



**HAL**  
open science

# Caractérisation fonctionnelle et évolution moléculaire des gènes codant pour les facteurs d'initiation de la traduction eIF4E : des facteurs clés dans la résistance des plantes aux Potyvirus

Carine Charron

► **To cite this version:**

Carine Charron. Caractérisation fonctionnelle et évolution moléculaire des gènes codant pour les facteurs d'initiation de la traduction eIF4E : des facteurs clés dans la résistance des plantes aux Potyvirus. Sciences du Vivant [q-bio]. Université de la Méditerranée (Aix Marseille 2), 2007. Français. NNT : . tel-02824610

**HAL Id: tel-02824610**

**<https://hal.inrae.fr/tel-02824610>**

Submitted on 6 Jun 2020

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

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



Université Aix-Marseille II - Faculté des Sciences de Luminy

Ecole Doctorale des Sciences de la Vie et de la Santé

## THESE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE AIX-MARSEILLE II

Spécialité Microbiologie Moléculaire, Biologie Végétale et Biotechnologies

présentée et soutenue publiquement par

**Carine CHARRON**

le 21 décembre 2007

---

# Caractérisation fonctionnelle et évolution moléculaire des gènes codant pour les facteurs d'initiation de la traduction eIF4E : des facteurs clés dans la résistance des plantes aux *Potyvirus*

---

## JURY

M. Christophe ROBAGLIA, Professeur, Université Aix-Marseille II	Président
M. Stéphane BLANC, Directeur de Recherche, INRA	Rapporteur
M. Thierry LANGIN, Directeur de Recherche, CNRS	Rapporteur
Mme Nathalie CHANTRET, Chargé de Recherche, INRA	Examineur
M. Miguel ARANDA, Chercheur, CSIC, Espagne	Examineur
Mme Carole CARANTA, Directeur de Recherche, INRA	Directeur de Thèse



Unité de Génétique et Amélioration des Fruits et Légumes,  
Centre INRA Avignon - Montfavet - France



# Sommaire

<b>Production scientifique .....</b>	<b>6</b>
<b>Liste des abréviations et acronymes .....</b>	<b>8</b>
<b>Introduction générale .....</b>	<b>10</b>

## Chapitre I. Synthèse bibliographique

<b>I.1. Polymorphisme de séquence et détection de traces de sélection .....</b>	<b>13</b>
I.1.1. Définitions .....	13
I.1.2. Théorie neutraliste de l'évolution moléculaire de Kimura .....	14
I.1.3. Signatures de sélection au niveau moléculaire .....	15
I.1.4. Méthodes de détection de la sélection naturelle à l'échelle moléculaire .....	16
I.1.5. Détection de sélection positive chez les plantes : exemple de la résistance aux herbicides.....	18
<b>I.2. Evolution moléculaire des gènes dominants de résistance aux bioagresseurs chez les plantes : coévolution entre hôtes et agents pathogènes.....</b>	<b>19</b>
I.2.1. Mécanismes de résistance des gènes <i>R</i> .....	20
I.2.2. Scénarios d'évolution des gènes <i>R</i> : course aux armements <i>versus</i> guerre des tranchées .....	21
I.2.3. Exemples de traces de sélection détectées sur les gènes <i>R</i> de résistance aux agents pathogènes chez les plantes.....	23
I.2.4. Impact du mode de reconnaissance <i>R</i> / <i>Avr</i> sur l'évolution des gènes <i>R</i> .....	25
I.2.5. Conclusion .....	26
<b>I.3. Résistance récessive aux virus à ARN et complexe d'initiation de la traduction.....</b>	<b>27</b>
I.3.1. Variation génétique et évolution des virus à ARN de plante .....	27
I.3.2. Gènes récessifs de résistance aux virus ARN .....	29
I.3.3. La famille multigénique <i>eIF4E</i> .....	30
I.3.3.1. <i>eIF4E</i> et l'initiation de la traduction des ARNm cellulaires .....	30
I.3.3.2. Paralogs et orthologues <i>eIF4E</i> .....	31
I.3.3.3. Polymorphisme de séquence des gènes <i>eIF4E</i> .....	33
I.3.4. Rôle clé des protéines <i>eIF4E</i> dans l'interaction plante-virus à ARN et dans la résistance des plantes	34
I.3.4.1. Interactions entre protéines <i>eIF4E</i> et VPg .....	34
I.3.4.2. Approches génétiques .....	36
I.3.4.3. Utilisation sélective des membres de la famille <i>eIF4E</i> par les potyvirus .....	37
I.3.4.4. La VPg des potyvirus est le déterminant de virulence vis-à-vis des résistances contrôlées par <i>eIF4E</i> .....	38
<b>I.4. Interactions potyvirus-Solanacées et potyvirus-<i>A.thaliana</i> .....</b>	<b>39</b>
I.4.1. Les potyvirus .....	39
I.4.2. Les potyvirus des Solanacées .....	41
I.4.3. Gènes récessifs de résistance aux potyvirus chez les Solanacées .....	42
I.4.4. Le couple <i>A. thaliana</i> -potyvirus : un pathosystème de laboratoire.....	44
<b>I.5. Objectif de la thèse .....</b>	<b>45</b>

<b>Chapitre II. Impact du polymorphisme des gènes <i>eIF4E</i> sur l'interaction <i>Capsicum annuum</i> - PVY et TEV .....</b>	<b>48</b>
--	-----------

<b>II.1. - Article 1 - Natural variation and functional analyses provide evidence for coevolution between plant <i>eIF4E</i> and potyviral VPg.....</b>	<b>50</b>
---	-----------

<b>II.2. Impact du cumul de substitutions en acides aminés dans 2 régions distinctes de la protéine eIF4E1 de <i>C. annuum</i> sur le spectre de résistance aux potyvirus .....</b>	<b>73</b>
II.2.1. Contexte et objectif .....	73
II.2.2. Matériel et méthodes .....	74
II.2.3. Résultats .....	75
II.2.3. Discussion .....	76
<b>II.3. Rôle des autres isoformes eIF4E dans le contournement de la résistance eIF4E1-pvr2 par les potyvirus.....</b>	<b>78</b>
II.3.1. Contexte et objectif .....	78
II.3.2. Résultats .....	80
II.3.3. Discussion .....	80

### **Chapitre III. Analyse comparée du polymorphisme des gènes *eIF4E* chez 2 Solanacées (piment et tomate) et *A. thaliana***

<b>III.1. Introduction.....</b>	<b>84</b>
<b>III.2. Matériel et méthodes.....</b>	<b>85</b>
III.2.1. Matériel végétal .....	85
III.2.2. Amplification et séquençage des gènes <i>eIF4E</i> .....	86
III.2.3. Amplification et séquençage des locus indépendants "non-eIF4E" .....	87
III.2.4. Analyse des données de séquence .....	88
III.2.4.1. Alignement et détection de polymorphisme.....	88
III.2.4.2. Tests statistiques et détection de signatures de sélection.....	88
III.2.5. Test de résistance aux potyvirus .....	89
<b>III.3. Polymorphisme des gènes <i>eIF4E</i> chez <i>Capsicum</i> spp.....</b>	<b>90</b>
III.3.1. Contexte et objectif .....	90
III.3.2. Recherche des pressions de sélection s'exerçant sur le gène <i>Ca-eIF4E1</i> par l'analyse de son polymorphisme nucléotidique.....	92
III.3.2.1. Polymorphisme intraspécifique du gène <i>Ca-eIF4E1</i> .....	92
III.3.2.2. Structure haplotypique de <i>Ca-eIF4E1</i> .....	93
III.3.2.3. Recherche de sites sous sélection positive .....	94
III.3.3. Recherche des pressions de sélection qui s'exercent sur les gènes <i>Ca-eIF4E2</i> et <i>Ca-eIF(iso)4E</i> par l'analyse de leur diversité nucléotidique .....	95
III.3.3.1. Polymorphisme intraspécifique des gènes <i>Ca-eIF4E2</i> et <i>Ca-eIF(iso)4E</i> .....	95
III.3.3.2. Recherche de sites sous sélection positive .....	96
III.3.4. Polymorphisme interspécifique des 3 gènes <i>Ca-eIF4E</i> chez 51 accessions de <i>Capsicum</i> spp. ....	96
III.4.4. Conclusions.....	98
<b>III.4. Polymorphisme des gènes <i>eIF4E</i> chez <i>Lycopersicon</i> spp.....</b>	<b>101</b>
III.4.1. Contexte et objectif .....	101
III.4.2. Analyse du polymorphisme intraspécifique des gènes <i>Ly-eIF4E</i> .....	102
III.4.3. Phylogénie des gènes <i>Ly-eIF4E</i> .....	103
III.4.4. Polymorphisme interspécifique et résistance aux potyvirus chez <i>Lycopersicon</i> spp. ....	104
III.4.5. Recherche de sites sous sélection positive .....	105
III.4.6. Conclusions.....	105
<b>III.5. Polymorphisme des gènes <i>eIF4E</i> chez <i>A. thaliana</i>.....</b>	<b>106</b>
III.5.1. Contexte .....	106
III.5.2. Polymorphisme intraspécifique des gènes <i>At-eIF4E</i> .....	107
III.5.2.1. Pression de sélection purificatrice sur les gènes <i>At-eIF4E1</i> , <i>At-eIF4E3</i> , <i>At-eIF(iso)4E</i> et <i>At-nCBP</i> .....	108
III.5.2.2. Pression de sélection balancée sur le gène <i>At-eIF4E2</i> .....	109
III.5.3. Recherche de sites sous sélection positive .....	110
III.5.4. Substitutions en acides aminés identifiées dans les 5 protéines <i>At-eIF4E</i> .....	110
III.5.5. Recherche de combinaisons d'allèles <i>At-eIF4E</i> de résistance aux potyvirus .....	111
III.5.6. Phylogénie et polymorphisme interspécifique des gènes <i>At-eIF4E</i> .....	113
III.5.7. Conclusions.....	114

## Chapitre IV. Vers la caractérisation du rôle fonctionnel des différents facteurs eIF4E

<b>IV.1. Introduction</b> .....	<b>115</b>
<b>IV. 2. Matériel et méthodes</b> .....	<b>118</b>
IV.2.1. RT-PCR semi-quantitative.....	118
IV.2.2. RT-PCR quantitative .....	118
IV.2.3. Test de complémentation fonctionnelle dans la levure.....	120
<b>IV. 3. Résultats</b> .....	<b>120</b>
IV.3.1. Etude de l'accumulation des transcrits <i>eIF4E</i> chez <i>A. thaliana</i> par RT-PCR semi-quantitative....	120
IV.3.2. Etude de l'accumulation des transcrits <i>eIF4E</i> chez <i>A. thaliana</i> par RT-PCR quantitative.....	121
IV.3.2.1. Gènes témoins et contrôles de PCR .....	121
IV.3.2.1. Comparaison des niveaux d'accumulation des transcrits <i>eIF4E</i> entre organes .....	122
IV.3.2.2. Comparaison des niveaux d'accumulation des 5 transcrits <i>eIF4E</i> .....	122
IV.3.3. Complémentation fonctionnelle dans la levure JO55 .....	123
<b>IV.4. Discussion</b> .....	<b>123</b>

## Chapitre V. Discussion générale et perspectives .....

126

<b>V.1. Rôle de la redondance des gènes <i>eIF4E</i> dans les fonctions cellulaires de la plante et dans l'interaction plante-potyvirus</b> .....	<b>127</b>
V.1.1. A la quête de l'ensemble des membres de la famille multigénique eIF4E .....	127
V.1.2. Redondance <i>versus</i> spécialisation des protéines eIF4E.....	130
V.1.2.1. Arguments en faveur de la redondance fonctionnelle des protéines eIF4E.....	130
V.1.2.2. Arguments en faveur de la spécialisation des protéines eIF4E .....	131
V.1.3. Utilisation sélective des différentes protéines eIF4E par les potyvirus .....	134
<b>V.2. Vers l'élucidation du/des rôle(s) de eIF4E dans l'interaction plante-potyvirus</b> .....	<b>135</b>
V.2.1. Impact de l'infection virale sur l'expression de eIF4E.....	135
V.2.2. Vers l'identification des partenaires de eIF4E au cours de l'infection virale .....	136
<b>V.3. Evolution des gènes <i>eIF4E</i> et résistance aux virus à ARN</b> .....	<b>138</b>
V.3.1. Co-évolution entre le facteur eIF4E1 du piment et la VPg des potyvirus .....	138
V.3.2. Comparaison du profil d'évolution des gènes <i>eIF4E</i> chez différentes familles végétales.....	140
V.3.2.1. Le profil d'évolution du gène <i>eIF4E1</i> n'est pas conservé chez les Solanacées .....	140
V.3.2.2. <i>A. thaliana</i> ne constitue pas un modèle adapté à l'étude de la coévolution plante-potyvirus ..	142
V.3.2.3. Le profil d'évolution des gènes <i>eIF4E</i> n'est pas conservé chez les Légumineuses.....	142
V.3.2.4. Une pression de sélection positive s'exerce sur le gène <i>eIF4E</i> chez la vigne.....	143
V.3.2.5. Le polymorphisme du gène <i>eIF4E</i> chez le blé se distingue du polymorphisme détecté chez l'orge.....	144
V.3.2.6. Vers un modèle d'évolution convergente des gènes <i>eIF4E</i> de résistance .....	145
<b>V.4. Diversité des gènes <i>eIF4E</i> et gestion des résistances aux potyvirus</b> .....	<b>146</b>
V.4.1. Paramètres influençant l'efficacité et de la durabilité des résistances médiées par <i>eIF4E</i> .....	146
V.4.2. Un coût à la résistance contrôlée par les gènes <i>eIF4E</i> ?.....	149

## Références bibliographiques.....

152

### Annexes

Annexe 1 : Conditions de culture des plantes.....	170
Annexe 2 : Conservation des souches virales .....	170
Annexe 3 : Inoculation manuelle des plantes.....	170
Annexe 4 : Détection de l'accumulation virale par DAS-ELISA ou RT-PCR.....	171
Annexe 5 : Bactéries et levures.....	171

Annexe 6 : Extraction des acides nucléiques .....	172
Annexe 7 : Amplification des séquences nucléotidiques.....	173
Annexe 8 : Techniques de clonage .....	174
Annexe 9 : Système de double hybride dans la levure .....	175
Annexe 10 : Western blot .....	176
Annexe 11 : Bases de données et liens internet .....	178
<b>Liste des tableaux et figures.....</b>	<b>180</b>

# Production scientifique

## Publications

Charron, C., Nicolai, M., Gallois, J.L., Robaglia, C., Moury, B., Palloix, A. and Caranta, C. Natural variation and functional analyses provide evidence for coevolution between plant eIF4E and potyviral VPg. *The Plant Journal*, in press.

Charron, C., De Mita, S., Sanchez, P., Moretti, A., Brunel, D., Santoni, S., Moury, B., David, J. and Caranta, C. Comparative analysis of the DNA polymorphism at the *eIF4E* genes: conservation versus positive selection exerted by potyviral infection. *En préparation*.

## Communications orales

Charron, C., Sanchez, P., Moretti, A., Palloix, A., De Mita, S., David, J. et Caranta C. (2007) Contrasted evolution of *eIF4E* genes: conservation in *A. thaliana* versus selection in pepper. 4<sup>ème</sup> Journée de la Biologie Végétale. Montpellier, France, 24 mai 2007.

Charron, C., Sanchez, P., Moretti, A., Palloix, A., De Mita, S., David, J. et Caranta C. (2007) Evolution contrastée des gènes *eIF4E* : conservation chez *A. thaliana* versus sélection chez le piment (exercée par les virus à ARN ?). 11<sup>èmes</sup> Rencontres de Virologie Végétale. Aussois, France, 28 janvier-1er février 2007.

**Prix de la meilleure communication orale des doctorants remis par la Société Française de Phytopathologie.**

Ruffel, S., Gallois, J.L., Charron, C., Moury, B., Palloix, A., Caranta C. (2005) La résistance totale du piment au *Pepper veinal mottle virus* résulte de la combinaison de mutations dans les gènes *eIF4E1* et *eIF(iso)4E*. 10<sup>èmes</sup> Rencontres de Virologie Végétale. Aussois, France, 6-10 mars 2005.

Caranta, C., Charron, C., Nicolai, M., Sanchez, P., Moretti, A., Robaglia, C. and Palloix, A. (2007) Amino acid diversity and functional analysis of eIF4E translation initiation factors in both model and cultivated species: a comparative study. *The 6th symposium on Post-Transcriptional Regulation of Plant Gene Expression*, Carry-le-Rouet, France, 10-13 mai 2007.

Caranta, C., German-Retana, S., Charron, C., Gallois, J.L., Ruffel, S., Sanchez, P., Walter, J., Doublet, B., Tavert Roudet, G., Nicaise, V., Michon, T., Le Gall, O., Mazier, M., Robaglia, C. (2006) NEWVIR - New strategy to control RNA virus infections targeted on host factors required for virus life cycle. *Séminaire Génoplante*, Arles, France, 3-5 avril 2006.

## Posters

Charron, C., Nicolai, M., Gallois, J.L., Ruffel, S., Moury, B., Sanchez, P., Moretti, A. et Caranta, C. (2007) Natural diversity and functional analysis of translation initiation factor eIF4E involved in resistance against RNA viruses. *XIII International Congress on Molecular Plant-Microbe Interactions*, Sorrento, Italia, 21-27 July 2007.

Charron, C., Nicolai, M., Gallois, J.L., Ruffel, S., Moury, B., Sanchez, P., Moretti, A. et Caranta, C. (2006) Caractérisation du polymorphisme des gènes *eIF4E* impliqués dans la résistance aux virus et impact sur l'interaction avec ces virus. *Rencontres du DGAP*, Batz-sur-mer, France, 25-27 septembre 2006.

Charron, C., Ruffel, S., Gallois, J.L., Sanchez, P. et Caranta, C. (2006) Characterization of the polymorphism at the eukaryotic initiation factor *4E* genes: analysis of its impact on plant-potyvirus interactions. *XIV colloque de l'Ecole Doctorale des Sciences de la Vie et de la Santé des Université d'Aix-Marseille*, Marseille, France, 8-9 juin 2006.

# Liste des abréviations et acronymes

## Abréviations

ACCase	Acetyl CoenzymeA Carboxylase
ADN	Acide désoxyribonucléique
ADNc	ADN complémentaire
ARN	Acide ribonucléique
ARNm	ARN messenger
ATP	Adénosine tri-phosphate
BET	Bromure d'éthidium
CC	Coiled-coil
Ct	Cycle threshold
CTAB	Cetyl triméthylammonium bromide
DAS-ELISA	Double Antibody Sandwich Enzyme-Linked Immuno Sorbent Assay
dNTP	désoxyribonucléotide tri-phosphate
DIECA	diéthylthiocarbonate de sodium
EBI	European Bioinformatics Institute
eIF	eukaryotic Initiation Factor
EvoluTIF	Evolutionary dynamics of Translation Initiation Factors in plants
EMS	Ethyl Methyl Sulfonate
EST	Expressed Sequence Tag
GAPDH	Glyceraldéhyde-3-phosphate désydrégénase
GTP	Guanosine tri-phosphate
HKA	Hudson Kreitman Aguade
IRES	Internal Ribosome Entry Site
LB	Luria Bertani
LRR	Leucine Rich Repeat
MOPS	Acide 3-(N-morpholino)propanesulfonique
NBS	Nucleotide Binding Site
NBT-BCIP	Nitro blue tetrazolium chloride - 5-Bromo-4-chloro-3-indolyl phosphate
NCBI	National Center for Biotechnology Information
nCBP	new Cap Binding Protein
OGM	Organisme Génétiquement Modifié
ORF	Open Reading Frame
PABP	Poly(A) Binding Protein
PCR	Polymerase Chain Reaction
PMSF	phenylmethyl-sulfonyl fluoride
pvr	potyvirus resistance
QTL	Quantitative Trait Loci
RT	Reverse Transcription
SDS	Sodium Dodecyl Sulfate
SGN	Sol Genomics Networks
SNP	Single Nucleotide Polymorphism
SSAP	Sequence-Specific Amplification Polymorphisms
TAIR	The Arabidopsis Information Resource

TC	Tentative Consensus
TIGR	The Institute for Genomic Research
TILLING	Targeting Induced Local Lesions IN Genomes
TIR	Toll-Interleukin Receptor
UAS	Upstream Activating Sequences
VPg	Viral Protein genome-linked
YPD	Yeast extract Peptone Dextrose

### Acronymes virus végétaux

<b>Abréviation</b>	<b>Nom scientifique</b>	<b>Genre</b>
ArMV	<i>Arabidopsis mosaic virus</i>	<i>Nepovirus</i>
BaMMV	<i>Barley mild mosaic virus</i>	<i>Bymovirus</i>
BaYMV	<i>Barley yellow mosaic virus</i>	<i>Bymovirus</i>
BMMV	<i>Bean mild mosaic virus</i>	<i>Carmovirus</i>
BYMV	<i>Bean yellow mosaic virus</i>	<i>Potyvirus</i>
BYMV-W	<i>White lupin mosaic virus</i>	<i>Potyvirus</i>
CIYVV	<i>Clover yellow vein virus</i>	<i>Potyvirus</i>
CMV	<i>Cucumber mosaic virus</i>	<i>Cucumovirus</i>
CVMV	<i>Chili veinal mottle virus</i>	<i>Potyvirus</i>
GFLV	<i>Grapevine fanleaf virus</i>	<i>Nepovirus</i>
LMV	<i>Lettuce mosaic virus</i>	<i>Potyvirus</i>
MNSV	<i>Melon necrotic spot virus</i>	<i>Carmovirus</i>
PepMoV	<i>Pepper mottle virus</i>	<i>Potyvirus</i>
PoMV	<i>Poplar mosaic virus</i>	<i>Carlavirus</i>
PPV	<i>Plum pox virus</i>	<i>Potyvirus</i>
PRV	<i>Papaya ringspot virus</i>	<i>Potyvirus</i>
PSbMV	<i>Pea seed-borne mosaic virus</i>	<i>Potyvirus</i>
PTV	<i>Peru tomato virus</i>	<i>Potyvirus</i>
PVMV	<i>Pepper veinal mottle virus</i>	<i>Potyvirus</i>
PVA	<i>Potato virus A</i>	<i>Potyvirus</i>
PVV	<i>Potato virus V</i>	<i>Potyvirus</i>
PVX	<i>Potato virus X</i>	<i>Potyvirus</i>
PVY	<i>Potato virus Y</i>	<i>Potyvirus</i>
RYMV	<i>Rice yellow mosaic virus</i>	<i>Sobemovirus</i>
TEV	<i>Tobacco etch virus</i>	<i>Potyvirus</i>
TMV	<i>Tobacco mosaic virus</i>	<i>Tobamovirus</i>
ToRSV	<i>Tomato ringspot virus</i>	<i>Nepovirus</i>
TuMV	<i>Turnip mosaic virus</i>	<i>Potyvirus</i>
TVMV	<i>Tobacco vein mottling virus</i>	<i>Potyvirus</i>
WMV	<i>Watermelon mosaic virus</i>	<i>Potyvirus</i>
WSMV	<i>Wheat streak mosaic virus</i>	<i>Tritimovirus</i>
WSSMV	<i>Wheat spindle streak mosaic virus</i>	<i>Bymovirus</i>
ZYMV	<i>Zucchini yellow mosaic virus</i>	<i>Potyvirus</i>

# Introduction générale

Les plantes cultivées sont soumises à l'infection par de nombreux bioagresseurs (virus, bactéries, champignons, insectes, nématodes) qui les affaiblissent et compromettent les récoltes en qualité et/ou en quantité, parfois dans des proportions considérables. Les traitements phytosanitaires répétés posent des problèmes de sécurité alimentaire et de préservation de l'environnement. La lutte génétique contre les agents pathogènes, basée sur l'utilisation de gènes de résistance présents dans la diversité naturelle des espèces, présente l'avantage d'être simple à mettre à œuvre, peu coûteuse pour l'agriculteur et respectueuse de l'environnement. La lutte chimique contre les virus est indirecte car elle cible les vecteurs de transmission de ces agents pathogènes (pucerons, insectes...). Elle est d'autant plus inefficace dans le cas des virus transmis très rapidement selon le mode nonpersistant (c'est-à-dire courte durée de rétention et non circulant dans le vecteur) : aucun insecticide n'est actuellement capable de tuer un puceron en un laps de temps assez court (quelques dizaines de secondes) pour éviter la transmission de ces virus. Pour lutter contre les maladies virales, des gènes de résistance identifiés dans la diversité naturelle des espèces peuvent être introduits par croisements dans des variétés performantes au niveau agronomique. La diversité des résistances disponibles dans la variabilité naturelle est loin d'être exploitée au maximum ; ainsi, de nombreux travaux de recherche ont pour objectif la caractérisation de nouveaux systèmes génétiques contrôlant une résistance efficace et durable. Déjà utilisée avec succès, la lutte génétique trouve cependant ses limites dans l'adaptation progressive des pathogènes aux résistances variétales lors de la mise en culture à grande échelle.

Dans ce contexte, de nombreux travaux de recherche conduits ces 20 dernières années ont permis d'avancer dans la compréhension des mécanismes d'interaction entre les agents pathogènes et leurs hôtes et de caractériser au niveau moléculaire les gènes de résistance impliqués. Le clonage et l'analyse fonctionnelle de nombreux gènes de résistance ont apporté des informations sur les différents mécanismes de résistance (McDowell and Woffenden, 2003; Maule *et al.*, 2007).

En parallèle à ces travaux de caractérisation fonctionnelle, des analyses de l'évolution moléculaire des gènes de résistance aux agents pathogènes ont été mises en oeuvre afin de caractériser les processus d'évolution à l'origine de la variation nucléotidique de ces gènes. On peut ainsi déterminer si le gène subit une pression de sélection de type coévolution caractéristique des relations hôtes/parasites (Tiffin and Moeller, 2006). Bien que peu d'études regroupent sur un même pathosystème des analyses fonctionnelles des interactions facteurs de résistance de l'hôte-facteurs de virulence du pathogène et des analyses d'évolution, il apparaît pertinent de coupler ces approches. En effet, comprendre les raisons du maintien de la diversité allélique dans la variabilité naturelle sera crucial pour identifier les conditions dans lesquelles la résistance est efficace et durable en conditions de culture (de Meaux and Mitchell-Olds, 2003).

## **Chapitre I**

# **S**ynthèse bibliographique

## **I.1. Polymorphisme de séquence et détection de traces de sélection**

Analyser les profils de polymorphisme de séquence et la divergence d'un gène donné entre espèces permet de caractériser les forces évolutives agissant sur ce gène. En effet, différentes formes de sélection produisent des profils spécifiques de diversité de séquence (Charlesworth *et al.*, 2001). Afin de bien comprendre quelles sont les pressions de sélection susceptibles de s'exercer sur les gènes de résistance aux bio-agresseurs, les grands principes d'analyse du polymorphisme en vue de la détection de traces de sélection seront décrits dans une première partie de ce chapitre.

