monocots and dicots was built using the information available from *The Plant List* (http://www.theplantlist.org) (The Angiosperm Phylogeny Group 2009). A complete taxonomic description of these 13,551 genera is available upon request. To establish whether our sample of 40 species is representative of the diversity of these taxa of flowering plants, we carried out simulations of the representativity of randomly drawn sets of species to correct for any bias. We distinguish four types of comparison: inter-monocot-dicot comparison, named BC (between class); between-order (within class, BO); between-family (within order, BF); and between-genera (within family, BG). Our sample of 40 species necessitated $(40 \times 39)/2 = 780$ pairwise comparisons, from which we ignored four that correspond to intra-genus comparisons (melon/cucumber; *Arabidopsis thaliana*/*Arabidopsis lyrata*; clementine/orange; and tomato/potato). The remaining 776 comparisons could be classified into 256 BC, 467 BO, 12 BF, and 41 BG comparisons. We randomly drew 1000 samples of 36 genera from the complete list of the 13,551 genera. For each draw, we computed the number of BC, BO, BF, and BG comparisons and plotted their distributions for the 1000 samples (Fig. 4). Values for our sample were compared with this distribution. These values fall within the mode of the distribution for BC and BO. We therefore consider that our sample is not biased for these two types of comparisons, and a direct extrapolation of the total number of HTTs among monocots and dicots can be made (see below). In the case of BG, our sample is clearly biased, because it contains more species of the same family than if randomly drawn, certainly because the first plant genome projects concerned cereal crops (that belong to the *Poaceae* family). We therefore had to correct our estimation accordingly: The peak of

the distribution corresponds to 25 BG comparisons. Our sample contains 41 BG comparisons. We therefore corrected the total number of BG HTTs by multiplying it by a factor of 25/41. The 13,551 genera of monocots and dicots would necessitate 13,551 × 13,550/2 = 91,808,025 comparisons. These can be declined into: 31,288,248 BC; 54,582,625 BO; 2,475,511 BF; and 3,461,641 BG comparisons. We found 1 BC, 8 BO, and 23 BG HTTs. Our estimation for the total number of HTTs among monocots and dicots is 1 × (31,288,248/256) + 8 × (54,582,625/467) + 23 × (25/41) × (3,461,641/41) = 2,241,337 HTTs.

## Acknowledgments

This work was supported by the University of Perpignan, CNRS, and IRD; a PhD fellowship from the CNRS and Region Languedoc-Rousillon to M.E.B.; an Agence Nationale de la Recherche grant (AZORIZ) to N.P.; and a Labex grant from the Agropolis Fondation (RETROCROP 1202-041) to M.M. and O.P. We thank N. Billotte, A. D'Hont, E. Guiderdoni, C. Roux, P. This, and J. Wood for providing DNA material; and M.A. Grandbastien and T. Wicker for their useful comments on the manuscript.

## References

The Angiosperm Phylogeny Group. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot J Linn Soc* **161:** 105–121.

Bennetzen JL. 2000. Transposable element contributions to plant gene and genome evolution. *Plant Mol Biol* **42:** 251–269.

Diao X, Freeling M, Lisch D. 2006. Horizontal transfer of a plant transposon. *PLoS Biol* **4:** e5.

Eickbush TH, Malik HS. 2002. Origin and evolution of retrotransposons. In *Mobile DNA II* (ed. Craig NL, et al.), pp. 1111–1144. ASM Press, Washington DC.

El Baidouri M, Panaud O. 2013. Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. *Genome Biol Evol* **5:** 954–965.

Ellinghaus D, Kurtz S, Willhoeft U. 2008. *LTRharvest*, an efficient and flexible software for *de novo* detection of LTR retrotransposons. *BMC Bioinformatics* **9:** 18.

Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* **9:** 397–405.

Finnegan DJ. 1985. Transposable elements in eukaryotes. *Int Rev Cytol* **93:** 281–326.

Galtier N, Gouy M, Gautier C. 1996. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci* **12:** 543–548.

Gilbert C, Schaack S, Pace JK, Brindley PJ, Feschotte C. 2010. A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Nature* **464:** 1347–1350.

Kobayashi S, Goto-Yamamoto N, Hirochika H. 2004. Retrotransposon-induced mutations in grape skin color. *Science* **304:** 982.

Kuraku S, Qiu H, Meyer A. 2012. Horizontal transfers of Tc1 elements between teleost fishes and their vertebrate parasites, lampreys. *Genome Biol Evol* **4:** 929–936.

Ma J, Bennetzen JL. 2004. Rapid recent growth and divergence of rice nuclear genomes. *Proc Natl Acad Sci* **101:** 12404–12410.

Miele V, Penel S, Duret L. 2001. Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC Bioinformatics* **12:** 116.

Mower JP, Stefanović S, Young GJ, Palmer JD. 2004. Plant genetics: gene transfer from parasitic to host plants. *Nature* **432:** 165–166.

Piegu B, Guyot R, Picault N, Roulin A, Sanyal A, Kim H, Collura K, Brar DS, Jackson S, Wing RA, et al. 2006. Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res* **16:** 1262–1269.

Prak ET, Kazazian HH Jr. 2000. Mobile elements and the human genome. *Nat Rev Genet* **1:** 134–144.

Rebollo R, Romanish MT, Mager DL. 2012. Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu Rev Genet* **46:** 21–42.

Rigal M, Mathieu O. 2011. A "mille-feuille" of silencing: epigenetic control of transposable elements. *Biochim Biophys Acta* **1809:** 452–458.

Rocha EPC. 2013. With a little help from prokaryotes. *Science* **8:** 1154–1155.

Roulin A, Piegu B, Fortune PM, Sabot F, D'Hont A, Manicacci D, Panaud O. 2009. Whole genome surveys of rice, maize and sorghum reveal multiple horizontal transfers of the LTR-retrotransposon *Route66* in *Poaceae*. *BMC Evol Biol* **9:** 58.

Schaack S, Gilbert C, Feschotte C. 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol Evol* **25:** 537–546.

Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8:** 272–285.

Sun BF, Xiao JH, He S, Liu L, Murphy RW, Huang DW. 2013. Multiple interkingdom horizontal gene transfers in *Pyrenophora* and closely related species and their contributions to phytopathogenic lifestyles. *PLoS ONE* **8:** e60029.

Vitte C, Panaud O. 2005. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenet Genome Res* **110:** 91–107.

Wallau GL, Ortiz MF, Loreto EL. 2012. Horizontal transposon transfer in eukarya: detection, bias, and perspectives. *Genome Biol Evol* **4:** 689–699.

Walsh AM, Kortschak RD, Gardner MG, Bertozzi T, Adelson DL. 2013. Widespread horizontal transfer of retrotransposons. *Proc Natl Acad Sci* **110:** 1012–1016.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. 2007. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* **8:** 973–982.

Zhang Z, Li J, Zhao XQ, Wang J, Wong GKS, Yu J. 2006. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics* **4:** 259–263.

Zhang N, Zeng L, Shan H, Ma H. 2012. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytol* **195:** 923–937.

*Received August 1, 2013; accepted in revised form January 24, 2014.*

L'approche bio-informatique a permis de rechercher des rétrotransposons à LTR hautement conservés parmi les 40 génomes végétaux séquencés. Nous avons identifié 26 génomes présentant au moins un cas de transfert horizontal d'un rétrotransposon à LTR. Ces transferts peuvent concerner des espèces très éloignées telles que le palmier et la vigne, la tomate et le haricot, le peuplier et la pêche. Au total, 32 cas de transferts horizontaux ont été identifiés, ce qui pourraient se traduire par plus de 2 millions de cas parmi les 13 551 monocotylédones et dicotylédones. De plus, nous avons montré que ces éléments restent fonctionnels après leur transfert, entraînant dans certaines espèces, comme le maïs, une amplification génomique significative.

Ces résultats suggèrent que les plantes peuvent fréquemment échanger du matériel génétique par les transferts horizontaux et que ce mécanisme pourrait être important dans l'impact des éléments transposables sur l'évolution des génomes.

Les mécanismes de ces transferts ne sont pas encore bien connus, mais les hypothèses actuelles privilégient, les relations sympatriques, le parasitisme, les pathogènes. Les transferts horizontaux pourraient fournir aux éléments transposables une voie leur permettant d'échapper au silencing et à leur élimination dans les génomes et seraient essentiels à leur survie dans les génomes.

## 4. Activité de recherche en parallèle

J'ai co-encadré Christophe Calvayrac qui a réalisé sa thèse à l'Université de Perpignan Via Domitia. Il a effectué celle-ci en même temps que son emploi d'enseignant dans un BTS (ce qui explique la durée de sa thèse, 2006 à 2011). Lors de ce co-encadrement avec Jean-François Cooper (PR, Université de Perpignan Via Domitia), j'ai principalement supervisé les expériences de biologie moléculaire. Son travail s'est principalement concentré sur la dégradation biologique de la sulcotrione dans un sol agricole : recherche d'une éventuelle biodégradation accélérée et caractérisation de souches bactériennes potentiellement dégradantes (Calvayrac et al., 2012).

La présence de résidus de composés xénobiotiques dans les principaux compartiments de l'environnement est avérée depuis plusieurs décennies. Les diverses activités anthropiques (développement urbain, pollutions industrielles, agriculture, épandage des boues de stations d'épuration, ...) exercent des pressions environnementales variables, que les institutions nationales et/ou supranationales s'efforcent d'identifier et de limiter (plan national Ecophyto 2018, législation Européenne REACH,...).

L'agriculture intensive conventionnelle participe de manière notable à la dégradation de la qualité des eaux de surface et souterraines, du sol et de l'atmosphère, consécutivement à l'utilisation d'intrants agricoles tels que les engrais, le lisier et les produits phytopharmaceutiques.

Les herbicides de la famille chimique des tricétones, mis sur le marché dans les années 90, ont été homologués dans certains itinéraires techniques agricoles, notamment en maïsiculture, « bénéficiant » du retrait de substances actives comme l'atrazine. Ces molécules de nouvelle génération (mésotrione et sulcotrione), employées à doses inférieures à celles de l'atrazine (150 - 450 g.Ha -1 vs 1200 g.Ha -1 ) ont été qualifiées de composés « présentant des risques minimes pour l'environnement » (Mitchell et al., 2001).

Le sol de Perpignan, présentant une réactivité biologique notable et ayant reçu antérieurement des traitements répétés par la sulcotrione lors des travaux de H. Chaabane, a constitué un outil précieux pour essayer de provoquer puis de caractériser une éventuelle biodégradation accélérée (BDA) en conditions contrôlées, consécutivement à une possible adaptation de la microflore tellurique par acquisition de gènes codant les enzymes cataboliques responsables de la dégradation.

Par ailleurs, nous avons souhaité utiliser la réactivité avérée de ce sol de Perpignan afin d'en isoler des microorganismes dégradants spécifiques, capables d'utiliser la sulcotrione comme seule source de carbone et /ou d'énergie. La culture de ces microorganismes au laboratoire, et le souhait de les caractériser de la manière la plus formelle possible, nous ont amené à proposer une stratégie expérimentale comprenant plusieurs étapes successives :

- sélection de milieux adaptés à la culture, au laboratoire, de souches telluriques, généralement réputées comme difficilement cultivables sur milieux synthétiques ;
- sélection des souches d'intérêt ;
- évaluation de leur potentiel biodégradant ;
- caractérisations phylogénétiques et fonctionnelles des isolats retenus.

L'ensemble de ces données a pour objectif la perspective d'une valorisation biotechnologique des isolats présentant le plus fort potentiel de dégradation. En outre, ce travail visait à appréhender certains mécanismes évolutifs conduisant à l'adaptation des microorganismes en réponse aux changements de leur environnement.

Research Article

Received: 30 March 2011 Revised: 24 May 2011 Accepted: 24 June 2011 Published online in Wiley Online Library:

(wileyonlinelibrary.com) DOI 10.1002/ps.2263

# Isolation and characterisation of a bacterial strain degrading the herbicide sulcotrione from an agricultural soil

Christophe Calvayrac,[a,b] Fabrice Martin-Laurent,[c] Alexia Faveaux,[a] Nathalie Picault,[b] Olivier Panaud,[b] Camille-Michel Coste,[a] Hanene Chaabane[d] and Jean-François Cooper[a]*

## Abstract

**BACKGROUND: The dissipation kinetics of the herbicide sulcotrione sprayed 4 times on a French soil was studied using a laboratory microcosm approach. An advanced cultivation-based method was then used to isolate the bacteria responsible for biotransformation of sulcotrione. Chromatographic methods were employed as complementary tools to define its metabolic pathway.**

**RESULTS: Soil microflora was able quickly to biotransform the herbicide ($DT_{50}$ ≈ 8 days). 2-Chloro-4-mesylbenzoic acid, one of its main metabolites, was clearly detected. However, no accelerated biodegradation process was observed. Eight pure sulcotrione-resistant strains were isolated, but only one (1OP) was capable of degrading this herbicide with a relatively high efficiency and to use it as a sole source of carbon and energy. In parallel, another sulcotrione-resistant strain (1TRANS) was shown to be incapable of degrading the herbicide. Amplified ribosomal restriction analysis (ARDRA) and repetitive extragenic palendromic PCR genomic (REP-PCR) fingerprinting of strains 1OP and 1TRANS gave indistinguishable profiles.**

**CONCLUSION: Sequencing and aligning analysis of 16S rDNA genes of each pure strain revealed identical sequences and a close phylogenetic relationship (99% sequence identity) to *Pseudomonas putida*. Such physiological and genetic properties of 1OP to metabolise sulcotrione were probably governed by mobile genetic elements in the genome of the bacteria.**


**Keywords:** sulcotrione; biodegradation; soil; *Pseudomonas putida*; metabolites

## 1 INTRODUCTION

Sulcotrione [2-(2-chloro-4-methylsulfonylbenzoyl)-1,3-cyclohexanedione] is a triketone herbicide applied at 300–450 g $ha^{-1}$ on corn crop to control the development of a wide range of annual and perennial broadleaf weeds.

A large proportion of this herbicide reaches the soil, where both abiotic (i.e. sorption, chemical degradation) and biotic (biodegradation) processes control the fate and the activity of this herbicide. 2-Chloro-4-mesylbenzoic acid (CMBA) and 1,3-cyclohexanedione (CHD) have been described as biodegradation and hydrolysis transformation products.[1–8] A derivative of phenylheptanoic acid was also described under special conditions.[6,7,9] Hitherto, no report has shown the occurrence of accelerated biodegradation of sulcotrione as a result of repeated applications. Soil microorganisms able to transform sulcotrione are yet to be described, in contrast to mesotrione, another herbicide belonging to the triketone family.[10–14] In this context, the aim of the present work was to evaluate the response of an agricultural soil regularly treated with sulcotrione, and possibly adapted to its biodegradation, in order to isolate novel and efficient sulcotrione degraders. Sulcotrione transformation was monitored in the soil using chromatographic methods. Sulcotrione-degrading bacterial strain was isolated by conducting an adaptation culture to the sulcotrione. The isolate was characterised by ARDRA, REP-PCR and by cloning/sequencing of the partial 16S rDNA sequence. Its capacity to degrade aerobically sulcotrione was estimated.

* *Correspondence to: Jean-François Cooper, Laboratoire de Chimie des Biomolécules et de l'Environnement (LCBE, EA 4215), Université de Perpignan Via Domitia (UPVD), 52 avenue Paul Alduy, 66860 Perpignan, France. E-mail: cooper@univ-perp.fr*

a *Laboratoire de Chimie des Biomolécules et de l'Environnement, Université de Perpignan Via Domitia, Perpignan, France*

b *Laboratoire Génome et Développement des Plantes, Université de Perpignan Via Domitia, Perpignan, France*

c *INRA, Université de Bourgogne, UMR1229 Microbiologie du Sol et de l'Environnement, Dijon, France*

d *Unité de Recherche Protection Intégrée des Cultures, Institut National Agronomique de Tunisie, Tunis Mahrajène, Tunisia*

## 2 MATERIALS AND METHODS

### 2.1 Experimental design

#### 2.1.1 Experimental field site

The experimental field site was set up at the University of Perpignan, France (42° 40′ 55″ N, 2° 53′ 49″ E). The field was divided into two plots each of 184 $m^2$, separated by a buffer strip. One plot (4S) was treated with 300 g $ha^{-1}$ of sulcotrione (Mikado®) 4 times at 1 month intervals. After this, plot 4S was neither pesticide treated nor cultivated for 24 months. The second plot (C), used as control, was never treated with sulcotrione. According to FAO classification, the soil is a clayey sandy loam soil. The soil physicochemical characteristics were: clay 13.9%, silt 60.5%, sand 25.6%, soil moisture 25%, organic carbon 0.9%, organic nitrogen 0.98 g $kg^{-1}$, C/N ratio 9.64, cation exchange capacity (CEC) 15.5 meq 100 $g^{-1}$, $Ca^{2+}$/CEC 214% and pH in water 8.1. Soil samples were collected from the surface layer (0–7 cm) of the experimental field. Composite samples were sieved to 2 mm and stored in the dark at 4 °C until use.