### **I.1.1. Définitions**

La variabilité au sein des populations est générée par 3 types d'évènements aléatoires : des recombinaisons, des duplications de gènes et des mutations. La recombinaison génétique consiste en un véritable réarrangement du génome puisque de l'ADN est échangé soit entre chromosomes homologues soit entre des chromosomes non-homologues *via* des crossing-over (enjambements) au cours de la méiose. Ce brassage génétique conduit donc à l'apparition de nouvelles combinaisons de gènes qui seront ensuite soumises à la sélection naturelle. De récentes études se sont intéressées aux points chauds de recombinaison chez les plantes, avec notamment l'identification de taux de recombinaison 70 fois plus élevés dans des régions hautement recombino-gènes par rapport au reste du génome chez *A. thaliana* (Li *et al.*, 2007). Les évènements de duplication sont à l'origine de l'accroissement de la taille des génomes, des familles de gènes et des locus de répétitions en tandem. Malgré cela, le rôle et les caractéristiques évolutives, les raisons de leur existence et les mécanismes qui contrôlent leur développement sont encore peu connus. D'après les premières données issues du séquençage des génomes entiers d'*A. thaliana*, du riz et du peuplier, il semblerait que de nombreux gènes aient été dupliqués puis maintenus par la sélection naturelle au cours de l'évolution des angiospermes (Sterck *et al.*, 2007). La mutation est une erreur qui survient lors de la réplication du matériel génétique. Par cycle de réplication, l'ADN polymérase fait environ une erreur tous les  $10^9$  nucléotides, taux assez constant chez l'ensemble des organismes, des bactéries aux eucaryotes supérieurs. Lorsqu'une mutation apparaît dans la séquence codante d'un gène, elle va générer un nouvel allèle. Les mutations touchant les séquences codantes

sont de 2 types : synonymes (ou silencieuses), si elles ne modifient pas la séquence en acides aminés, ou non-synonymes (ou codantes) dans le cas contraire. Deux types de forces évolutives peuvent s'exercer sur une population constituée des individus possédant ce nouvel allèle et l'ancien : la dérive génétique, aléatoire, et la sélection, orientée. Ces forces évolutives font varier les fréquences alléliques au sein de la population. Une fluctuation aléatoire de la fréquence des gènes au fil des générations est appelée dérive génétique. Les populations naturelles ont rarement des effectifs constants. Ainsi, la dérive va jouer surtout dans les périodes où les effectifs sont faibles. Si à une époque de son histoire, une population est passée pendant plusieurs générations par un effectif très faible, elle a pu perdre par dérive pendant cette période de "goulot démographique" une grande part de sa diversité allélique. La sélection naturelle engendre l'accumulation de génotypes ayant un fort pouvoir adaptatif (ou fitness). La sélection permet donc d'adapter une population à son milieu en augmentant ou en maintenant dans le patrimoine génétique de cette population la fréquence des allèles favorables. La sélection naturelle influence de 3 manières le devenir des mutations codantes : on parle de mutations avantageuses (ou adaptatives) si le nouvel allèle procure un avantage sélectif, de mutations délétères dans le cas inverse (éliminées par la sélection) ou de mutations neutres si le nouvel allèle ne procure ni avantage ni désavantage adaptatif.

### **I.1.2. Théorie neutraliste de l'évolution moléculaire de Kimura**

Le modèle de Motoo Kimura (Kimura, 1983), maintenant connu sous le nom de "théorie neutraliste de l'évolution moléculaire", propose que la plus grande partie de la variation génétique (à la fois le polymorphisme intra-espèce et la divergence entre espèces) est neutre et que les polymorphismes sont éliminés ou fixés dans les populations sous l'influence des effets stochastiques de la dérive génétique. Pour mettre en évidence des mécanismes de sélection au niveau moléculaire, il est donc important de définir auparavant ce que l'on attend en absence de sélection (modèle nul). Si les observations ne peuvent être expliquées sous un modèle d'évolution neutre, alors seulement on pourra invoquer l'action de la sélection.

### I.1.3. Signatures de sélection au niveau moléculaire

A l'inverse des facteurs démographiques (goulots d'étranglement, expansions démographiques, mélanges de populations) qui affectent la fréquence allélique de tous les gènes d'un individu de manière similaire, la sélection naturelle n'affecte que certains gènes. Mais comment détecter cette sélection naturelle ? Une approche directe serait de suivre expérimentalement une population au cours du temps, mais ce mode de détection nécessite l'étude de populations de taille très importante sur de très longues périodes de temps. Pour s'affranchir de telles contraintes, une approche alternative indirecte consiste à détecter les signatures moléculaires de la sélection, en comparant les données observées (distribution des fréquences alléliques, diversité nucléotidique, taux de mutations synonymes et non synonymes) aux données théoriques attendues dans le cas d'absence de sélection (modèle nul neutraliste).

Lorsqu'une mutation avantageuse apparaît dans un génome, elle peut être fixée et donc remplacer toute autre variation présente à ce locus. La région où la mutation est apparue se retrouve alors dépourvue des mutations neutres qui s'accumulent normalement aux sites silencieux de l'ADN. Ce phénomène est appelé balayage sélectif ou « selective sweep ». Les régions voisines du chromosome se retrouvent également dépourvues d'une part de leur variation neutre comme simple conséquence de leur liaison génétique; on parle alors d'effet d'auto-stop ou « genetic hitchhiking ». A plus longue distance, l'effet de balayage s'estompe progressivement, jusqu'à ne plus être repérable. Dans ce cas, la signature moléculaire de la sélection naturelle est donc un "trou" de variation neutre dans le génome. L'amplitude de l'effet d'auto-stop augmente avec la force de la sélection et décroît avec celle du taux de recombinaison, qui tend à découpler l'association entre régions voisines. Si l'on connaît le taux de recombinaison, on peut en déduire l'intensité de la sélection. Le retour progressif de la variation neutre nous renseigne sur l'âge de l'événement. En effet, le balayage sélectif est surtout sensible juste après la survenue de l'événement sélectif. Ensuite, les mutations recommencent à s'accumuler sur le chromosome et restaurent progressivement la variabilité. La sélection positive, également appelée diversifiante ou adaptative, peut être repérée *via* l'identification de balayages sélectifs. Ce type de sélection fait augmenter la fréquence d'un allèle avantageux, contrairement à la sélection négative, ou purificatrice, qui tend à diminuer la fréquence d'un allèle désavantageux *via* l'élimination des mutations délétères. Mais la sélection naturelle ne fait pas toujours augmenter ou diminuer la fréquence d'un seul allèle à

un locus. Quelquefois, elle tend à maintenir 2 allèles ou plus à un locus dans une population. Ceci est connu sous le nom de sélection balancée, ou sélection équilibrante. La sélection balancée peut maintenir un excès d'allèles à des fréquences intermédiaires pendant de très longues périodes de temps. La signature moléculaire de la sélection balancée n'est détectable que quelques centaines de paires de bases autour du site sélectionné et pas au delà. Deux situations biologiques peuvent expliquer le maintien de la sélection balancée : (i) lorsque les hétérozygotes ont une meilleure fitness que les homozygotes, donc le maintien des deux allèles est un avantage et (ii) lorsque l'allèle donne un avantage quand il est rare, on parle alors de sélection fréquence-dépendante. Chez les plantes et les animaux, la sélection balancée semble impliquée dans le maintien de la diversité aux locus qui coordonnent la reconnaissance entre le soi et le non-soi (Richman and Kohn, 1999). Le tableau I.1 récapitule les sélections naturelles qui peuvent être observées et leurs effets sur la généalogie des gènes et les fréquences des allèles d'une population de 12 individus haploïdes.

#### **I.1.4. Méthodes de détection de la sélection naturelle à l'échelle moléculaire**

Toutes les méthodes présentées dans cette partie visent à examiner l'adéquation d'un jeu de données de séquences au modèle d'évolution neutre. La quasi-totalité des tests statistiques utilisés ont été construits sous l'hypothèse que les populations étudiées sont conformes au modèle nul, c'est-à-dire que la distribution observée des fréquences n'a subi au fil des générations aucune force sélective ou autre facteur environnemental pouvant affecter cette distribution (populations de taille constante à l'équilibre mutation-dérive, appelées populations de Wright-Fisher). Un certain nombre d'auteurs a développé des méthodes permettant de mesurer la neutralité des populations. L'hypothèse du modèle à nombre infini de sites s'applique lorsque le taux de mutation est suffisamment faible par rapport à la taille de la séquence considérée pour que les mutations apparaissent de façon unique (pas de mutation récurrente). Sous cette hypothèse, la diversité nucléotidique d'une population ( $\theta$ ) est proportionnelle à la taille efficace de la population ( $N_e$ , nombre d'individus d'une population idéale pour lequel on aurait un degré de dérive génétique équivalent à celui de la population réelle) et au taux de mutation par gène et par génération ( $\mu$ ) :  $\theta = 4N_e\mu$  (pour des individus diploïdes). Les 2 estimateurs de cette diversité  $\theta$  (valeurs calculées sur un échantillon et que l'on espère très proches des valeurs que l'on aurait calculées sur la population totale) les plus fréquemment utilisés sont  $\pi$  (Nei and Li, 1979) et  $\theta_w$  (Watterson, 1975). La diversité

nucléotidique  $\pi$  est définie comme le nombre moyen de différences nucléotidiques par base entre 2 séquences prises au hasard dans un échantillon. L'indice  $\theta_w$  est le nombre de sites en ségrégation par base, corrigé par un facteur basé sur la taille de l'échantillon. Ces 2 estimateurs diffèrent en ce que seul  $\pi$  prend en compte la fréquence des allèles. La définition de ces indices est récapitulée dans le tableau I.2.

A partir de ces estimateurs, Tajima (Tajima, 1983) a développé un des premiers tests de neutralité sélective basé sur la diversité nucléotidique. L'ensemble des tests de neutralité utilisés couramment pour détecter de la sélection à partir des séquences d'ADN génomique est répertorié dans le tableau I.3. Le calcul de l'indice  $D = \pi - \theta_w$  permet de tester l'hypothèse nulle. Sous cette hypothèse nulle,  $\pi = \theta_w$  donc  $D = 0$ . Si  $D$  est différent de 0, il y a rejet de l'hypothèse nulle (Tajima, 1989). Une valeur de  $D$  négative peut être la conséquence soit d'une fixation récente d'un allèle au locus considéré (sélection positive ou négative) soit d'une expansion démographique. A l'inverse, une valeur de  $D$  positive indiquera soit une sélection balancée soit une réduction de la taille de la population. Pour distinguer les effets de la sélection des effets démographiques, il est nécessaire de séquencer plusieurs gènes indépendants chez une même espèce, puisque les facteurs démographiques touchent l'ensemble des gènes d'un génome et non pas un seul locus comme le fait la sélection. Les tests de Fu et Li (1993) comparent aussi ces estimateurs mais prennent en compte les variants alléliques présents une seule fois (la mutation est alors appelée singleton). Un excès de singletons caractérise par exemple un balayage sélectif récent. Le test H de Fay et Wu (2000) est une généralisation du test de Fu and Li mais diffère par l'importance accordée aux fréquences alléliques intermédiaires.

D'autres tests sont basés sur l'hétérogénéité entre plusieurs locus chez plusieurs espèces. Le test HKA développé par Hudson, Kreitmann et Aguadé (Hudson *et al.*, 1987) permet de comparer les niveaux de variation intra- et inter-spécifique pour le locus d'intérêt et pour un autre locus supposé neutre. Sous l'hypothèse nulle, les locus les plus variables au sein d'une espèce sont ceux qui montrent le plus de variation entre espèces (la diversité est générée par la mutation uniquement). Le test de McDonald et Kreitman (1991) compare polymorphisme/divergence des sites synonymes et non-synonymes (les synonymes étant pris comme référence neutre puisqu'ils ne conduisent pas à des changements en acides aminés); sous l'hypothèse de neutralité, les 2 ratios devraient être égaux. Enfin, on peut comparer les

taux de substitutions : l'indice  $\omega$  correspondant au rapport du taux de substitutions non synonymes (dN) sur le taux de substitutions synonymes (dS) indique le régime de sélection auquel est soumise la protéine.  $\omega$  est égal à 1 en absence de sélection. Si  $\omega$  est inférieur à 1, cela signifie que les changements en acides aminés sont délétères (sélection négative). Dans le cas d'un  $\omega$  supérieur à 1, les substitutions non synonymes ont été plus fréquentes dans l'histoire de l'échantillon, indiquant la possibilité d'une sélection positive. Il est très intéressant de comparer les résultats de ces différents tests pour essayer de mettre en évidence non seulement un écart à la neutralité, mais également proposer un modèle de sélection.

### **I.1.5. Détection de sélection positive chez les plantes : exemple de la résistance aux herbicides**

Lorsque l'on étudie l'adaptation des plantes avec les outils moléculaires de génétique des populations, 2 principaux types de sélection naturelle sont mis en évidence : la sélection balancée décrite précédemment qui permet le maintien sélectif à long terme de plusieurs allèles et la sélection positive où il y a fixation d'une mutation favorable (Wright and Gaut, 2005). Un exemple remarquable de sélection positive est celui des gènes de résistance à un stress abiotique, les herbicides. En effet, la résistance aux herbicides est un caractère de choix pour étudier les pressions de sélection qui s'exercent sur les gènes de résistance dans les populations naturelles de plantes annuelles : la sélection est forte et l'adaptation est rapide (Delye *et al.*, 2004). La plupart des herbicides agissent en inhibant des enzymes impliquées dans les processus métaboliques vitaux chez les mauvaises herbes. En réponse à ce nouveau stress, des mutations dans les enzymes cibles des herbicides chez la plante sont apparues. Certains herbicides ciblent le domaine carboxyl-transférase (CT) de l'Acetyl CoenzymeA Carboxylase (ACCCase) chloroplastique; ils bloquent la synthèse des acides gras et entraînent la mort de la plante. Récemment, des allèles mutants d'ACCCase avec une sensibilité réduite aux herbicides ont été identifiés dans des populations naturelles d'*Alopecurus myosuroides* (mauvaise herbe des céréales de l'Europe du Nord-Ouest), suggérant que, dans les champs, une forte pression de sélection peut conduire à une adaptation rapide de ces populations. Parmi les 13 substitutions non-synonymes détectées, 6 d'entre elles, connues pour être impliquées dans la résistance aux herbicides, se regroupent dans une région de 70 acides aminés dans le domaine CT de l'ACCCase (Delye *et al.*, 2004); ces mutations sont maintenues

par sélection positive. A l'inverse, le reste de la séquence de l'ACCAsé est soumise à de la sélection purificatrice, ce qui est cohérent avec la nécessité de maintenir intacte la fonction de la protéine ACCase, une enzyme métabolique vitale pour la plante. L'avantage dans cette étude est l'existence d'une population d'*A. myosuroides* collectée avant la mise en place de ce type de traitement herbicide. Cette population peut donc être considérée comme représentative de la diversité allélique d'ACCAsé "pré-herbicide", donc "pré-sélection". Grâce à cette population, il a été montré que la sélection positive qui cible les quelques mutations codantes impliquées dans la résistance aux herbicides apparaît environ 15 générations après le début des traitements (Menchari *et al.*, 2006). L'utilisation des herbicides, bien que très récente, a donc clairement un effet visible sur la variation génétique du gène codant pour l'ACCAsé. La réponse adaptative à une pression de sélection drastique est survenue en très peu de générations.

## **I.2. Evolution moléculaire des gènes dominants de résistance aux bioagresseurs chez les plantes : coévolution entre hôtes et agents pathogènes**

Une adaptation diversificatrice au niveau moléculaire, telle que décrite précédemment dans le cas de la résistance aux herbicides chez les mauvaises herbes, a également été retrouvée pour les gènes de résistance aux agents pathogènes (Bergelson *et al.*, 2001). Les interactions entre organismes vivants représentent un facteur écologique majeur entraînant une évolution non-neutre. Le processus de coévolution entre les plantes et leurs bioagresseurs - virus, bactéries, nématodes, champignons, insectes - serait, d'après de nombreux biologistes, à l'origine de la plupart de la diversité naturelle observée sur Terre (Rausher, 2001). Ainsi, des pressions de sélection fortes sur les gènes impliqués dans les mécanismes de résistance aux pathogènes chez les plantes sont attendues car l'histoire évolutive de ces gènes est souvent modelée par la nécessité soit de reconnaître les pathogènes soit d'échapper à la reconnaissance, ce qui induit des profils d'évolution rapide. Ce paragraphe est consacré aux connaissances acquises sur l'évolution des gènes dominants de résistance aux agents pathogènes de type R chez les plantes.

### **I.2.1. Mécanismes de résistance des gènes *R***

La majorité des résistances aux agents pathogènes caractérisés chez les plantes est contrôlée par des gènes dominants. Ce type de résistance implique des mécanismes de reconnaissance du pathogène conduisant à l'induction d'un vaste ensemble de réactions de défense. Le concept de Flor (1971), appelé "modèle gène à gène", permet d'expliquer la grande majorité des résistances monogéniques dominantes et a été confirmé au niveau moléculaire : les mécanismes de résistance sont induits si et seulement si la protéine codée chez la plante par un gène de résistance *R* "reconnaît" la protéine codée chez l'agent pathogène par un gène dominant dit "d'avirulence" (gène *Avr*). De nombreux gènes *R* ont été caractérisés au niveau moléculaire chez différentes espèces végétales, telles que les céréales (riz, orge, maïs), les dicotylédones cultivées (lin, tomate, tabac, laitue) ou l'espèce modèle *A. thaliana*. Ces gènes contrôlent une résistance vis-à-vis de bioagresseurs aussi différents que des bactéries, des virus, des champignons, des nématodes et même des insectes (McDowell and Woffenden, 2003). La plupart des protéines codées par les gènes *R* possèdent 2 domaines caractéristiques : un domaine LRR (pour Leucine-Rich Repeats) qui apparaît être impliqué de façon centrale dans la reconnaissance de l'agresseur (Jones, 2001) et un domaine NBS (pour Nucleotide Binding Site) plutôt impliqué dans la transduction du signal de perception (Dangl and Jones, 2001; Holub, 2001) (Figure I.1). Les domaines LRR et NBS contiennent des motifs très conservés qui ont permis d'identifier et de classer les gènes *R* (Meyers *et al.*, 1999). D'autres domaines sont retrouvés chez les protéines *R* (Dangl and Jones, 2001) : (i) le domaine TIR (Toll-Interleukin Receptor) qui, chez les animaux, se retrouve dans les protéines impliquées dans l'immunité innée ou qui participent à la transmission d'un signal permettant l'activation de la mort cellulaire programmée, (ii) le domaine CC (Coiled-Coil) est connu pour permettre les interactions protéine-protéine et (iii) les domaines transmembranaires, où dans ce cas le domaine LRR se retrouve à l'extérieur de la cellule, architecture qui constituerait un système de surveillance externe des agresseurs.

L'analyse des génomes végétaux révèle que les gènes *R* sont fréquemment regroupés sous forme de grands clusters (Young, 2000) et représentent une part substantielle du contenu en gènes du génome. Chez l'accession Columbia (Col-0) d'*A. thaliana*, les gènes de résistance représentent de l'ordre de 0,5 à 1% des gènes de la plante et la majorité des 149 gènes identifiés codant des protéines de résistance de type NBS-LRR sont regroupés en clusters de 2 à 9 locus (Arabidopsis Genome Initiative, 2000; Meyers *et al.*, 2003). Les analyses

phylogénétiques montrent que les clusters sont le résultat à la fois de duplications de segments anciennes et de réarrangements de gènes récents (recombinaisons inter- et intra-géniques) (Richly *et al.*, 2002). Les travaux menés sur le génome du riz confirment les données obtenues chez *A. thaliana* : 535 séquences codant pour des gènes de type NBS ont été identifiées, 76% de ces gènes sont regroupés en clusters et 16 duplications de gènes ont été détectées (Zhou *et al.*, 2004).

Il existe 2 modes de reconnaissance entre les protéines R et Avr : (i) une détection directe, on parle de modèle "éliciteur-récepteur" ou (ii) une détection indirecte dans laquelle la protéine R n'intervient que comme "senseur" d'une interaction entre la protéine Avr et une autre protéine de la plante; dans ce dernier cas, on parle de modèle "de garde" (Figure I.2) (McDowell and Woffenden, 2003). L'interaction physique directe entre les 2 partenaires n'a pu être démontrée que dans un nombre extrêmement limité de cas, comme par exemple celui du couple AvrPto de *Pseudomonas syringae* / Pto de tomate (Scofield *et al.*, 1996) ou tout récemment AvrL567 de la rouille du lin / L du lin (Dodds *et al.*, 2006). Même si la plupart des résistances dominantes caractérisées au niveau moléculaire confèrent la résistance à des bactéries ou à des champignons, il existe à ce jour une douzaine de gènes dominants impliqués dans la résistance aux virus identifiés chez les plantes cultivées (pomme de terre, tomate, tabac, soja et pois) et *A. thaliana* (Maule *et al.*, 2007). Dix de ces gènes appartiennent à la classe des NBS-LRR. Aucune protéine codée par ces gènes de résistance ne possède de domaine transmembranaire, ce qui est cohérent avec le cycle de vie intracellulaire des virus et la présence des protéines Avr exclusivement dans les cellules de la plante hôte. Une interaction directe entre le produit du gène de résistance *N* chez le tabac et la réplicase p50 du *Tobacco mosaic virus* (TMV) a été identifiée (Ueda *et al.*, 2006), mais il semble qu'une interaction indirecte entre ces 2 protéines, *via* l'intervention d'un troisième partenaire, NIP-1, puisse également avoir lieu (Maule *et al.*, 2007).

### **I.2.2. Scénarios d'évolution des gènes R : course aux armements *versus* guerre des tranchées**

Les gènes de résistance chez les plantes impliqués dans les interactions gène pour gène avec les agents pathogènes sont attendus pour être soumis à des compétitions de type

coévolution (Figure I.3). La coévolution antagoniste entre une plante et son agent pathogène est un processus en plusieurs étapes : (i) l'agent pathogène attaque et exploite la plante, (ii) l'attaque de l'agent pathogène réduit la fitness de la plante, (iii) sélection d'un système de résistance qui peut se répandre dans l'ensemble de la population de la plante, (iv) diminution de la fitness de l'agent pathogène dans les plantes ayant acquis ce système de résistance et (v) sélection d'un génotype de l'agent pathogène capable de contourner la résistance qui va ensuite se répandre dans la population des agents pathogènes. Une interprétation moléculaire à cette coévolution a récemment été proposée par Jones et Dangl (2006) ; il s'agit du « modèle en zig-zag » : (i) les PAMPs (Pathogen-associated molecular patterns, c'est-à-dire des motifs caractéristiques des pathogènes, comme la flagelline de certaines bactéries) sont reconnus par des récepteurs PRRs (Pattern recognition receptors) chez la plante, ce qui entraîne une première immunité appelée PTI (PAMP-Triggered Immunity), (ii) les pathogènes déploient des effecteurs de virulence spécifiques (ETS, Effector Triggered Susceptibility) pour contrer cette première barrière immunitaire et permettre la maladie en inhibant la PTI, (iii) des protéines codées par des gènes de résistance chez la plante reconnaissent ces facteurs de virulence (ETI, Effector Triggered Immunity), ce qui induit une réponse d'hypersensibilité (HR) au site d'infection et enfin (iv) des isolats virulents de l'agent pathogène surmontent l'ETI en développant de nouveaux effecteurs. Dans ce système, il existe une véritable "interconnexion" entre le pathogène et la plante hôte, illustrant l'existence de coévolutions à tous les niveaux de l'infection (Jones and Dangl, 2006).

Des analyses d'évolution moléculaire ont permis de caractériser les processus d'évolution qui façonnent la variation nucléotidique des gènes *R* impliqués dans la résistance des plantes aux agents pathogènes. Bien que le nombre de ces études reste encore limité, 3 types de scénarios évolutifs ont été décrits : la course aux armements, la guerre des tranchées et le modèle mixte.

Le mode de coévolution appelé "course aux armements" (Dawkins and Krebs, 1979) ou "polymorphisme transitoire" (Holub, 2001) entraîne une série de balayages sélectifs : les anciens allèles *R* sont successivement conduits à l'extinction dans une population suite à l'émergence de nouvelles formes de virulence du pathogène. Ces anciens allèles sont remplacés à chaque cycle par un nouvel allèle qui a soit divergé à partir d'un allèle existant par mutation ou recombinaison, soit qui vient d'une autre population (Figure I.4.a). Le modèle "course aux armements" prédit que la variation pour la résistance est transitoire et que les

populations hôtes sont généralement monomorphes au locus du gène de résistance. Les allèles de résistance sont jeunes et leur séquence est quasi-identique aux allèles de sensibilité (Bergelson *et al.*, 2001). Et pourtant, certaines études de gènes *R* chez *A. thaliana* et *Lycopersicon* spp. montrent un polymorphisme considérable à certains locus *R* impliqués dans la reconnaissance des agents pathogènes (Kunkel, 1996) et révèlent que ces gènes *R* existaient avant la différenciation des espèces possédant ces gènes (Riely and Martin, 2001; Stahl *et al.*, 1999). Une autre hypothèse sur l'évolution des gènes *R* a donc été avancée : il s'agit de la théorie du polymorphisme balancé (Hudson and Kaplan, 1988) qui prédit que les allèles de résistance et de sensibilité qui coexistent dans les populations sont anciens et qu'ils diffèrent fortement dans leur séquence ADN, spécialement à proximité du site sous sélection. Cette sélection balancée peut maintenir les couples Avr/R sur des périodes de temps infinies, d'où l'utilisation de l'expression "guerre des tranchées" (Stahl *et al.*, 1999) ou "polymorphisme de recyclage" (Holub, 2001) pour décrire ce type de sélection (Figure I.4.b). Cependant, il est important de remarquer que le modèle "guerre des tranchées" ne permet pas d'expliquer l'apparition de nouveaux allèles de résistance. Les 2 modèles "course aux armements" et "guerre des tranchées" ne sont donc pas exclusifs, et un modèle "mixte" d'évolution des gènes *R* a été proposé. Ce nouveau modèle propose que les gènes *R* contournés sont soit éliminés soit diminuent en fréquence pour réapparaître au moment opportun (Figure I.4.c). Il est donc difficile de formuler des modèles généraux d'évolution des gènes *R*.

### **I.2.3. Exemples de traces de sélection détectées sur les gènes *R* de résistance aux agents pathogènes chez les plantes**

Des exemples illustrant la diversité des chemins évolutifs suivis par les gènes *R* de résistance sont récapitulés dans le tableau I.4.

Chez *A. thaliana*, les analyses moléculaires de génétique des populations montrent que la sélection balancée est la force prédominante qui agit pour maintenir à la fois les allèles de résistance et de sensibilité. Les locus *Rpp1* (Botella *et al.*, 1998), *Rpp5* (Noel *et al.*, 1999), *Rpp8* (McDowell *et al.*, 1998) et *Rpp13* (Ding *et al.*, 2007; Rose *et al.*, 2004) pour la résistance à l'oomycète *Peronospora parasitica* et *Rpm1* (Shen *et al.*, 2006; Stahl *et al.*, 1999), *Rps2* (Caicedo *et al.*, 1999; Mauricio *et al.*, 2003) et *Rps5* (Shen *et al.*, 2006; Tian *et*

*al.*, 2002) impliqués dans la résistance à la bactérie *Pseudomonas syringae* sont sous sélection balancée. Ces locus partagent différentes signatures de variation moléculaire entre les allèles de résistance et de sensibilité, avec des allèles hautement divergents à des fréquences intermédiaires, un niveau élevé de polymorphisme silencieux et/ou un excès relatif de substitutions non-synonymes ( $dN/dS > 1$ ) (Bakker *et al.*, 2006). Plusieurs auteurs ont attribué le maintien des polymorphismes observés à de la sélection fréquence dépendante. En effet, dans le cas des interactions hôtes-parasites, la fréquence de l'allèle de résistance dépend de la fréquence de l'allèle d'avirulence, donc de la fréquence de certains génotypes dans la population du pathogène et *vice versa* (Ellis *et al.*, 1999; Stahl *et al.*, 1999; Tian *et al.*, 2003).

De manière inattendue, il n'existe qu'un seul exemple de gène de résistance chez *A. thaliana* qui présente les caractéristiques d'un balayage sélectif récent typique d'une "course aux armements"; il s'agit du locus *Rps4*, pour lequel les allèles de résistance et de sensibilité sont quasiment identiques au niveau nucléotidique, signe d'un ancêtre commun récent (Bergelson *et al.*, 2001). Cependant, le fait que la sélection balancée soit prédominante par rapport à la sélection directionnelle dans l'évolution des gènes *R* est discutable car les gènes de résistance étudiés à ce jour ont été initialement identifiés par des études de cartographie de locus ségrégant pour des allèles de résistance et sensibilité. Ainsi, les gènes *R* avec peu de variation génétique n'ont pas été identifiés. Mais, on peut penser qu'une faible variation génétique associée à une faible variation phénotypique sera moins soumise à la sélection.

La résistance à *P. syringae* chez la tomate est sous le contrôle du gène *R* de résistance *Pto*. Les analyses de polymorphisme intraspécifique, chez des populations naturelles de plusieurs espèces de *Lycopersicon*, montrent que *Pto* n'évolue pas de la même façon que les autres locus "non-R" analysés chez ces espèces. Le niveau de substitutions non-synonymes est beaucoup plus élevé pour *Pto* que pour les autres gènes analysés. Cependant, la distribution de la variation n'est pas en accord avec les prédictions d'un polymorphisme balancé franc comme ce qui a été rapporté pour les autres locus de résistance (*Rpp1*, *Rpp5*, *Rpp8*, *Rpp13*, *Rpm1*, *Rps2* et *Rps5* chez *A. thaliana*) : l'évolution de *Pto* semble être influencée à la fois par de la sélection purificatrice et de la sélection balancée (Rose *et al.*, 2007). Un scénario proposé par l'auteur pour expliquer ce mélange de sélection purificatrice et de maintien du polymorphisme est le suivant : des périodes pendant lesquelles une pression de sélection forte du pathogène s'exerce sur *Pto* alternent avec des périodes où les populations de pathogènes sont constituées de génotypes ne possédant pas les gènes *Avr* adéquats, relâchant ainsi la

contrainte sur *Pto* et favorisant l'accumulation des substitutions neutres ou délétères. Chaque tour de "sélection relâchée" pourrait créer de nouveaux allèles. La ré-invasion de pathogènes exprimant des protéines Avr fonctionnelles favorise les individus avec des protéines Pto fonctionnelles, expliquant ainsi les signatures paradoxales de sélection purificatrice et de polymorphisme balancé. De plus, la petite protéine Pto est capable d'interagir avec au moins 2 protéines Avr différentes. Cette reconnaissance de plusieurs protéines par Pto pourrait expliquer la nécessité du maintien de la différenciation en acides aminés des allèles Pto afin de reconnaître différentes molécules effectrices du pathogène.

Le locus *L* chez le lin *Linum usitatissimum*, très polymorphe avec au moins 12 variants alléliques connus, code pour des protéines de type TIR-NBS-LRR qui interagissent directement et spécifiquement avec les protéines Avr de différentes souches de *Melampsora lini*, champignon responsable de la rouille du lin. De récentes études originales, car portant à la fois sur la diversité des gènes *R* et des gènes *Avr*, ont montré que les locus *L* et *Avr* correspondant étaient tous les deux sous pression de sélection diversificatrice (Dodds *et al.*, 2006; Dodds *et al.*, 2000; Luck *et al.*, 2000). Les différences de spécificité de reconnaissance des différents variants *Avr* et l'évidence d'une sélection positive agissant sur les gènes *L* et *avr* suggèrent qu'ils sont impliqués dans une « course aux armements ». La reconnaissance par interaction directe entre les protéines *L* et *Avr* argumente en faveur de ce type de coévolution rapide à la fois chez l'hôte et chez le pathogène (Dodds *et al.*, 2006; Ellis *et al.*, 2007). Un pré-requis pour une telle issue évolutive est que les protéines *Avr* puissent accumuler des mutations qui affectent la reconnaissance par les protéines *R* sans imposer un coût de fitness; c'est le cas ici, puisque les variants *Avr* qui échappent à la reconnaissance ont une structure protéique conservée et stable et ont maintenu une fonction effectrice (encore inconnue).

#### **I.2.4. Impact du mode de reconnaissance R /Avr sur l'évolution des gènes R**

Dodds *et al.* (2006) proposent que le mode de reconnaissance R/Avr est un paramètre majeur influençant l'évolution des gènes *R*. La reconnaissance directe entre protéines *R* et *Avr* conduirait à une évolution rapide de type "course aux armements", où l'on retrouve une forte diversité des gènes *R* et *Avr* (*i.e.*, des séries multialléliques présentant différentes spécificités)

contrairement à la définition initiale du modèle "course aux armements" qui prédit le maintien d'un seul allèle sensible et d'un seul allèle résistant dans les plantes hôtes. La reconnaissance indirecte, faisant intervenir une protéine de garde, serait plutôt associée à des polymorphismes balancés pour le maintien d'une résistance stable à long terme, comme c'est le cas pour les locus *Rpm1* et *Rps2* chez *A. thaliana*. L'adoption de l'un ou l'autre de ces 2 modes de reconnaissance pourrait être guidée par une histoire évolutive propre à chaque pathosystème. Les agents pathogènes comme *P. syringae* qui peuvent infecter un large spectre de plantes hôtes peuvent déployer beaucoup de protéines effectrices. A l'inverse, le spectre d'hôte étroit et le parasitisme obligatoire de certains pathogènes, comme le champignon responsable de la rouille du lin, impliqueraient que les interactions R/Avr coévoluent dans un système clos qui pourrait imposer plus de contraintes comparé au contexte dans lequel les molécules de *P. syringae* ont évolué. Au niveau mécanistique, des différences dans la fonction de pathogénicité des protéines Avr peuvent aussi influencer le mécanisme de reconnaissance adopté. En effet, les effecteurs de *P. syringae* génèrent des modifications des protéines de garde de l'hôte (phosphorylations ou clivages protéolytiques) qui sont reconnues par les protéines R correspondantes. D'autres effecteurs pourraient toucher la fonction de la protéine cible de la plante sans induire pour autant une forme modifiée de cette protéine cible facilement identifiable (par exemple en agissant comme des inhibiteurs compétitifs); dans ce cas, le mode de reconnaissance directe entre protéines R et Avr apparaît comme la seule option possible pour détecter la protéine du pathogène.

### **I.2.5. Conclusion**

Des polymorphismes alléliques anciens, qui dominent le paysage sélectif, et des balayages sélectifs récents, sous-représentés comparé au reste du génome, sont donc retrouvés à différents locus *R* de résistance chez plusieurs espèces de plante. Cependant, il ne faut pas oublier que les tests de neutralité utilisés sont peu efficaces pour la détection de régimes de sélection qui n'appartiennent pas à ces 2 histoires extrêmes d'évolution. Il faut également bien garder à l'esprit qu'il est toujours difficile de faire la distinction entre la signature de sélection naturelle et les évènements démographiques. Or ces facteurs démographiques ont des effets importants sur l'évolution des résistances puisque les épidémies induisent généralement des réductions de population, conduisant à des goulots d'étranglement et/ou une dérive génétique. Dans le futur, il sera intéressant de coupler la génétique moléculaire des populations non

seulement avec les études fonctionnelles mais aussi avec les études écologiques. Un échantillonnage écologique pertinent de la variation moléculaire aux locus neutres et liés à la résistance permettrait d'élucider les causes de sélection et les profils de diversité résultants. Comprendre les raisons du maintien de cette diversité allélique dans la variabilité naturelle sera crucial pour identifier les conditions dans lesquelles la résistance est durable en agriculture (de Meaux and Mitchell-Olds, 2003).

Il est à noter qu'aucune étude d'évolution moléculaire des gènes dominants de résistance aux virus n'a été décrite à ce jour.

### **I.3. Résistance récessive aux virus à ARN et complexe d'initiation de la traduction**

#### **I.3.1. Variation génétique et évolution des virus à ARN de plante**

Les virus à ARN, majoritaires chez les plantes, peuvent évoluer rapidement. En effet, au cours du processus de réplication de leur génome, des mutations et des recombinaisons peuvent se produire. Le génome viral est répliqué par une ARN polymérase ARN dépendante qui ne possède pas de système de contrôle et de correction des erreurs de réplication. Des taux de mutation de  $10^{-4}$  et  $10^{-5}$  erreur d'incorporation par base et par cycle de réplication ont été estimés pour les virus à ARN lytiques chez les mammifères, soit environ une erreur par génome et par cycle de réplication (Drake and Holland, 1999). Chez le TMV, virus à ARN positif de plante, la fréquence de mutation est sensiblement la même, avec un taux de mutation estimé à  $10^{-5}$  substitutions par nucléotide par cycle de réplication (Malpica *et al.*, 2002). Au cours de la réplication, la polymérase peut également "déraper" de la matrice en cours de lecture, provoquant le détachement du complexe de réplication qui se retrouve alors sur une autre matrice pendant la synthèse du brin complémentaire. Les recombinaisons apparues lors de ce saut du complexe de réplication consistent donc en un échange génétique entre différents variants viraux. Il semblerait que des événements de recombinaison se produisent assez fréquemment entre certaines souches de potyvirus, un genre de phytovirus à ARN (Glais *et al.*, 2002; Moury *et al.*, 2002; Revers *et al.*, 1996). La recombinaison peut fortement modifier les propriétés biologiques du virus, avec des conséquences

épidémiologiques majeures comme l'apparition de souches capables de contourner les gènes de résistance ou l'acquisition d'un spectre d'hôtes plus large.

En dépit de ce fort potentiel de variation génétique possible chez les virus à ARN, relativement peu de modifications profondes sont retrouvées dans les génomes viraux, indiquant qu'il existe des restrictions de la variabilité engendrée. Deux processus majeurs vont gouverner l'évolution des variants : la dérive génétique et la sélection. On peut compter jusqu'à  $10^{12}$  particules virales dans une plante infectée du fait du taux de réplication élevé des génomes viraux (Malpica *et al.*, 2002). Cependant, ce nombre ne représente pas la taille efficace de la population car une grande fraction de la population est constituée de variants incapables d'accomplir leur cycle infectieux. En ce qui concerne la dérive, plusieurs études ont montré que des goulets d'étranglement importants avaient lieu pendant le développement de l'infection dans la plante hôte et lors de la transmission du virus d'un hôte à un autre : en moyenne, seulement 4 génomes viraux de *Wheat streak mosaic virus* (WSMV) participent à l'infection systémique d'une talle de blé (French and Stenger, 2003) et entre 2 à 20 particules virales de TMV participent à l'infection d'une feuille inoculée et de 2 feuilles systémiques de tabac (Sacristan *et al.*, 2003). Ces effets d'échantillonnage aléatoire apparaissent donc dès les premières étapes de l'infection de la plante par les virus. La sélection va favoriser les variants les mieux adaptés. Toutes les composantes de l'environnement d'un virus à ARN constituent des facteurs de sélection. L'organisation même du génome viral constitue un premier niveau de sélection, pour le maintien des structures et de la fonctionnalité des protéines. Il est nécessaire pour les virus d'être transmis par des vecteurs biologiques, qui exercent donc un second niveau de sélection sur les motifs protéiques impliqués dans la transmission. Enfin, la plante hôte exerce elle-aussi une pression de sélection sur les virus à ARN puisque des interactions ARN polymérase virale-machinerie cellulaire de l'hôte sont indispensables pour l'accomplissement du cycle infectieux.

Les virus à ARN ont donc une capacité d'évolution très rapide. De plus, leur petit génome codant pour un nombre très limité de protéines (entre 4 et 12) fait qu'ils dépendent totalement des protéines de la plante hôte pour accomplir leur cycle infectieux : réplication et traduction de leur génome, mouvements de cellule à cellule et à longue distance. Ces caractéristiques orientent très certainement le mécanisme de reconnaissance plante hôte-virus ainsi que l'évolution des gènes de résistance et des gènes d'avirulence.

### I.3.2. Gènes récessifs de résistance aux virus ARN

Contrairement aux résistances contre les bactéries, les champignons, les nématodes ou insectes qui sont majoritairement dominantes, près de la moitié des résistances monogéniques vis-à-vis des virus de plantes sont récessives (Diaz-Pendon *et al.*, 2004; Kang *et al.*, 2005b). Cette caractéristique est particulièrement marquée pour les potyvirus; En effet, près de 50% des gènes de résistance aux potyvirus sont récessifs, alors que cette fréquence atteint rarement 20% pour les autres agents pathogènes (Provvidenti and Hampton, 1992).

Les résistances contrôlées par des gènes récessifs présentent différents phénotypes, allant de la résistance totale au niveau cellulaire, ou pendant le mouvement de cellule à cellule ou le mouvement à longue distance, jusqu'à différents niveaux de résistance partielle, qui se traduisent par une accumulation réduite du virus, une diminution du mouvement de cellule à cellule et/ou du mouvement systémique. Contrairement à la plupart des résistances dominantes aux virus qui entraînent une réponse hypersensible ou un phénotype de résistance extrême, les résistances récessives montrent donc une importante diversité de phénotypes de résistance.

Pour expliquer les résistances récessives aux virus, l'hypothèse la plus communément admise et proposée par Fraser (1990) est que l'allèle dominant de sensibilité code pour une protéine nécessaire à l'accomplissement du cycle du pathogène dans la plante. Le clonage moléculaire du premier gène récessif *pvr2* chez le piment, impliqué dans la résistance à 2 potyvirus (*Potato virus Y*, PVY et *Tobacco etch virus*, TEV), a permis de valider cette hypothèse. Le gène *pvr2* code pour une protéine eIF4E (eukaryotic Initiation Factor 4E) et des mutations dans cette protéine sont responsables de la résistance (Ruffel *et al.*, 2002). Depuis, toutes les études de clonage des gènes récessifs de résistance chez différentes espèces végétales ont montré que des mutations dans les facteurs d'initiation de la traduction eIF4E ou eIF4G sont responsables de la résistance vis-à-vis d'un large spectre de virus à ARN de plantes (Robaglia and Caranta, 2006).

### I.3.3. La famille multigénique eIF4E

#### I.3.3.1. eIF4E et l'initiation de la traduction des ARNm cellulaires

Dans toutes les cellules eucaryotes, l'initiation de la traduction des ARN messagers représente une étape majeure de régulation de la synthèse protéique. C'est l'étape la plus limitante et la plus complexe : plus d'une dizaine de facteurs d'initiation eIF (eukaryotic Initiation Factors), très régulés et conservés entre eucaryotes, interviennent dans cette étape de la traduction (Browning, 1996; Pain, 1996) (Figure I.5). Le recrutement de la machinerie traductionnelle va permettre la fixation de la sous-unité 40S du ribosome au codon d'initiation AUG du transcrit. Les facteurs d'initiation restent présents pendant le "scanning" jusqu'à ce que la sous-unité 60S se place sur le codon initiateur (Kozak, 1989). La sous-unité 60S peut alors s'associer à la 40S et ainsi former un ribosome 80S fonctionnel capable d'effectuer la synthèse des polypeptides.

Le facteur eIF4E intervient au tout début de la phase d'initiation. Son rôle est très important puisqu'il reconnaît et se lie spécifiquement à la coiffe des ARNm (résidu 7-méthyl guanosine triphosphate). Il correspond à la petite sous-unité du complexe eIF4F. Chez les mammifères, le complexe eIF4F est composé des facteurs eIF4E, eIF4G et eIF4A. Chez les plantes, le facteur eIF4A, bien que présent, n'a pas été identifié dans le complexe eIF4F. eIF4A possède une activité d'hydrolyse de l'ATP dépendante de l'ARN et participe au déroulement de l'ARN *via* une activité ARN-hélicase dépendante de l'ATP. Le facteur eIF4G possède de multiples sites de liaison à eIF4E, eIF4A, eIF3 et à la PABP (Poly(A) Binding Protein), qui interagit avec la queue de poly(A) des ARNm. La protéine eIF4G sert donc "d'échaffaudage" entre les différents composants de la machinerie d'initiation. Sa liaison avec la PABP va permettre la circularisation de l'ARNm et son interaction avec eIF3 va entraîner le recrutement de la sous-unité 40S (associé à d'autres facteurs eIF et à l'ARN de transfert initiateur ARNt-méthionine) (Luo and Goss, 2001). Chez les plantes, la PABP interagit également avec eIF4B (Le *et al.*, 1997). eIF4B est une protéine de liaison à l'ARN qui favorise l'activité ATPase et ARN-hélicase de eIF4A (Bi *et al.*, 2000; Metz *et al.*, 1999). L'efficacité de la traduction des ARNm dépend du rapprochement physique de la coiffe et de la queue de poly(A) (Gallie *et al.*, 1998). Le complexe de pré-initiation 43S (sous-unité ribosomique 40S, eIF2-GTP-ARNt<sup>met</sup>, eIF3, eIF5) est positionné dans la partie 5' de l'ARNm et va « scanner » l'ARNm dans la direction 5'→3' jusqu'au codon d'initiation AUG avec

lequel il se lie par l'intermédiaire de l'ARNt<sup>met</sup>. La reconnaissance du site d'initiation est assistée par les facteurs eIF1 et eIF1A (Pestova *et al.*, 1998). Enfin, grâce à l'intervention de eIF5B, la sous-unité ribosomale 60S s'associe à la 40S pour former le ribosome entier au niveau du codon initiateur du cadre de lecture. La phase d'initiation est alors suivie des phases d'élongation et de terminaison (Gallie, 2002).

### ***1.3.3.2. Paralogues et orthologues eIF4E***

Plus de 400 séquences nucléotidiques codant pour des homologues eIF4E chez les eucaryotes et les procaryotes ont été identifiées (Joshi *et al.*, 2005; Tableau I.5). Récemment, un gène *eIF4E* a même été découvert chez les *Mimivirus*, les plus grands virus d'eucaryotes connus à ce jour avec un génome de 1,2 mégabases (Raoult *et al.*, 2004). Il existe plusieurs protéines eIF4E chez chacune des espèces étudiées à ce jour (entre 2 et 8 protéines) sauf chez la levure *S. cerevisiae* qui ne possède qu'une seule protéine eIF4E (Joshi *et al.*, 2005). Chez la drosophile, l'existence des 8 protéines eIF4E serait due à des mécanismes d'épissages alternatifs (Hernandez *et al.*, 2005; Lavoie *et al.*, 1996).

Chez les plantes, eIF4E appartient à une petite famille multigénique : chez *A. thaliana*, 3 gènes codent pour des protéines eIF4E (eIF4E1 -At4g18040-, eIF4E2 -At1g29590- et eIF4E3 -At1g29550), 1 pour l'isoforme eIF(iso)4E -At5g35620-, protéine spécifique des plantes supérieures, et enfin un gène code pour une autre protéine de liaison à la coiffe, la nCBP -At5g18110- (new Cap Binding Protein). eIF4E1 est capable de se lier à la protéine eIF4G pour former le complexe eIF4F et eIF(iso)4E interagit préférentiellement avec la protéine eIF(iso)4G pour former le complexe eIF(iso)4F (Browning *et al.*, 1987; Browning *et al.*, 1992; Lax *et al.*, 1985). Chez le blé, le complexe eIF(iso)4F a les mêmes propriétés fonctionnelles que eIF4F : tous les 2 sont capables d'initier la traduction *in vitro* (Browning, 1996), même si, dans ces conditions expérimentales, eIF4F permet préférentiellement la traduction d'ARNm présentant des structures secondaires en 5' et des ARNm non coiffés et/ou dicistroniques alors que eIF(iso)4F traduirait préférentiellement des ARNm avec des structures linéaires et des coiffes hyperméthylées (Gallie and Browning, 2001). Des expériences *in vitro* et *in vivo* chez le maïs montrent que eIF(iso)4F se localise au niveau des microtubules, suggérant qu'il pourrait être impliqué dans la traduction de protéines spécialisées (Bokros *et al.*, 1995). Les gènes *eIF4E2* et *eIF4E3*, identifiés *via* le séquençage

complet du génome d'*A. thaliana*, ne sont séparés que de 7 kb sur le chromosome 1 et seraient donc issus d'une duplication récente. Le rôle des protéines codées par ces 2 gènes, dans l'initiation de la traduction ou dans d'autres processus cellulaires, n'a pour l'instant jamais été mis en évidence. La nCBP est présente chez tous les eucaryotes supérieurs. Elle est capable d'interagir *in vitro* avec eIF(iso)4G pour initier la traduction d'ARNm coiffés (Ruud *et al.*, 1998). En revanche, son rôle *in vivo* dans la traduction n'a jamais été démontré. Ruud (1998) suggère que la nCBP pourrait avoir un rôle de séquestration des ARNm afin de ralentir ou d'empêcher leur traduction.

Les protéines codées par les gènes *eIF4E* orthologues (gènes homologues appartenant à des espèces différentes) présentent un pourcentage d'identité plus fort que les protéines codées par les gènes *eIF4E* paralogues (gènes homologues au sein d'une même espèce). (Tableau I.6). Par exemple, chez le riz (*Oryza sativa*), le pourcentage d'identité en acides aminés entre eIF4E1 et eIF(iso)4E est de 41% alors que les eIF4E1 du riz et du blé (*Triticum aestivum*) sont identiques à 80%.

Le nombre de copies des gènes *eIF4E* diffère entre monocotylédones et dicotylédones. En effet, le blé, le riz et l'orge ne comptent qu'une copie *eIF4E* et une copie *eIF(iso)4E* alors que plusieurs copies de ces gènes sont présentes chez différentes dicotylédones : 3 *eIF4E* chez *A. thaliana* et 2 *eIF(iso)4E* chez le tabac et le tournesol (Combe *et al.*, 2005). Une seule copie *nCBP* est retrouvée chez toutes les espèces végétales. Trois gènes appartenant à la famille multigénique *eIF4E* ont été clonés et séquencés chez le piment et la tomate, grâce aux séquences EST disponibles : 2 gènes codant pour la protéine eIF4E (eIF4E1 et eIF4E2) et un gène codant pour l'isoforme eIF(iso)4E (Ruffel, 2004). Cependant, la cartographie génétique des marqueurs RFLP obtenus à partir de sondes EST des gènes *eIF4E1* et *eIF(iso)4E* suggère qu'il existerait au moins 2 copies supplémentaires de ces gènes chez le piment et la tomate (Ruffel, 2004).

En terme d'évolution, il semblerait qu'un seul gène *eIF4E* ancestral ait subi des duplications générant de multiples classes structurales (comme eIF4E, eIF(iso)4E et nCBP chez les plantes). En accord avec cette hypothèse, un seul gène *eIF4E* a été identifié chez *Ostreococcus tauri*, algue planctonique unicellulaire apparue il y a 1,5 milliards d'années qui figure donc parmi les membres les plus anciens de la lignée végétale et dont le génome vient d'être entièrement séquencé ([http://bioinformatics.psb.ugent.be/genomes/ostreococcus\\_tauri/](http://bioinformatics.psb.ugent.be/genomes/ostreococcus_tauri/)).

Les évènements de duplication auraient eu lieu après la divergence entre monocotylédones et dicotylédones.

### ***1.3.3.3. Polymorphisme de séquence des gènes eIF4E***

L'alignement des séquences en acides aminés des facteurs d'initiation de la traduction eIF4E chez de nombreuses espèces suggère la présence d'une région centrale conservée au cours de l'évolution (Joshi *et al.*, 2005). Sur l'alignement représenté à la figure I.6, cette région essentielle pour la fonction des eIF4E couvre approximativement 160-170 résidus, de l'histidine en position 62 à l'histidine en position 228 et comprends les tryptophanes indispensables au bon fonctionnement de la protéine. L'importance fonctionnelle de cette région des protéines eIF4E a été validée par des expériences de délétion (Robalino *et al.*, 2004; Vasilescu *et al.*, 1996). Les régions N- et C-terminales ne sont pas conservées chez toutes les protéines eIF4E. Ces régions pourraient être impliquées dans la régulation de l'activité des eIF4E (Gross *et al.*, 2003; Scheper *et al.*, 2002) ou affecter la stabilité des interactions, comme c'est le cas pour la partie N-terminale de eIF4E chez *Saccharomyces cerevisiae* qui interagit avec eIF4G et stabilise l'interaction (Gross *et al.*, 2003). Une protéine sera classée dans la famille eIF4E si sa séquence s'aligne avec la séquence consensus de la région cœur conservée.

La structure tridimensionnelle de eIF4E chez le blé vient tout juste d'être résolue par l'équipe de Karen Browning (Monzingo *et al.*, 2007). Il s'agit de la première description d'une structure d'une protéine eIF4E chez les plantes. Comme attendu, la protéine eIF4E du blé est très similaire aux protéines eIF4E de levure, drosophile, souris et homme précédemment résolues (Marcotrigiano *et al.*, 1997; Matsuo *et al.*, 1997; Tomoo *et al.*, 2002; Tomoo *et al.*, 2003; Volpon *et al.*, 2006). Environ 55 résidus sont totalement conservés pour des raisons architecturales et fonctionnelles évidentes, formant le cœur hydrophobe de la protéine et d'autres interactions tertiaires (encadrés en noir sur la figure I.6). Les acides aminés impliqués dans la fixation à la coiffe sont aussi conservés entre espèces (surlignés en orange sur la figure I.6). Onze résidus sont très bien conservés au sein du règne végétal uniquement (surlignés en vert sur la figure I.6). Certains de ces résidus "plante-spécifiques" sont facilement accessibles à la surface de la protéine, suggérant qu'ils sont impliqués dans des interactions fonctionnelles de eIF4E avec d'autres protéines spécifiques chez les plantes. La présence de ces acides

aminés spécifiques des plantes n'est pas surprenante puisqu'on sait que la traduction chez les plantes présente certaines caractéristiques propres à ce règne, comme l'existence de 2 complexes eIF4F et eIF(iso)4F (Browning, 2004). Treize acides aminés sont "isoforme-spécifique" (surlignés en bleu sur la figure I.6). Ces résidus, conservés à plus de 95% pour chaque isoforme eIF4E et eIF(iso)4E, se retrouvent majoritairement à des positions relativement accessibles à la surface de la protéine, avec notamment 5 acides aminés "isoforme-spécifique" dans les boucles à proximité du site de fixation à la coiffe. L'existence de telles spécificités étaye l'idée que eIF4E et eIF(iso)4E auraient des spécificités d'interactions différentes avec la coiffe des ARNm et donc des fonctions distinctes. Bien qu'on ne comprenne pas encore très bien le rôle des différentes isoformes eIF4E chez les plantes, le maintien de ces protéines homologues chez l'ensemble des monocotylédones et des dicotylédones conforte l'hypothèse d'une fonction "isoforme-spécifique" (Duprat *et al.*, 2002; Rodriguez *et al.*, 1998).