#### 2.1.2 Sulcotrione biodegradation

The ability of the microflora of the soil samples collected on the experimental field (4S and C) to degrade sulcotrione was tested under laboratory conditions. Each microcosm, prepared with 10 g of moist soil placed in petri dishes (5.5 cm ID), were treated with 550 µL of sterile aqueous sulcotrione solution at 22 mg $L^{-1}$ (i.e. 1.2 µg of sulcotrione per g of soil) and homogenised. Treatments were applied 4 times at days 0, 30, 60 and 90 ($T1_{D0}$, $T2_{D30}$, $T3_{D60}$ and $T4_{D90}$), necessitating the preparation of 40 microcosms for each soil type. All analyses were made in duplicate. In addition, autoclaved controls were set up to estimate the occurrence of abiotic biodegradation (*St*4S and *St*C). All samples were kept in the dark to avoid photodegradation of the sulcotrione. Throughout the incubation period, soil moisture content was kept at 25% (moist weight), making it comparable with field moisture. After each treatment, each entire soil aliquot (10 g) was regularly sampled (i.e. after 0, 2, 7, 15 and 30 days) to monitor sulcotrione dissipation.

### 2.2 Chemical analysis

#### 2.2.1 Reagents

A standard of sulcotrione (98.8% purity) – a weak acidic herbicide, pKa value 2.87, MW 328.77 g $mol^{-1}$ and water solubility 165 mg $L^{-1}$ at 25 °C – was purchased from Sigma-Aldrich, France. Standards of 1,3-cyclohexanedione (CHD) (97.0% purity) and 2-chloro-4-mesylbenzoic acid (CMBA) (95.0% purity) were purchased from Fluka and Acros Organics respectively.

Methanol and acetonitrile (HPLC quality) were purchased from Carlo Erba, dichloromethane for pesticide analysis from Riedel-deHaën GmbH, trifluoroacetic acid (TFA) (99.0% purity) from Aldrich and hydrochloric acid (38.0%) from Prolabo, and water was Milli-Q quality. (Trimethylsilyl) diazomethane 2.0 M in diethyl ether was supplied by Sigma-Aldrich. Supelclean™ ENVI-Chrom P SPE cartridges were supplied by Supelco, France.

#### 2.2.2 Analytical procedure

##### 2.2.2.1 Degradation of sulcotrione in soil samples

- *Extraction step.* Soil samples (10 g) were extracted twice with 60 and 30 mL of an acetonitrile/0.1 M HCl (90/10; v/v) mixture over 30 and 15 min periods, and then filtered on a Whatman filter GF/A. The organic filtrate was evaporated at 30 °C, and the acidic aqueous solution was then extracted twice with 5.0 mL of dichloromethane. The organic phase was evaporated to dryness, and the extract was solubilised in 0.10 M HCl. It was then purified on an EnviChrom P cartridge, eluted with methanol (5 mL) and concentrated to 1 mL. The extract was finally analysed by HPLC. One blank (sulcotrione-free soil) was systematically included and analysed under the same conditions.
- *Chromatographic analysis.* Soil extracts were analysed using Shimadzu HPLC apparatus equipped with a Supelco ODS Hypersil $C_{18}$ column (5 µm, 250 mm × 4.6) and an SPD-10 A*vp* UV/Vis detector set at 265 nm. The mobile phase consisted of a mixture of water acidified by 0.3% trifluoroacetic acid (AW) and acetonitrile (ACN), delivered at a flow rate of 1 mL $min^{-1}$ and using a gradient system [70/30 (AW/ACN) at $t = 0$ min to 50/50 (AW/ACN) at 8 min, then maintained for 10 min].
- *Analytical performance.* This analytical method was an adaptation of the procedure previously published by Chaabane *et al.*[9] The quantification limit was estimated to be 0.05 µg $g^{-1}$ for sulcotrione herbicide, with a mean recovery rate of 87 ± 10%.

##### 2.2.2.2 Degradation of sulcotrione by the bacterial culture

- *Extraction step.* Aliquots (1 mL) of the bacterial culture (selected among sulcotrione-resistant ones) were centrifuged (15 min, 1100 × *g*) at room temperature. The supernatant was recovered, acidified with 0.3 mL of 0.16 M HCl and brought to 2.5 mL with distilled water. Then, the acidic aqueous solution was extracted twice with 2.5 mL of dichloromethane. The organic phase was evaporated to dryness and solubilised in 1 mL of methanol before analysis. No adsorption of sulcotrione was observed on the centrifuge tubes.
- *High-performance liquid chromatography (HPLC/UV) of sulcotrione and its main metabolites.* Culture medium extracts were analysed using Jasco HPLC apparatus equipped with a Supelco ODS Hypersil $C_{18}$ column (5 µm, 250 mm × 4.6) and a UV/visible detector Jasco 875-UV set at 265 nm. The mobile phase consisted of a mixture of water acidified by 0.3% trifluoroacetic acid (AW) and acetonitrile (ACN), delivered at a flow rate of 1 mL $min^{-1}$. To improve the separation of the active ingredient and its metabolites during degradation kinetics, a gradient system [70/30 (AW/ACN) from 0 to 4 min, evolving to 40/60 (AW/ACN) at 7 min, then maintained for 15 min] was used. Detection of sulcotrione and its main metabolites (1,3-cyclohexanedione and 2-chloro-4-mesylbenzoic acid) was done at 265 nm. Retention times were estimated as follows: sulcotrione 11.1 min, CMBA 5.1 min and CHD 4.5 min.
- *High-performance liquid chromatography/mass spectrometry (HPLC/ESI/MS) analysis of metabolite 1,3-cyclohexanedione (CHD).* LC/MS analysis was carried out on a LCQ-Fleet (Thermo Fisher Scientific) equipped with an electrospray ionisation (ESI) source and a 3D ion-trap analyser. LC analysis was achieved using an Accela Thermo Fisher chromatograph equipped with an Accela 600 pump, an Accela autosampler and a diode array detector (Accela PDA detector). Full-scan spectra from *m/z* 60 to 300 in both positive (source voltage 4 kV) and negative (source voltage 3.5 kV) ion modes were recorded. Separation was carried out at room temperature using a Phenomenex Gemini $C_6$-phenyl column (150 × 3.0 mm; 5 µm). The mobile phase consisted of a mixture of 0.1% formic acid water (A-W) and 0.1% formic acid acetonitrile (A-ACN) delivered at a flow rate of 0.4 mL $min^{-1}$, with a gradient system: 90/10 (A-W/A-ACN) from 0 to 2 min, evolving to 10/90 (A-W/A-ACN) at 20 min. Under

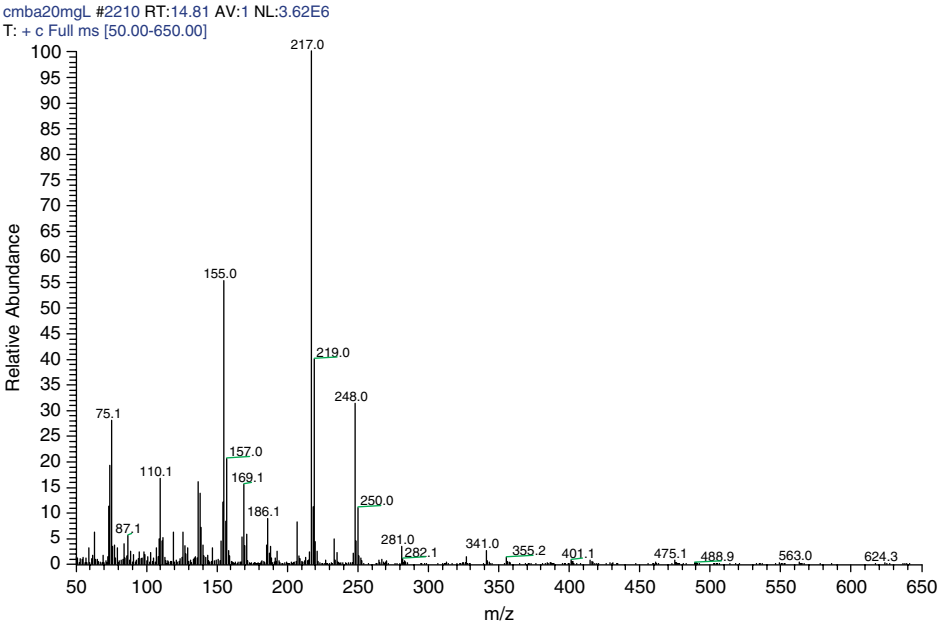



**Figure 1.** Electronic impact/mass spectrometry (EI/MS) spectrum of methylated CMBA analytical standard (20 mg $L^{-1}$).

these conditions, the retention time of 1,3-CHD was 4.51 min. The quantification limit was estimated in positive mode to be 0.5 mg $L^{-1}$.

- *Gas chromatography/mass spectrometry (GC/MS) analysis of metabolite 2-chloro-4-mesylbenzoic acid (CMBA) after derivatisation.* In the presence of methanol, trimethylsilyldiazomethane (TMS) gives methyl esters in good yields at room temperature with various carboxylic acids including aromatic, heteroaromatic, alicyclic and aliphatic ones. The esterification proceeds instantaneously and quantitatively when an excess of TMS is used, and the reaction can be easily monitored by the disappearance of the yellow colour.[15] TMS (25 µL) prepared in diethyl ether solution was added to 0.5 mL of culture medium extract mixed with 0.5 mL of methanol. After a 2 h reaction time, the extract was analysed by GC/MS. GC/MS analysis was carried out on a Focus GC connected to a mass spectrometric detector DSQII (Thermo Fisher Scientific) equipped with a G-SQC GC column, 15 m × 0.25 mm × 0.25 µm (5% phenyl methylpolysiloxane), initial oven temperature 80 °C for 1 min, raised at 10 °C $min^{-1}$ to 260 °C for 15 min; carrier gas helium 6.0; electron impact mode 70 eV; full scan mode; split injection mode; transfer line temperature 300 °C. Under such conditions, the retention time of methylated CMBA was 14.81 min and characteristic *m/z* ions were 155, 217 and 248 (Fig. 1). A calibration curve was established on *m/z* 217 for the concentration range 0.4–20 mg $L^{-1}$ ($y = 364x - 62$, $r^2 = 0.996$).
- *Analytical performance.* Culture medium samples were spiked with known amounts of analytical standards of sulcotrione, CHD and CMBA (2–30 mg $L^{-1}$). Calculations were made by HPLC/UV and/or GC/MS, after analysis in duplicate, using calibration curves. The mean recoveries were estimated as 75 ± 6%, 55 ± 8% and 75 ± 10% for sulcotrione, CHD and CMBA respectively. The limits of quantification, defined as the sample concentration required to give a signal-to-noise ratio of 5 : 1, were estimated to be 0.5–1.0 mg $L^{-1}$.

### 2.3 Isolation of the sulcotrione degrader

#### 2.3.1 *Culture media and cell preparation*

A culture medium (NS), where the nutrient source was the control soil C, was prepared and used to promote the growth of bacteria previously described as hardly cultivable on conventional culture media.[16] Briefly, soil was sieved through a 2 mm mesh and dried at 65 °C for 3 days. The culture medium was prepared by adding 40 g of soil and 2 g of agar in a final volume of 100 mL of deionised water. After sterilisation, the medium was supplemented with sterilised solutions of sulcotrione (30 mg $L^{-1}$) and of cycloheximide (50 mg $L^{-1}$) and poured into petri dishes to solidify.

A sterile mineral salt medium (MS), modified from Rousseaux *et al.*,[17] was used for both the adaptation culture and the biodegradation assays of sulcotrione-degrading strains. This medium contained per litre: 1.6 g of $K_2HPO_4$, 0.4 g of $KH_2PO_4$, 0.2 g of $MgSO_4.7H_2O$, 0.1 g of NaCl, 0.002 g of $CaCl_2$, 3.5 g of $(NH_4)_2SO_4$, 30 mg of sulcotrione, 1 mL of salt stock solution, 1 mL of vitamin stock solution and 1 mL of $FeSO_4.6H_2O$ stock solution (5 g $L^{-1}$). The salt stock solution contained 2 g of boric acid, 1.8 g of $MnSO_4.H_2O$, 0.2 g of $ZnSO_4$, 0.1 g of $CuSO_4$ and 0.25 g of $Na_2MoO_4$ (per litre). The vitamin stock solution contained 100 mg $L^{-1}$ of thiamine-HCl and 40 mg $L^{-1}$ of biotin. After sterilisation, 0.2 µm filter sterilised sulcotrione, $FeSO_4.6H_2O$, and vitamin stock solutions were added to the medium. The pH was adjusted to 7.7 with aqueous 0.1 M NaOH solution. A solid sulcotrione-MS medium obtained by adding 15 g $L^{-1}$ of agar (Biokar Diagnostics, France) was used to test the capacity of the strains to grow on such a solid medium.

Triptic soy broth medium (TS) obtained by adding 15 g $L^{-1}$ of agar (Biokar Diagnostics, France) supplemented with sulcotrione (30 mg $L^{-1}$) was used to purify bacterial strains by successive stricking. A sulcotrione-(TS) medium (BioMérieux, France) was also used to improve microbial biomass culture during batch cultures for degradation assays. Microbial biomass was harvested by centrifugation (15 min, 1600 × *g* at 4 °C), and the pellet was washed twice with Knapp buffer (1 g of $KH_2PO_4$; 1 g of $K_2HPO_4$; 4 mg of $FeCl_3$; 40 mg of $MgSO_4.7H_2$, per litre of deionised water, pH

6.6) as described earlier[18] before being resuspended in sulcotrione-MS medium to obtain an absorbance of ~0.3 at 600 nm.

#### 2.3.2 *Adaptation cultures to the sulcotrione*

Soil 4S (10 g) was suspended in 90 mL of 0.9% NaCl and placed on an orbital shaker at 200 rpm for 30 min. Duplicates were carried out. Soil suspension aliquots (500 µL) were plated on NS-sulcotrione or NS medium. Plates were incubated in the dark at 28 °C for 7 days. Growing colonies identified as sulcotrione-resistant strains and/or possible degraders were selected, grown in 100 mL of Triptic soy broth (TS) medium supplemented with sulcotrione for 24 h at 30 °C under agitation. They were then plated on TS agar with sulcotrione.

For each isolated bacterial strain, a selective adaptation step was then carried out by performing five successive cultures ($A_1$ to $A_5$) in 100 mL of sulcotrione-MS medium at 28 °C under agitation at 150 rpm for 20 days. Subsequent studies were performed using $A_5$ adaptation solutions.

#### 2.3.3 *Estimation of the sulcotrione-degrading capability of the bacterial isolates*

The $A_5$ bacterial culture (1 mL) was grown in TS supplemented with sulcotrione. After 15 h of culture, bacterial cells were harvested in the late exponential growth phase by centrifugation (15 min, 1100 × *g*, 4 °C). The bacterial pellets were washed twice with Knapp buffer and was resuspended in sulcotrione-MS medium to obtain an absorbance of ~0.3 at 600 nm. They were then incubated at 28 °C in the dark on an orbital shaker (200 rpm). An aliquot (1 mL) of each bacterial suspension was then regularly sampled (0, 6, 15 and 27 days) and analysed to establish the kinetics of degradation of sulcotrione and the appearance of its main metabolites. These degradation tests were realised in triplicate. One control without bacterial inoculums was also included.

#### 2.3.4 *ARDRA, REP-PCR and partial 16S rRNA cloning and sequencing*

The extraction of bacterial genomic DNA was done following the protocol previously described by Cheneby *et al.*[19] and purified using the GENECLEAN® Turbo kit according to the manufacturer's instructions (MP Biomedicals, Europe). DNA quality was checked by 1% (w/v) agarose gel in 1× TBE buffer and quantified by spectrophotometry (NanoDrop® 2000; Thermo Fisher Scientific).