### **I.3.4. Rôle clé des protéines eIF4E dans l'interaction plante-virus à ARN et dans la résistance des plantes**

#### ***I.3.4.1. Interactions entre protéines eIF4E et VPg***

Des méthodes biochimiques, utilisant le système de double hybride dans la levure, des tests immuno-enzymatiques (ELISA), des tests de pull-down, de la chromatographie d'affinité ou de la spectrométrie ont mis en évidence des interactions entre isoformes eIF4E de blé, tabac, tomate, laitue et *A. thaliana* et la protéine virale appelée VPg (Viral protein genome-linked) de plusieurs potyvirus (PVY, TEV, TuMV (*Turnip mosaic virus*) et LMV (*Lettuce mosaic virus*) (Tableau I.7). Pour le pathosystème *A. thaliana* -TuMV, il a été montré que cette interaction VPg-eIF4E était nécessaire pour l'infection virale et avait un effet positif sur la réplication du virus (Leonard *et al.*, 2000).

Une étude récente s'est focalisée sur la localisation subcellulaire des interactions eIF(iso)4E d'*A. thaliana* et les différentes formes précurseurs de la VPg, VPg-Pro et 6K-VPg-Pro du TuMV (Beauchemin *et al.*, 2006). Les expériences de BiFC (Bimolecular Fluorescence Complementation) montrent que les interactions eIF(iso)4E/VPg-Pro ont lieu

dans le noyau des cellules alors que les interactions eIF(iso)4E/6K-VPg-Pro sont observées dans des vésicules cytoplasmiques bourgeonnant à partir du réticulum endoplasmique. Ces données suggèrent que les associations eIF4E/VPg pourraient être impliquées à différentes étapes du cycle infectieux viral, dépendant de la forme précurseur de la VPg impliquée dans l'interaction. L'interaction eIF(iso)4E/VPg-Pro perturberait des fonctions nucléaires de la plante hôte tandis que l'interaction eIF(iso)4E/6K-VPg-Pro pourrait jouer un rôle dans la réplication et/ou la traduction de l'ARN viral.

En accord avec ces données de biologie cellulaire, plusieurs auteurs proposent différents modèles pour expliquer le rôle biologique des interactions VPg/eIF4E dans l'infection virale. Logiquement, au vu de la fonction d'initiation de la traduction des protéines eIF4E, on pourrait penser que cette interaction permet l'initiation de la traduction de l'ARN viral (Figure I.7.a, Lellis *et al.*, 2002). En effet, la VPg pourrait fonctionner comme la structure coiffe en 5' de l'ARNm et recruter les facteurs d'initiation de la traduction et la sous-unité 40S du ribosome. Cependant, cette hypothèse est fragilisée par le fait que la traduction de certains potyvirus peut être "coiffe-indépendante" *via* un IRES (Internal Ribosome Entry Site) qui fait intervenir directement eIF4G (Carrington and Freed, 1990). La seconde hypothèse est que eIF4E pourrait stabiliser le génome des potyvirus *via* son interaction avec la VPg, de façon similaire à l'interaction eIF4E/coiffe qui permet la stabilité des ARNm chez la levure en le protégeant des enzymes de dégradation de l'ARN (Schwartz and Parker, 2000) (Figure I.7.b) (Lellis *et al.*, 2002). Une troisième hypothèse propose l'implication de eIF4E dans la localisation intracellulaire et le transport de l'ARN des potyvirus (Gao *et al.*, 2004). La protéine eIF(iso)4G du blé, à laquelle est liée eIF(iso)4E, interagit avec les microtubules dans des cellules de blé (Bokros *et al.*, 1995). Les génomes des potyvirus pourraient donc interagir avec les microtubules pour leur mouvement intracellulaire *via* des interactions eIF(iso)4E-eIF(iso)4G/VPg (Figure I.7.c). Une dernière hypothèse, en accord avec le système de réplication décrit chez les poliovirus (Herold and Andino, 2001), serait que eIF4E appartient à un complexe protéique impliqué dans la circularisation de l'ARN viral nécessaire à l'initiation de la synthèse du brin négatif (Robaglia and Caranta, 2006) (Figure I.7.d). La PABP pourrait faire le lien entre eIF4F (*via* eIF4G) et la queue polyadénylée des potyvirus. Le fait que la polymérase des potyvirus soit capable d'interagir avec la PABP et avec le domaine VPg (Hong *et al.*, 1995; Wang *et al.*, 2000) appuie cette hypothèse; il est possible que le complexe protéique impliquant eIF4E et la VPg pour la circularisation de l'ARN viral permette le

positionnement correct de la polymérase et agisse ainsi directement sur l'efficacité de réplication du génome.

#### ***I.3.4.2. Approches génétiques***

Plusieurs approches génétiques ont confirmé l'importance des gènes du complexe d'initiation de la traduction dans le cycle infectieux viral et dans la résistance aux virus à ARN (Tableau I.8). La première a été la caractérisation moléculaire du locus de résistance *pvr2* du piment par une approche gène candidat. Cette approche a permis de valider le fait que *pvr2* codait pour le gène *eIF4E1* (Ruffel *et al.*, 2002). Seulement deux substitutions en acides aminés distinguent l'allèle dominant de sensibilité des 2 allèles récessifs de résistance, *pvr2*<sup>1</sup> et *pvr2*<sup>2</sup>. Depuis, la caractérisation d'autres gènes récessifs met en avant le rôle générique des gènes *eIF4E* dans l'interaction plante-virus et dans la résistance. Par une approche gène candidat similaire à celle conduite pour *pvr2*, il a été montré que le gène *pot1* de la tomate, contrôlant la résistance à de nombreux isolats de PVY et TEV, correspondait également au gène *eIF4E1* (Ruffel *et al.*, 2005). Comme pour *pvr2*, très peu de mutations différencient les allèles récessifs de résistance des allèles dominants de sensibilité. Le gène récessif *pvr6* du piment de résistance au *Pepper veinal mottle virus* (PVMV) a lui aussi été caractérisé; il correspond au gène *eIF(iso)4E* (Ruffel *et al.*, 2006). Le gène récessif *pvr6* présente une délétion de 82 nucléotides dans sa séquence codante entraînant l'apparition d'un codon stop conduisant à la synthèse d'une protéine tronquée non fonctionnelle.

De même, le clonage des allèles récessifs *mo1*<sup>1</sup> et *mo1*<sup>2</sup> conférant la résistance au LMV chez la laitue a montré que ces gènes correspondaient à *eIF4E* (Nicaise *et al.*, 2003). Le rôle clé des facteurs eIF4E dans la résistance aux virus à ARN chez les plantes a été par la suite confirmé par la caractérisation d'autres gènes récessifs de résistance : *sbm1* chez le pois pour la résistance aux potyvirus, *rym4/5* chez l'orge pour la résistance à plusieurs souches de *Bymovirus* et *nsv* chez le melon pour la résistance au MNSV (*Melon necrotic spot virus*, genre *Carmovirus*) (Gao *et al.*, 2004; Nieto *et al.*, 2006; Stein *et al.*, 2005). La résistance résulte toujours de modifications au niveau de la séquence protéique de eIF4E. La position de la plupart de ces substitutions "signatures" des allèles de résistance apparaît conservée entre espèces et restreinte à 2 régions adjacentes (nommées régions I et II), localisées en surface de la protéine eIF4E, à proximité de la zone d'interaction avec la coiffe (Figure I.8). Cependant,

dans le cas de la résistance au MNSV, l'unique mutation se trouve à l'extrémité C-terminale de la protéine eIF4E (en position 228) et non dans les régions I et II en surface de la protéine décrites précédemment, suggérant un mécanisme de résistance différent.

En parallèle à ces études sur les gènes récessifs de résistance présents dans la variabilité naturelle des espèces cultivées, la caractérisation de mutants d'*A. thaliana* pour les différents gènes *eIF4E* a permis de confirmer l'importance de ces gènes dans le cycle viral. Il apparaît que des mutations induites par EMS (Ethyl Methyl Sulfonate) ou T-DNA dans les différents gènes *eIF4E* affectent la sensibilité aux potyvirus. Des mutations knock-out du gène *eIF(iso)4E* chez *A. thaliana* n'entraînent pas de modifications morphologiques ou de développement de la plante mais confère une résistance à l'infection par les potyvirus TEV, TuMV, LMV et PPV (*Plum pox virus*) (Decroocq *et al.*, 2006; Duprat *et al.*, 2002; Lellis *et al.*, 2002) (Tableau I.8). Le mutant EMS *At-eIF4E1 (cum1)* est résistant au CIYVV (*Clover yellow vein virus*), alors que ce virus est capable de se multiplier dans les plantes mutées pour le gène *eIF(iso)4E* (Sato *et al.*, 2005). Enfin, la mutation *cum1*, qui provoque l'apparition d'un codon stop dans la région codante du gène *eIF4E1*, réduit fortement l'accumulation du CMV (*Cucumber mosaic virus*) chez *A. thaliana* (Yoshii *et al.*, 2004).

D'autres travaux ont évalué la sensibilité de plantes mutées dans les gènes *eIF4G* vis-à-vis de plusieurs potyvirus. Chez *A. thaliana*, 3 gènes codent pour des membres de la famille eIF4G : *eIF4G*, *eIF(iso)4G1* et *eIF(iso)4G2*. Il apparaît que le mutant *eIF4G* est résistant au CIYVV, le mutant *eIF(iso)4G1* est résistant au PPV et au LMV et enfin le double mutant *eIF(iso)4G1 x eIF(iso)4G2* est résistant au TuMV (Nicaise *et al.*, 2007). Si on fait un parallèle entre ces résultats et ceux obtenus avec les mutants eIF4E-eIF(iso)4E énoncés précédemment, il apparaît que le CIYVV utilise le complexe eIF4F tandis que les autres potyvirus semblent plutôt interagir avec le complexe eIF(iso)4F. Ces données argumentent donc en faveur d'un rôle de eIF4E et eIF4G sous forme de complexe eIF4F dans le cycle infectieux des potyvirus.

#### ***1.3.4.3. Utilisation sélective des membres de la famille eIF4E par les potyvirus***

Les facteurs eIF4E et eIF4G ne semblent pas être utilisés par les différents potyvirus de façon stochastique. Ils recrutent de façon spécifique les isoformes eIF4E : TuMV, LMV et

PPV semblent utiliser eIF(iso)4E tandis que le CIYVV utilise eIF4E1 chez *A. thaliana*. Le fait que les mutants *At-eIF(iso)4E* soient résistants aux TEV, TuMV, LMV et PPV suggère que les autres protéines eIF4E, toujours fonctionnelles, ne sont pas capables de remplacer eIF(iso)4E dans le cycle infectieux de ces virus et *vice versa* pour le CIYVV chez les mutants *At-eIF4E1*. En revanche, le PVMV semble capable d'utiliser plusieurs facteurs eIF4E chez un même hôte puisqu'on observe de la résistance uniquement lorsqu'il y a combinaison de mutations à la fois dans eIF4E (*pvr2*) et dans eIF(iso)4E (*pvr6*) (Ruffel *et al.*, 2006). Le TuMV semble lui capable d'utiliser plusieurs formes eIF(iso)4G puisque seuls les doubles mutants eIF(iso)4G1 x eIF(iso)4G2 chez *A. thaliana* sont résistants à ce virus.

Très peu de choses sont connues sur l'implication des formes eIF4E2, eIF4E3 et nCBP dans la résistance aux potyvirus car ces facteurs sont très rarement pris en compte dans les études. Les simples mutants ont été récemment obtenus et sont en cours d'analyse au laboratoire (collaboration K. Browning). De plus, jamais un gène récessif de résistance aux potyvirus n'a été montré comme correspondant à l'un de ces 3 facteurs.

#### ***1.3.4.4. La VPg des potyvirus est le déterminant de virulence vis-à-vis des résistances contrôlées par eIF4E***

Plusieurs travaux ont montré que des mutations dans le domaine central de la protéine virale VPg sont responsables du contournement des gènes récessifs de résistance, qui codent pour la plupart pour des protéines eIF4E (Borgstrom and Johansen, 2001; Moury *et al.*, 2004; Nicolas *et al.*, 1997; Rajamäki and Valkolen, 2002). Les substitutions de 5 acides aminés dans la partie centrale de la VPg du PVY sont responsables du contournement de la résistance contrôlée par les allèles *pvr2<sup>1</sup>* et *pvr2<sup>2</sup>* chez *C. annuum* et la substitution d'un seul acide aminé dans cette même région suffit au contournement du gène récessif de résistance *pot1* chez la tomate *L. esculentum* (Moury *et al.*, 2004) (Figure I.9). Des substitutions dans la VPg sont également responsables des contournements des résistances contrôlées par *sbml* chez le pois *Pisum sativum* (Borgstrom and Johansen, 2001) et *va* chez le tabac *Nicotiana tabacum* (Nicolas *et al.*, 1997) (Figure I.9).

Le contournement des résistances récessives *via* des mutations dans la zone centrale de la VPg ne concerne pas que les potyvirus; en effet, une seule mutation dans la VPg du *Rice*

*Yellow Mosaic Virus* (RYMV, genre *Sobemovirus*) entraîne le contournement du gène récessif *rymv-1* correspondant au gène eIF(iso)4G chez le riz (Hebrard *et al.*, 2006). De la même façon, le déterminant du contournement des gènes récessifs de résistance *rym4* et *rym5* de l'orge par le *Barley yellow mosaic virus* (BaYMV, genre *Bymovirus*) se situe vraisemblablement dans la zone centrale de la VPg (Kanyuka *et al.*, 2004; Kuhne *et al.*, 2003).

L'étude de la variabilité de la séquence VPg du PVY et du PVA (*Potato virus A*) montre que les acides aminés impliqués dans le contournement des résistances sont sous pression de sélection positive : le ratio substitutions non-synonymes/substitutions synonymes  $\omega$  est égal à 2,84 pour la VPg du PVA et 1,88 pour la VPg du PVY. Les substitutions en acides aminés pourraient induire des changements dans la conformation tridimensionnelle de la VPg. Ces changements seraient responsables de l'infection ou non des plantes possédant différents allèles *pvr2/pot1* en modifiant les interactions physiques avec une ou plusieurs protéines de l'hôte et/ou du virus (Moury *et al.*, 2004).

D'autres protéines que la VPg sont impliquées dans le contournement des résistances médiées par eIF4E. En effet, la région 3'NTR du MNSV est impliquée dans le contournement du gène récessif *nsv-eIF4E* chez le melon (Diaz *et al.*, 2004) ; les protéines CI-6K2 du LMV sont impliquées dans le contournement de *mol-eIF4E* chez la laitue (Redondo *et al.*, 2001) et, finalement, les protéines P3-6K1 du *Pea seed-borne mosaic virus* (PSbMV) dans celui de *sbm2* chez le pois (Johansen *et al.*, 2001).

## **I.4. Interactions potyvirus-Solanacées et potyvirus-*A.thaliana***

### **I.4.1. Les potyvirus**

Le genre *Potyvirus* représente un quart des virus de plantes et inclut quelques uns des virus les plus communs et ravageurs de nombreuses plantes cultivées à travers le monde, dont le virus Y de la pomme de terre *Potato virus Y* (PVY) (Shukla *et al.*, 1994) ou le *Plum pox virus* (PPV) responsable de la sharka chez les *Prunus*. Les potyvirus appartiennent à la famille des *Potyviridae*, du super-groupe des virus "Picorna-like", où l'on retrouve des *Picornaviridae*

animaux. La famille des *Potyviridae* regroupe également des virus de plante appartenant aux genres *Comovirus* et *Nepovirus* (Riechmann *et al.*, 1992). Les espèces végétales cibles des potyvirus sont nombreuses : les Solanacées, famille comprenant les espèces piment, tomate, pomme de terre, tabac, aubergine, d'autres espèces maraîchères comme le haricot, le pois, la laitue ou le melon, des espèces fruitières du genre *Prunus*, mais aussi de grandes cultures comme la betterave, le soja et certaines céréales (Figure I.10). Cette large gamme d'hôtes associée aux dégâts causés par ces virus justifient les recherches visant à comprendre les bases moléculaires des interactions plantes-potyvirus.

La symptomatologie et les dégâts causés par les potyvirus sont variables en fonction des espèces hôtes. Les symptômes causés par les potyvirus sur le piment sont des anomalies de coloration des feuilles, se traduisant par des marbrures plus ou moins marquées, des jaunissements le long des nervures ou entre ces dernières. Des nécroses se manifestent parfois, en partant de la nervure principale et/ou des nervures secondaires, pouvant conduire à la nécrose de toute la plante. Les plantes atteintes sont souvent de taille réduite et présentent des feuilles plus petites, crispées et enroulées. Chez la tomate, les potyvirus sont responsables de mosaïques ou de filiformisme mais n'ont pas de conséquence majeure sur les récoltes. Le PVY et le TEV sont des composants majeurs des infections mixtes avec d'autres genres viraux comme le *Cucumber mosaic virus* (CMV, genre *Cucumovirus*) ou le *Tobacco mosaic virus* (TMV, genre *Tobamovirus*) (Nelson *et al.*, 1982; Morel, 2000).

La transmission des potyvirus se fait par l'intermédiaire des pucerons, et en particulier *Myzus persicae*, selon le mode non persistant (Raccah *et al.*, 1985). Le cycle des potyvirus débute par la pénétration, *via* une blessure provoquée par le stylet des pucerons, d'une ou plusieurs particules virales dans une cellule épidermique de l'hôte. L'ARN viral est ensuite décapsidé pour être traduit et répliqué dans le cytoplasme des cellules de l'hôte (Astier *et al.*, 2001). Le matériel viral migre ensuite de cellule à cellule par les plasmodesmes. Ce mouvement de cellule à cellule permet aux virus d'entrer dans les tissus vasculaires : à partir du mésophylle, le virus infecte le parenchyme périvasculaire, le parenchyme phloémien, les cellules compagnes et enfin les tubes criblés. Le virus migre vers les autres feuilles *via* un transport passif phloème-dépendant. Cette migration à longue distance entraîne l'infection systémique à l'origine de l'apparition des symptômes. L'encapsidation des nouveaux acides nucléiques viraux conduit à la formation de nouveaux virions prêts pour la vécution vers d'autres plantes (Carrington *et al.*, 1996).

Les potyvirus sont des virus à ARN simple brin de polarité positive. Leur génome, d'environ 10Kb, est encapsidé. L'unique ORF code pour une polyprotéine qui sera clivée par des protéases virales en 10 protéines fonctionnelles. L'ARN est polyadénylé à l'extrémité 3' et l'extrémité 5' est liée de façon covalente à la VPg (24 kDa). L'organisation génomique, les processus de maturation des potyvirus et les fonctions attribuées à chaque protéine sont décrites sur la figure I.11 (Urcuqui-Inchima *et al.*, 2001).

#### I.4.2. Les potyvirus des Solanacées

Neuf potyvirus majeurs (PVY, TEV, PVA, PVV - *Potato virus V-*, PVMV, PepMoV, CVMV - *Chili veinal mottle virus -*, TVMV - *Tobacco vein mottling virus -* et PTV - *Peru tomato virus -*) infectent les Solanacées, principalement la pomme de terre (*Solanum tuberosum*), le piment (*Capsicum spp.*), la tomate (*Lycopersicon spp.*), et le tabac (*Nicotiana tabacum*) (Morel, 2000). L'aubergine (*Solanum melongena*) est également un hôte potentiel pour certains potyvirus.

Le PVY est le potyvirus majeur des cultures de piment, de pomme de terre et à un moindre niveau des cultures de tomate. Les souches de PVY isolées sur piment ont été classées en pathotypes (entités de sous-espèces virales qui sont contrôlées par une série allélique de résistance spécifique de l'hôte) appelés (0), (0,1) et (0,1,2) en fonction de leur comportement vis-à-vis de génotypes de piments portant les allèles de sensibilité  $pvr2^+$  et de résistance  $pvr2^1$  et  $pvr2^2$ . Les souches qui affectent seulement les génotypes chez *C. annuum* dépourvus d'allèles de résistance ( $pvr2^+$ ) sont du pathotype 0. Les souches de pathotype (0,1) infectent les plantes possédant l'allèle  $pvr2^1$  et le pathotype (0,1,2) est attribué aux souches qui surmontent les allèles de résistance  $pvr2^1$  et  $pvr2^2$  (Gébré Sélassié *et al.*, 1985). La plupart des isolats de PVY collectés dans les champs de piment appartiennent aux pathotypes (0) ou (0,1) (Palloix *et al.*, 1994). Les isolats collectés en France, en Italie et en Espagne sont majoritairement de pathotype (0) (Arteaga and Ortega, 1986). Au Sud de la Méditerranée, en Turquie et dans les régions chaudes du continent américain, les souches de PVY de pathotype (0,1) sont plus fréquentes (Palloix *et al.*, 1994). La présence ponctuelle d'isolats de pathotype (0,1,2) a été rapportée au Brésil (Boiteux *et al.*, 1996). Cette classification, qui ne repose que

sur quelques différences au niveau génomique (Moury *et al.*, 2004) ne reflète cependant pas l'histoire évolutive et les relations de filiation entre souches.

Le TEV est un pathogène important pour les cultures de piment, de tomate et de tabac, principalement en Amérique du Nord, en Amérique Centrale et dans les Caraïbes (Ariyaratne *et al.*, 1996). Le PVMV a été décrit seulement dans quelques pays de l'Afrique de l'Ouest et infecte principalement le piment, la tomate, le tabac et l'aubergine (Moury *et al.*, 2006). Le PVA et le PVV constituent des potyvirus d'importance majeure mais exclusivement sur pomme de terre. Le PVA est trouvé dans toutes les zones de culture de la pomme de terre tandis que le PVV n'est présent qu'à l'Est de l'Europe et au Pérou. Le PepMoV et le CVMV ont également une gamme d'hôtes restreinte puisqu'ils n'infectent que les cultures de piment. Le PepMoV est présent principalement en Amérique du Sud et Centrale, où on le retrouve en infections mixtes avec le PVY et le TEV (Nelson *et al.*, 1982). Le CVMV est présent en Asie du Sud-Est, constituant le principal potyvirus pathogène du piment en Thaïlande (Chiemsoombat *et al.*, 1998). Le TVMV n'a été décrit que dans les pays du Sud-Est des Etats-Unis, par des producteurs de tabac *cv.* Burley. Le PTV n'a été décrit qu'au Pérou sur des cultures de tomate et de piment (Fribourg, 1979).

#### **I.4.3. Gènes récessifs de résistance aux potyvirus chez les Solanacées**

Plusieurs facteurs récessifs de résistance, gènes majeurs et QTL (Quantitative Trait Loci), ont été identifiés et certains localisés sur le génome du piment (Figure I.12). Le locus majeur *pvr2*, correspondant au gène *eIF4E1* de *C. annuum* situé sur le chromosome 4, est impliqué dans la résistance au PVY et au TEV. Trois allèles ont été initialement identifiés au locus *pvr2* : *pvr2*<sup>+</sup> issu de Yolo Wonder (sensible), *pvr2*<sup>1</sup> issu des génotypes YoloY et YoloRP10 et *pvr2*<sup>2</sup> de Florida VR2 qui confère la résistance au PVY et aux souches communes de TEV (Kyle and Palloix, 1997). Chez *C. chinense*, le gène *pvr1* confère la résistance aux souches communes de TEV, aux pathotypes (0) et (0,1) du PVY et au PepMoV (Kyle and Palloix, 1997). Des expériences récentes ont montré que ce gène est un allèle de *pvr2* (Kang *et al.*, 2005b; Ruffel *et al.*, 2004). Le locus *pvr6* localisé sur le chromosome 3 et issu de la lignée indienne *C. annuum* Perennial, correspond au gène *eIF(iso)4E*. Les allèles *pvr6* et *pvr2*<sup>2</sup> sont nécessaires pour contrôler une résistance totale à un autre potyvirus, le PVMV. De façon intéressante, *pvr2*<sup>2</sup> et *pvr6* seuls n'ont aucun effet sur ce virus (Caranta *et*

*al.*, 1996). Finalement, le locus *pvr3*, mis en évidence chez la lignée *C. annuum* Avelar, contrôle une résistance totale à la migration à longue distance du *Pepper mottle virus* (PepMoV) (Guerini and Murphy, 1999); sa position sur la carte génétique du piment reste à déterminer. En plus de ces gènes majeurs, plusieurs QTL de résistance aux potyvirus ont été cartographiés sur le génome du piment (Caranta *et al.*, 1997; Caranta and Palloix, 1996).

Chez la tomate, la résistance à de nombreux isolats de PVY et TEV est contrôlée par le gène récessif *pot1*, identifié chez l'espèce sauvage *L. hirsutum*. *pot1* code, comme *pvr2*, pour une protéine eIF4E1 (Ruffel *et al.*, 2005). De précédents travaux présageaient de ce résultat : il avait été montré que le gène *pot1* se situait dans une région du génome de la tomate colinéaire à celle où se trouve *pvr2* chez le piment. Il s'agissait d'un des rares cas de conservation de position de carte pour des gènes de résistance présentant des phénotypes identiques (Parrella *et al.*, 2002).

Chez l'aubergine, une quasi-absence de sensibilité aux potyvirus a été identifiée. On parle dans ce cas de résistance non-hôte, c'est-à-dire lorsque la plupart des génotypes d'une espèce végétale donnée est résistante à la plupart des souches d'un pathogène donné. Le gène récessif de résistance *ptv1* confère une résistance totale à plusieurs souches de PVY chez l'accession MM738 de l'espèce cultivée *Solanum melongena* (Deulvot, 2002). Ce gène n'est pas localisé dans la région génomique colinéaire à celle de la tomate et du piment où ont été cartographiés les gènes *pot1* et *pvr2* et les analyses de séquences ont permis d'exclure le rôle du gène *eIF4E1* de l'aubergine dans la résistance contrôlée par *ptv1* (Ruffel, 2004).

Le gène récessif *va* contrôle la résistance au TVMV, au PVY et au TEV chez le tabac. De plus, une résistance au mouvement à longue distance de la souche TEV-HAT serait sous le contrôle de 2 gènes récessifs indépendants chez *N. tabacum* (Schaad and Carrington, 1996). Chez la pomme de terre, la résistance au PVA est digénique : (i) le gène  $Na_{adg}$  confère une résistance de type HR (Hypersensitive response) et est localisé dans la même région génomique que le locus  $Ry_{adg}$  de résistance au PVY, (ii) le gène récessif *ra* contrôle un blocage du transport vasculaire du PVA (sa localisation sur le génome de la pomme de terre n'a pas encore été déterminée) (Hamalainen *et al.*, 2000). A ce jour, la structure moléculaire de ces gènes de résistance chez le tabac et la pomme de terre n'a pas été identifiée.

#### I.4.4. Le couple *A. thaliana*-potyvirus : un pathosystème de laboratoire

L'espèce *A. thaliana* a été adoptée par la communauté scientifique en tant qu'espèce modèle pour les études de génétique. La séquence complète du génome de l'accession Columbia a été publiée en 2000 (Arabidopsis Genome Initiative, 2000). A ce jour, *A. thaliana* est la seule dicotylédone pour laquelle on connaît toutes les séquences des gènes codant pour des protéines de liaison à la coiffe, grâce au séquençage complet de son génome.

Les outils disponibles pour étudier les gènes *eIF4E* chez *A. thaliana* sont nombreux et facilement accessibles : (i) données de séquences disponibles sur le TAIR (The Arabidopsis Information Resource - <http://www.arabidopsis.org/>), (ii) banques de mutants (<http://signal.salk.edu/> - <http://www-ijpb.versailles.inra.fr/fr/sgap/equipements/crg/collections/index.htm>), (iii) structuration des populations en « collection cœur » ou « core collection » constituée d'un sous-échantillon d'individus représentant au mieux la diversité totale de l'espèce. Appliquée à *A. thaliana*, cette stratégie a permis de définir un lot de 48 lignées homozygotes fixes (accessions) utilisé comme échantillon de référence pour étudier la diversité naturelle de l'espèce (Projet AraCore-INRA; (McKhann *et al.*, 2004) (Figure I.13). Ces lignées sont maintenues et multipliées par l'INRA de Versailles et distribuées à la communauté internationale.

Des efforts importants de description de la diversité naturelle basés sur l'analyse des séquences de gènes ont été entrepris sur cette espèce. Actuellement, dans le cadre du projet "Arabidopsis program 2010" de la National Science Foundation (NSF), les laboratoires de Nordborg, Kreitman et Bergelson (Université de Caroline du Sud, Los Angeles, USA; Université de Washington, St Louis, USA; Université de Chicago, Chicago, USA) séquencent 1500 courts fragments de gènes sur un échantillon de 96 individus issus de populations d'origines géographiques distinctes (Europe, Asie Centrale, Amérique du Nord) afin de déterminer le polymorphisme global du génome de cette espèce (<http://walnut.usc.edu/2010>). Au préalable, 876 fragments de gènes chez ces mêmes 96 accessions et 334 régions distribuées de façon aléatoire sur le génome chez 12 accessions d'*A. thaliana* avaient été séquencés (Nordborg *et al.*, 2005; Schmid *et al.*, 2005). Ces études sur le polymorphisme global du génome de l'espèce végétale modèle montrent que, bien qu'*A. thaliana* soit autogame, le profil de polymorphisme observé concorde avec ce qui est attendu pour les espèces allogames sauvages. Les distributions de D de Tajima et les statistiques F et D de Fu

and Li s'écartent des distributions théoriques attendues dans le cas du modèle neutre de Kimura, à cause d'un excès d'allèles rares (valeurs négatives). Les distributions des valeurs H de Fay and Wu théoriques et observées sont équivalentes, suggérant que les processus démographiques et non la sélection à plusieurs locus sont responsables de la déviation par rapport au modèle neutre standard.

Le TuMV est le seul potyvirus inféodé à la famille des Brassicacées dont fait partie *A. thaliana*. D'autres potyvirus inféodés à d'autres familles végétales sont pourtant capables d'infecter la plante modèle en conditions de laboratoire. Il s'agit du CIYVV, du TEV, du PPV, du LMV et du WMV (*Watermelon mosaic virus*). Plusieurs études ont conduit à l'identification de gènes récessifs de résistance aux potyvirus chez différentes accessions d'*A. thaliana* mais jusqu'à présent aucun ne code pour une protéine du complexe d'initiation de la traduction. Deux gènes récessifs de résistance ont été identifiés chez l'accession Cvi d'*A. thaliana* : le gène récessif *rlm1* de résistance à plusieurs souches de LMV (Revers *et al.*, 2003) et le gène récessif *rpv1*, impliqué dans la restriction du mouvement à longue distance d'un isolat de PPV (Decroocq *et al.*, 2006). Des études de cartographie localisent ces gènes dans des régions génomiques différentes de celle où se trouvent les gènes *eIF4E*. De plus, les comparaisons de séquences des gènes *eIF4E* révèlent une complète identité entre les allèles *eIF4E* de plantes sensibles et de plantes résistantes.

## **I.5. Objectif de la thèse**

Les recherches menées dans l'équipe de l'unité de Génétique et Amélioration des Fruits et Légumes de l'INRA d'Avignon où j'ai effectué ce travail de thèse visent à mieux comprendre le fonctionnement des gènes de résistance des plantes et les interactions plantes-pathogènes, connaissances indispensables pour mieux choisir les gènes de résistance à introduire dans les programmes d'innovation variétale, dans un contexte d'une agriculture durable et respectueuse de l'environnement, et d'élaborer de nouvelles stratégies de lutte contre les agents pathogènes. L'équipe s'intéresse en particulier à la caractérisation des facteurs de la plante hôte nécessaires au cycle infectieux des virus à ARN, dont les virus du genre *Potyvirus*. Les études portent principalement sur les plantes maraîchères, telles que les Solanacées, pour lesquelles nous disposons d'importantes données phénotypiques relatives

aux bases génétiques de la résistance aux virus. Du fait des données acquises montrant la conservation des mécanismes d'interaction avec les virus au sein du règne végétal, des études sont également conduites sur l'espèce modèle *Arabidopsis thaliana*, espèce hôte pour la plupart des virus des plantes maraîchères. Mon équipe a été la première à cloner un gène récessif de résistance aux virus à ARN chez les plantes : le gène *pvr2* chez le piment code pour un facteur d'initiation de la traduction eIF4E et des mutations dans ce gène sont responsables de la résistance à 2 potyvirus (Ruffel *et al.*, 2002).

Au niveau international, les travaux sur les gènes du complexe d'initiation de la traduction et leurs rôles dans les interactions plantes-virus portent principalement sur des aspects fonctionnels, notamment *via* la caractérisation de mutants d'*A. thaliana*. Au début de ma thèse, il n'existait pas de projet visant à caractériser et à comparer la variabilité naturelle des gènes *eIF4E* dans différents complexes d'espèces, et à mettre en relation cette variabilité avec la résistance aux virus à ARN.

Dans ce contexte, ce projet de thèse a été initié pour décrire puis comparer le polymorphisme de séquence des gènes récessifs *eIF4E* de résistance aux virus chez les plantes et mettre en relation ce polymorphisme avec la fonction de la protéine eIF4E, et notamment son rôle générique dans les interactions plantes-virus. Le polymorphisme a été analysé à 2 niveaux :

(i) celui du gène pour déterminer, chez des espèces végétales présentant des différences d'interactions avec les virus, quelles sont les pressions de sélection qui s'exercent sur les gènes *eIF4E*, en relation avec leur double fonction (dans les processus cellulaires et dans l'interaction plante-virus à ARN)

(ii) celui de la protéine, afin de déterminer l'impact des mutations identifiées dans les allèles codant pour les protéines de résistance sur les interactions entre eIF4E et VPg et sur la capacité de traduction des ARNm.

La stratégie employée repose sur l'association d'analyses fonctionnelles et d'études d'évolution moléculaire des gènes *eIF4E* chez 2 Solanacées (piment et tomate) infectées par les potyvirus et chez lesquelles des gènes récessifs de résistance correspondant aux facteurs eIF4E avaient été identifiés et chez l'espèce modèle *A. thaliana*, seule espèce où tous les gènes *eIF4E* sont connus. *In fine*, l'objectif est de mieux comprendre le "paradoxe eIF4E" : ce gène est essentiel pour la plante car il intervient dans de nombreux processus cellulaires liés

au métabolisme des ARNm; il ne doit donc pas accumuler de mutations pour maintenir l'intégrité fonctionnelle de la protéine. Et pourtant, ce gène est impliqué dans la résistance aux phytovirus *via* des substitutions en acides aminés, ce qui implique l'apparition de nouveaux allèles par mutations.

La suite du manuscrit s'articule autour de 4 chapitres : dans les chapitres 2, 3, et 4, nous aborderons respectivement les résultats concernant la caractérisation fonctionnelle de l'impact du polymorphisme de séquence des gènes *eIF4E* chez le piment sur l'interaction avec les potyvirus et sur les fonctions cellulaires, l'analyse comparée du polymorphisme des gènes *eIF4E* chez 2 Solanacées (piment et tomate) et *A. thaliana* afin de déterminer si des pressions de sélection s'exercent sur les gènes *eIF4E* de différentes familles végétales et finalement, une étude moléculaire préliminaire du rôle fonctionnel des paralogues *eIF4E*. Une discussion générale portant sur l'ensemble de mes résultats et les perspectives de ce travail constituent le cinquième chapitre.

## Chapitre II

# **I**mpact du polymorphisme des gènes *eIF4E* sur l'interaction *Capsicum annuum* - PVY et TEV

Des substitutions en acides aminés dans la protéine eIF4E1, codée par les allèles récessifs *pvr2*<sup>1</sup> et *pvr2*<sup>2</sup> chez *C. annuum*, sont responsables de la résistance à des virus à ARN du genre *Potyvirus*, tels que le PVY ou le TEV. Ces substitutions sont localisées dans 2 régions à la surface de la protéine eIF4E1 (nommées régions I et II), à proximité de la poche de fixation à la coiffe des ARNm (Robaglia and Caranta, 2006). Ces résistances récessives contrôlées par eIF4E1 peuvent par ailleurs être contournées par des mutations dans la protéine VPg des potyvirus.

Plusieurs études ont mis en évidence une interaction *in vitro* et *in vivo* entre les protéines eIF4E et VPg de différents pathosystèmes. Cette interaction est nécessaire pour l'infection virale et a un effet positif sur la réplication du virus. Par ailleurs, des travaux récents conduits avec les protéines codées par 2 allèles de résistance au locus *pvr2* (*pvr1* et *pvr2*<sup>2</sup>) montrent que les substitutions en acides aminés abolissent l'interaction avec la VPg du TEV (Yeam *et al.*, 2007).

Ce chapitre, divisé en 3 parties, présente une étude de l'impact du polymorphisme des gènes *eIF4E* sur l'interaction *C. annuum* - PVY et TEV. Une première partie, présentée sous forme d'article soumis à *The Plant Journal*, a pour objectif, *via* des études de diversité du gène *eIF4E1-pvr2* couplées à des analyses fonctionnelles, d'étudier l'impact des substitutions en acides aminés dans la protéine eIF4E1 sur le phénotype de résistance/sensibilité aux potyvirus, sur l'interaction physique avec les protéines VPg de souches avirulentes et virulentes de PVY et de TEV, et sur la capacité des protéines eIF4E1 à conserver leur fonction cellulaire chez la levure. La seconde partie vise à déterminer l'impact des combinaisons de substitutions en acides aminés dans 2 régions distinctes de la protéine eIF4E1 sur la résistance à large spectre. Enfin, la troisième partie a pour objectif d'étudier le rôle des autres protéines eIF4E *i.e.*, eIF4E2 et eIF(iso)4E, notamment dans le contournement des résistances contrôlées par eIF4E1-*pvr2*.

## **II.1. - Article 1 - Natural variation and functional analyses provide evidence for coevolution between plant eIF4E and potyviral VPg**

Carine Charron, Maryse Nicolai, Jean-Luc Gallois, Christophe Robaglia, Benoît Moury, Alain Palloix and Carole Caranta. *The Plant Journal*, soumis le 10 octobre 2007 et accepté avec corrections mineures le 12 novembre 2007.

Cet article décrit le polymorphisme observé au locus *eIF4E1-pvr2* chez 25 accessions de *C. annuum* représentatives de la diversité de cette espèce. Dix variants alléliques comprenant différentes combinaisons de mutations ont été identifiés. De façon très intéressante, toutes les substitutions nucléotidiques identifiées entraînent des changements en acides aminés dans la protéine eIF4E1. Ces analyses révèlent donc une importante diversité au locus *eIF4E1-pvr2* associée à un excès de substitutions non-synonymes, indiquant que ce gène est sous sélection diversifiante. A partir de ces données, la seconde partie de l'article porte sur la caractérisation fonctionnelle de l'impact de ces différentes substitutions en acides aminés sur (i) le phénotype de résistance aux PVY et TEV, (ii) l'interaction physique avec la protéine VPg de souches avirulentes et virulentes de PVY et TEV et (iii) la capacité à assurer la fonction de traduction des ARNm chez la levure. Les tests phénotypiques de résistance et les analyses de complémentation génétique montrent que 9 des 10 variants alléliques (*pvr2*<sup>1</sup> à *pvr2*<sup>9</sup>) correspondent à des allèles de résistance aux potyvirus. Tous ces allèles contrôlent la résistance au PVY et seul le locus *pvr2*<sup>2</sup> identifié chez les accessions Florida VR2 et Chay Angolano confère une résistance à la souche commune TEV-HAT. Chaque protéine eIF4E codée par les 9 variants alléliques porte différentes combinaisons de substitutions en acides aminés, principalement localisées au niveau des régions I et II à la surface de la protéine eIF4E1, précédemment montrées comme étant impliquées dans la résistance aux potyvirus. Par la méthode de double hybride dans la levure, nous avons pu montrer que ces substitutions en acides aminés abolissent l'interaction avec les protéines VPg des souches avirulentes de PVY et TEV alors que cette interaction est rétablie avec les VPg des souches virulentes. Enfin, les expériences *in vivo* de complémentation fonctionnelle dans une souche de levure délétée de son gène *eIF4E* ont permis de vérifier que les mutations dans les protéines eIF4E1 de résistance n'affectaient pas la fonction cellulaire de eIF4E1 (dont la fonction de traduction des ARNm). Ces résultats suggèrent donc que la protéine eIF4E1 peut accumuler des

mutations qui affectent la fixation de la VPg des potyvirus sans imposer un coût de fitness significatif puisque sans incidence sur les fonctions de eIF4E1 dans la traduction des ARNm et/ou dans le cycle et la croissance cellulaire.

Les résultats de ce travail argumentent donc en faveur d'une coévolution de type course aux armements entre le facteur eIF4E1 de *C. annuum* et la VPg des potyvirus. Comme dans le cas des résistances médiées par *Pto* chez la tomate et *L* chez le lin contre *P. syringae* et *M. lini* respectivement (Dodds *et al.*, 2006; Rose *et al.*, 2007), l'interaction directe entre les protéines d'avirulence et de résistance se traduit par le maintien d'un grand nombre de variants alléliques par le biais d'une sélection diversifiante.

# Natural variation and functional analyses provide evidence for coevolution between plant eIF4E and potyviral VPg

---

Carine Charron<sup>1</sup>, Maryse Nicolai<sup>1</sup>, Jean-Luc Gallois<sup>1</sup>, Christophe Robaglia<sup>3</sup>, Benoît Moury<sup>2</sup>, Alain Palloix<sup>1</sup> and Carole Caranta<sup>1\*</sup>

<sup>1</sup> INRA-UR1052, Genetics and Breeding of Fruits and Vegetables, Dom. St Maurice, BP94, F-84143 Montfavet, France,

<sup>2</sup> INRA, Plant Pathology, Dom. St Maurice, BP94, F-84143 Montfavet, France, and

<sup>3</sup> Laboratoire de Génétique et Biophysique des Plantes, CEA-CNRS-Université Aix-Marseille II, Faculté des Sciences de Luminy, F-13009 Marseille, France

**\* Corresponding author:**

Carole CARANTA,

INRA-UR1052, Genetics and Breeding of Fruits and Vegetables, Dom. St Maurice, BP 84, F-84143 Montfavet, France.

Phone : 334 327 227 26, Fax : 334 327 227 02,

e-mail : carole.caranta@avignon.inra.fr

**Running title:** Coevolution between eIF4E and VPg

**Key words:** eukaryotic translation initiation factor 4E, VPg, *Potato virus Y*, *Tobacco etch virus*, *Capsicum annuum*, recessive virus resistance

**GenBank accession numbers:** AY122052, AF521964, AF5211965, AY723737, AY723738, AY723739, AY723740, EU106863, EU106864 and EU106865

---

**Word count:** TOTAL: 6718 - Abstract: 239 - Introduction: 1278 - Results: 2048 -

Discussion: 1909 - Experimental procedures: 1238 - References: 1415 –

With 3 Tables and 6 Figures

## Summary

Amino acid substitutions in the eukaryotic translation initiation factor 4E (eIF4E) result in recessive resistance to potyviruses in a range of plant species, including *Capsicum* spp. Correspondingly, amino acid changes in the central part of the viral genome-linked protein (VPg) are responsible for the potyvirus's ability to overcome eIF4E-mediated resistance. A key observation was that the physical interaction between eIF4E and the VPg is required for viral infection while eIF4E mutations causative of resistance prevent VPg binding and inhibit the viral cycle. In this study, polymorphism analysis of the *pvr2-eIF4E* coding sequence in a worldwide sample of twenty five *C. annuum* accessions identified ten allelic variants with exclusively non-synonymous variations clustered in two surface loops of eIF4E. Resistance and genetic complementation assays demonstrated that *pvr2* variants, each with signature amino acid changes, corresponded to potyvirus resistance alleles. Systematic analysis of the interactions between eIF4E proteins encoded by the ten *pvr2* alleles and VPgs of virulent and avirulent *Potato virus Y* (PVY) and *Tobacco etch virus* (TEV) strains demonstrated that resistance phenotypes arose from the disruption of the interaction between eIF4E and VPg, and that viral adaptation to eIF4E-mediated resistance resulted from the restored interaction with the resistance protein. Complementation of an eIF4E knock-out yeast strain by *C. annuum* eIF4E proteins further shows that amino acid changes did not impede essential eIF4E functions. Altogether, these results argue in favour of a coevolutionary arms race between *Capsicum* eIF4E and potyviral VPg.

## Introduction

In recent years, several studies have implicated the eukaryotic translation initiation factors in resistance against specific RNA viruses in both monocots and dicots, leading to the definition of a new class of natural virus resistance genes (Maule *et al.*, 2007; Robaglia and Caranta, 2006). They differ from the major class of disease resistance genes encoding nucleotide binding site-leucine rich repeat proteins (R genes), by their recessive inheritance, a feature related to the obligate parasite nature of viruses. Because the small genome of plant viruses encodes generally less than a dozen proteins, the viral infectious cycle largely relies on the use of cellular factors, making the completion of the viral cycle a complex interplay between

virus-encoded and host-encoded factors. Therefore, the absence or the inadequacy of a single host factor leads to full or partial resistance to viruses, and these non-functional versions of host factors result in recessive inheritance.

Although many host factors are required for plant virus infection (Kushner *et al.*, 2003), molecular cloning of recessive resistance genes identified in the natural diversity of crops ranging from pepper, lettuce, pea, melon, barley and rice has so far only revealed a group of proteins linked to the translation machinery : the eukaryotic initiation factor 4E (eIF4E) and to a lesser extent, the eukaryotic initiation factor 4G (eIF4G). eIF4E recognizes and binds to the 5' cap structure of mRNA (m<sup>7</sup>GpppN) and also to eIF4G to form the eIF4F complex. Additional translation initiation factors and the 40S ribosomal subunit are then recruited to initiate mRNA translation (Browning, 2004). In plants, several genes encode for two protein subfamilies, eIF4E and eIF(iso)4E, having both overlapping and isoform-specific biological roles (Combe *et al.*, 2005; Gallie and Browning, 2001; Rodriguez *et al.*, 1998).

The majority of recessive resistances controlled by eIF4E functions against potyvirus infection, although eIF4E has also been implicated in natural resistance to a bymovirus (genus related to *Potyvirus*) and a carmovirus (Nieto *et al.*, 2006; Stein *et al.*, 2005). The genus *Potyvirus* is the largest among plant viruses and includes the most common and destructive viruses for a number of crops worldwide. The potyviral RNA is polyadenylated at its 3' end and is covalently linked to a 25 kD virus-encoded protein named VPg, replacing the cap structure of mRNAs at the 5' end (Riechmann *et al.*, 1992). VPg is cleaved from the NIa protein, composed of N-terminal VPg and C-terminal protease domains, and plays central roles in replicative and proteolytic functions (Revers *et al.*, 1999). A key observation is the ability of the VPg (or the NIa precursor) to bind eIF4E isoforms in yeast two-hybrid and *in vitro* binding assays (Leonard *et al.*, 2000; Leonard *et al.*, 2004; Schaad *et al.*, 2000; Wittmann *et al.*, 1997).

eIF4E-mediated recessive resistances against potyviruses all result from a small number of amino acid changes, most of which being non-conservative (Gao *et al.*, 2004; Kang *et al.*, 2005a; Nicaise *et al.*, 2003; Ruffel *et al.*, 2002; Ruffel *et al.*, 2005). Structural models predicted that these mutations are on the surface of eIF4E and clustered in two neighbouring regions, one near the cap-binding pocket (hereafter named region I), and another rotated 90° from the cap-binding pocket (named region II) (Monzingo *et al.*, 2007; Robaglia and Caranta, 2006). Recent analysis of the expression level of eIF4E mRNAs and proteins from various pepper genotypes confirmed that phenotypic differences in potyviral infection are determined by the amino acid changes themselves rather than by other components regulating expression

or accumulation of eIF4E (Yeam *et al.*, 2007). A further common feature linking eIF4E-mediated resistance to potyviruses is that VPg was identified as the virulence determinant. Amino acid mutations in the central part of the protein are responsible for the virus's ability to overcome recessive resistances controlled by eIF4E and cause a compatible interaction (Moury *et al.*, 2004). The exact mechanism by which eIF4E mutations control resistance remains to be elucidated, but recent protein-protein interaction studies carried out with yeast two-hybrid, GST pull-down and bimolecular fluorescence complementation (BiFC) assays argue in favour of disruption or impairing of the direct interaction between eIF4E and VPg proteins. Therefore, a physical interaction between eIF4E and viral VPg is necessary for viral infection and amino acid substitutions induce altered function with respect to VPg binding (Kang *et al.*, 2005a).

Host-pathogen interactions are an important force shaping organism diversity and, according to the extreme intimacy between plant and virus life cycles, co-evolution with complex and dynamic selection pressures on the genes involved in the interaction is probable. In this respect, the interaction between *Capsicum* spp. and potyviruses provides an excellent system for functional and evolutionary studies of eIF4E-mediated resistance because potyvirus infection is one of the major constraints to pepper cultivation worldwide and both the resistance protein and the viral ligand (avirulence protein) have been characterized at the molecular level. Among the five potyviruses reported to infect peppers, *Potato virus Y* (PVY) is widespread throughout most of the cultivated areas while *Tobacco etch virus* (TEV) occurs mainly in North and Central America and in the Caribbean (Green and Kim, 1991). Facing this pathogen diversity, resistance appears to be a usual outcome of the interaction between *Capsicum* and potyviruses. Approximately 40% of *Capsicum* accessions are resistant to common strains of PVY and a number of recessive genetic factors have been described and confer different spectra and levels of resistance against potyviruses (Kyle and Palloix, 1997; Palloix *et al.*, pers. com.). Among these, the *pvr2* gene has been successfully used for decades in crop breeding programs as an effective and stable source of resistance (Greenleaf, 1986). *pvr2* was demonstrated to encode for an *eIF4E* gene, and up to now, two resistance alleles, *pvr2*<sup>1</sup> and *pvr2*<sup>2</sup>, were identified in *C. annuum* and one, named *pvr1*, in *C. chinense* (Kang *et al.*, 2005a; Ruffel *et al.*, 2004; Ruffel *et al.*, 2002). Interestingly, these alleles differ not only by signature amino acid substitutions localized to regions I and II of eIF4E but also by their resistance spectrum. *pvr2*<sup>1</sup> is only effective against common strains of PVY (*i.e.*, pathotype PVY-0) while *pvr2*<sup>2</sup> is effective against PVY-0 and -0,1 and also controls common strains of TEV. Finally, the *pvr1* allele from *C. chinense* confers broad-spectrum resistance to common

strains of PVY, TEV and *Pepper mottle virus* (Kyle and Palloix, 1997). These allelic variants lead to a common mechanism of resistance, blocking viral accumulation in inoculated leaves (Deom and He, 1997; Ruffel *et al.*, 2002). From the virus side, amino acid substitutions in the central part of VPg were shown to determine the ability of PVY to infect pepper genotypes homozygous for the *pvr2*<sup>1</sup> or *pvr2*<sup>2</sup> resistance alleles (Ayme *et al.*, 2006; Moury *et al.*, 2004). Similarly, the avirulence determinant of TEV has been identified as the VPg (Kang *et al.*, 2005a).

In this article, we examine the pattern of polymorphism of the *pvr2-eIF4E* coding sequence among a set of twenty five *C. annuum* accessions representing a worldwide sample from diverse origins. Ten different allelic variants were found among which all nucleotide polymorphisms resulted in amino acid changes, mostly clustered in region I and II of the eIF4E protein. We performed a systematic functional analysis of the impact of these distinct combinations of amino acid substitutions on (i) the resistance spectrum against PVY and TEV, (ii) the physical interaction with the VPg of avirulent and virulent PVY and TEV strains, and (iii) the ability to complement an eIF4E-knockout yeast strain. Our results show that this extensive non-synonymous variation modulates interactions with potyviruses, and also provide functional evidence for coevolution between *C. annuum* eIF4E and potyviral VPg.

## RESULTS

### **Sequence analysis at *pvr2-eIF4E* identified an extensive non-synonymous variation**

Coding sequence variability at *pvr2-eIF4E* was surveyed in twenty five *C. annuum* accessions representing a worldwide sample from diverse geographic regions (Figure 1). The sample included YW, YY and F for which resistance phenotypes against potyviruses and eIF4E cDNA sequences were previously determined and assigned to *pvr2*<sup>+</sup>, *pvr2*<sup>1</sup> and *pvr2*<sup>2</sup> alleles, respectively. *pvr2*<sup>1</sup> and *pvr2*<sup>2</sup> encode for eIF4E resistance proteins carrying signature amino acid changes at positions 67, 79 and 109 (Ruffel *et al.*, 2002).