Partial 16S rRNA sequences were amplified by PCR in a 50 µL volume reaction containing 1 µM of the universal primers 27f (5′-AGA GTT TGA TCM TGG CTC AG-3′) and 1492r (5′-TAC GGH TAC CTT GTT ACG ACT T-3′), 200 µM of each dNTP (Promega, USA), 2.5 U of Go*Taq* DNA polymerase (Promega, USA) and 25 ng of DNA.[20] PCR reactions were carried out in a PTC-200 gradient cycler (MJ Research) under the following conditions: 5 min at 94 °C, 35 cycles of 1 min at 94 °C, 1 min at 55 °C and 2 min at 72 °C, plus an additional 15 min cycle at 72 °C. PCR products were then separated by electrophoresis on a 1% agarose gel, purified using the GENECLEAN® Turbo kit and quantified by spectrophotometry.

ARDRA profiles were obtained by digesting 16S rRNA PCR fragments (7 µL) with the restriction endonucleases *Alu*I and *Hae*III (New England Biolabs®) overnight at 37 °C. The obtained restriction fragments for each isolate were separated on a 2% (v/w) Metaphor agarose gel in 1× TBE buffer (80 V for 3.5 h).

Repetitive extragenic palendromic PCR (REP-PCR) was carried out with REP1R-I and REP2-1 primers, targeting the conserved stem of each REP-like sequence in the genome.[21] DNA amplification was carried out in a 25 µL mixture using 50 ng of extracted DNA as template in 5 µL of 5× Giescher buffer [83 mM of $(NH_4)_2SO_4$; 335 mM of Tris HCl, pH 8.8; 150 mM of $\beta$-mercaptoethanol; 33.5 µM EDTA, pH 8.8], 6.5 mM of $MgCl_2$, 0.16 mg mL$^{-1}$ of bovine serum albumin (BSA) (New England Biolabs®), 1.25 mM of each dNTP, 0.1 µL mL$^{-1}$ of dimethyl sulfoxide, 2 µM of each of the two primers (REP1R-I and REP2-1) and 1 U of Go*Taq* DNA polymerase. Amplifications were carried out in a PTC-200 gradient cycler (MJ Research): 1 cycle at 95 °C for 6 min, 30 cycles at 94 °C for 1 min, 40 °C for 1 min and 65 °C for 4 min, plus an extension cycle at 65 °C for 16 min.[17] Amplified DNA fragments were separated by electrophoresis on 1% agarose gel. PCR reactions were performed independently in duplicate to ensure the consistency of the REP profiles obtained.

16S rRNA PCR products were cloned into the pGEM-T Easy vector according to the provider's recommendation (Promega, USA). Recombinant plasmids of interest were then purified by using Wizard® Plus Minipreps DNA Purification Systems kit (Promega). Both DNA strands of the insert were sequenced with an Applied Biosystems Hitachi 3130*xl* genetic analyser by using SP6 and T7 primers. Sequences of recombinant clones from the degrading and the non-degrading strains were treated by SeqManNGen® software (DNASTAR, USA) to provide consensus sequences, and were then compared with known sequences found in the GenBank database using Blast.[22] Sequences were deposited in GenBank under the accession numbers JF303891 (1TRANS strain) and JF303892 (1OP strain). Multiple alignments were realised using the ClustalX programme.[23] The phylogenetic tree was elaborated using the NJ plot programme.[24]

## 3 RESULTS AND DISCUSSION

### 3.1 Sulcotrione degradation kinetics in Perpignan soils

The microflora of the Perpignan soil was previously shown to transform sulcotrione to CHD and CMBA by cleaving the bond between 1,3-cyclohexanedione (CHD) and the benzoic ring.[3] Here, the transformation of sulcotrione only in 4S and C soils and not in the two sterilised soils (*St*C and *St*4S) is reported, further confirming the predominance of the biotic pathway for sulcotrione dissipation in the soil of Perpignan. Similar dissipation kinetics was observed in C and 4S soils, which differed in herbicide treatment, the former never being treated with it and the latter having been treated 4 times at the agronomic dose 24 months ago. Indeed, after 30 days of incubation (treatment 1), almost 90% of the applied sulcotrione was dissipated, with half-lives averaging 8 days, without significant differences between C and 4S plots. For both soils a no-lag-phase period was observed, which suggests that a cometabolic process is responsible for sulcotrione dissipation. By fitting sulcotrione dissipation kinetics with $R^2$ values close to 0.9, it was possible to observe similar half-life values ($DT_{50} \approx 8$ days) and dissipation rates ($k \approx 0.08$ day$^{-1}$) in C and 4S soils (Table 1). Compared with the literature, sulcotrione dissipation recorded in the soil of Perpignan is rather rapid when considering $DT_{50}$ values.[3,4,6–8,25] In addition, as plots C and 4S showed a similar capacity to dissipate sulcotrione, besides different herbicide treatments, the possible accelerated biodegradation (ABD) of sulcotrione seems unlikely. To explore this possibility further, C and 4S soils were submitted to three additional sulcotrione treatments under laboratory conditions ($T2_{D30}$, $T3_{D60}$ and $T4_{D90}$). Similar dissipation kinetics of sulcotrione and kinetic parameters ($k$ and $DT_{50}$) were observed in these two plots which differed in scenario of exposure to sulcotrione (Table 1). In addition, no correlation could be found between the rate of degradation ($k$) and the number of sulcotrione treatments. These results further

**Table 1.** Degradation coefficients ($k$) and half-live values ($DT_{50}$) for all soils

| Plot | Treatment | $k$ (days$^{-1}$) | $DT_{50}$ (days) | $R^2$ |
|---|---|---|---|---|
| C plot | Treatment 1 ($T1_{D0}$) | 0.083 | 8 | 0.894 |
| | Treatment 2 ($T2_{D30}$) | 0.060 | 11.5 | 0.955 |
| | Treatment 3 ($T3_{D60}$) | 0.100 | 7 | 0.989 |
| | Treatment 4 ($T4_{D90}$) | 0.100 | 7 | 0.962 |
| | Mean value | $0.085 \pm 0.020$ | $8 \pm 2.5$ | – |
| 4S plot | Treatment 1 ($T1_{D0}$) | 0.088 | 8 | 0.872 |
| | Treatment 2 ($T2_{D30}$) | 0.093 | 7 | 0.936 |
| | Treatment 3 ($T3_{D60}$) | 0.068 | 10 | 0.955 |
| | Treatment 4 ($T4_{D90}$) | 0.089 | 8 | 0.828 |
| | Mean value | $0.084 \pm 0.012$ | $8 \pm 2.0$ | – |
| *St*C plot | Treatment 1 ($T1_{D0}$) | 0.0004 | – | – |
| *St*4S plot | Treatment 1 ($T1_{D0}$) | 0.0004 | – | – |

confirm that sulcotrione degradation most likely results from a cometabolic process, and that accelerated biodegradation did not occur in the soil of Perpignan. This feature was also observed for several other herbicides belonging to *s*-triazine or substituted urea families, for which degradation for a long period resulted from cometabolic activity of the soil microflora prior to the adaptation of the soil microflora to metabolic degradation relying on specific catabolic pathways. The switch between cometabolic and metabolic degradation of these two classes of herbicides occurred after almost 30 years of repeated exposure to these compounds of the agricultural fields. In order to gain a better insight into the processes involved in the dissipation of sulcotrione from the soil of Perpignan, an adaptation culture already shown to be successful in isolating different pesticide-degrading strains, including atrazine, isoproturon, carbofuran and mesotrione, was established.[10,11,26–29]

### 3.2 Isolation of sulcotrione-degrading bacterial cultures

The first step of the isolation process relied on the use of a soil agar medium NS supplemented with sulcotrione, adapted from Ferreira *et al.*,[16] and which was shown to promote the growth of pesticide-degrading bacterial strains known to be difficult to cultivate on conventional media, like most of the soil microbes. Starting from soil suspensions plated on NS-sulcotrione medium, numerous bacterial colonies showing different morphotypes grew. Eight colonies of the dominant morphotype, forming little white colonies, showing viscid consistency without mycelial growth, were further purified by successive plating. Except for their opaque or translucent appearance, these isolates showed comparable morphology, forming white colonies with a diameter of ~2.5 mm and the following characteristics: smooth, circular form, convex elevation. Each of the eight colonies was then cultivated repeatedly on MS-sulcotrione ($A_1 \rightarrow A_5$), where they were able to grow. Nevertheless, as agar can also represent a source of carbon, the aptitude of these bacterial isolates to degrade sulcotrione was checked in MS-sulcotrione liquid medium.

### 3.3 Estimation of the sulcotrione-degrading capacity of bacterial isolates

To get enough microbial biomass to set up a batch experiment to follow sulcotrione transformation, each of the eight cultures was grown for 15 h in TS-herbicide medium. They were then diluted to

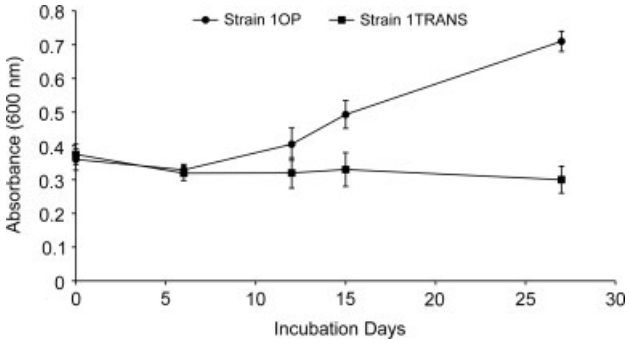



**Figure 2.** Growth measurement of strain 1OP versus 1TRANS on mineral salt medium (MS) supplemented with sulcotrione (100 µM).

0.3 OD at 600 nm in MS herbicide-supplemented medium. Dissipation of herbicide and concomitant accumulation of CMBA and CHD in the culture medium were monitored by high-performance liquid chromatography. Among the eight isolates tested here, only 1OP showed an increased OD from 0.3 to ~0.7 after 30 days of incubation (Fig. 2), reflecting a bacterial growth using sulcotrione as carbon source. 1TRANS, one of the seven non-growing strains, was chosen as a non-degrading control. No significant growth was measured either for 1OP or for 1TRANS in MS sulcotrione-free medium under the same incubation conditions. It could be concluded that no interfering carbon sources are contained in MS medium, and that sulcotrione most likely constitutes the sole carbon and energy source for the growth of 1OP strain. HPLC/UV monitoring clearly showed the dissipation of sulcotrione in 1OP culture and not in the two controls – 1TRANS, the non-degrading isolate, and the uninoculated MS-sulcotrione medium (Fig. 3). Sulcotrione dissipation was clearly effective after 6 days lag phase, which was in accordance with the lag phase observed for the biomass development (Fig. 2). It could be hypothesised that this lag phase is the time necessary to induce the expression of the genes and the synthesis of the catabolic enzymes responsible for sulcotrione transformation. From the seventh day of incubation, sulcotrione dissipation kinetics could be reasonably described by a first-order kinetics ($C = 121\,e^{-0.081t}$, $R^2 = 0.993$, $n = 3$). $DT_{50}$ could be estimated to be ~15 days after the beginning of the incubation.

During the time course of sulcotrione transformation, the appearance of 2-chloro-4-mesylbenzoic acid (CMBA), known as the main metabolite of sulcotrione herbicide (Fig. 3), could be observed. It has to be noted that this metabolite was observed in neither of the two controls – 1TRANS, the non-degrading isolate, and the uninoculated MS-sulcotrione medium. CMBA was detected from the sixth day of incubation, which is in accordance with the lag phase observed for the initiation of sulcotrione disappearance. After 1 month of incubation, the concentration of CMBA reached $12 \pm 1$ mg L$^{-1}$ ($48 \pm 3$ µM) in the culture medium. This result was confirmed by GC/MS analysis after derivatisation of CMBA to give the corresponding methyl ester. In addition, molar recovery, taking into account sulcotrione and CMBA, was seen to decrease slightly between day 15 ($65 \pm 7$µmol L$^{-1}$ + $25 \pm 2$ µmol L$^{-1}$ respectively) and day 27 ($17 \pm 2$ µmol L$^{-1}$ + $48 \pm 3$ µmol L$^{-1}$ respectively), suggesting a possible degradation of benzoic acid derivative CMBA by 1OP. Finally, based on the results presented here, it could be concluded that 1OP strain was capable of using sulcotrione as a carbon source for its growth by degrading sulcotrione to CMBA.

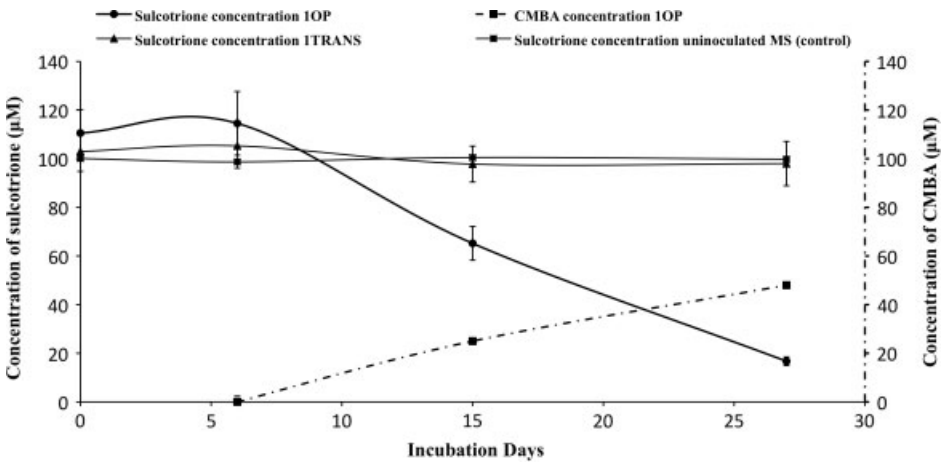



**Figure 3.** Dissipation of sulcotrione and accumulation of CMBA by strain 1OP versus control and strain 1TRANS in liquid mineral salt medium MS supplemented with sulcotrione (100 µM). Errors bars indicate standard deviation.

1,3-Cyclohexanedione (CHD), the other degradation product of sulcotrione, previously identified at very low concentrations in the Perpignan soil,[3] was not detected ($<0.2$ mg $L^{-1}$) in the culture media of strain 1OP either by HPLC/UV or by HPLC/MS/MS. This discrepancy could be attributed to chemical transformation by hydrolysis of CHD to 5-oxohexanoic acid at pH $>7$, or to an oxidation mechanism catalysed by enzymatic cleavage of C–C bonds in the $\beta$-diketone before entering the tricarboxylic acid cycle (TCA), as previously reported.[30–32] It is noteworthy that, for mesotrione, another herbicide belonging to the triketone family but with a nitro radical, a minor metabolic route catalysed by a *Bacillus* sp. with a comparable mechanism has been reported.[11,12]

### 3.4 Characterisation of sulcotrione-degrading (1OP) and non-degrading (1TRANS) strains by ARDRA, REP-PCR and 16S rDNA sequencing

The amplified ribosomal rDNA restriction analysis technique (ARDRA) has been reported to be a reliable and valuable tool for phylogenetic and taxonomic studies of large sets of cultured or uncultured microorganisms from different habitats.[33–36] In this study, 1OP and 1TRANS isolate strains were characterised by ARDRA with *Alu*I and *Hae*III restriction endonucleases. 1OP and 1TRANS bacterial strains showed similar ARDRA profiles made up of four major bands of approximately 550, 380, 210 and 70 bp length. In order to obtain more taxonomical information than the genus level offered by ARDRA fingerprinting, 1OP and 1TRANS isolates were analysed by repetitive extragenic palindromic PCR (REP-PCR), which provides a rapid identification of bacterial isolates and is useful for the analysis of prokaryotic genomes.[18,21] Again, 1OP and 1TRANS bacterial strains showed similar REP-PCR profiles made up of five major bands of approximately 550, 400, 270, 150 and 100 bp length (Fig. 4). This observation suggests that 1OP and 1TRANS belong to the same bacterial species. 16S rRNA sequencing further revealed that these two isolates showed identical 16S rRNA sequences. It is concluded that these isolates were different strains from the same species. Comparison of the 16S rRNA sequence with those found in GenBank showed highest level of similarity (99% similarity) to *Pseudomonas* sp. Furthermore, multiple alignments revealed that 1OP and 1TRANS 16S rRNA sequences were closely related to *Pseudomonas putida* strains both for 1OP and for 1TRANS (Fig. 5). Based on 16S rDNA phylogenetic data, it is suggested that the present sulcotrione-degrading isolate be named *Pseudomonas* sp. 1OP. It was classified as Bacteria, phylum *Proteobacteria*, class *Gammaproteobacteria*, order *Pseudomonadales*, family *Pseudomonaceae* and genus *Pseudomonas*, with the following characteristics: rod shaped, gram-negative, one or more flagella providing motility, positive catalase test and non-spore-forming bacteria. Several strains belonging to the *Pseudomonas* genus and able to degrade various pesticides have been described earlier, including *Pseudomonas* sp. ADP degrading atrazine,[37] *Pseudomonas cepacia* DBO1 degrading 2,4-dichlorophenoxyacetic[38] and *Pseudomonas* sp. degrading propoxur (2-isopropoxyphenyl-*N*-methylcarbamate).[39] Interestingly, the isolation of two strains identical from the taxonomical point of view but differing in their sulcotrione-degrading capacity highlights the versatility of the degrading character. It could be hypothesised that such versatility might be due to mobile genetic elements (MGEs) surrounding catabolic genes in the genome of 1OP. MGEs such as plasmids, transposons, integrons, genomic islands and bacteriophages acted as a powerful evolutionary engine and regulator on microbial populations and were collectively referred to as drivers or vectors of 'the horizontal gene pool'.[40,41] Such properties are known to be associated with various strains of genus *Pseudomonas* and its relatives, and were often specified by large ($>50$ kb) catabolic plasmids, although the catabolic genes could also be loaded on transposable elements.[42–44] Further work will aim differentially to characterise the genetic determinants of the sulcotrione-degrading ability of *Pseudomonas* sp. 1OP.