The sequenced coding regions span 687 bp from start to stop codons and encode for a predicted protein of 228 amino acids. The eIF4E coding sequences are intact in all genotypes with no indel polymorphism. Ten allelic variants were identified based on nine polymorphic nucleotide sites localized in exon 1 (6 single nucleotide polymorphisms -SNPs- between nucleotides 196 and 236), exon 2 (2 SNPs at positions 319 and 325) and exon 4 (1 SNP at

position 614). Only two different nucleotides were present at each polymorphic site. To compare with the polymorphism in non-coding regions, sequences of eIF4E introns 2, 3 and 4 (representing 1347 bp) were compared between the twenty five genotypes. Interestingly, there was a significantly higher proportion of polymorphic sites in the coding regions in comparison with non-coding regions (9 SNPs within 687 bp *versus* 6 SNPs within 1347 bp; chi-square test of homogeneity,  $\chi^2=4.65$ ;  $P=0.03$ ).

A major observation from this study was that all SNPs lead to amino acid changes in the deduced eIF4E proteins (Table 1). Ten accessions from diverse geographic origins shared the same nucleotide and amino acid sequences with YW, previously assigned to *pvr2*<sup>+</sup>. Each new allelic variant was designated *pvr2* with a numerical superscript indicating the new allele following the chronological precedent (*pvr2*<sup>2</sup>), according to the nomenclature proposed by Kyle and Palloix (1997). This nomenclature will be used throughout the paper. In comparison with *pvr2*<sup>+</sup>, the nine other allelic variants presented from one to four amino acid substitutions and all nine shared substitutions at positions 66-68 of the protein. The *pvr2*<sup>3</sup> allele was observed three times; *pvr2*<sup>1</sup>, *pvr2*<sup>2</sup> and *pvr2*<sup>6</sup> were present twice and *pvr2*<sup>4</sup>, *pvr2*<sup>5</sup>, *pvr2*<sup>7</sup>, *pvr2*<sup>8</sup> and *pvr2*<sup>9</sup> were observed only once (Table 1).

In comparison with previous studies (Ruffel *et al.*, 2002; Kang *et al.*, 2005), this survey identified 7 new allelic variants within *C. annuum* and 6 new polymorphic sites at positions 66, 68, 73, 74, 107 and 205 of the eIF4E protein. None of the amino acid substitutions, except D109N, targeted amino acids involved in m<sup>7</sup>GTP binding, amino acids important for stabilizing the structure of the protein or amino acids conserved in plants (boxed in red, black and green, respectively Figure 2). Based on sequence alignments, the D109N substitution from *pvr2*<sup>1</sup> and *pvr2*<sup>2</sup> corresponds to amino acid D96 of the wheat eIF4E which is involved in binding the cap-structure through the stabilization of amino acid R158 by a salt bridge (Protein Data Bank accession 2IDR). According to the 3D model for plant eIF4E (Monzingo *et al.*, 2007), all substitutions are localized at, or near, the surface of the eIF4E protein, where mutations are tolerated and do not destabilize the core structure. Moreover, all except D109N corresponded to non-conservative substitutions in the respective proteins and to a change of a hydrophobic amino acid into a hydrophilic one, but none are predicted to have a major effect on the overall folding of the protein (data not shown). One important observation is that all amino acid changes, except D205G, occurred within regions I and II on different facets of the eIF4E protein where residues were demonstrated to be involved in potyvirus or bymovirus resistance (Robaglia and Caranta, 2006). Amino acid substitutions P66T, V67E, A68E,

A73D, A74D and L79R localized in region I, near the cap-binding site, and substitutions G107R and D109N in region II, rotated 90° from the cap-binding pocket.

### ***pvr2<sup>3</sup>* - to *pvr2<sup>9</sup>* -*eIF4E* corresponds to potyvirus resistance alleles**

To test if amino acid changes identified in the new eIF4E allelic variants could be related to potyvirus resistance, a systematic study to characterize infectivity of PVY and TEV strains was conducted (Table 2). As expected, all plants from the genotype YW mechanically inoculated with PVY-LYE84, PVY-SON41, TEV-HAT or TEV-CAA10 presented obvious typical mosaic symptoms in apical non-inoculated leaves at 7 days post-inoculation (dpi) and exhibited high DAS-ELISA values at 30 dpi. The ten genotypes belonging to *pvr2<sup>+</sup>*-*eIF4E* presented a phenotype similar to YW *i.e.*, susceptible to all virus strains. The other genotypes were resistant to PVY-LYE84 with no systemic symptoms and no virus detected by DAS-ELISA at 30 dpi. Six genotypes belonging to *pvr2<sup>5</sup>*, *pvr2<sup>6</sup>*, *pvr2<sup>7</sup>*, *pvr2<sup>8</sup>* and *pvr2<sup>9</sup>* variants were also resistant to PVY-SON41. For TEV, all accessions were susceptible to TEV-CAA10 and only accessions F and Chay Angolano with the *pvr2<sup>2</sup>* allele were resistant to TEV-HAT (Table 2).

Genetic complementation analyses were then conducted to determine whether PVY and TEV resistance was controlled by *pvr2* allele(s) or not. In a first step, the recessive nature of the resistance was checked by resistant X susceptible crosses and, in a second step, allelism relationships were evaluated by resistant X resistant crosses. In this case, if resistance is controlled by distinct loci, then only susceptible offspring should be recovered in F<sub>1</sub> progeny; on the contrary, if they are allelic, the F<sub>1</sub> should be uniformly resistant. Allelic relationships between PVY resistant genotypes (YY known to carry the *pvr2<sup>1</sup>* resistance allele and Perennial, HD-C69, PI322719, SC81, Maroc1, Serrano Vera Cruz, PI195301 and Chile de Arbol) were evaluated by inoculation with PVY-LYE84 and allelic relationships between TEV resistant genotypes (F carrying the *pvr2<sup>2</sup>* resistance allele and Chay Angolano) were evaluated by inoculation with TEV-HAT. In each assay, plants of the susceptible control YW developed bright systemic mosaic symptoms characteristic of PVY or TEV on newly emerged non-inoculated leaves and accumulated PVY and TEV to high levels in both inoculated and non-inoculated leaves. F<sub>1</sub> hybrids between the PVY-LYE84 resistant genotypes Perennial, HD-C69, PI322719, SC81, Maroc1, Serrano Vera Cruz, PI195301 and Chile de Arbol and the susceptible genotype YW (*pvr2<sup>+</sup>*) were all susceptible to PVY-LYE84 demonstrating fully recessive inheritance for resistance to PVY-LYE84 (Figure 3). F<sub>1</sub> hybrids between the same genotypes and the PVY-LYE84 resistant genotype YY (*pvr2<sup>1</sup>*) were uniformly resistant to

PVY-LYE84 showing that resistance is controlled by alleles of *pvr2* (Figure 3). Similarly, the F<sub>1</sub> between YW and Chay Angolano was susceptible to TEV-HAT indicating recessive inheritance, and the F<sub>1</sub> between the TEV-HAT resistant genotypes F (*pvr2*<sup>2</sup>) and Chay Angolano was resistant to TEV-HAT indicating that resistance of Chay Angolano is controlled by *pvr2*<sup>2</sup>.

Genetic analysis demonstrated that PVY and TEV resistances identified in genotypes belonging to allelic variants *pvr2*<sup>3-</sup>, *pvr2*<sup>4-</sup>, *pvr2*<sup>5-</sup>, *pvr2*<sup>6-</sup>, *pvr2*<sup>7-</sup>, *pvr2*<sup>8-</sup> and *pvr2*<sup>9-eIF4E</sup> were controlled by alleles at the *pvr2* locus. In addition to distinct combinations of amino acid changes, these alleles differ by their resistance spectrum against PVY and TEV strains. All alleles except *pvr2*<sup>+</sup> control resistance to PVY-LYE84, belonging to pathotype 0. In contrast, PVY-SON41 classified as pathotype 0,1,2 due to its capacity to break down the *pvr2*<sup>1</sup> and *pvr2*<sup>2</sup> resistance alleles (Moury *et al.*, 2004), infect plants homozygous for *pvr2*<sup>3</sup> and *pvr2*<sup>4</sup> but not those homozygous for *pvr2*<sup>5</sup>, *pvr2*<sup>6</sup>, *pvr2*<sup>7</sup>, *pvr2*<sup>8</sup> and *pvr2*<sup>9</sup>. Concerning TEV resistance, only the *pvr2*<sup>2</sup> allele controls TEV-HAT in addition to PVY-LYE84. This resistance allele is overcome by the TEV-CAA10 strain.

### **Amino acid substitutions at *pvr2-eIF4E* impair physical interactions with VPg of avirulent PVY and TEV strains**

With the dual objective to test the impact of the distinct combinations of amino acid changes identified in eIF4E proteins on the interaction with the viral VPg, and to get insights into molecular mechanisms underlying viral adaptation to eIF4E-mediated resistance (*i.e.*, resistance breaking), we performed a systematic study of the physical interaction between the ten eIF4E proteins encoded by alleles at the *pvr2* locus and the VPg of the two PVY and two TEV strains. Recent studies demonstrated that viral adaptation of PVY-SON41 to *pvr2*<sup>1</sup> and *pvr2*<sup>2</sup> resistance alleles results from five amino acid changes in the central part of the VPg cistron of PVY (Table 3; Moury *et al.*, 2004). Similarly, VPg of the virulent strain TEV-CAA10 differs from those of TEV-HAT by 13 amino acids substitutions, 4 of them occurring in the 23-codon-long region in the central part of the protein which is known to be involved in resistance breaking (Table 3). Since previous results showed that a physical interaction between eIF4E and VPg seems necessary for viral infection and that eIF4E resistance proteins show altered function with respect to VPg binding (Kang *et al.*, 2005; Yeam *et al.*, 2007), it is tempting to speculate that viral adaptation to eIF4E-mediated resistance results in a restoration of the interaction with the resistance protein.

Expression of the ten *pvr2-eIF4E* coding sequences in yeast was checked by western blot analysis and showed that all proteins were stably expressed at similar levels (Figure 4). In three independent protein-protein interaction experiments, each conducted in triplicate, the fusion protein translated from *pvr2*<sup>+</sup> (from YW) interacted in yeast with VPg proteins from the two PVY and TEV strains. *pvr2*<sup>+</sup>-eIF4E interactions with TEV VPg were stronger than those with PVY VPg as yeast transformants complemented both adenine and histidine auxotrophy (Figure 5A). On the contrary, the pattern of interaction of eIF4E proteins translated from *pvr2*<sup>1</sup> to *pvr2*<sup>9</sup> alleles obviously differed depending on the VPg used (Figure 5). eIF4E encoded by *pvr2*<sup>1</sup>, *pvr2*<sup>3</sup> and *pvr2*<sup>4</sup> interacted with VPg of PVY-SON41, TEV-HAT and TEV-CAA10 but failed to interact with VPg from PVY-LYE84. eIF4E encoded by *pvr2*<sup>2</sup> interacted with PVY-SON41 VPg and TEV-CAA10 VPg but not with PVY-LYE84 VPg and TEV-HAT VPg. Finally, eIF4E proteins encoded by *pvr2*<sup>5</sup>, *pvr2*<sup>6</sup>, *pvr2*<sup>7</sup>, *pvr2*<sup>8</sup> and *pvr2*<sup>9</sup> interacted with VPgs from the two TEV strains but not with VPgs of the two PVY strains. Results of the interaction assays between *pvr2*<sup>1</sup>-, *pvr2*<sup>2</sup>-, *pvr2*<sup>3</sup>- and *pvr2*<sup>4</sup>-eIF4E and VPg of PVY-LYE84 and PVY-SON41, and between *pvr2*<sup>2</sup>-eIF4E and VPg of TEV-HAT and TEV-CAA10 are of special interest. These proteins control resistance to PVY-LYE84 and to TEV-HAT for *pvr2*<sup>2</sup>, but are overcome by PVY-SON41 and TEV-CAA10. In all cases, the VPg of avirulent strains failed to interact with the eIF4E resistance proteins encoded by *pvr2*<sup>1</sup>, *pvr2*<sup>2</sup>, *pvr2*<sup>3</sup> and *pvr2*<sup>4</sup> whereas the VPg of virulent strains interacted with these resistance proteins.

In summary, these results identified four patterns of interaction between the ten eIF4E proteins and the VPg proteins from four strains that perfectly match the resistance/susceptibility phenotypes controlled by alleles at the *pvr2* locus: susceptibility correlated with interaction with the VPg of virulent PVY and TEV strains whereas resistance correlated with impaired/disrupted interaction with the VPg of avirulent strains (Figure 5B). This systematic analysis of eIF4E/VPg interactions confirmed and expanded the observation that eIF4E proteins corresponding to resistance alleles showed altered function with respect to VPg binding. These data also supported the hypothesis that the overcoming of eIF4E-mediated resistance arose from the restoration of the interaction with VPg.

### ***pvr2*-eIF4E recapitulates essential eIF4E functions in yeast**

To address the functional consequences of amino acid substitutions identified in *pvr2*-eIF4E proteins encoded by the different allelic variants, we analysed whether these proteins were able to rescue the growth of the haploid yeast strain JO55 in which the sole chromosomal copy of *eIF4E* is deleted and replaced by human *eIF4E* (Altmann *et al.*, 1989). Because the

expression of the human *eIF4E* cDNA required for yeast growth is galactose-dependent, this system provides an *in vivo* functional assessment of any *eIF4E* coding sequence under the control of a glucose-inducible promoter (Hughes *et al.*, 1999) and has been successfully used to address the function(s) of plant eIF4E isoforms (Rodriguez *et al.*, 1998).

Empty yeast strains JO55 or JO55 transformed with an empty p424GBP/TRP1 vector were used as negative controls. The positive control was JO55 transformed with p424GBP/TRP1:*At-eIF4E* (At4g18040). The ten *pvr2-eIF4E* coding sequences from susceptible (*pvr2*<sup>+</sup>) and resistant (*pvr2*<sup>1</sup> to *pvr2*<sup>9</sup>) genotypes allowed the growth of the yeast strain in medium containing glucose in a similar manner as *A. thaliana eIF4E* cDNA (Figure 6). The ability of these exogenous eIF4Es to complement the eIF4E-knockout yeast strain indicated that *C. annuum* eIF4E shares functions with yeast and human eIF4Es, and that amino acid substitutions identified in resistance proteins do not impede essential eIF4E functions including those linked to mRNA translation.

## DISCUSSION

This work which combines genetic diversity and functional analyses, provides evidence for coevolution between *C. annuum* eIF4E and potyviral VPg. The nucleotide variation analysis at the *pvr2-eIF4E* locus identified solely non-synonymous mutations and a significantly higher proportion of polymorphic sites in coding regions in comparison with non-coding regions (*i.e.*, introns) of the gene. These two features strongly suggest that this gene is subjected to diversifying selection. Resistance assays and genetic complementation analysis demonstrated that nine among the ten identified allelic variants correspond to potyvirus resistance alleles each with signature amino acid substitutions. The majority of these alleles controls PVY although *pvr2*<sup>2</sup> also confers resistance to the common strain of TEV. This is in agreement with previous surveys for potyvirus resistance in *Capsicum* spp. which show a high proportion of resistance to PVY and a much lower proportion of resistance to TEV.

Each resistance protein encoded by the nine allelic variants carried a distinct combination of non-conservative amino acid changes largely localized on the surface of eIF4E in regions I and II, demonstrated to be targets in virus resistance (Robaglia and Caranta, 2006). Previous studies identified one or two eIF4E resistance alleles per species with signature amino acid substitutions in region I or II (in pepper, lettuce and pea) or both (in pepper, tomato and pea) (Gao *et al.*, 2004; Kang *et al.*, 2005; Nicaise *et al.*, 2003; Ruffel *et al.*, 2002; Ruffel *et al.*,

2005). Our data illustrate the occurrence and maintenance of a great diversity of combinations of amino acid substitutions leading to resistance and also emphasize the importance of amino acid changes in region I with respect to resistance in *C. annuum*, at least against PVY, since resistance alleles mainly control this potyvirus. Indeed, all resistance proteins shared, within region I, one or two (*pvr2*<sup>5</sup>) amino acid substitutions at positions 66-68 and for some alleles additional changes at positions 73-74 (in *pvr2*<sup>5</sup>, *pvr2*<sup>6</sup>, *pvr2*<sup>9</sup>) or 79 (in *pvr2*<sup>1</sup>, *pvr2*<sup>2</sup>, *pvr2*<sup>7</sup>). Interestingly, the allelic variant *pvr2*<sup>4</sup> identified in a single accession presented a single amino acid substitution (V67E) indicating that this change alone is sufficient to compromise PVY infection. Two variants (*pvr2*<sup>2</sup> and *pvr2*<sup>8</sup>) showed an additional amino acid change in region II of the protein (G107R or D109N), and only *pvr2*<sup>2</sup> was involved in resistance to TEV in addition to PVY. Finally, four variants shared the D205G mutation in the C-terminal region of eIF4E. Although the N- and C-termini of eIF4E are more polymorphic than the core of eIF4E (represented by approx. 170 amino acids; Joshi *et al.*, 2005), amino acid changes in the C-terminal region have never been described in potyvirus resistance proteins from other species. However, a single amino acid change mapped to the C-terminal part of eIF4E was recently demonstrated to control recessive resistance of melon to *Melon necrotic spot virus* (genus *Carmovirus*) through a mechanism distinct from the one acting against potyviruses (Nieto *et al.*, 2006). The conservation between unrelated plant species of the position of mutations leading to potyvirus resistance together with the fact that carmoviruses can infect peppers could suggest an involvement of D205G in another eIF4E-mediated resistance system targeted against a distinct viral genus.

Large-scale analysis of the interactions between eIF4Es and VPgs from two potyviruses demonstrates that the resistance phenotypes against PVY and TEV arise from disruption of the direct interaction between eIF4E and VPg, obviously linking amino acid mutations and altered function with respect to VPg binding. The occurrence of mutations in two regions on different facets of the eIF4E molecule suggests that VPg may have two binding sites for optimal interaction. However, the results of interaction studies obtained with the eIF4E protein encoded by the *pvr2*<sup>4</sup> resistance allele demonstrate that disruption of a single contact in region I of the protein is sufficient to abolish binding ability of the VPg of PVY. These data together with the systematic occurrence of amino acid changes in region I support the key role of mutations in this particular region with respect to VPg binding and to the outcome of the interaction with PVY. A recent functional analysis conducted in *Capsicum* using eIF4E constructs with induced point mutations affecting various amino acids has identified the amino acid change G107R in region II (initially identified in the *C. chinense pvr1* resistance

allele) as the sole change responsible for abolishing binding ability of the TEV-HAT VPg (Yeam *et al.*, 2007). It is striking that in our study, the *pvr2*<sup>8</sup> allele was shown to carry amino acid changes A68E and G107R but was fully susceptible to TEV-HAT; we also observed in yeast two hybrid experiments that the eIF4E protein encoded by *pvr2*<sup>8</sup> binds to the VPg of TEV-HAT. These contradictory observations have two possible explanations: (i) amino acid change A68E modulates the property of G107R or (ii) TEV resistance mediated by the G107R change also involves another ligand which is functional (or present) in *C. chinense* but non-functional (or absent) in *C. annuum*. Among naturally-occurring eIF4E resistance alleles against potyviruses, *pot1* from tomato and *sbm1*<sup>1</sup> from pea presented amino acid substitutions in both region I and II, *mol*<sup>2</sup> from lettuce and *sbm1*<sup>2</sup> from pea presented mutations only in region I and *mol*<sup>1</sup> from lettuce presented mutations only in region II of the protein (Gao *et al.*, 2004; Nicaise *et al.*, 2003; Ruffel *et al.*, 2005). Altogether, these data strongly indicate that the precise contact point between eIF4E and potyviral VPg is optimized for each potyvirus, even within a single plant species since, in pepper, the mutation V67E in region I is responsible for the disruption of binding of the PVY VPg whereas the G107R in region II was demonstrated to be responsible for the disruption of binding of the TEV VPg.

One important result from the present study is that viral adaptation to eIF4E-mediated resistance through amino acid changes in the central part of the VPg results from the restoration of the physical interaction with the resistance protein. Since two *eIF4E* genes and one *eIF(iso)4E* homolog were identified in the pepper genome (Kang *et al.*, 2005; Ruffel *et al.*, 2002; Ruffel *et al.*, 2006) and since analyses conducted in pepper and *A. thaliana* showed that potyviruses differ in their ability to use eIF4E isoforms (Duprat *et al.*, 2002; Lellis *et al.*, 2002; Ruffel *et al.*, 2006; Sato *et al.*, 2005), two functional models for virulence can be proposed. Amino acid changes in the VPg of PVY and TEV can allow for the overcoming of resistance alleles at the *pvr2* locus either (i) by restoring the interaction with a mutated eIF4E resistance protein or (ii) from *de novo* interaction with another eIF4E isoform, or with an alternative partner. Our protein-protein interaction experiments obviously argue in favour of the first hypothesis, because the VPg of avirulent PVY-LYE84 and TEV-HAT strains failed to interact with the eIF4E resistance proteins encoded by *pvr2*<sup>1</sup>, *pvr2*<sup>2</sup>, *pvr2*<sup>3</sup> and *pvr2*<sup>4</sup> whereas the VPg of virulent PVY-SON41 and TEV-CAA10 strains interacted with these resistance proteins. This is also in agreement with the demonstration that transient expression of *pvr2*<sup>1</sup>-eIF4E in *pvr2*<sup>2</sup> plants promoted the accumulation of PVY pathotype 0,1 (Ruffel *et al.*, 2002). However, recent data obtained with the *A. thaliana*-*Turnip mosaic virus* pathosystem showed that a single amino acid change in the central domain of the TuMV VPg

permits the overcoming of the complete resistance of an *A. thaliana* mutant bearing a transposon-induced null allele for the *eIF(iso)4E* gene (German-Retana, Revers and Caranta, *unpublished data*). Since this *A. thaliana* line completely lacks both eIF(iso)4E mRNA and protein (Duprat *et al.*, 2002), and given the hypothesis that eIF4E is required for potyvirus infection, these results argue in favour of mutations conferring to VPg a new or increased affinity to another eIF4E isoform or alternative susceptibility factors. Altogether, this indicates that potyviruses have probably developed different strategies to interact with the host's translational apparatus and to counter eIF4E-mediated resistance through mutations in VPg.

The main function of eIF4E is cap binding during the initiation of host protein synthesis although other functions in cell growth and cell cycle have been discovered such as the regulation of export of some mRNA from the nucleus (Culjkovic *et al.*, 2007). Therefore, a prerequisite for the maintenance of amino acid variations leading to resistance is that the resistance gene can accumulate mutations that affect the recognition of the viral VPg without imposing a significant fitness cost by impairing mRNA translation and/or other eIF4E function. Complementation of the eIF4E knock-out yeast strain by *C. annuum* eIF4E proteins showed that amino acid changes identified in resistance proteins do not drastically impair mRNA translation in yeast. Our results therefore extend those of Kang *et al.* (2005) which show that resistance proteins encoded by *pvr2*<sup>1</sup> and *pvr2*<sup>2</sup> maintain an *in vitro* cap-binding activity. However, the eIF4E proteins encoded by the *pvr1* allele from *C. chinense* and the *sbm1* allele from pea were demonstrated to abolish cap binding (Gao *et al.*, 2004; Kang *et al.*, 2005) and the *C. annuum pvr6* resistance locus was found to be a null allele of *eIF(iso)4E* (Ruffel *et al.*, 2006). Altogether this suggests that a trade-off between potyvirus resistance and eIF4E functions can result in defective eIF4E proteins. A recent study of *in vitro* interactions indicated that the cap analogue m<sup>7</sup>GDP and the VPg of LMV bind to lettuce eIF4E at two partially overlapping sites, however the binding of one ligand reduces the affinity for the other about 15-fold (Michon *et al.*, 2006). Therefore, a partial overlap between the VPg binding domain and the cap-binding domain together with the observation that the precise contact point between eIF4E and potyviral VPg is optimized for each potyvirus may explain these differences in the maintenance of cap-binding activities. Nevertheless, the occurrence of several eIF4E proteins in plants and their ability to compensate for one another (Combe *et al.*, 2005; Duprat *et al.*, 2002; Gallie and Browning, 2001; Rodriguez *et al.*, 1998) probably modulate the overall impact of the decreased cap-binding ability of some eIF4E resistance proteins on plant fitness and allow, indirectly, for the evolution of resistance.

In conclusion, there are a limited number of studies which combine polymorphism analysis and functional analysis and these are of particular interest in obtaining insights into antagonistic relationships between plants and pathogens. In this study, the observed exclusive non-synonymous variation, associated with their great diversity, provides evidence for diversifying selection acting on the *pvr2-eIF4E* locus. In addition, the established functional role of this variation in potyvirus resistance suggests that this selective pressure is exerted, at least partially if not totally, by phytopathogenic potyviruses. From the virus side, amino acids from the central domain of the VPg avirulence protein were also demonstrated to be subjected to positive selection (Moury *et al.*, 2004). Large-scale yeast two-hybrid assays indicate that the outcome of the eIF4E-VPg interaction is a key determinant of potyviral infection and that virulence arises from the restoration of the interaction with the mutated resistance protein. Altogether, our results argue in favour of a coevolutionary arms race between the eIF4E resistance protein and the VPg avirulence protein. Although we are referring to a particular class of resistance genes distinct from classical dominant R genes, our results fit with the coevolution model proposed for the flax *L* gene – *Melampsora lini* (Dodds *et al.*, 2006) and the tomato *Pto* gene – *Pseudomonas syringae* (Rose *et al.*, 2007) systems. In these cases, direct interaction between resistance and avirulence proteins is the driving force for the coevolution between resistance and avirulence which leads to the diversification of both genes.

## EXPERIMENTAL PROCEDURES

### Plant germplasm and progenies

Twenty five accessions of *C. annuum* were chosen to create a worldwide sample from diverse geographic regions (Figure 1). Efforts were made to sample broadly and to avoid relationships between genotypes. The sample includes the homozygous cultivars Yolo Wonder (YW), YoloY (YY) and Florida VR2 (F) carrying respectively the *pvr2*<sup>+</sup>, *pvr2*<sup>1</sup> and *pvr2*<sup>2</sup> alleles, and local populations from primary and secondary diversification centres. For local populations, seeds from single individuals were harvested to create single-seed stocks for producing the plant material used.

F<sub>1</sub> hybrids between YW and PVY-0 resistant genotypes representing the different eIF4E allelic variants (Perennial, HD-C69, PI322719, SC81, Maroc1, Serrano Vera Cruz, PI195301 and Chile de Arbol) were produced to determine the recessive *vs.* dominant nature of

resistance. F<sub>1</sub> hybrids between the same genotypes and YY were produced to determine if PVY-0 resistance was controlled by an allele of *pvr2*. Allelism for TEV-HAT resistance was assessed by using the F<sub>1</sub> hybrid between F and Chay Angolano.