## 4 CONCLUSION

This work confirms the aptitude of the soil microflora of Perpignan to transform sulcotrione, although no enhanced biodegradation was observed after repeated exposure to this herbicide. The isolation of a sulcotrione-degrading strain, identified as *Pseudomonas putida* strain 1OP, has been reported for the first time. This bacterial strain was shown to transform sulcotrione to 2-chloro-4-mesylbenzoic acid (CMBA). Nevertheless, 1,3-cyclohexanedione (CHD), another known sulcotrione metabolite, was not detected. Based on the evolution of the molar recovery of sulcotrione and CMBA during the time course of dissipation kinetics, it has been hypothesised that *Pseudomonas putida* strain 1OP may start degrading the benzoic acid derivative (CMBA) several days after its formation in the liquid medium.

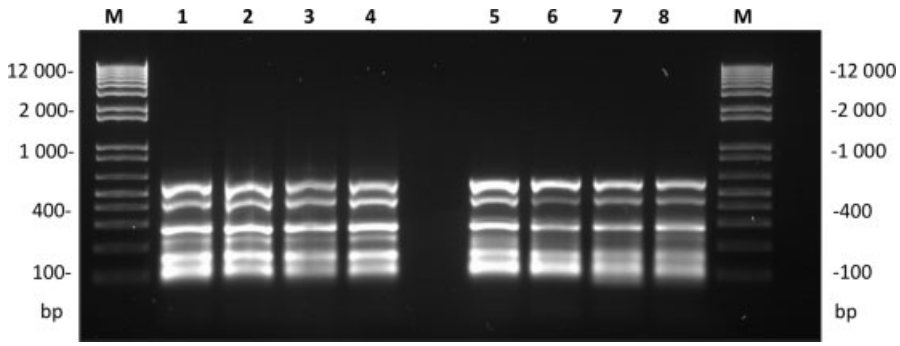



**Figure 4.** REP-PCR analysis of the isolates 1OP and 1TRANS. REP-PCR was performed in duplicate for each isolate with 1 or 2 ng $\mu L^{-1}$ of DNA. Lane M represents the molecular weight marker 1 Kb Plus DNA Ladder (sizes indicated in base pairs). Lanes 1 and 2: 1TRANS (1 ng $\mu L^{-1}$); lanes 3 and 4: 1TRANS (2 ng $\mu L^{-1}$); lanes 5 and 6: 1OP (1 ng $\mu L^{-1}$); lanes 7 and 8: 1OP (2 ng $\mu L^{-1}$).

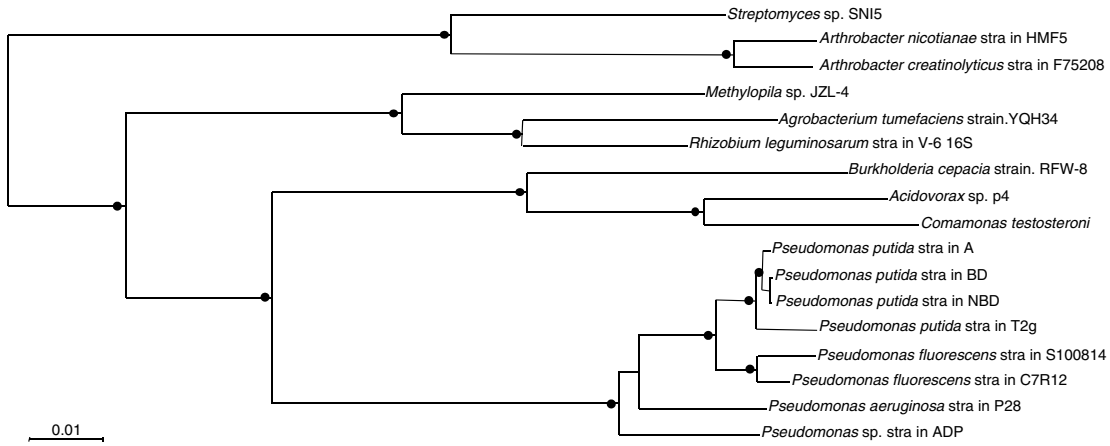



**Figure 5.** Neighbour-joining phylogenetic analysis resulting from the multiple alignment of 16S rRNA gene sequences of *Pseudomonas putida* strains 1TRANS (NBD) and 1OP (BD) with those retrieved from the GenBank database: *Arthrobacter creatinolyticus* strain F75208 (HQ908759), *Arthrobacter nicotianae* strain HMF5 (GU220490), *Streptomyces* sp. SNI5 (GU320191), *Rhizobium leguminosarum* strain V-6 (GU306144), *Agrobacterium tumefaciens* strain YQH34 (HQ143653), *Methylopila* sp. JZL-4 (FJ502233), *Burkholderia cepacia* strain RFW-8 (HQ650586), *Comamonas testosteroni* strain H18 (EU887829), *Acidovorax* sp. p4(HQ652596), *Pseudomonas putida* strain A (HQ697262), *Pseudomonas putida* strain T2g (AB539810), *Pseudomonas aeruginosa* strain P28 (HQ697285), *Pseudomonas fluorescens* strain C7R12 (AM229082), *Pseudomonas fluorescens* strain (HQ880245) and *Pseudomonas* sp. strain ADP (AM088478). Bootstrap values greater than 90% are marked as black circles. The phylogenetic distance is shown on the scale bar.

## ACKNOWLEDGEMENTS

The PhD work of Christophe Calvayrac was funded by Bayer Cropscience. The authors also acknowledge the members of laboratories of UPVD (LCBE, LGDP, IMAGE and IUT), INSA Toulouse and INRA Dijon, France.

## REFERENCES

1 Mamy L, Gabrielle B and Barriuso E, Comparative environmental impacts of glyphosate and conventional herbicides when used with glyphosate-tolerant and non-tolerant crops. *Environ Pollut* **158**:3172–3178 (2010).
2 Chaabane H, Vulliet E, Joux F, Lantoine F, Conan P, Cooper JF, *et al*, Photodegradation of sulcotrione in various aquatic environments and toxicity of its photoproducts for some marine micro-organisms. *Water Res* **41**:1781–1789 (2007).
3 Chaabane H, Vulliet E, Calvayrac C, Coste CM and Cooper JF, Behaviour of sulcotrione and mesotrione in two soils. *Pest Manag Sci* **64**:86–93 (2008).
4 Cherrier R, Perrin-Ganier C and Schiavon M, Degradation of sulcotrione in a brown soil with various organic matters. *Agronomie* **24**:29–33 (2004).
5 Maeghe L, Desmet EM and Bulcke R, Soil activity and persistence of sulcotrione and mesotrione. *Commun Agric Appl Biol Sc* **69**:41–48 (2004).
6 Rouchaud J, Thirion A, Callens D and Bulcke R, Soil dissipation of the post-emergence herbicide sulcotrione in maize crops treated with organic fertilizers. *Bull Environ Contam Toxicol* **57**:398–405 (1996).
7 Rouchaud J, Neus O, Bulcke R, Cools K and Eelen H, Sulcotrione soil metabolism in summer corn crops. *Bull Environ Contam Toxicol* **61**:669–676 (1998).
8 Rouchaud J, Neus O, Callens D and Bulcke R, Sulcotrione soil persistence and mobility in summer maize and winter wheat crops. *Weed Res* **38**:361–371 (1998).
9 Chaabane H, Cooper JF, Azouzi L and Coste CM, Influence of soil properties on the adsorption-desorption of sulcotrione and its hydrolysis metabolites on various soils. *J Agric Food Chem* **53**:4091–4095 (2005).
10 Durand S, Amato P, Delort AM, Combourieu B and Besse-Hogan P, First isolation and characterization of a bacterial strain that biotransforms the herbicide mesotrione. *Lett Appl Microbiol* **43**:222–228 (2006).
11 Durand S, Legeret B, Martin A, Sancelme M, Delort A, Besse-Hoggan P, *et al*, Biotransformation of the triketone herbicide mesotrione by a *Bacillus* strain. Metabolite profiling using liquid chromatography/electrospray ionization quadrupole time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* **20**:2603–2613 (2006).
12 Durand S, Sancelme M, Besse-Hoggan P and Combourieu B, Biodegradation pathway of mesotrione: complementarities of NMR, LC-NMR and LC-MS for qualitative and quantitative metabolic profiling. *Chemosphere* **81**:372–380 (2010).

13 Batisson I, Crouzet O, Besse-Hoggan P, Sancelme M, Mangot JF, Mallet C, *et al*, Isolation and characterization of mesotrione-degrading *Bacillus* sp. from soil. *Environ Pollut* **157**:1195–1201 (2009).

14 Crouzet O, Batisson I, Besse-Hoggan P, Bonnemoy F, Bardot C, Poly F, *et al*, Response of soil microbial communities to the herbicide mesotrione: a dose-effect microcosm approach. *Soil Biol Biochem* **42**:193–202 (2010).

15 Mashimoto N, Aoyama T and Shiori T, New methods and reagents in organic synthesis. A simple efficient preparation of methyl esters with trimethylsilyldiazomethane ($TMSCHN_2$) and its application to gas chromatographic analysis of fatty acids. *Chem Pharm Bull* **29**:1475–1478 (1981).

16 Ferreira EPDB, Dusi AN, Costa JR, Xavier GR and Rumjanek NG, Assessing insecticide and fungicide effects on the culturable soil bacterial community by analyses of variance of their DGGE fingerprinting data. *Eur J Soil Biol* **45**:466–472 (2009).

17 Rousseaux S, Hartmann A and Soulas G, Isolation and characterisation of a new Gram-negative and Gram-positive atrazine degrading bacteria from different French soils. *FEMS Microbiol Ecol* **36**:211–222 (2001).

18 Devers M, Henry S, Hartmann A and Martin-Laurent F, Horizontal gene transfer of atrazine-degrading genes (*atz*) from *Agrobacterium tumefaciens* St96-4 Padp::Tn5 to bacteria of maize-cultivated soil. *Pest Manag Sc* **61**:870–880 (2005).

19 Cheneby D, Perrez S, Devroe C, Hallet S, Couton Y, Bizouard F, *et al*, Denitrifying bacteria in bulk and maize-rhizospheric soil: diversity and $N_2O$-reducing abilities. *Can J Microbiol* **50**:469–474 (2004).

20 Gurtler V and Stanisich VA, New approaches to typing and identification of bacteria using the 16S-23S rDNA spacer region. *Microbiology* **142**:3–16 (1996).

21 Versalovic J, Koeuth TR and Lupski J, Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Res* **19**:6823–6831 (1991).

22 *GenBank*®. [Online]. Available: http://blast.ncbi.nlm.nih.gov/Blast.cgi (2011).

23 Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F and Higgins DG, The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **24**:4876–4882 (1997).

24 Perriere G and Gouy M, WWW-Query: an on-line retrieval system for biological sequence banks. *Biochimie* **78**:364–369 (1996).

25 Cherrier R, Boivin A, Perrin-Ganier C and Schiavon M, Sulcotrione versus atrazine transport and degradation in soil columns. *Pest Manag Sci* **61**:899–904 (2005).

26 Govantes F, García-Gonzalez V, Porrua O, Platero AI, Jimenez-Fernandez A and Santero E, Regulation of the atrazine-degradative genes in *Pseudomonas* sp. strain ADP. *FEMS Microbiol Lett* **310**:1–8 (2010).

27 El-Sebai T, Lagacherie B, Soulas G and Martin-Laurent F, Isolation and characterisation of an isoproturon-mineralising *Methylopila* sp. TES from French agricultural soil. *FEMS Microbiol Lett* **239**:103–110 (2004).

28 Hussain S, Devers-Lamrani M, El Azhari N and Martin-Laurent F, Isolation and characterization of an isoproturon mineralizing *Sphingomonas* sp. strain SH from a French agricultural soil. *Biodegradation* **22**:637–650 (2011).

29 Yano H, Garruto CE, Sota M, Ohtsubo Y, Nagata Y, Zylstra GJ, *et al*, Complete sequence determination combined with analysis of transposition/site-specific recombination events to explain genetic organization of IncP-7 TOL plasmid pWW53 and related mobile genetic elements. *J Mol Biol* **369**:11–26 (2007).

30 Dangel W, Tschech A and Fuchs G, Enzyme reactions involved in anaerobic cyclohexanol metabolism by a denitrifying *Pseudomonas* species. *Arch Microbiol* **62**:3245–3250 (1989).

31 Grogan G, [beta]-Diketone hydrolases. International Symposium on Biocatalysis and Biotransformations No. 5, Darmstadt, Germany. *J Mol Catalysis. B, Enzymatic* **19–20**:73–82 (2002).

32 Grogan G, Emergent mechanistic diversity of enzyme-catalysed [beta]-Diketone cleavage. *Biochem J* **388**:721–730 (2005).

33 Vaneechoutte M, Rossau R, de Vos P, Gillis M, Janssens D, Paepe N, *et al*, Rapid identification of bacteria of the Comamonadaceae with amplified ribosomal DNA-restriction analysis (ARDRA). *FEMS Microbiol Lett* **72**:227–233 (1992).

34 Vaneechoutte M, De Beenhouwer H, Claeys G, Verschraegen G, De Rouck A, Paepe N, *et al*, Identification of *Mycobacterium* species by using amplified ribosomal DNA-restriction analysis. *J Clin Microbiol* **31**:2061–2065 (1993).

35 Martinez-Murcia AJ, Acinas SG and Rodriguez-Valera F, Evaluation of prokaryotic diversity by restrictase digestion of 16S rDNA directly amplified from hypersaline environments. *FEMS Microbiol Ecol* **36**:247–256 (1995).

36 Martin-Laurent F, Philippot L, Hallet S, Chaussod R, Germon JC, Soulas G, *et al*, DNA extraction from soils: old bias for new microbial diversity analysis method. *Appl Environ Microbiol* **67**:2354–2359 (2001).

37 Mandelbaum R, Allan D and Wackett LP, Isolation and characterization of a *Pseudomonas* sp. that mineralizes the s-triazine herbicide atrazine. *Appl Environ Microbiol* **61**:1451–1457 (1995).

38 Daugherty DD and Karel SF, Degradation of 2,4-dichloro-phenoxyacetic acid by *Pseudomonas cepacia* DBO1(pRO101) in a dual-substrate chemostat. *Appl Environ Microbiol* **60**:3261–3267 (1994).

39 Kamanavalli CM and Ninnekar HZ, Biodegradation of propoxur by *Pseudomonas* species. *World J Microbiol Biotechnol* **16**:329–331 (2000).

40 Thomas CM and Nielsen KM, Mechanisms and barriers to horizontal gene transfer between bacteria. *Nat Rev Microbiol* **3**:711–721 (2005).

41 Frost LS, Leplae R, Summers AO and Toussaint A, Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* **3**:722–732 (2005).

42 Ogawa N, Chackrabarty AM and Zaborina O, Degradative plasmid, in *Plasmid Biology*, ed. by Funnell BE and Philipps GJ. ASM Press, Washington, DC, pp. 341–392 (2004).

43 Williams PA, Jones RM and Zylstra GJ, Genomics of catabolic plasmids, in *Pseudomonas. Vol. 1*, ed. by Ramos JL. Kluwer Academic/Plenum Publishers, New York, NY, pp. 165–195 (2004).

44 Yano H, Garruto CE, Sota M, Ohtsubo Y, Nagata Y, Zylstra GJ, *et al*, Complete sequence determination combined with analysis of transposition/site-specific recombination events to explain genetic organization of IncP-7 TOL plasmid pWW53 and related mobile genetic elements. *J Mol Biol* **369**:11–26 (2007).

La découverte d'un phénotype dégradant la sulcotrione a mis en lumière l'extraordinaire plasticité génétique d'une fraction de la microflore du sol génétiquement apte à dégrader certains produits phytopharmaceutiques. La caractérisation de deux souches appartenant à la même espèce, mais présentant des phénotypes différents en ce qui concerne le caractère dégradant vis-à-vis de l'herbicide, souligne le caractère versatile de ce potentiel de biodégradation. Ce résultat nous amène à proposer l'intervention d'éléments génétiques mobiles régulant ces mécanismes moléculaires adaptatifs, *via* la présence de plasmides conjugatifs et/ou d'éléments transposables (séquences d'insertion ou transposons) codant les enzymes cataboliques chez *Pseudomonas* sp. 1OP, intervention déjà décrite pour diverses souches dans le cas de l'atrazine. Cette perspective évolutive, en attente de confirmation par séquençage, va cependant dans le sens des premières données recueillies lors de notre étude fonctionnelle, effectuée sur les isolats 1OP et 1TRANS, par ciblage des gènes codant les enzymes impliquées dans le métabolisme de la voie catécholique. La mise en évidence, au niveau génomique, d'une réponse différente entre l'isolat 1OP et l'isolat 1TRANS lors du ciblage du gène codant la catéchol 2,3 dioxygénase, pourrait étayer cette hypothèse de travail. De plus, les gènes codant la voie de catabolisme catécholique méta sont présents chez *Pseudomonas* sp. et sont, dans la majorité des cas, portés par des plasmides conjugatifs de taille supérieure à 50 kb, eux-mêmes situés à l'intérieur d'éléments transposables.

# Discussion

La réactivation des ETs pendant la culture *in vitro* soulève une question essentielle : pourquoi certaines familles d'ETs redeviennent soudainement actives dans ces conditions ?

Pour l'instant nous ne disposons pas de réponse claire à cette question. Plusieurs études ont montré que dans les cultures de cals, certains ETs sont hypométhylés et donc fortement transcrits. Par exemple, l'élément *Tos17* est hypométhylé au niveau des sites CG et CHG (Liu et al., 2004). Dans ces mêmes conditions, l'élément *mPing* montre une altération de son profil de méthylation (Ngezahayo et al., 2009). Le LINE *Karma* est aussi hypométhylé dans les cals et les plantes régénérées, comparé aux plantes sauvages (Komatsu et al., 2003).

Toutes ces observations laissent supposer que pendant la culture *in vitro*, différents types de stress pourraient altérer le paysage épigénétique du génome d'une manière directe ou indirecte, au niveau global ou plus ciblé, entraînant ainsi la réactivation de quelques familles d'ETs. Il serait donc intéressant d'utiliser les nouvelles technologies, telle que le méthylome, séquençage des petits ARN dans les différents stress que nous avons étudiés, afin de faire un lien avec les mécanismes épigénétiques impliqués.

L'étude de l'activité transpositionnelle du mutant de riz à l'aide du re-séquençage nous a révélé que plusieurs familles d'ETs (appartenant à différentes classes) sont toujours actives et capables de générer de nouvelles insertions dans le génome. Cela nous questionne sur la fréquence et l'importance de telles activations *in planta* dans les populations naturelles. Le contrôle épigénétique de l'activité transpositionnelle est très efficace pour lutter contre la prolifération des ETs, car on observe rarement des événements de transposition *in planta* et sans utilisation de mutants de silencing ou l'application de stress.

Nous avons pu montrer que l'utilisation du séquençage NGS est un outil efficace pour l'analyse de l'activité transpositionnelle chez un mutant de riz. La diminution de plus en plus drastique du coût de séquençage va rendre accessible l'utilisation de cette stratégie à grande échelle. L'application du séquençage NGS pour l'analyse de l'activité transpositionnelle au niveau des populations ou pour l'étude de la réponse des plantes aux stress pourrait nous permettre d'avoir une vision assez globale de la dynamique des ETs en conditions naturelles.

Comme nous avons pu le voir, certaines familles possèdent des LTR qui répondent spécifiquement à certains types de stress. Il est assez facile d'imaginer qu'une espèce qui subit plus souvent dans son milieu naturel un type de stress environnemental va connaître la réactivation transpositionnelle des mêmes familles d'ETs au cours de son évolution.

L'étude de l'activité transpositionnelle et de sa régulation dans les conditions naturelles à l'aide des nouvelles techniques de séquençage à haut débit permettra dans un avenir proche l'évaluation globale de la dynamique transpositionnelle au niveau des populations. Ceci pourrait apporter un nouvel éclairage sur l'impact des ETs sur l'évolution des génomes.

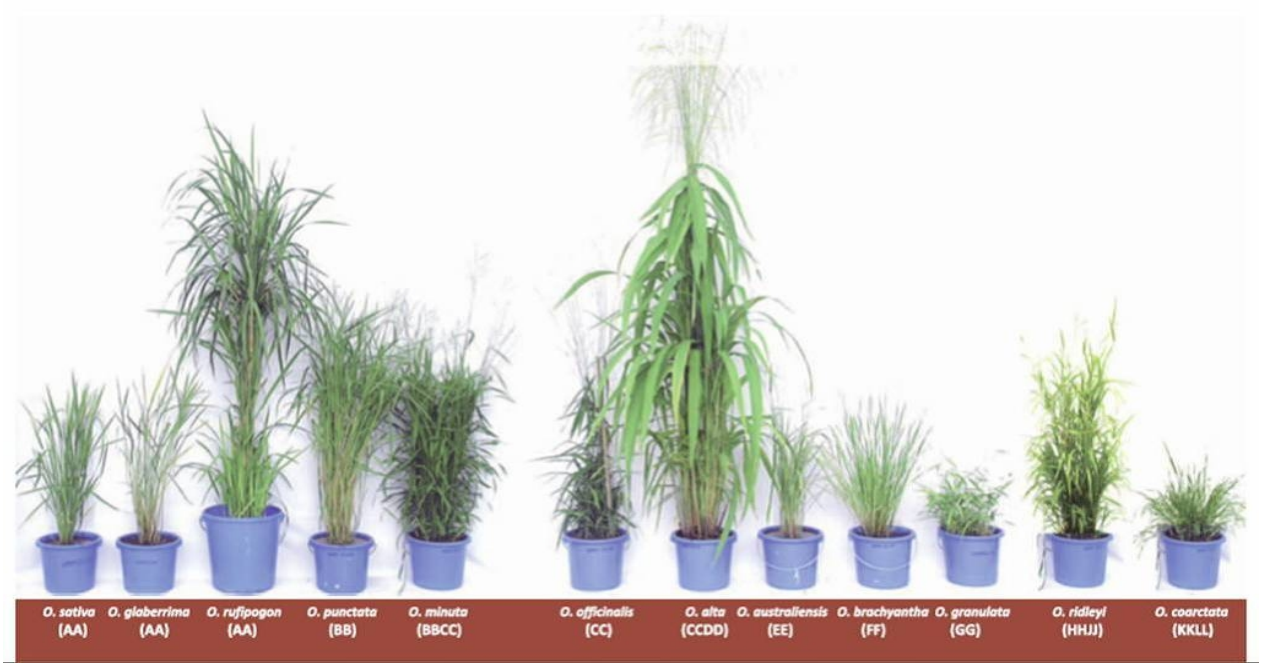



**Figure 15 : Comparaison morphologique des 12 espèces de riz dans le genre *Oryza* au stade de floraison (d'après D.S. Brar).**

# Projets de recherche

L'axe de recherche que je souhaite développer concerne l'étude de l'impact biologique des ETs au sein du genre *Oryza*. Les deux principaux objectifs de mon projet de recherche seront (1) l'étude de la domestication des ETs et (2) de caractériser l'impact des ETs sur les modifications d'expression des gènes à grande échelle, avec potentiellement un impact sur la création de nouvelles fonctions biologiques.

Actuellement, j'approfondis mes compétences en bio-informatique, plus particulièrement la programmation en langage Perl et Python (les plus utilisés dans cette discipline). En effet, le domaine de la génomique est un secteur qui change très rapidement et donc face à la forte augmentation des ressources génomiques disponibles, il est nécessaire d'adapter mes compétences scientifiques.

**Le genre *Oryza*, un modèle de choix pour la génomique comparative chez les plantes**

Le riz est l'espèce cultivée qui possède le plus de ressources génomiques. En effet, peu de temps après la publication du génome du riz, la projet OMAP (Oryza Map Alignement Project), financé par la NSF (National Science Foundation), a été lancé avec l'objectif de générer des ressources génomiques pour mieux exploiter la diversité génétique des espèces de riz sauvages. Ce projet a été à l'origine d'un consortium international, iOMAP (International Oryza Map Alignement Project) avec différents groupes de travail à travers le monde visant à séquencer entièrement le génome de toutes les espèces du genre *Oryza*.

Les espèces du genre *Oryza* présentent une grande variabilité au niveau de leur morphologie (Figure 15) et des caractères physiologiques associés à l'adaptation à leurs divers habitats, comme une bonne tolérance à la chaleur, sécheresse, salinité, et aux métaux lourds du sol. Le nombre de gènes et l'ordre des gènes le long des chromosomes (synténie) restent très conservés chez toutes les espèces (Kim et al., 2008). Ce groupe taxonomique constitue un matériel idéal pour étudier l'impact fonctionnel des ETs à travers des études de génomiques comparatives.

L'équipe à laquelle j'appartiens fait partie du consortium iOMAP depuis sa création. La disponibilité de ces 10 génomes fournit une opportunité pour étudier l'évolution du génome des plantes sur une période de 15 millions d'années et d'une résolution sans précédent (Figure 16). Nous disposons également des transcriptomes de différents organes pour ces plantes ainsi que 3000 génomes de différentes variétés cultivées de riz.

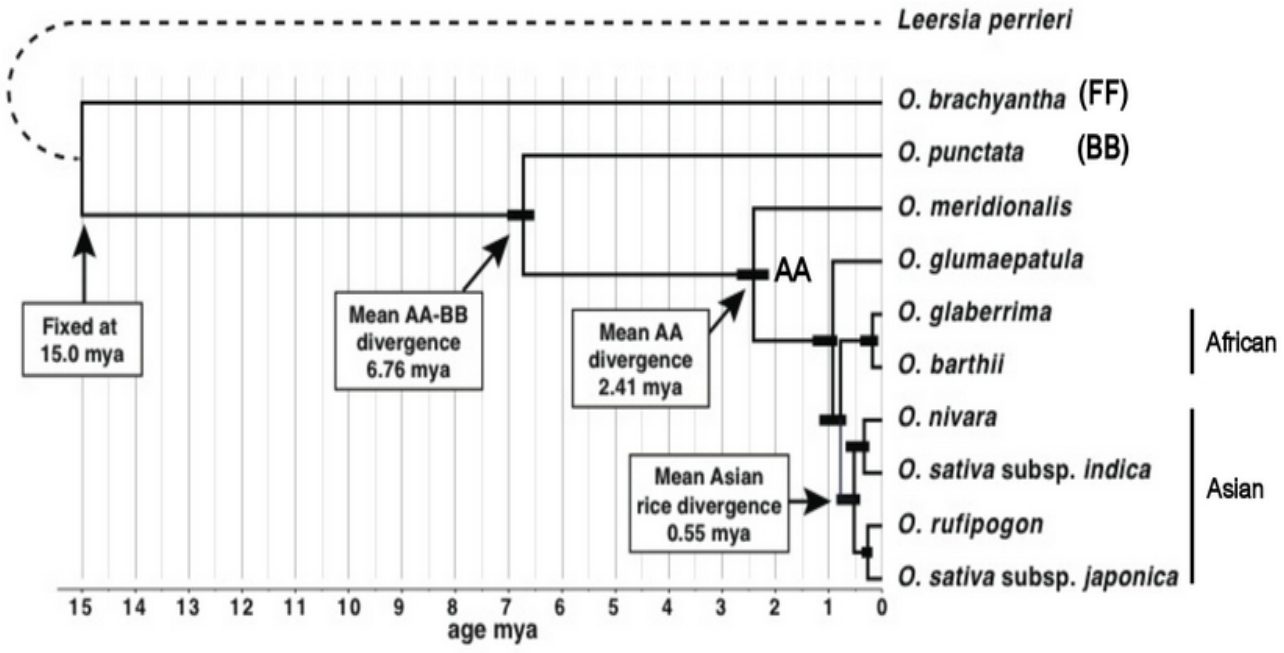



**Figure 16 : Arbre phylogénétique du genre *Oryza* présentant les différentes espèces (d'après D. Zwickl).** Les barres au niveau des nœuds indiquent le temps de divergence (en millions d'années) entre deux clades.

**Une nouvelle vision de l'impact des ETs**

Pendant de longues années, les ETs étaient décrits comme des éléments parasites des génomes entraînant des mutations délétères dans leur génome hôte, par leur capacité à se déplacer. En effet, beaucoup d'exemples montrent que les ETs entraînent des mutations, soit directement en s'insérant dans ou près des gènes, ou soit indirectement *via* des échanges génétiques inégaux entre des ETs dupliqués qui entraînent des duplications, inversions, translocations ou délétions (Craig et al., 2002).

Quelques exemples, mais de plus en plus nombreux, montrent que les ETs peuvent être bénéfiques pour l'hôte. Ils pourraient avoir un rôle dans l'évolution des génomes et/ou dans l'augmentation du pouvoir adaptatif. En effet, parmi les nombreuses copies d'éléments transposables insérées dans les génomes, certaines sont sélectivement neutres, c'est-à-dire qu'elles sont sans effet détectable ; d'autres, par contre, ont été « domestiquées » par les génomes et sont devenues des gènes ou des éléments de régulation des gènes.

La première observation d'un impact positif des ETs vient de la drosophile, où Biessmann et al., 1990 ont mis en évidence l'implication du rétrotransposon *HeT1* dans la fonction des télomères. Plusieurs études ont aussi mis en évidence l'implication de familles d'ETs dans la structure des chromosomes, comme le rétrotransposon *LAVA* dans la formation des centromères chez les Gibbons (Carbone et al., 2012). Il est intéressant de noter que la famille *LAVA* est chimérique et formée à partir de recombinaisons de familles d'éléments très répétés dans les génomes des primates, comme *L1*, *Alu* et *SVA*. Ceci suggère que l'acquisition de nouvelles fonctions pour les éléments peut avoir lieu après leur activité transpositionnelle. Le génome humain contient plusieurs centaines de gènes qui trouvent leur origine dans les éléments transposables. Le cas le plus typique est fourni par les gènes *Rag1* et *Rag2* du système immunitaire des mammifères, qui dérivent d'un élément transposable de la famille des transposons à ADN d'*hAT* (Biémont and Vieira, 2006; Zhou et al., 2004). Les mécanismes responsables de la diversité immunologique sont alors directement liés aux mécanismes enzymatiques associés à la transposition de ces anciens éléments transposables. Plusieurs études globales des génomes ont démontré que les ETs sont souvent recrutés comme des régulateurs clés dans l'expression des gènes et sont, entre autres, à l'origine d'un grand nombre de promoteurs chez les mammifères (Jacques et al., 2013). Cela a aussi été montré chez la drosophile, avec les multiples exemples du recrutement des éléments *PIF-like* pour former de nouveaux gènes (Casola et al., 2007). Il a été de même observé que l'enzyme transposase, qui est codée par les transposons à ADN, était essentielle au développement de la plante *Arabidopsis thaliana* (Bundock and Hooykaas, 2005) et à sa réponse à certains stimuli lumineux (Lin et al., 2007), ce qui suggère

que les protéines codées par les éléments transposables ont un impact sur le fonctionnement des génomes.