### **Potyvirus strains and disease resistance evaluation**

All plants were grown under greenhouse conditions and transferred into growth chambers before inoculation. The reactions of all *C. annuum* accessions to PVY and TEV were determined by mechanical inoculation of 20 plants per genotypes at the cotyledon stage with PVY-LYE84 (pathotype PVY-0) (Moury *et al.*, 2004), PVY-SON41 (pathotype PVY-0,1,2) (Moury *et al.*, 2004), TEV-HAT (Schaad *et al.*, 1997) and TEV-CAA10 (a gift from B. Moury, INRA). PVY and TEV strains were maintained on *C. annuum* Yolo Wonder and *Datura stramonium* respectively, and transferred every 4-8 weeks. Inoculum and mechanical inoculation procedures were as described previously (Caranta and Palloix, 1996). Thirty days post-inoculation (dpi), systemic infection was assayed by presence/absence of symptoms on non-inoculated leaves and confirmed by DAS-ELISA using PVY or TEV antibodies. For genetic complementation analysis, F<sub>1</sub> hybrids were evaluated for resistance to PVY-LYE84 and TEV-HAT.

### **Amplification, sequencing and sequence analysis of *eIF4E* cDNAs and partial gDNAs**

Total RNAs were isolated from pepper leaf tissues using TRI-Reagent (Sigma-Aldrich, St Louis). *eIF4E* cDNAs from genotypes YW, YY and F (GenBank accessions n° AY122052, AF521964, AF521965) were re-sequenced as controls. The *eIF4E* cDNAs from the 25 accessions were obtained by RT-PCR using primers defined on the 5' and 3' NTR regions of the YW *eIF4E* cDNA (forward primer 5'-A AAAGCACACAGCACCAACA-3' and reverse primer 5'-GATTAGAAGTGCAAACACCAATAC-3') and the following PCR conditions, 94°C for 30s, 53°C for 30s, 68°C for 1min, for 30 cycles. In order to compare polymorphism between coding and non coding regions, genomic DNA was isolated from leaf tissues using CTAB extraction method (Fulton *et al.*, 1995) and partial *eIF4E* gDNA from the 25 accessions were amplified by PCR using primers defined in exon 2 and exon 5 (forward primer 5'-GAAGATCCTGTATGTGCCAATG-3' and reverse primer 5'-CTGTGTAACGATTCTTTGC-3'). The PCR conditions were: 94°C for 3 min, 35x (94°C for 30s, 55°C for 30s, 72°C for 1min) followed by 72°C for 5 min.

Each amplification product was sequenced twice by the INRA sequencing platform of Montpellier using an ABI 3130XL POP7 sequencer (S. Santoni, INRA Montpellier, France).

Nucleotide sequences were analysed with Genalys Win3.3.24a software (<http://software.cng.fr>) and BioEdit (Hall, 1999). The nucleotide and amino-acid sequences were aligned using ClustalW (Thompson *et al.*, 1994). *eIF4E* cDNA sequences correspond to GenBank accessions n°AY122052 (*pvr2*<sup>+</sup>), AF521964 (*pvr2*<sup>1</sup>), AF5211965 (*pvr2*<sup>2</sup>), AY723737 (*pvr2*<sup>3</sup>), AY723738 (*pvr2*<sup>4</sup>), AY723739 (*pvr2*<sup>5</sup>), AY723740 (*pvr2*<sup>6</sup>), EU106863 (*pvr2*<sup>7</sup>), EU106864 (*pvr2*<sup>8</sup>) and EU106865 (*pvr2*<sup>9</sup>).

A predicted 2D structure of the pepper YW *eIF4E* protein was obtained using as template the known structure of wheat (*Triticum aestivum*) *eIF4E* (Protein Data Bank accession 2IDR) (Monzingo *et al.*, 2007). It was used to map all amino acid changes observed between resistant and susceptible genotypes.

### **Yeast two-hybrid analysis**

*eIF4E*, NIa, VPg and N1b (the RNA-dependent RNA polymerase) coding sequences were amplified by PCR using High Fidelity Platinum Taq polymerase (GIBCO/BRL, Life technologies) from oligodT-primed reverse transcription products. Gene-specific primers were designed to introduce restriction enzyme sites. The VPg cistron was amplified from the NIa PCR product using a reverse primer incorporating a stop codon at the end of the coding sequence. PCR products were cloned into pGEMT-easy vectors (Promega, Madison, WI, USA) and sequenced. Coding sequences were then subcloned in translational frame with the GAL4 Activation Domain or GAL4 Binding domain into the pGADT7 or pGBKT7 vectors, respectively (Clontech, Palo Alto, CA, USA). All pGADT7- and pGBKT7-derived vectors were sequenced with primer T7 to check orientation and frame.

*eIF4E* coding sequences corresponding to *pvr2*<sup>+</sup>, *pvr2*<sup>1</sup>, *pvr2*<sup>2</sup>, *pvr2*<sup>3</sup>, *pvr2*<sup>4</sup>, *pvr2*<sup>5</sup>, *pvr2*<sup>6</sup>, *pvr2*<sup>7</sup>, *pvr2*<sup>8</sup> and *pvr2*<sup>9</sup> alleles were amplified from respectively YW, YY, F, Perennial, PI322719, SC81, Maroc1, Serrano Vera Cruz, PI195301, and Chile de Arbol. NIa and VPg cistrons were amplified from PVY-LYE84, PVY-SON41, TEV-HAT and TEV-CAA10; the N1b cistrons were amplified from PVY-LYE84 and TEV-HAT.

The Matchmaker GAL4 two-hybrid system 3 (Clontech, Palo Alto, CA, USA) was used according to protocols described in the Clontech Yeast Protocol Handbook. pGADT7- and pGBKT7-derived vectors were transformed into respectively AH109 and Y187 yeast strains, which contain two independent reporter genes HIS3 and ADE2 (to confer respectively histidine and adenine auxotrophies) driven by hybrid GAL4 promoters. After yeast mating, large double-transformed yeast colonies were resuspended in 100 µL sterile water and 10 µL were spotted on several selective media including synthetic medium lacking leucine and

tryptophan (hereafter named -LW), medium lacking leucine, tryptophan and histidine (-LWH) and medium lacking leucine, tryptophan, histidine and adenine (-LWHA). Strong interactions are detected on medium lacking both histidine and adenine whereas weak interactions are detected on medium lacking histidine. This feature permits a robust and reliable semi-quantitative assay of the strength of the interaction using selective media. Plates were kept at 30°C. Growth was checked daily from 2 to 7 days after spotting. For negative controls, pGADT7 and pGBKT7 plasmids were used for prey and bait, respectively. Interaction between murine p53 and SV40 large T antigen (controls from the Matchmaker GAL4 two-hybrid system 3) and interaction between N1a and N1b of TEV-HAT (yet demonstrated by (Altmann *et al.*, 1989; Schaad *et al.*, 2000) were used as positive controls.

For Western blots, yeast total protein extractions were as described in the Clontech Yeast Protocol Handbook (Urea/SDS method). Equal loads of protein extracts were electrophoresed on a sodium dodecyl sulfate 12% polyacrylamide gel and blotted to Hybond ECL nitrocellulose membranes (Amersham Biosciences). The membrane was first incubated with HA monoclonal antibody (12CA5, Roche) at a dilution of 1/1000, washed and then incubated with alkaline phosphatase-conjugated rabbit anti-mouse serum (Sigma-Aldrich, St Louis, MO) at a dilution of 1/2000. After washing, the reactivity was visualized with nitroblue tetrazolium. Pictures were taken with a Nikon coolpix camera and figures were processed using Adobe Photoshop 7.0.

### **Complementation assays in *Saccharomyces cerevisiae***

*S. cerevisiae* strain JO55 (*cdc33-Δ LEU2 leu 2 ura3 his3 trp1 ade2 [YCp33supex-h4E URA3]*), a gift from J.M.X. Hughes, (Hughes *et al.*, 1999) carries a disrupted endogenous *eIF4E* gene (*cdc33*). Its survival depends on the presence of plasmid YCp33supex-h4E URA3 containing a copy of the human *eIF4E* cDNA under the control of the galactose-dependent GAL promoter. The coding sequences of each of the *pvr2-eIF4E* alleles were cloned into p424GBP/TRP1 glucose-dependent vector and independently used to transform *S. cerevisiae* strain JO55. After transformation, yeast cells were grown in appropriate selective nutrient drop-out media containing 2% glucose and tested at 30°C for their capacity to complement the lack of endogenous *eIF4E* (Altmann *et al.*, 1989). Control transformations were done with no DNA (empty JO55), plasmids p424GBP/TRP1 (empty)(negative controls), and with p424GBP/TRP1:*At-eIF4E* (At4g18040, positive control).

## Acknowledgements

We thank A. Labay for very helpful comments on the manuscript, S. Santoni and A. Weber for sequencing, A.M. Sage-Palloix for availability of *Capsicum* germplasm collection and P. Sanchez, T. Phally and G. Nemouchi for their excellent assistance. This work was supported by grants from GENOPLANTE (GNP05003G) and the Bureau des Ressources Génétiques (AO 2005-2007). C. Charron was supported by a doctoral fellowship from the French Ministry of Research.

## REFERENCES

- Altmann, M., Sonenberg, N. and Trachsel, H.** (1989) Translation in *Saccharomyces cerevisiae*: initiation factor 4E-dependent cell-free system. *Mol. Cell Biol.*, **9**, 4467-4472.
- Borgstrøm, B. and Johansen, I.E.** (2001) Mutations in *Pea seedborne mosaic virus* genome-linked protein VPg alter pathotype-specific virulence in *Pisum sativum*. *Mol. Plant Microbe Interact.*, **14**, 707-714.
- Browning, K.S.** (2004) Plant translation initiation factors: it is not easy to be green. *Biochem Soc Trans*, **32**, 589-591.
- Caranta, C. and Palloix, A.** (1996) Both common and specific genetic factors are involved in polygenic resistance of pepper to several potyviruses. *Theor. Appl. Genet.*, **92**, 15-20.
- Combe, J.P., Petracek, M.E., van Eldik, G., Meulewaeter, F. and Twell, D.** (2005) Translation initiation factors eIF4E and eIFiso4E are required for polysome formation and regulate plant growth in tobacco. *Plant Mol. Biol.*, **57**, 749-760.
- Culjkovic, B., Topisirovic, I. and Borden, K.L.** (2007) Controlling gene expression through RNA regulons: the role of the eukaryotic translation initiation factor eIF4E. *Cell Cycle*, **6**, 65-69.
- Deom, C.M. and He, X.Z.** (1997) Second-site reversion of a dysfunctional mutation in a conserved region of the *Tobacco mosaic tobamovirus* movement protein. *Virology*, **232**, 13-18.
- Dodds, P.N., Lawrence, G.J., Catanzariti, A.M., Teh, T., Wang, C.I., Ayliffe, M.A., Kobe, B. and Ellis, J.G.** (2006) Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proc Natl Acad Sci U S A*, **103**, 8888-8893.
- Duprat, A., Caranta, C., Revers, F., Menand, B., Browning, K.S. and Robaglia, C.** (2002) The *Arabidopsis* eukaryotic initiation factor (iso)4E is dispensable for plant growth but required for susceptibility to potyviruses. *Plant J.*, **32**, 927-934.
- Fulton, T.M., Chunwongse, J. and Tanksley, S.** (1995) Microprep protocol for extraction of DNA from tomato and other Herbaceous plants. *Plant Molecular Biology Reporter*, **13**, 207-209.
- Gallie, D. and Browning, K.** (2001) eIF4G functionally differs from eIF(iso)4G in promoting internal initiation, cap-independent translation and translation of structured mRNAs. *J. Biol. Chem.*, **276**, 36951-36960.

- Gao, Z., Johansen, E., Eyers, S., Thomas, C., Ellis, T. and Maule, A.** (2004) The potyvirus recessive resistance gene, *sbm1*, identifies a novel role for translation initiation factor eIF4E in cell-to-cell trafficking. *Plant J.*, **40**, 376-385.
- Green, S.K. and Kim, J.S.** (1991) Characteristics and control of viruses infecting peppers: a literature review. *Asian Vegetable and Development Center. Technical Bulletin*, Vol. **18**, 60 p.
- Greenleaf, W.H.** (1986) Pepper breeding. In Basset, M.J. (ed.), *Breeding vegetable crops*. AVI Publishing Co. Inc., Westport Connecticut, pp. 67-134.
- Hall, T.A.** (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucl. Acids Symp. Ser.*, **41**, 95-98.
- Hughes, J.M., Ptushkina, M., Karim, M.M., Koloteva, N., von der Haar, T. and McCarthy, J.E.** (1999) Translational repression by human 4E-BP1 in yeast specifically requires human eIF4E as target. *J Biol Chem*, **274**, 3261-3264.
- Joshi, B., Lee, K., Maeder, D.L. and Jagus, R.** (2005) Phylogenetic analysis of eIF4E-family members. *BMC Evol. Biol.*, **5**, 48 doi:10.1186/1471-2148-5-48.
- Kang, B.C., Yeam, I., Frantz, J.D., Murphy, J.F. and Jahn, M.M.** (2005) The *pvr1* locus in *Capsicum* encodes a translation initiation factor eIF4E that interacts with *Tobacco etch virus* VPg. *Plant J.*, **42**, 392-405.
- Kushner, D., Lindenbach, B., Grzelishvili, V., Noueir, A., Paul, S. and Ahlquist, P.** (2003) Systematic, genome-wide identification of host genes affecting replication of a positive-strand RNA virus. *Proc. Natl. Acad. Sci. U S A*, **100**, 15764-15769.
- Kyle, M.M. and Palloix, A.** (1997) Proposed revision of nomenclature for potyvirus resistance genes in *Capsicum*. *Euphytica*, **97**, 183-188.
- Lellis, A., Kasschau, K., Whitham, S. and Carrington, J.** (2002) Loss-of-susceptibility mutants of *Arabidopsis thaliana* reveal an essential role for eIF(iso)4E during potyvirus infection. *Curr. Biol.*, **12**, 1046-1051.
- Leonard, S., Plante, D., Wittmann, S., Daigneault, N., Fortin, M.G. and Laliberte, J.F.** (2000) Complex formation between potyvirus VPg and translation eukaryotic initiation factor 4E correlates with virus infectivity. *J Virol*, **74**, 7730-7737.
- Leonard, S., Viel, C., Beauchemin, C., Daigneault, N., Fortin, M.G. and Laliberte, J.F.** (2004) Interaction of VPg-Pro of *Turnip mosaic virus* with the translation initiation factor 4E and the poly(A)-binding protein *in planta*. *J Gen Virol*, **85**, 1055-1063.
- Maule, A., Caranta, C. and Boulton, M.** (2007) Sources of natural resistance to plant viruses: status and prospects. *Molecular Plant Pathology*, **8**, 223-231.
- Michon, T., Estevez, Y., Walter, J., German-Retana, S., Le Gall, O.** (2006) The potyviral virus genome-linked protein VPg forms a ternary complex with the eukaryotic initiation factors eIF4E and eIF4G and reduces eIF4E affinity for a mRNA cap analogue. *FEBS J.*, **273**, 1312-1322.
- Monzingo, A.F., Dhaliwal, S., Dutt-Chaudhuri, A., Lyon, A., Sadow, J.H., Hoffman, D.W., Robertus, J.D. and Browning, K.S.** (2007) The structure of translation initiation factor eIF4E from wheat reveals a novel disulfide bond. *Plant Physiol.*, **143**, 1504-1518.
- Moury, B., Morel, C., Johansen, E., Guilbaud, L., Souche, S., Ayme, V., Caranta, C., Palloix, A. and Jacquemond, M.** (2004) Mutations in *Potato virus Y* genome-linked protein determine virulence toward recessive resistances in *Capsicum annuum* and *Lycopersicon hirsutum*. *Mol. Plant Microbe Interact.*, **17**, 322-329.
- Nicaise, V., German-Retana, S., Sanjuan, R., Dubrana, M.P., Mazier, M., Maisonneuve, B., Candresse, T., Caranta, C. and Le Gall, O.** (2003) The eukaryotic translation initiation factor 4E controls lettuce susceptibility to the potyvirus *Lettuce mosaic virus*. *Plant Physiol.*, **132**, 1272-1282.

- Nieto, C., Morales, M., Orjeda, G., Clepet, C., Monfort, A., Sturbois, B., Puigdomenech, P., Pitrat, M., Caboche, M., Dogimont, C., Garcia-Mas, J., Aranda, M.A. and Bendahmane, A.** (2006) An *eIF4E* allele confers resistance to an uncapped and non-polyadenylated RNA virus in melon. *Plant J*, **48**, 452-462.
- Revers, F., Le Gall, O., Candresse, T. and Maule, A.** (1999) New advances in understanding the molecular biology of plant / potyvirus interactions. *Mol. Plant Microbe Interact.*, **12**, 367-376.
- Riechmann, J.L., Lain, S. and Garcia, J.A.** (1992) Highlights and prospects of potyvirus molecular biology. *J Gen Virol*, **73**, 1-16.
- Robaglia, C. and Caranta, C.** (2006) Translation initiation factors: a weak link in plant RNA virus infection. *Trends Plant Sci*, **11**, 40-45.
- Rodriguez, C., Freire, M., Camilleri, C. and Robaglia, C.** (1998) The *Arabidopsis thaliana* cDNAs coding for eIF4E and eIF(iso)4E are not functionally equivalent for yeast complementation and are differentially expressed during plant development. *Plant J*, **13**, 465-473.
- Rose, L.E., Micheltore, R.W. and Langley, C.H.** (2007) Natural variation in the *Pto* disease resistance gene within species of wild tomato (*Lycopersicon*). II. Population genetics of *Pto*. *Genetics*, **175**, 1307-1319.
- Ruffel, S., Gallois, J.L., Lesage, M.L. and Caranta, C.** (2005) The recessive potyvirus resistance gene *pot-1* is the tomato orthologue of the pepper *pvr2-eIF4E* gene. *Mol. Genet. Genomics*, **274**, 346-353.
- Ruffel, S., Gallois, J.L., Moury, B., Robaglia, C., Palloix, A. and Caranta, C.** (2006) Simultaneous mutations in translation initiation factors eIF4E and eIF(iso)4E are required to prevent pepper veinal mottle virus infection of pepper. *J Gen Virol*, **87**, 2089-2098.
- Ruffel, S., Dussault, M.H., Palloix, A., Moury, B., Bendahmane, A., Robaglia, C. and Caranta, C.** (2002) A natural recessive resistance gene against *Potato virus Y* in pepper corresponds to the eukaryotic initiation factor 4E (eIF4E). *Plant J*, **32**, 1067-1075.
- Ruffel, S., Dussault, M.H., Palloix, A., Moury, B., Revers, F., Bendahmane, A., Robaglia, C. and Caranta, C.** (2004) The key role of the eukaryotic initiation factor 4E (eIF4E) in plant-potyvirus interactions. In *Biology of Plant-Microbe Interactions*, Vol. 4, *Molecular Plant-Microbe Interaction: New Bridges between past and future*. Edited by I. Tikhonovich, B. Lugtenberg & N. Provorov. St Paul, MN: International Society for Molecular Plant-Microbe Interactions, pp. 81-83.
- Sato, M., Nakahara, K., Yoshii, M., Ishikawa, M. and Uyeda, I.** (2005) Selective involvement of members of the eukaryotic initiation factor 4E family in the infection of *Arabidopsis thaliana* by potyviruses. *FEBS Lett*, **579**, 1167-1171.
- Schaad, M., Anderberg, R. and Carrington, J.** (2000) Strain-specific interaction of the *Tobacco etch virus* NIa protein with the translation initiation factor eIF4E in the yeast two-hybrid system. *Virology*, **273**, 300-306.
- Stein, N., Perovic, D., Kumlehn, J., Pelli, B., Stracke, S., Streng, S., Ordon, F. and Graner, A.** (2005) The eukaryotic translation initiation factor 4E confers multiallelic recessive *Bymovirus* resistance in *Hordeum vulgare* (L.). *Plant J*, **42**, 912-922.
- Thompson, J., Higgins, D. and Gibson, T.** (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673-4680.

- Wittmann, S., Chatel, H., Fortin, M.G. and Laliberte, J.F.** (1997) Interaction of the viral protein genome linked of *Turnip mosaic potyvirus* with the translational eukaryotic initiation factor (iso)4E of *Arabidopsis thaliana* using the yeast two-hybrid system. *Virology*, **234**, 84-92.
- Yeam, I., Cavatorta, J.R., Ripoll, D., Kang, B.C., Jahn, MM.** (2007) Functional dissection of naturally occurring amino acid substitutions in eIF4E that confers recessive potyvirus resistance in plants. *Plant Cell*, 10.1105/tpc.107.050997

## **II.2. Impact du cumul de substitutions en acides aminés dans 2 régions distinctes de la protéine eIF4E1 de *C. annuum* sur le spectre de résistance aux potyvirus**

### **II.2.1. Contexte et objectif**

La position de la plupart des substitutions "signatures" des protéines de résistance eIF4E apparaît conservée entre espèces (piment, tomate, laitue et pois) et restreinte à 2 régions adjacentes, localisées en surface de la protéine, à proximité de la zone d'interaction avec la coiffe (Monzingo *et al.*, 2007; Robaglia and Caranta, 2006). Plusieurs études montrent que des mutations dans une seule de ces 2 régions sont associées à la résistance à un seul potyvirus (voire à une seule souche) tandis que le cumul de substitutions dans les 2 régions est systématiquement associé à la résistance à plusieurs potyvirus. Par exemple, chez le pois, la protéine eIF4E codée par l'allèle *sbm1*<sup>1</sup> et présentant uniquement des substitutions dans la région I, contrôle la résistance au pathotype P1 de PSbMV, alors que celle codée par l'allèle *sbm1* présentant plusieurs substitutions dans les régions I et II, confère la résistance à 2 pathotypes P1 et P4 de PSbMV et au BYMV-W (*White lupin mosaic virus*) (Gao *et al.*, 2004b). De même, chez la tomate, la protéine eIF4E1, codée par l'allèle *pot1* de l'accession *L. hirsutum* PI247087, présentant des substitutions en acides aminés à la fois dans les régions I et II de la protéine, confère une résistance à large spectre *i.e.*, vis-à-vis de plusieurs souches de PVY et de TEV.

La position des substitutions en acides aminés détectées chez eIF4E1-pvr2 de *C. annuum* (article 1) a été reportée sur la structure 3D prédite à partir de la structure obtenue récemment chez le blé (Monzingo *et al.*, 2007) (Figure II.1) : P66T, V67E, A68E, A73D, A74D et L79R se situent dans la région I, tandis que G107R et D109N se positionnent dans la région II. A partir de l'analyse fonctionnelle de la variabilité allélique au locus *pvr2-eIF4E1*, nous avons montré que la présence d'une ou plusieurs substitutions dans la région I de la protéine eIF4E1 suffisait à expliquer la résistance restreinte au PVY et dans le cas de *pvr2*<sup>4</sup>, une seule substitution (V67E) suffit à la résistance. De façon plus générale, et si l'on considère la variabilité naturelle du genre *Capsicum*, la fréquence de résistance au TEV est faible par rapport au PVY : de l'ordre de quelques % pour la résistance au TEV alors qu'elle atteint les 40 % pour la résistance au PVY (Palloix & Daubèze, com. pers.). De plus, la

résistance au TEV est systématiquement associée à la résistance au PVY. Dans notre étude, les protéines eIF4E1 des accessions FloridaVR2 et Chay Angolano (*pvr2*<sup>2</sup>), les seuls génotypes résistants au TEV et au PVY, possèdent à la fois des substitutions en acides aminés dans la région I (V67E et L79R) et une substitution dans la région II (D109N). De même, chez *C. chinense*, l'allèle *pvr1* contrôlant une résistance à large spectre vis-à-vis du PVY et du TEV cumule des mutations dans les régions I (T51A et P66T) et II (G107R) (Kang *et al.*, 2005a). D'après ces observations, nous pouvons émettre l'hypothèse que la résistance à large spectre (*i.e.*, plusieurs souches de PVY et de TEV) résulterait de l'effet complémentaire, plutôt qu'individuel, de substitutions en acides aminés dans les régions I et II de la protéine eIF4E1.

Afin de tester cette hypothèse, un ADNc *eIF4E1*, codant pour une protéine ne comportant que la substitution D109N dans la région II, a été générée par mutagenèse dirigée, à partir de l'ADNc *eIF4E1-pvr2*<sup>+</sup> de Yolo Wonder. Des expériences de complémentation fonctionnelle par expression transitoire d'ADNc *eIF4E* via le vecteur viral PVX (Baulcombe *et al.*, 1995) chez le piment, ont été réalisées dans l'objectif de tester la capacité de cet allèle à restaurer la sensibilité au TEV chez un génotype résistant.

## II.2.2. Matériel et méthodes

Le PVX utilisé comme vecteur viral dans les expériences de complémentation fonctionnelle par expression transitoire des gènes *eIF4E* a été fourni par D.C. Baulcombe (Sainsbury Laboratory, UK). Les expériences d'expression transitoire d'ADNc *eIF4E* via le vecteur viral PVX ont été adaptées du protocole de Baulcombe *et al.* (1995). Les constructions eIF4E-PVX étaient déjà disponibles au laboratoire : pPVXeYW, vecteur PVX avec l'ADNc *eIF4E1-pvr2*<sup>+</sup> de Yolo Wonder; pPVXeF, vecteur PVX avec l'ADNc *eIF4E1-pvr2*<sup>2</sup> de Florida VR2; pPVXeYY, vecteur PVX avec l'ADNc *eIF4E1-pvr2*<sup>1</sup> de Yolo Y et pPVXeD109N, vecteur PVX avec l'ADNc *eIF4E1-pvr2*<sup>+</sup> de Yolo Wonder portant la mutation D109N. Les expériences ont été réalisées selon le protocole décrit par Ruffel *et al.* (2002) en chambre climatisée dans des conditions de confinement OGM du groupe II. Les vecteurs d'expression PVX-eIF4E ont été multipliés sur *N. benthamiana* (inoculation au stade 4 feuilles étalées avec 2,5 µg d'ADNc/feuille). Dix jours après, les feuilles de *N. benthamiana*

présentant des symptômes de l'infection par le PVX (mosaïques) ont été utilisées comme source d'inoculum pour l'inoculation des feuilles de piment, selon le protocole décrit à l'annexe 3. Dix jours après l'inoculation (jpi) de la construction PVX-eIF4E, les feuilles inoculées avec les vecteurs PVX-eIF4E (présentant des points nécrotiques sur les feuilles inoculées et les feuilles systémiques, symptômes caractéristiques de l'infection par le PVX) ont été inoculées avec du PVY ou du TEV. Les feuilles doublement inoculées PVX-potyvirus ont été prélevées 10 jours plus tard pour mesurer l'accumulation des potyvirus par DAS-ELISA (cf annexe 4) et RT-PCR avec des amorces spécifiques de la capsidie des virus.