**Objectif 1 : Etude de la domestication des ETs chez les plantes**

La contribution bénéfique des ETs la plus saisissante est illustrée par le processus de la « domestication moléculaire », qui est un phénomène par lequel une séquence codante dérivée d'un ET donne naissance à un gène hôte fonctionnel. Ainsi, les ETs pourraient servir de réservoir dynamique pour créer de nouvelles fonctions cellulaires. Il faut rappeler que les ETs possèdent une machinerie attractive et élaborée. Ils codent des transposases, intégrases, reverses transcriptases, mais aussi des protéines structurales et de l'enveloppe, et peuvent donc être recrutés par leur hôte au cours de l'évolution. La domestication moléculaire est un processus évolutif qui conduit un ET à devenir un composant fonctionnel stable du génome associé à une fonction biologique (Miller et al., 1992).

Plusieurs critères ont été proposés pour déterminer les cas où les gènes dérivent de transposons et seraient en faveur d'une domestication moléculaire (Feschotte and Pritham, 2007). Contrairement à la nature des ETs, les gènes domestiqués existent en une seule copie dans le génome et sont conservés au cours de l'évolution, c'est-à-dire que des orthologues sont détectables dans les régions synténiques dans des espèces proches. Structurellement, ces gènes ont perdu les marques moléculaires de la transposition, comme les TIRs et les TSDs. Les protéines des gènes domestiqués sont phylogénétiquement liées aux protéines codées par l'élément dont elles dérivent. Les gènes domestiqués sont transcrits, contrairement aux ETs qui sont généralement inactivés. Ces gènes jouent un rôle biologique important « *in vivo* », mais en général, ils ont perdu leur capacité à transposer.

Les premiers gènes dérivés de transposons ont été identifiés chez la drosophile à partir de l'élément *P* (Miller et al., 1992). D'autres exemples ont ensuite été caractérisés dans les génomes des plantes et des animaux (Volff, 2006).

À chaque fois, ce sont des exemples ponctuels qui sont mis en évidence. Je propose d'utiliser une approche bio-informatique pour réaliser une étude systématique globale de la domestication des ETs. Je pourrai ainsi examiner de manière exhaustive la rétention synténique de toutes les séquences d'ETs sur diverses échelles de temps dans une variété de clades. L'avantage de cette approche est que je vais utiliser des données déjà existantes et donc avec un coût très réduit.

Mon premier objectif concerne l'étude de la domestication de l'ensemble des ETs chez les plantes. Ce projet est réalisé en collaboration avec Damon Lisch (Purdue University), spécialiste de

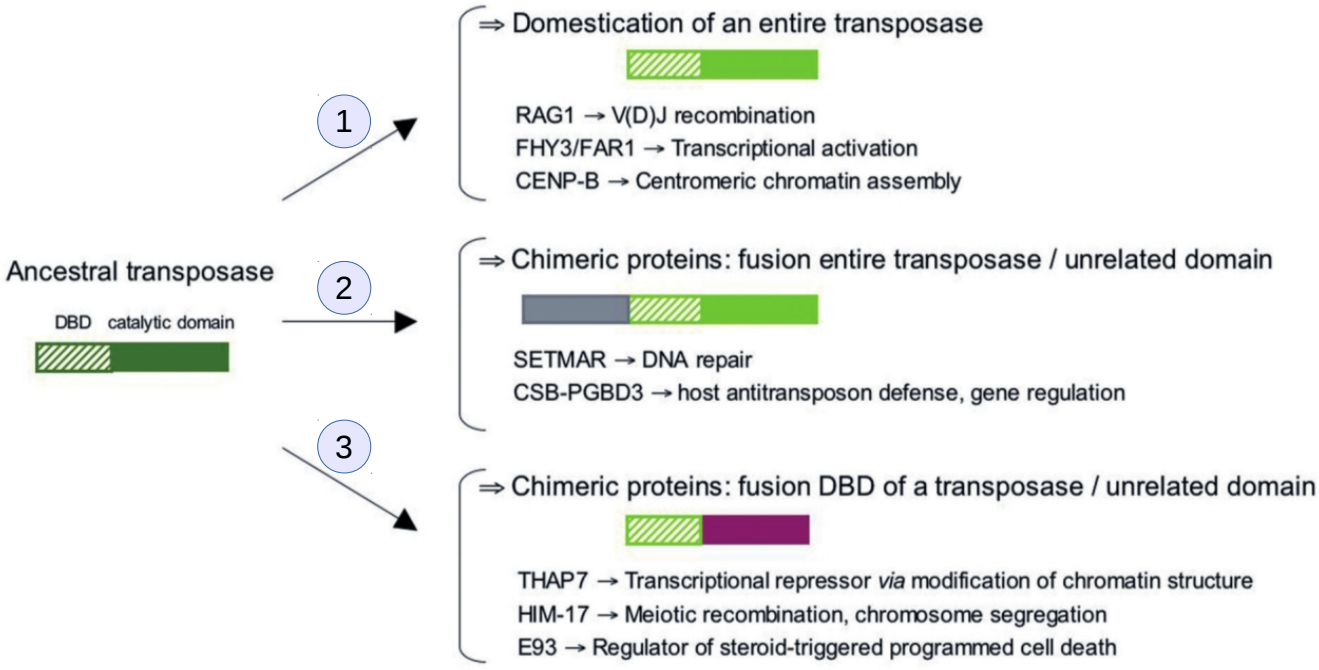



**Figure 17 : Diversité structurale de protéines domestiquées (d'après Sinzelle et al., 2009).** Schéma d'une transposase classique contenant un domaine de fixation à l'ADN (DBD) (rectangle vert hachuré) et le domaine catalytique (rectangle vert). Les événements de domestication de la transposase donnent diverses protéines : (1) domestication du gène entier de la transposase, (2) gènes chimériques formés par le domaine complet de la transposase et d'un domaine fonctionnel additionnel, (3) gènes chimériques formés à partir du domaine DBD de la transposase et d'un domaine fonctionnel additionnel. Pour chacun de ces trois cas, des protéines domestiquées et leurs rôles fonctionnels sont fournis comme exemple.

la domestication des éléments MULEs (classe II). Nous commencerons donc notre étude sur les éléments de classe II, puisque Damon a déjà identifié un exemple lors de la caractérisation de la domestication d'un gène de transposase (Hudson et al., 2003). Il s'agit des gènes *FAR1* et *FHY3*, des facteurs de transcription impliqué dans la réponse à la lumière rouge et dérivés d'un gène de transposase d'un transposon MULE (*Mutator*-like element ; (Hudson et al., 2003). Notre approche sera ensuite appliquée aux éléments de classe I.

**1ère partie : Domestication des transposases**

Le nombre d'exemples de nouveaux gènes domestiqués découverts augmente régulièrement et met en évidence leur diversité structurale. Certains gènes proviennent de la séquence codante entière de la transposase, mais il y a aussi des gènes chimériques dans lesquels la séquence codante de la transposase est fusionnée avec un domaine fonctionnel pré-existant (Figure 17 ; (Sinzelle et al., 2009)). De plus, la diversité structurale est renforcée par le fait que de nombreux gènes domestiqués ont seulement conservé le domaine DBD (DNA-binding domain) de la protéine ancestrale codée par le transposon (Figure 17). Cela peut expliquer pourquoi un grand nombre de gènes dérivés de transposons sont connus pour être recrutés par l'hôte pour fonctionner comme régulateurs de la transcription. Cependant, le rôle biologique de la majorité d'entre eux reste inconnu.

Les transposases possèdent deux domaines essentiels : un domaine de fixation à l'ADN (DBD) spécifique et un domaine catalytique responsable du clivage de l'ADN et des réactions de jonctions. Bien qu'un grand nombre des protéines domestiquées aient évoluées à partir du gène de la transposase en entier, et donc contenant les deux domaines fonctionnels, le DBD apparaît pour être préférentiellement co-opté par l'hôte. Il n'est pas surprenant que la majorité des protéines domestiquées fonctionnent comme des régulateurs transcriptionnels (activateurs ou répresseurs).

Lorsque des transposons sont conservés à une position donnée sur une période supérieure à 10 millions d'années, ils sont certainement domestiqués. En effet, on sait que pour un transposon donné, une fois qu'il s'est inséré, il aura tendance à muter, dégénérer et disparaître puisqu'il n'y a normalement pas de sélection positive pour ce type d'élément. Cela suggère que les transposases synténiques devraient mettre en évidence de la pseudogénisation (c'est-à-dire mutations ponctuelles et petites délétions), car elles ne sont pas soumises à la sélection même si elles sont encore synténiques. Au contraire, un transposon dans les premières étapes de la domestication doit être encore synténique et montrer la preuve de la sélection sur la protéine.

Nous allons utiliser ces différentes caractéristiques des transposases pour développer notre approche. Parmi les différents éléments de classe II, nous allons tout d'abord étudier les MULEs. Nous pourrons ainsi à chaque étape tester nos résultats par rapport à ceux déjà obtenu par Damon. Puis, lorsque notre approche sera validée sur ces éléments, nous pourrons élargir notre recherche aux autres éléments de classe II, comme les *hAT*, *CACTA*, *Pif/Pong*, *Mariner…* L'objectif étant de développer des outils bio-informatiques pour conduire des analyses exhaustives de génomiques comparatives, ce qui nous permettra de déterminer le répertoire complet des gènes qui sont potentiellement dérivés d'éléments de classe II dans les différentes espèces d'*Oryza*.

**Approche expérimentale :**

Tout d'abord, je vais rechercher tous les gènes liés aux transposases provenant des MULEs dans le génome de Nipponbare. Pour cela, j'utiliserai tblastn et Psi-Blast en parallèle. Je réaliserai la caractérisation des signatures structurales de ces séquences (conservation des domaines des transposases, présence des TIRs…) puis une classification phylogénétique avec les gènes identifiés et les transposases pour remettre les événements de domestication dans leur contexte. Lors de ces analyses, j'utiliserai un jeu de données d'ETs non domestiqués qui me servira de contrôle. Ces éléments devraient présenter des traces de pseudogénisation *via* des délétions.

Ensuite les deux gènes flanquants (le plus proches du coté 5' et 3' sur les pseudomolécules) seront extraits pour chaque gène de transposase. Je rechercherai les relations de synténie dans les différents génomes d'*Oryza* et *sorghum,* et je déclarerai une synténie conservée si au moins un des gènes flanquants est conservé. Quand j'aurai identifié de nouveaux candidats, Damon m'aidera à les vérifier en utilisant GeVo, un outil très puissant permettant d'identifier rapidement des d'erreurs (comme les faux positifs) à travers une observation directe des régions synténiques.

Les signatures de sélection seront ensuite étudiée avec le ratio de substitution non synonyme par substitutions synonymes (Ka/Ks). Cela permettra de voir, par exemple, s'il existe une sélection positive sur les domaines de fixation à l'ADN.

De plus, comme nous possédons les données RNAseq des différents génomes d'*Oryza* et *sorghum*, je pourrai les analyser pour les candidats retenus. La transcription étant une caractéristique importante des ETs domestiqués car requise dans les fonctions cellulaires de l'hôte, en regardant l'activité transcriptionnelle et en combinant avec les différentes approches, des gènes candidats dérivés de la domestication d'ETs pourront être identifiés.

Il serait aussi intéressant de pouvoir combiner des analyses de méthylomes et de petits ARNs pour obtenir une vue globale de ces éléments domestiqués. Je pourrai aussi étudier les séquences *cis*-régulatrices présentes dans les promoteurs de ces gènes. De plus, j'analyserai des mutants de ces gènes afin de voir si je peux mettre en évidence des phénotypes particuliers. Ainsi, nous pourrons observer l'ensemble des mécanismes génétiques et épigénétiques intervenant dans la domestication des ETs et peut être identifier des gènes domestiqués qui seraient à l'origine de changement phénotypique au cours de l'évolution des plantes.

Ces différentes étapes seront réalisées sur les espèces suivantes, représentant une divergence allant jusqu'à 50 millions d'années : *Oryza glaberrima*, *O. meridionalis*, *O. punctata*, *O. brachyantha* et *sorghum* avec 0.5, 2.5, 7, 15 et 50 millions d'années de divergence avec *O. sativa,* respectivement. Avec cette échelle de temps, nous pourrons détecter des événements de domestication très récents.

**2ème partie :**

J'appliquerai ensuite ce protocole aux éléments de classe I. En effet, quelques exemples de domestication concernant les différentes protéines des éléments de classe I ont été mis en évidence : l'intégrase chez les mammifères (Lloréns and Marín, 2001), la reverse transcriptase chez la drosophile, qui serait à l'origine des télomérases (Arkhipova et al., 2003), la Gag, dont les gènes dérivés auraient un rôle essentiel dans le développement des mammifères (Brandt et al., 2005), et l'enveloppe, dont les gènes dérivés joueraient un rôle dans la formation du placenta (Mi et al., 2000). J'analyserai donc ces différents candidats dans le genre *Oryza*, en utilisant la même approche que celle décrite pour les éléments de classe II.

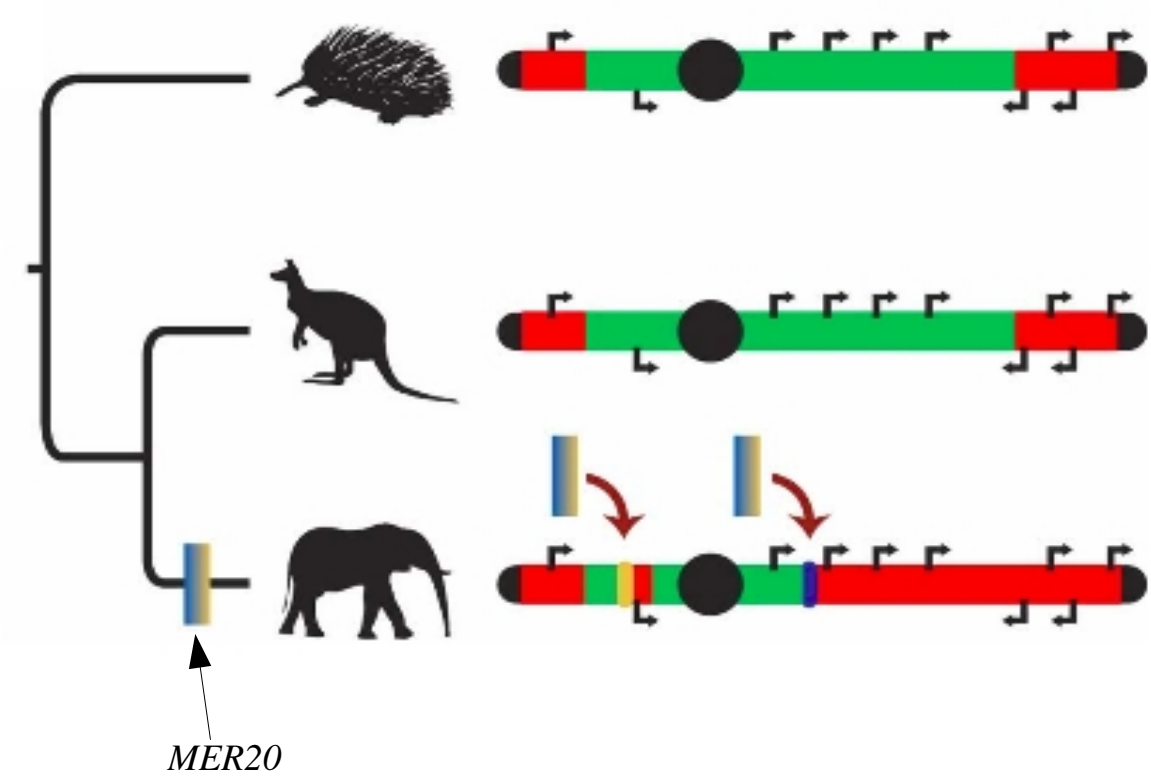



**Figure 18 : Modèle de la reprogrammation de gènes régulateurs par *MER20* (d'après Lynch et al., 2012 ).** Ancestralement, certains gènes (flèches noires) ne sont pas exprimé dans les cellules endométriales car réprimé par des modifications épigénétiques de la chromatine. Lorsque MER20 s'insère dans le génome des lignées des mammifères (blocs bleus et jaunes sur la phylogénie), ils empêchent la diffusion de l'état de la chromatine, établissant de nouvelles frontières entre la chromatine active (rouge) et inactive (vert).

**Reprogrammation des réseaux de gènes à l'échelle génomique**

En ce qui concerne l'impact des ETs sur l'expression des gènes, la majorité des exemples qui ont été identifiés concernait des interactions gène-ET à un seul locus. Sachant que les ETs peuvent subir des bursts de transposition donnant des centaines ou des milliers de répétitions et que beaucoup d'entre eux peuvent s'insérer près des gènes, on se pose la question de l'impact des ETs à l'échelle génomique. En effet, pourraient-ils modifier l'expression d'un grand nombre de gènes au niveau de l'ensemble du génome ?

Plusieurs études récentes sur les animaux suggèrent que cela pourrait être le cas et surtout montrent que les ETs seraient des agents puissants dans l'évolution de réseaux de régulation génique et seraient impliqués dans l'apparition de nouveaux caractères. Par exemple, le burst de transposition de l'élément *MER20* serait à l'origine de la formation du placenta chez les mammifères en reprogrammant un réseau de gènes régulateurs intervenant au cours du développement (Lynch et al., 2011). En explorant le paysage transcriptionnel des gènes dans les cellules endométriales des mammifères, ces auteurs ont mis en évidence que l'évolution de la grossesse était associée à une reprogrammation à grande échelle d'un réseau de régulation génique et qu'environ 13 % de ces gènes étaient dans une région de 200 kb de l'élément *MER20*.

Les résultats de Lynch suggèrent que l'insertion de *MER20* dans le génome ancestral des mammifères établirait de nouvelles limites entre la chromatine active et inactive dans les cellules endométriales et permettrait ainsi aux gènes préalablement inactifs d'être disponibles pour l'activation (Figure 18).

D'autres études sur les génomes ont démontré que les ETs sont souvent associés avec des séquences régulatrices des gènes (Hénaff et al., 2014; Jacques et al., 2013), suggérant que les bursts de transposition peuvent avoir un impact très fort sur la régulation du génome.

Certaines publications ont décrit que l'activation des éléments était réalisée par la présence de sites de fixation aux facteurs de transcription dans les régions promotrices des éléments et peuvent ainsi réguler l'expression des gènes proches (Mhiri et al., 1997). Si cette observation est généralisée au règne végétal, elle pourrait ouvrir de nouvelles perspectives pour l'impact des ETs au niveau fonctionnel sur tout le génome. Ce processus, particulièrement pour les éléments de classe I, pourrait mener à la diffusion de régions promotrices de réponse aux stress dans le génome qui pourrait former un réseau de gènes répondant aux stress.

**Objectif 2 : L'impact des ETs sur la reprogrammation des réseaux de gènes chez les plantes et potentiellement sur de nouvelles fonctions biologiques**

Mon deuxième projet concerne donc l'étude de l'impact des ETs sur la régulation de l'expression des gènes à l'échelle du génome et, à plus long terme peut-être, de leur impact sur de nouvelles fonctions biologiques.

La plupart des analyses sont basées sur des études de génomique comparative dans le but d'établir une corrélation entre l'insertion des ETs et la fonction biologique. Chez les animaux, de telles comparaisons concernent des espèces assez distantes (comme le marsupial et les mammifères, Lynch et al., 2011). L'exemple du placenta des mammifères concerne un événement majeur qui a eu lieu il y a plus de 150 millions d'années et a donné naissance à un large groupe taxonomique. Le travail d'Hénaff et collaborateurs sur les plantes concernait la famille des brassicacées avec seulement un cas de comparaison intra-genre (*i.e. Arabidopsis thaliana* et *A. lyrata*). Cependant, la dynamique du processus par lequel les ETs mettent en place les réseaux de régulation pourrait être découverte avec une étude comparative utilisant des espèces plus proches. Notre étude peut paraître moins ambitieuse que l'analyse menée chez les mammifères, mais le turn-over étant beaucoup plus rapide chez les plantes, je vais réaliser, dans un premier temps, mon étude au niveau d'un genre. Actuellement, le matériel pour lequel nous disposons des ressources génomiques appropriées est le genre *Oryza*. En effet, nous disposons de 3000 génomes de différentes variétés cultivées de riz et de 10 génomes de riz sauvages. Ainsi, l'activité transpositionnelle des ETs pourra être analysée au sein d'un groupe variétal et permettra de réaliser un inventaire exhaustif des éléments actifs et d'identifier les bursts de transposition, ainsi que les relations de polymorphisme.

Le groupe taxonomique à notre disposition constitue donc un matériel idéal pour étudier l'impact fonctionnel du burst des ETs à travers des études de génomiques comparatives, car les variations transcriptomiques, que nous pourrons observer parmi les espèces, pourront être corrélées avec la propagation des familles d'ETs très répétés. Des analyses de RNAseq seront réalisées à cet effet. Les variations dans l'expression des gènes seront confrontées avec la présence d'une famille d'ETs pour chaque variété et/ou espèce. Cela consiste à détecter des insertions des ETs dans le voisinage des gènes différentiellement exprimés. Pour les cas où des corrélations seront établies, des caractérisations plus approfondies des interactions gènes-ETs seront réalisées afin de déchiffrer les mécanismes par lesquels les ETs ont un impact sur l'expression des gènes. En particulier, je testerai si les insertions des ETs induisent des changements dans l'état de méthylation et/ou si ils agissent comme des activateurs de la transcription, à travers des signaux de régulation qu'ils peuvent abriter dans leur propre promoteur. Je mènerai aussi des recherches d'ontologie des gènes

identifiés afin d'identifier les processus biologiques et les différentes voies recrutées dans le phénotype étudié. Je pourrai aussi réaliser des constructions avec la luciférase afin de valider l'impact d'un ET sur l'expression de gènes adjacents. Toutes ces analyses nous permettront de découvrir si les bursts de transposition observés ont un impact sur la reprogrammation de réseaux de régulation génique.

L'un des défis fondamentaux en biologie est d'essayer d'expliquer l'origine de nouveaux caractères phénotypiques, comme par exemple la formation de nouveaux types cellulaires, mais les mécanismes moléculaires qui provoquent ces nouveautés sont encore peu clairs.

À plus long terme, j'aimerais développer mon approche pour étudier une fonction biologique innovante chez les plantes comme, par exemple, le développement des carpelles chez les dicotylédones, ou la photosynthèse C4, ou la fixation de l'azote chez les légumineuses. Ce projet se ferait en collaboration avec Scott Jackson et Doug Soltis (USA). Nous sommes actuellement en train de discuter pour élaborer ce projet de l'étude de l'innovation biologique dans l'évolution des plantes. Nous faisons actuellement le point sur ce dont nous disposons sur chacune de ces fonctions biologiques, comme les différentes ressources génomiques disponibles, les réseaux de gènes connus… Pour pouvoir choisir celle qui paraît le plus judicieux à analyser.

Au vue, de la capacité des ETs à répondre à des signaux spécifiques, combinés à leur nature répétitive et répandue dans les génomes, ces éléments peuvent engendrer une régulation fine de l'expression des gènes et les variations ainsi générées peuvent étendre la diversification de l'activité des gènes, entraînant des conséquences importantes comme la création de diversité phénotypique. Il y a donc actuellement un intérêt grandissant par la communauté scientifique dans cette vue « fonctionnelle » des éléments mobiles, par laquelle les ETs peuvent représenter des éléments clés au cœur de réseaux régulateurs, menant à la reprogrammation d'une batterie de gènes en répondant par exemple à un stimuli spécifique. C'est dans ce domaine d'analyse des réseaux d'interactions gènes-éléments transposables, responsables de divers phénotypes que les découvertes les plus stimulantes sont attendues.

# Références bibliographiques


Anca, I.-A., Fromentin, J., Bui, Q.T., Mhiri, C., Grandbastien, M.-A., and Simon-Plas, F. (2014). Different tobacco retrotransposons are specifically modulated by the elicitor cryptogein and reactive oxygen species. J. Plant Physiol. *171*, 1533–1540.

Arkhipova, I.R., Pyatkov, K.I., Meselson, M., and Evgen'ev, M.B. (2003). Retroelements containing introns in diverse invertebrate taxa. Nat. Genet. *33*, 123–124.

Banks, J.A., Masson, P., and Fedoroff, N. (1988). Molecular mechanisms in the developmental regulation of the maize Suppressor-mutator transposable element. Genes Dev. *2*, 1364–1380.

Beguiristain, T., Grandbastien, M.A., Puigdomènech, P., and Casacuberta, J.M. (2001). Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. Plant Physiol. *127*, 212–221.

Bennetzen, J.L., Ma, J., and Devos, K.M. (2005). Mechanisms of recent genome size variation in flowering plants. Ann. Bot. *95*, 127–132.

Biémont, C., and Vieira, C. (2006). Genetics: junk DNA as an evolutionary force. Nature *443*, 521–524.

Biessmann, H., Mason, J.M., Ferry, K., d' Hulst, M., Valgeirsdottir, K., Traverse, K.L., and Pardue, M.L. (1990). Addition of telomere-associated HeT DNA sequences "heals" broken chromosome ends in Drosophila. Cell *61*, 663–673.

Bingham, P.M., Kidwell, M.G., and Rubin, G.M. (1982). The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family. Cell *29*, 995–1004.

Blewitt, M.E., Vickaryous, N.K., Paldi, A., Koseki, H., and Whitelaw, E. (2006). Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice. PLoS Genet. *2*, e49.

Brandt, J., Schrauth, S., Veith, A.-M., Froschauer, A., Haneke, T., Schultheis, C., Gessler, M., Leimeister, C., and Volff, J.-N. (2005). Transposable elements as a source of genetic innovation: expression and evolution of a family of retrotransposon-derived neogenes in mammals. Gene *345*, 101–111.

Bucher, E., Reinders, J., and Mirouze, M. (2012). Epigenetic control of transposon transcription and mobility in Arabidopsis. Curr. Opin. Plant Biol. *15*, 503–510.

Bui, Q.T., and Grandbastien, M.A. (2012) LTR retrotransposons as controlling elements of genome response to stress ? In : Grandbastien, M.A., Casacuberta, J.M. (Eds.) Plant Transposable Elements : impact on Genome Structure and Function, Topics in Current Genetics, 24, Springer Berlin, Heidelberg, pp. 273296.

Bundock, P., and Hooykaas, P. (2005). An Arabidopsis hAT-like transposase is essential for plant development. Nature *436*, 282–284.

Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., Reforgiato-Recupero, G.,

and Martin, C. (2012). Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. Plant Cell *24*, 1242–1255.

Calvayrac, C., Martin-Laurent, F., Faveaux, A., Picault, N., Panaud, O., Coste, C.-M., Chaabane, H., and Cooper, J.-F. (2012). Isolation and characterisation of a bacterial strain degrading the herbicide sulcotrione from an agricultural soil. Pest Manag. Sci. *68*, 340–347.

Capy, P. (2005). Classification and nomenclature of retrotransposable elements. Cytogenet. Genome Res. *110*, 457–461.

Carbone, L., Harris, R.A., Mootnick, A.R., Milosavljevic, A., Martin, D.I.K., Rocchi, M., Capozzi, O., Archidiacono, N., Konkel, M.K., Walker, J.A., et al. (2012). Centromere remodeling in Hoolock leuconedys (Hylobatidae) by a new transposable element unique to the gibbons. Genome Biol. Evol. *4*, 648–658.

Casola, C., Lawing, A.M., Betrán, E., and Feschotte, C. (2007). PIF-like transposons are common in drosophila and have been repeatedly domesticated to generate new host genes. Mol. Biol. Evol. *24*, 1872–1888.

Chan, S.W.-L., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. (2004). RNA silencing genes control de novo DNA methylation. Science *303*, 1336.

Chaparro, C., Guyot, R., Zuccolo, A., Piégu, B., and Panaud, O. (2007). RetrOryza: a database of the rice LTR-retrotransposons. Nucleic Acids Res. *35*, D66–D70.

Chomet, P.S., Wessler, S., and Dellaporta, S.L. (1987). Inactivation of the maize transposable element Activator (Ac) is associated with its DNA modification. EMBO J. *6*, 295–302.

Chopra, S., Brendel, V., Zhang, J., Axtell, J.D., and Peterson, T. (1999). Molecular characterization of a mutable pigmentation phenotype and isolation of the first active transposable element from Sorghum bicolor. Proc. Natl. Acad. Sci. U. S. A. *96*, 15330–15335.

Cordaux, R., and Batzer, M.A. (2009). The impact of retrotransposons on human genome evolution. Nat. Rev. Genet. *10*, 691–703.

Craig, N.L., Craigie R., Gellert M, and Lambowitz, A.M. (2002) Mobile DNA II. ASM Press, Washington, DC.

Daniels, S.B., Chovnick, A., and Boussy, I.A. (1990). Distribution of hobo transposable elements in the genus Drosophila. Mol. Biol. Evol. *7*, 589–606.

Devos, K.M. (2010). Grass genome organization and evolution. Curr. Opin. Plant Biol. *13*, 139–145.

Drogue, B., Sanguin, H., Chamam, A., Mozar, M., Llauro, C., Panaud, O., Prigent-Combaret, C., Picault, N., and Wisniewski-Dyé, F. (2014). Plant root transcriptome profiling reveals a strain-dependent response during Azospirillum-rice cooperation. Front. Plant Sci. *5*, 607.

Duan, K., Ding, X., Zhang, Q., Zhu, H., Pan, A., and Huang, J. (2008). AtCopeg1, the unique gene originated from AtCopia95 retrotransposon family, is sensitive to external hormones and abiotic stresses. Plant Cell Rep. *27*, 1065–1073.

Ebbs, M.L., and Bender, J. (2006). Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. Plant Cell *18*, 1166–1176.

Ebina, H., and Levin, H.L. (2007). Stress management: how cells take control of their transposons. Mol. Cell *27*, 180–181.

El Baidouri, M., and Panaud, O. (2013). Comparative Genomic Paleontology across Plant Kingdom Reveals the Dynamics of TE-Driven Genome Evolution. Genome Biol. Evol. *5*, 954–965.

El Baidouri, M., Carpentier, M.-C., Cooke, R., Gao, D., Lasserre, E., Llauro, C., Mirouze, M., Picault, N., Jackson, S.A., and Panaud, O. (2014). Widespread and frequent horizontal transfers of transposable elements in plants. Genome Res. *24*, 831–838.

Estep, M.C., DeBarry, J.D., and Bennetzen, J.L. (2013). The dynamics of LTR retrotransposon accumulation across 25 million years of panicoid grass evolution. Heredity *110*, 194–204.

De Felice, B., Wilson, R.R., Argenziano, C., Kafantaris, I., and Conicella, C. (2009). A transcriptionally active copia-like retroelement in Citrus limon. Cell. Mol. Biol. Lett. *14*, 289–304.

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. Nat. Rev. Genet. *9*, 397–405.

Feschotte, C., and Pritham, E.J. (2007). DNA transposons and the evolution of eukaryotic genomes. Annu. Rev. Genet. *41*, 331–368.

Finatto, T., de Oliveira, A.C., Chaparro, C., da Maia, L.C., Farias, D.R., Woyann, L.G., Mistura, C.C., Soares-Bresolin, A.P., Llauro, C., Panaud, O., et al. (2015). Abiotic stress and genome dynamics: specific genes and transposable elements response to iron excess in rice. Rice N. Y. N *8*, 13.

Finnegan, D.J. (1989). Eukaryotic transposable elements and genome evolution. Trends Genet. TIG *5*, 103–107.

Gaudet, F., Rideout, W.M., Meissner, A., Dausman, J., Leonhardt, H., and Jaenisch, R. (2004). Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing. Mol. Cell. Biol. *24*, 1640–1648.

Grandbastien, M.-A. (2015). LTR retrotransposons, handy hitchhikers of plant regulation and stress response. Biochim. Biophys. Acta *1849*, 403–416.

Grandbastien, M.A., Spielmann, A., and Caboche, M. (1989). Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature *337*, 376–380.

Grandbastien, M.-A., Audeon, C., Bonnivard, E., Casacuberta, J.M., Chalhoub, B., Costa, A.-P.P., Le, Q.H., Melayah, D., Petit, M., Poncet, C., et al. (2005). Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. Cytogenet. Genome Res. *110*, 229–241.

Grandbastien, M.A. (1998) Activation of plant retrotransposons under stress conditions. Trends Plant Sci 3, 181187.

Haag, J.R., and Pikaard, C.S. (2011). Multisubunit RNA polymerases IV and V: purveyors of non-

coding RNA for plant gene silencing. Nat. Rev. Mol. Cell Biol. *12*, 483–492.

Hawkins, J.S., Kim, H., Nason, J.D., Wing, R.A., and Wendel, J.F. (2006). Differential lineage-specific amplification of transposable elements is responsible for genome size variation in Gossypium. Genome Res. *16*, 1252–1261.

Hénaff, E., Vives, C., Desvoyes, B., Chaurasia, A., Payet, J., Gutierrez, C., and Casacuberta, J.M. (2014). Extensive amplification of the E2F transcription factor binding sites by transposons during evolution of Brassica species. Plant J. Cell Mol. Biol. *77*, 852–862.

Hirochika, H., Sugimoto, K., Otsuki, Y., Tsugawa, H., and Kanda, M. (1996). Retrotransposons of rice involved in mutations induced by tissue culture. Proc. Natl. Acad. Sci. U. S. A. *93*, 7783–7788.

Huang, C.R.L., Burns, K.H., and Boeke, J.D. (2012). Active transposition in genomes. Annu. Rev. Genet. *46*, 651–675.

Hudson, M.E., Lisch, D.R., and Quail, P.H. (2003). The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway. Plant J. Cell Mol. Biol. *34*, 453–471.

Ishimaru, Y., Suzuki, M., Tsukamoto, T., Suzuki, K., Nakazono, M., Kobayashi, T., Wada, Y., Watanabe, S., Matsuhashi, S., Takahashi, M., et al. (2006). Rice plants take up iron as an Fe3+-phytosiderophore and as Fe2+. Plant J. Cell Mol. Biol. *45*, 335–346.

Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., and Paszkowski, J. (2011). An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature *472*, 115–119.

Ivashuta, S., Naumkina, M., Gau, M., Uchiyama, K., Isobe, S., Mizukami, Y., and Shimamoto, Y. (2002). Genotype-dependent transcriptional activation of novel repetitive elements during cold acclimation of alfalfa (Medicago sativa). Plant J. Cell Mol. Biol. *31*, 615–627.

Jacob, Y., Feng, S., LeBlanc, C.A., Bernatavichute, Y.V., Stroud, H., Cokus, S., Johnson, L.M., Pellegrini, M., Jacobsen, S.E., and Michaels, S.D. (2009). ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. Nat. Struct. Mol. Biol. *16*, 763–768.

Jacques, P.-É., Jeyakani, J., and Bourque, G. (2013). The majority of primate-specific regulatory sequences are derived from transposable elements. PLoS Genet. *9*, e1003504.

Jiang, N., Bao, Z., Zhang, X., Hirochika, H., Eddy, S.R., McCouch, S.R., and Wessler, S.R. (2003). An active DNA transposon family in rice. Nature *421*, 163–167.

Jones, J.M., and Gellert, M. (2004). The taming of a transposon: V(D)J recombination and the immune system. Immunol. Rev. *200*, 233–248.

Kalendar, R., Tanskanen, J., Immonen, S., Nevo, E., and Schulman, A.H. (2000). Genome evolution of wild barley (Hordeum spontaneum) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc. Natl. Acad. Sci. U. S. A. *97*, 6603–6607.

Kapitonov, V.V., and Jurka, J. (2001). Rolling-circle transposons in eukaryotes. Proc. Natl. Acad.

Sci. U. S. A. *98*, 8714–8719.

Kidwell, M.G., Kidwell, J.F., and Sved, J.A. (1977). Hybrid Dysgenesis in DROSOPHILA MELANOGASTER: A Syndrome of Aberrant Traits Including Mutation, Sterility and Male Recombination. Genetics *86*, 813–833.

Kim, H., Hurwitz, B., Yu, Y., Collura, K., Gill, N., SanMiguel, P., Mullikin, J.C., Maher, C., Nelson, W., Wissotski, M., et al. (2008). Construction, alignment and analysis of twelve framework physical maps that represent the ten genome types of the genus Oryza. Genome Biol. *9*, R45.

Kimura, Y., Tosa, Y., Shimada, S., Sogo, R., Kusaba, M., Sunaga, T., Betsuyaku, S., Eto, Y., Nakayashiki, H., and Mayama, S. (2001). OARE-1, a Ty1-copia retrotransposon in oat activated by abiotic and biotic stresses. Plant Cell Physiol. *42*, 1345–1354.

Kobayashi, S., Goto-Yamamoto, N., and Hirochika, H. (2004). Retrotransposon-induced mutations in grape skin color. Science *304*, 982.

Komatsu, M., Shimamoto, K., and Kyozuka, J. (2003). Two-step regulation and continuous retrotransposition of the rice LINE-type retrotransposon Karma. Plant Cell *15*, 1934–1944.

Korbel, J.O., Urban, A.E., Grubert, F., Du, J., Royce, T.E., Starr, P., Zhong, G., Emanuel, B.S., Weissman, S.M., Snyder, M., et al. (2007). Systematic prediction and validation of breakpoints associated with copy-number variants in the human genome. Proc. Natl. Acad. Sci. U. S. A. *104*, 10110–10115.

Kumar, A., and Bennetzen, J.L. (1999). Plant retrotransposons. Annu. Rev. Genet. *33*, 479–532.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Larmande, P., Gay, C., Lorieux, M., Périn, C., Bouniol, M., Droc, G., Sallaud, C., Perez, P., Barnola, I., Biderre-Petit, C., et al. (2008). Oryza Tag Line, a phenotypic mutant database for the Genoplante rice insertion line library. Nucleic Acids Res. *36*, D1022–D1027.

Lesage, P., and Todeschini, A.L. (2005). Happy together: the life and times of Ty retrotransposons and their hosts. Cytogenet. Genome Res. *110*, 70–90.

Lin, R., Ding, L., Casola, C., Ripoll, D.R., Feschotte, C., and Wang, H. (2007). Transposase-derived transcription factors regulate light signaling in Arabidopsis. Science *318*, 1302–1305.

Lippman, Z., May, B., Yordan, C., Singer, T., and Martienssen, R. (2003). Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. PLoS Biol. *1*, E67.

Lisch, D. (2013). How important are transposons for plant evolution? Nat. Rev. Genet. *14*, 49–61.

Lisch, D., and Slotkin, R.K. (2011). Strategies for silencing and escape: the ancient struggle between transposable elements and their hosts. Int. Rev. Cell Mol. Biol. *292*, 119–152.

Liu, Z.L., Han, F.P., Tan, M., Shan, X.H., Dong, Y.Z., Wang, X.Z., Fedak, G., Hao, S., and Liu, B.

(2004). Activation of a rice endogenous retrotransposon Tos17 in tissue culture is accompanied by cytosine demethylation and causes heritable alteration in methylation pattern of flanking genomic regions. TAG Theor. Appl. Genet. Theor. Angew. Genet. *109*, 200–209.

Lloréns, C., and Marín, I. (2001). A mammalian gene evolved from the integrase domain of an LTR retrotransposon. Mol. Biol. Evol. *18*, 1597–1600.

Lynch, V.J., Leclerc, R.D., May, G., and Wagner, G.P. (2011). Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nat. Genet. *43*, 1154–1159.

Martin, A., Troadec, C., Boualem, A., Rajab, M., Fernandez, R., Morin, H., Pitrat, M., Dogimont, C., and Bendahmane, A. (2009). A transposon-induced epigenetic change leads to sex determination in melon. Nature *461*, 1135–1138.

Matzke, M.A., and Mosher, R.A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat. Rev. Genet. *15*, 394–408.

McCLINTOCK, B. (1950). The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. U. S. A. *36*, 344–355.

McCLINTOCK, B. (1951). Chromosome organization and genic expression. Cold Spring Harb. Symp. Quant. Biol. *16*, 13–47.

McClintock, B. (1984). The significance of responses of the genome to challenge. Science *226*, 792–801.

Medstrand, P., Landry, J.R., and Mager, D.L. (2001). Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. J. Biol. Chem. *276*, 1896–1903.

Mhiri, C., Morel, J.B., Vernhettes, S., Casacuberta, J.M., Lucas, H., and Grandbastien, M.A. (1997). The promoter of the tobacco Tnt1 retrotransposon is induced by wounding and by abiotic stress. Plant Mol. Biol. *33*, 257–266.

Mi, S., Lee, X., Li, X., Veldman, G.M., Finnerty, H., Racie, L., LaVallie, E., Tang, X.Y., Edouard, P., Howes, S., et al. (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature *403*, 785–789.

Miller, W.J., Hagemann, S., Reiter, E., and Pinsker, W. (1992). P-element homologous sequences are tandemly repeated in the genome of Drosophila guanche. Proc. Natl. Acad. Sci. U. S. A. *89*, 4018–4022.

Mitchell, G., Bartlett, D.W., Fraser, T.E., Hawkes, T.R., Holt, D.C., Townson, J.K., and Wichert, R.A. (2001). Mesotrione: a new selective herbicide for use in maize. Pest Manag. Sci. *57*, 120–128.

Miura, A., Yonebayashi, S., Watanabe, K., Toyama, T., Shimada, H., and Kakutani, T. (2001). Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature *411*, 212–214.

Miyao, A., Iwasaki, Y., Kitano, H., Itoh, J.-I., Maekawa, M., Murata, K., Yatou, O., Nagato, Y., and

Hirochika, H. (2007). A large-scale collection of phenotypic data describing an insertional mutant population to facilitate functional analysis of rice genes. Plant Mol. Biol. *63*, 625–635.

Moon, S., Jung, K.-H., Lee, D.-E., Jiang, W.-Z., Koh, H.J., Heu, M.-H., Lee, D.S., Suh, H.S., and An, G. (2006). Identification of active transposon dTok, a member of the hAT family, in rice. Plant Cell Physiol. *47*, 1473–1483.

Muehlbauer, G.J., Bhau, B.S., Syed, N.H., Heinen, S., Cho, S., Marshall, D., Pateyron, S., Buisine, N., Chalhoub, B., and Flavell, A.J. (2006). A hAT superfamily transposase recruited by the cereal grass genome. Mol. Genet. Genomics MGG *275*, 553–563.

Neumann, P., Yan, H., and Jiang, J. (2007). The centromeric retrotransposons of rice are transcribed and differentially processed by RNA interference. Genetics *176*, 749–761.

Ngezahayo, F., Xu, C., Wang, H., Jiang, L., Pang, J., and Liu, B. (2009). Tissue culture-induced transpositional activity of mPing is correlated with cytosine methylation in rice. BMC Plant Biol. *9*, 91.

Picault, N., Chaparro, C., Piegu, B., Stenger, W., Formey, D., Llauro, C., Descombin, J., Sabot, F., Lasserre, E., Meynard, D., et al. (2009). Identification of an active LTR retrotransposon in rice. Plant J. Cell Mol. Biol. *58*, 754–765.

Rice Annotation Project, Tanaka, T., Antonio, B.A., Kikuchi, S., Matsumoto, T., Nagamura, Y., Numa, H., Sakai, H., Wu, J., Itoh, T., et al. (2008). The Rice Annotation Project Database (RAP-DB): 2008 update. Nucleic Acids Res. *36*, D1028–D1033.

Rigal, M., and Mathieu, O. (2011). A "mille-feuille" of silencing: Epigenetic control of transposable elements. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1809*, 452–458.

Rocheta, M., Carvalho, L., Viegas, W., and Morais-Cecílio, L. (2012). Corky, a gypsy-like retrotransposon is differentially transcribed in Quercus suber tissues. BMC Res. Notes *5*, 432.

Roudier, F., Teixeira, F.K., and Colot, V. (2009). Chromatin indexing in Arabidopsis: an epigenomic tale of tails and more. Trends Genet. TIG *25*, 511–517.

Sabot, F., and Schulman, A.H. (2006). Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. Heredity *97*, 381–388.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. Science *326*, 1112–1115.

Shapiro, J., Machattie, L., Eron, L., Ihler, G., Ippen, K., and Beckwith, J. (1969). Isolation of pure lac operon DNA. Nature *224*, 768–774.

Singer, T., Yordan, C., and Martienssen, R.A. (2001). Robertson's Mutator transposons in A. thaliana are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1). Genes Dev. *15*, 591–602.

Sinzelle, L., Izsvák, Z., and Ivics, Z. (2009). Molecular domestication of transposable elements: from detrimental parasites to useful host genes. Cell. Mol. Life Sci. CMLS *66*, 1073–1093.

Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nat. Rev. Genet. *8*, 272–285.

Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzić, M., Becker, J.D., Feijó, J.A., and Martienssen, R.A. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell *136*, 461–472.

Stewart, C., Kural, D., Strömberg, M.P., Walker, J.A., Konkel, M.K., Stütz, A.M., Urban, A.E., Grubert, F., Lam, H.Y.K., Lee, W.-P., et al. (2011). A comprehensive map of mobile element insertion polymorphisms in humans. PLoS Genet. *7*, e1002236.

Sugimoto, K., Takeda, S., and Hirochika, H. (2000). MYB-related transcription factor NtMYB2 induced by wounding and elicitors is a regulator of the tobacco retrotransposon Tto1 and defense-related genes. Plant Cell *12*, 2511–2528.

Takeda, S., Sugimoto, K., Otsuki, H., and Hirochika, H. (1999). A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon Tto1 is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. Plant J. Cell Mol. Biol. *18*, 383–393.

Teixeira, F.K., and Colot, V. (2009). Gene body DNA methylation in plants: a means to an end or an end to a means? EMBO J. *28*, 997–998.

Tsugane, K., Maekawa, M., Takagi, K., Takahara, H., Qian, Q., Eun, C.-H., and Iida, S. (2006). An active DNA transposon nDart causing leaf variegation and mutable dwarfism and its related elements in rice. Plant J. Cell Mol. Biol. *45*, 46–57.

Ungerer, M.C., Strakosh, S.C., and Zhen, Y. (2006). Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. Curr. Biol. CB *16*, R872–R873.

Vaucheret, H., and Fagard, M. (2001). Transcriptional gene silencing in plants: targets, inducers and regulators. Trends Genet. TIG *17*, 29–35.

Villasante, A., Abad, J.P., Planelló, R., Méndez-Lago, M., Celniker, S.E., and de Pablos, B. (2007). Drosophila telomeric retrotransposons derived from an ancestral element that was recruited to replace telomerase. Genome Res. *17*, 1909–1918.

Vitte, C., and Panaud, O. (2005). LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. Cytogenet. Genome Res. *110*, 91–107.

Voinnet, O. (2001). RNA silencing as a plant immune system against viruses. Trends Genet. TIG *17*, 449–459.

Volff, J.-N. (2006). Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. BioEssays News Rev. Mol. Cell. Dev. Biol. *28*, 913–922.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., et al. (2007). A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. *8*, 973–982.

Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J., and van der Knaap, E. (2008). A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit.

Science *319*, 1527–1530.

Yassi, M.N., Ritzenthaler, C., Brugidou, C., Fauquet, C., and Beachy, R.N. (1994). Nucleotide sequence and genome characterization of rice yellow mottle virus RNA. J. Gen. Virol. *75 ( Pt 2)*, 249–257.

Zedek, F., Smerda, J., Smarda, P., and Bureš, P. (2010). Correlated evolution of LTR retrotransposons and genome size in the genus Eleocharis. BMC Plant Biol. *10*, 265.

Zhang, J., Yu, C., Pulletikurti, V., Lamb, J., Danilova, T., Weber, D.F., Birchler, J., and Peterson, T. (2009). Alternative Ac/Ds transposition induces major chromosomal rearrangements in maize. Genes Dev. *23*, 755–765.

Zhou, L., Mitra, R., Atkinson, P.W., Hickman, A.B., Dyda, F., and Craig, N.L. (2004). Transposition of hAT elements links transposable elements and V(D)J recombination. Nature *432*, 995–1001.