### II.2.3. Résultats

Le principe de l'expérience est décrit sur la figure II.2. L'ADNc *eIF4E1* pour lequel on veut déterminer s'il s'agit d'un allèle de résistance ou de sensibilité vis-à-vis d'un potyvirus donné est exprimé chez un génotype de piment résistant ; s'il s'agit d'un allèle de sensibilité, les feuilles inoculées montrent une accumulation virale tandis que s'il s'agit d'un allèle de résistance, les feuilles inoculées ne présentent pas d'accumulation virale. Dans notre expérience, si la substitution D109N ne restaure pas la sensibilité au TEV, cela signifie qu'elle permet seule la résistance à ce potyvirus. En revanche, si elle restaure la sensibilité, nous pourrions conclure que la présence de substitutions à la fois dans les régions I et II est nécessaire pour la résistance à large spectre.

Sur les 55 plantes co-inoculées pPVXeYW et TEV, 28 sont sensibles au TEV (Figure II.3. barre 4). De même, toutes les plantes co-inoculées avec pPVXeYW et PVY sont sensibles (Figure II.3 barre 5). Ces résultats confirment que *pvr2<sup>+</sup>* est un allèle de sensibilité aux TEV et PVY. En revanche, aucune plante co-inoculée pPVXeF + TEV et pPVXeF + PVY ne présente d'accumulation virale de l'un ou l'autre de ces virus (Figure II.3 barres 6 et 7), ce qui montre que *pvr2<sup>2</sup>* est un allèle de résistance à ces 2 potyvirus. Les plantes co-inoculées pPVXeYY + PVY sont sensibles au PVY tandis qu'aucune plante co-inoculée pPVXeYY + TEV ne présente d'accumulation virale, ce qui confirme que *pvr2<sup>1</sup>* est un allèle de sensibilité au PVY mais un allèle de résistance au TEV (Figure II.2; données non montrées).

Sur les 38 plantes co-inoculées pPVXeD109N + TEV, 11 sont sensibles au TEV (Figure II.3. barre 8). De la même façon, 9 feuilles sur 22 co-inoculées pPVXeD109N-PVY présentent une accumulation virale de PVY (Figure II.3 barre 9). Le mutant *eIF4E-mutD109N* est donc un allèle de sensibilité aux PVY et TEV. La mutation D109N ne suffit pas à elle seule à contrôler la résistance aux PVY et TEV, suggérant que la résistance à large spectre contrôlée par *pvr2*<sup>2</sup> résulte de l'effet cumulé des substitutions en acides aminés dans les régions I et II. Ces résultats confortent également le rôle des substitutions en acides aminés de la région I dans la résistance au PVY.

### II.2.3. Discussion

Ces résultats indiquent que la substitution D109N seule ne suffit pas à contrôler la résistance au TEV chez le piment. Ces données sont en accord avec les récents travaux publiés par Yeam *et al.* (2007) qui montrent qu'une protéine mutante eIF4E1, ne possédant que la mutation D109N, est toujours capable d'interagir avec la protéine VPg de la souche TEV-HAT en système de double hybride dans la levure. Le cumul de substitutions en acides aminés dans 2 régions distinctes de la protéine eIF4E1 chez *C. annuum* apparaît donc nécessaire pour abolir l'interaction avec la VPg et ainsi induire la résistance au TEV.

D'un point de vue mécanistique, ces résultats impliquent que des mutations dans 2 régions distinctes sont nécessaires pour empêcher la protéine VPg du TEV de se fixer à eIF4E1, et ce contrairement à la protéine VPg du PVY où des substitutions dans la région I de la protéine eIF4E1 de la plante suffisent à bloquer l'interaction VPg/eIF4E1. La VPg du TEV apparaît donc capable d'interagir avec eIF4E au niveau de 2 sites de fixation, l'un dans la région I et l'autre dans la région II, la résistance étant le résultat de la rupture de ces 2 interactions, pour peut-être optimiser la stabilité de cette interaction. Ces travaux suggèrent donc qu'il existe différentes modalités d'interaction eIF4E-VPg, avec ici la mise en place de stratégies distinctes entre 2 virus du même genre infectant la même plante hôte.

Les VPg du TEV et du PVY utilisées dans notre étude présentent 47 % d'identité au niveau de leur séquence primaire en acides aminés. Au niveau de la zone centrale de la protéine où l'on retrouve les acides aminés impliqués dans le contournement des résistances, 13 résidus sur 27 diffèrent. Ainsi, ces différences d'acides aminés entre les 2 protéines

pourraient expliquer les différences d'interactions avec une même protéine eIF4E. Ces différences d'interaction pourraient également être liées à des différences au niveau de la conformation tridimensionnelle (3D) de ces VPg. Récemment, il a été montré que les VPg du RYMV (genre *Sobemovirus*) et du LMV (genre *Potyvirus*) étaient des protéines dites désordonnées (Hébrard, 2007). Contrairement à la plupart des protéines fonctionnelles, les protéines entièrement ou partiellement désordonnées ne se replient pas pour adopter une structure 3D unique (Uversky, 2002). En conditions physiologiques, ces protéines existent sous un ensemble dynamique de conformations ne comprenant que peu ou pas d'éléments de structure secondaire (hélices ou brins). Mais ce désordre structural n'empêche pourtant pas ces protéines, et en l'occurrence les VPg, d'interagir avec leurs partenaires. Au contraire, il joue un rôle essentiel dans leur fonction biologique, en rendant possibles de nombreuses interactions (Hébrard, 2007). On peut donc penser que les VPg du PVY et du TEV n'adoptent pas la même conformation structurale et n'interagissent donc pas avec exactement les mêmes acides aminés de la protéine eIF4E1. Chaque couple plante-potyvirus posséderait des sites spécifiques d'interaction, optimum d'un point de vue coévolutif.

A l'encontre de ces résultats, une étude récente montre que la substitution G107R dans la région II de la protéine eIF4E1, codée par l'allèle de résistance *pvr1* chez *C. chinense*, suffit à abolir l'interaction avec la VPg de la souche TEV-HAT et donc à induire la résistance (Yeam *et al.*, 2007). Une hypothèse avancée pour expliquer ces observations contradictoires est une différence d'effet des substitutions D109N et G107R sur la résistance mais aussi sur la fonction de traduction des ARNm. En effet, des expériences *in vitro* montrent que la mutation G107R abolit non seulement l'interaction eIF4E1-VPg mais aussi l'interaction eIF4E1-m<sup>7</sup>GTP, alors que la mutation D109N seule n'empêche ni l'une ni l'autre de ces interactions (Yeam *et al.*, 2007). Le changement G107R pourrait avoir plus de conséquences sur les fonctions des gènes *eIF4E1* que la substitution D109N. D'un point de vue biochimique, le changement de glycine en arginine à la position 107 introduit une charge positive supplémentaire et un acide aminé plus encombrant, ce qui pourrait provoquer une répulsion électrostatique forte avec les résidus chargés positivement à proximité et constituer un obstacle stérique (Yeam *et al.*, 2007). *A contrario*, la mutation D109N ne perturbe ni l'équilibre électrostatique ni l'encombrement stérique de la protéine eIF4E (l'acide aspartique et l'asparagine sont des acides aminés hydrophiles très proches). Cette observation couplée au fait que cette substitution supplémentaire conditionne la résistance au TEV suggère qu'un

changement sans conséquence apparente sur la biochimie de la protéine eIF4E, peut suffir à la résistance aux potyvirus.

Au vue de ces résultats, une analyse à plus grande échelle et chez plusieurs couples plantes-potyvirus de l'effet cumulé des substitutions en acides aminés sur le spectre de résistance s'avère nécessaire. Ces études constitueraient un élément crucial pour la création de matériel végétal résistant à plusieurs potyvirus (large spectre), à partir de l'analyse de la diversité présente dans la variabilité naturelle.

### **II.3. Rôle des autres isoformes eIF4E dans le contournement de la résistance eIF4E1-pvr2 par les potyvirus**

#### **II.3.1. Contexte et objectif**

Les souches SON41 et SON41p de PVY contournent les allèles de résistance *pvr2*<sup>1</sup> et *pvr2*<sup>2</sup>. Les substitutions en acides aminés identifiées dans la zone centrale de la VPg de ces souches sont responsables du contournement (Moury *et al.*, 2004; Ayme, 2005).

Les résultats présentés dans l'article 1 suggèrent que les mutations dans la région centrale de la VPg, impliquées dans le contournement de certains allèles de résistance au locus *pvr2-eIF4E1*, permettent de rétablir une interaction avec les protéines eIF4E1 mutées; les mutations dans la VPg ont donc un rôle d'adaptation aux mutations des protéines eIF4E1 de résistance. Cependant, de récents travaux ont montré qu'un seul changement en acide aminé dans la VPg des TuMV-CDN1 et TuMV-UK suffisait au contournement de la résistance d'un mutant T-DNA d'*A. thaliana* dont le gène *eIF(iso)4E* est éteint (German-Retana, Revers and Caranta, données non publiées). Ces résultats indiquent que, dans ce cas, le contournement ne se fait pas (et ne peut se faire du fait de l'absence de protéine eIF(iso)4E) par une restauration de l'interaction. Une hypothèse est donc que le TuMV pourrait contourner l'absence de eIF4E1 par l'utilisation d'une autre isoforme eIF4E.

Chez le piment, 3 gènes appartenant à la famille multigénique *eIF4E* ont été clonés et séquencés : 2 gènes codant pour la protéine eIF4E (eIF4E1 et eIF4E2) et 1 gène codant pour

l'isoforme eIF(iso)4E (Ruffel, 2004). Ces protéines comportent toutes les acides aminés importants pour la stabilité de la structure de la protéine et les acides aminés impliqués dans la liaison à la coiffe des ARNm (Figure II.4), suggérant qu'ils sont tous potentiellement capables d'initier la traduction. A ce jour, eIF4E2 apparaît n'avoir aucun rôle dans la résistance aux potyvirus malgré une forte homologie de séquence avec la forme eIF4E1 (Tableau II.1). En revanche, le facteur eIF(iso)4E présentant un pourcentage d'identité avec les autres protéines 4E plus faible (environ 47%, Tableau II.1) est impliqué dans la résistance au PVMV (genre *Potyvirus*) chez *C. annuum*. Une délétion de 82 nucléotides dans la séquence codante de *eIF(iso)4E* chez l'accession Perennial (locus *pvr6*) introduit un codon stop qui conduit à la synthèse d'une protéine tronquée non fonctionnelle. Mais une résistance totale au PVMV n'est observée que lorsqu'il y a combinaison de la délétion dans eIF(iso)4E et de substitutions en acides aminés dans la protéine eIF4E1-*pvr2*<sup>2</sup> indiquant que le PVMV est capable d'utiliser plusieurs protéines eIF4E chez un même hôte (Ruffel *et al.*, 2006).

A partir de ces éléments et afin de déterminer le rôle des protéines eIF4E2 et eIF(iso)4E, notamment dans le contournement des allèles de résistance *eIF4E1-pvr2*<sup>1</sup> et *eIF4E1-pvr2*<sup>2</sup>, des tests d'interaction protéine-protéine par le système de double hybride dans la levure ont été réalisés entre différentes VPg de souches de PVY et de TEV avirulentes et virulentes sur eIF4E1-*pvr2* et les protéines eIF4E2 et eIF(iso)4E de l'accession *C. annuum* Yolo Wonder (selon le même protocole que pour l'article 1; cf annexe 9). Les souches avirulentes testées sont PVY-LYE84 et TEV-HAT utilisées au § II.1. Les souches virulentes sont PVY-SON41, PVY-SON41p et TEV-CAA10. PVY-SON41 et TEV-CAA10 sont les mêmes souches que celles décrites dans l'article 1, pour lesquelles nous connaissons déjà le profil d'interaction avec les protéines eIF4E1-*pvr2* (Figure 5 article 1, § II.1). La troisième souche virulente est PVY-SON41p de pathotype (0,1,2), capable de contourner les allèles de résistance *pvr2*<sup>1</sup> et *pvr2*<sup>2</sup> chez le piment. PVY-SON41p est un variant naturel de PVY-SON41 et 4 substitutions en acides aminés différencient les VPg des deux souches. L'alignement des séquences en acides aminés de la partie centrale de la VPg des 5 souches testées au cours de ces expériences est présenté à la figure II.5.

### II.3.2. Résultats

Dans un premier temps, le profil d'interaction de la VPg PVY-SON41p avec les différentes protéines eIF4E1-pvr2 a été étudié. La VPg de la souche PVY-SON41p interagit bien avec la protéine eIF4E1 de l'accession sensible Yolo Wonder (*pvr2*<sup>+</sup>) mais n'interagit pas avec les protéines codées par les allèles de résistance *eIF4E1-pvr2*<sup>1</sup> et *eIF4E1-pvr2*<sup>2</sup> alors que cette souche contourne ces deux allèles (Figure II.6.A). Les mutations dans la VPg de la souche PVY-SON41p, responsables du contournement des résistances, ne permettent donc pas de rétablir une interaction avec les protéines eIF4E1 mutées. Cette VPg a donc un comportement différent de la VPg de la souche contournante PVY-SON41 qui interagit avec les protéines eIF4E1 mutées (cf article 1).

En revanche, une interaction forte entre la VPg de la souche PVY-SON41p et la protéine eIF4E2 de l'accession Yolo Wonder est observée, alors que cette interaction n'est jamais détectée avec les VPg des souches PVY-LYE84 et PVY-SON41 (Figure II.6.B). Ces résultats suggèrent donc que les mutations dans la VPg de la souche PVY-SON41p permettent une interaction avec eIF4E2.

En ce qui concerne le TEV, les VPg des souches de TEV-HAT et TEV-CAA10 interagissent fortement (sur milieu très stringent -LWHA) avec la protéine eIF4E2 de Yolo Wonder en système de double hybride dans la levure (Figure II.6.B). Ces 2 souches de TEV sont donc capables de recruter soit eIF4E1 (Figure II.6.A) soit eIF4E2.

La protéine eIF(iso)4E n'interagit avec aucune des VPg testées, suggérant que cette protéine n'est pas impliquée dans l'interaction avec le PVY et le TEV.

### II.3.3. Discussion

Les VPg des différentes souches de potyvirus testées varient dans leur capacité à interagir ou non non seulement avec les différentes protéines codées par les allèles *eIF4E1* mais également avec les différentes protéines codées par les paralogues *eIF4E*. Aucune interaction entre VPg du PVY et TEV et eIF(iso)4E n'a pu être détectée, suggérant que cette

protéine n'est pas impliquée dans l'interaction avec ces deux potyvirus. Par contre, les résultats récents obtenus au laboratoire par M. Nicolai montrent que la VPg du PVMV interagit avec eIF(iso)4E, résultats qui d'une part valident le fait que ce potyvirus utilise à la fois eIF4E1 et eIF(iso)4E et d'autre part illustrent la diversité d'utilisation des protéines eIF4E par les différents potyvirus infectant le piment.

Au regard des résultats d'interaction entre les isoformes eIF4E1 et eIF4E2 et les VPg des souches de PVY et TEV, 3 types d'interaction peuvent être distingués :

(i) les VPg des souches PVY-LYE84 et PVY-SON41 n'interagissent qu'avec la forme eIF4E1 et pas avec eIF4E2; dans ce cas, les résultats obtenus par le système double hybride dans la levure suggèrent que le contournement de eIF4E1-pvr2<sup>1</sup> et eIF4E1-pvr2<sup>2</sup> par PVY-SON41 résulte d'une restauration de l'interaction entre la VPg et les protéines eIF4E1 de résistance.

(ii) la VPg de la souche PVY-SON41p interagit uniquement avec eIF4E1-pvr2<sup>+</sup> et aussi avec la protéine eIF4E2; dans ce cas le contournement de eIF4E1-pvr2<sup>1</sup> et eIF4E1-pvr2<sup>2</sup> pourrait s'expliquer non pas par la restauration de l'interaction avec les protéines eIF4E1-pvr2<sup>1</sup> et eIF4E1-pvr2<sup>2</sup> mais par l'utilisation de la protéine eIF4E2. Dans ce cas, les 4 substitutions en acides aminés qui distinguent les VPg de PVY-SON41 et PVY-SON41p pourraient expliquer les différences d'interaction.

(iii) les VPg des souches TEV-HAT et TEV-CAA10 interagissent avec les protéines eIF4E1 et le contournement de eIF4E1-pvr2<sup>2</sup> par TEV-CAA10 peut s'expliquer par une restauration de l'interaction entre la VPg de TEV-CAA10 et la protéine eIF4E1-pvr2<sup>2</sup>. Cependant les résultats obtenus au cours de ces expériences montrent que les VPg de TEV-HAT et TEV-CAA10 interagissent également avec eIF4E2.

Ces résultats posent question en ce qui concerne la résistance des géotypes *pvr2*<sup>2</sup> à la souche de TEV-HAT sachant que les séquences en acides aminés eIF4E2 des géotypes *pvr2*<sup>+</sup>, *pvr2*<sup>1</sup> et *pvr2*<sup>2</sup> sont 100% identiques. Pourquoi, pour contourner les substitutions en acides aminés de la protéine eIF4E1, le virus n'est-il pas capable d'utiliser la protéine eIF4E2? Des tests *via* le système de double hybride dans la levure, utilisant la protéine eIF4E2 de ces 2 accessions résistantes, permettraient dans un premier temps de s'assurer de l'interaction entre cette forme et la VPg du TEV-HAT. Deux hypothèses peuvent être avancées pour expliquer ces résultats :

(i) l'absence de colocalisation de eIF4E2 et de la VPg du TEV en condition naturelle. Si les protéines eIF4E2 ne sont pas exprimées dans les cellules où se multiplient les souches de TEV, le virus ne peut pas recruter ce facteur pour accomplir son cycle infectieux. Des mesures quantitatives d'accumulation des protéines eIF4E2 chez *C. annuum* pourraient être mises en place pour vérifier cette hypothèse, en parallèle avec des expériences de visualisation de la co-localisation *in planta* des protéines eIF4E2 et VPg, telles que des expériences de BiFC (Bimolecular fluorescence assay) : cette technique a déjà permis de détecter des interactions *in planta* entre eIF4E1-pvr1<sup>+</sup> de *C. chinense* et la VPg du TEV-HAT (Yeam *et al.*, 2007). Cette hypothèse est cependant en contradiction avec l'hypothèse que la souche PVY-SON41p utilise la protéine eIF4E2 pour contourner la résistance.

(ii) la multiplication du TEV-HAT nécessite la présence d'un autre partenaire protéique, en plus de eIF4E2, et différent de celui utilisé par le PVY.

Ces expériences mettent en avant la grande diversité d'utilisation des allèles et des formes eIF4E par les potyvirus mais soulèvent au moins autant de questions qu'elles n'apportent de réponse. Il s'avère donc essentiel de mettre en place une approche globale, *in planta* et vis-à-vis d'une plus large gamme de potyvirus, pour étudier les interactions entre protéines de l'hôte et protéines virales.

## Chapitre III

# **A**analyse comparée du polymorphisme des gènes *eIF4E* chez 2 Solanacées (piment et tomate) et *A. thaliana*

### **III.1. Introduction**

Ce travail s'intègre dans un projet financé par le Bureau des Ressources Génétiques (BRG AO 2005-2007) visant à comparer les patrons de diversité des gènes *eIF4E* au sein de 5 familles végétales : Solanacées, Crucifères, Vitacées, Légumineuses et Triticées. A partir des connaissances acquises sur le rôle de eIF4E dans les interactions plantes-virus à ARN, ces analyses de polymorphisme ont pour objectifs (i) d'identifier de nouveaux allèles de résistance aux virus à ARN, (ii) de contribuer à la caractérisation du rôle de la redondance des gènes *eIF4E* chez les plantes et (iii) de réaliser une analyse comparative de l'évolution moléculaire de ces gènes chez plusieurs familles végétales afin de déterminer si certains de ces gènes ont subi une pression de sélection caractéristique des gènes de résistance aux pathogènes (i.e., sélection balancée *versus* diversifiante). En effet, dans le cas où les gènes subissent, ou ont subi dans un passé récent, une forme quelconque de sélection, des signatures spécifiques de la sélection peuvent être détectées par des analyses de polymorphisme.

Dans le cadre de ma thèse, je me suis focalisée sur l'analyse du polymorphisme des gènes *eIF4E* chez les Solanacées et *A. thaliana* afin de détecter des traces de sélection liées à l'interaction avec les virus à ARN du genre *Potyvirus*. Ces deux familles végétales présentent un comportement contrasté vis-à-vis de l'infection par les potyvirus. Dans la diversité naturelle du piment, la fréquence de résistance aux potyvirus contrôlée par les gènes *eIF4E* est élevée (cf article 1). Chez la tomate, autre Solanacée infectée par plusieurs potyvirus, un gène récessif de résistance au PVY et au TEV correspondant à eIF4E a été identifié chez l'espèce sauvage *L. hirsutum* (gène *pot1* orthologue de *pvr2*). Cependant, la fréquence de résistance aux potyvirus contrôlée par eIF4E apparaît beaucoup moins élevée que chez le piment. *A contrario*, aucune résistance contrôlée par les gènes *eIF4E* n'a encore été identifiée dans la variabilité naturelle d'*A. thaliana*, et ce malgré l'identification de plusieurs gènes récessifs de résistance.

Le premier objectif est de compléter l'analyse descriptive de la diversité du gène *Ca-eIF4E1* chez le genre *Capsicum*. Ces analyses seront étendues aux autres gènes *Ca-eIF4E* afin de caractériser les pressions de sélection qui s'exercent sur l'ensemble des gènes *eIF4E* connus chez *Capsicum*. Le second objectif est de comparer le profil d'évolution des gènes *eIF4E* entre des espèces présentant différentes fréquences de résistance aux potyvirus : la tomate,

une autre Solanacée infectée par les mêmes potyvirus que le piment mais pour laquelle très peu de résistances récessives aux potyvirus correspondant à *eIF4E* ont été identifiées et *A. thaliana*, où aucune résistance contrôlée par les gènes *eIF4E* n'a encore été trouvée. L'ensemble de ces analyses devrait permettre de faire la distinction entre les signatures résultant de la fonction cellulaire des *eIF4E*, les éventuelles signatures espèces-spécifiques et les signatures évolutives résultant de l'adaptation aux potyvirus.

## **III.2. Matériel et méthodes**

### **III.2.1. Matériel végétal**

Les 25 géotypes de *C. annuum* utilisées dans l'article 1 représentent notre échantillon intraspécifique. Pour les analyses de polymorphisme interspécifique, 26 accessions de 11 espèces ont été sélectionnées, représentatives des espèces de *Capsicum* domestiquées et des espèces sauvages apparentées aux espèces cultivées (Tableau III.1.A) : 7 accessions de *C. annuum glabrisuculum*, 3 *C. frutescens*, 2 *C. chinense*, 2 *C. chacoense*, 1 *C. galapogense*, 2 *C. baccatum*, 2 *C. pubescens*, 1 *C. microcarpum*, 1 *C. praetermissum*, 3 *C. cardenasii* et 2 *C. eximium*.

Vingt accessions de l'espèce cultivée *L. esculentum* ont été utilisées pour les analyses intraspécifiques de la variabilité des gènes *eIF4E*. Elles ont été choisies par rapport à la diversité de leurs caractères morphologiques (forme et grosseur du fruit ; Tam *et al.*, 2007). Pour les analyses interspécifiques, 30 accessions supplémentaires, appartenant aux 8 espèces sauvages apparentées à l'espèce cultivée, ont été sélectionnées (Tableau III.1.B) : 8 accessions de *L. pimpinellifolium*, 10 *L. hirsutum* (dont l'accession PI247087 possédant l'allèle *pot1-eIF4E1* de résistance aux potyvirus), 2 *L. pennellii*, 2 *L. chilense*, 2 *L. cheesmanii*, 2 *L. peruvianum*, 2 *L. chmielewskii* et 1 *L. parviflorum*. J'utiliserai dans ce manuscrit l'ancienne nomenclature des *Lycopersicon*, nomenclature qui a été changée au cours de ma thèse. La correspondance entre les deux nomenclatures, indiquée au tableau III.1.B, est également disponible sur le site du Solanaceae Genomics Network (SGN; [http://solddb.cit.cornell.edu/about/solanum\\_nomenclature.pl](http://solddb.cit.cornell.edu/about/solanum_nomenclature.pl)).

La caractérisation des patrons de diversité des différents membres de la famille multigénique eIF4E chez *A. thaliana* (*At-eIF4E*) a été effectuée chez 48 lignées homozygotes fixées (accessions) regroupées dans une core collection établie par McKhann *et al.* (2004). Cette core collection capture 93 % de la diversité observée au sein de l'espèce *A. thaliana* pour 10 locus séquencés (265 accessions analysées). Les accessions choisies couvrent tous les habitats géographiques et écologiques mondiaux. Les 48 accessions sont subdivisées en sous-core de 8 accessions chacune, les accessions de la core 8 étant les plus polymorphes (représentant à elles seules 73 % de la variabilité estimée à partir des 10 locus). A cause de problème d'amplification des gènes *At-eIF4E2* chez certaines accessions, mon échantillon d'étude a été réduit à 33 accessions parmi les 48 initialement prévues (Tableau III.1.C). Deux accessions supplémentaires viennent compléter cet échantillon : Col-0, accession de référence, et WS, accession résistante à plusieurs potyvirus, tel que le WMV (Pacaly, 2005) et plusieurs souches de PPV (Decroocq *et al.*, 2006). L'espèce *A. lyrata*, l'une des plus proches de l'espèce modèle *A. thaliana*, a été utilisée comme groupe externe afin d'estimer la divergence interspécifique des gènes *eIF4E*. Les ADN de l'accession North Carolina d'*A. lyrata ssp lyrata* nous ont été fournis par J. Leppälä (Université de Oulu, Finlande).

### III.2.2. Amplification et séquençage des gènes *eIF4E*

Pour toutes les espèces, l'ADNg a été extrait à partir de feuilles en croissance selon un protocole de microextraction au CTAB adapté de la méthode de Fulton *et al.* (1995) (annexe 6). Les ARN totaux ont été isolés à partir de feuilles en croissance en utilisant le kit TRI-Reagent® (Sigma-Aldrich, St Louis, USA) (annexe 6). Les ADNc ont été obtenus par transcription inverse (annexe 7). Des amorces spécifiques pour amplifier les 3 ADNc des *Ca-eIF4E* et *Ly-eIF4E* ont été définies dans les régions 5' et 3' UTR. Les ADNg partiels des 3 gènes *Ca-eIF4E* et *Ly-eIF4E* ont été amplifiés à partir d'amorces définies dans les exons 2 et 5 (Tableau III.2). Les ADNg des 5 gènes *At-eIF4E* ont été amplifiés à partir d'amorces de PCR définies dans les régions 5' et 3' UTR (Tableau III.2). L'ADNg de *At-eIF4E2* n'a pu être amplifié avec succès que chez les 35 accessions répertoriées dans le tableau III.1.

Les produits PCR issus de 2 réactions d'amplification indépendantes ont été poolés et séquencés directement par les amorces sens, internes et antisens, afin de recouvrir la totalité de la séquence du gène et d'obtenir des séquences chevauchantes. Les séquençages des gènes

de piment et de tomate ont été effectués sur la plateforme de génotypage et séquençage de Montpellier (S. Santoni et A. Weber, INRA-UMR Diversité et Génome des Plantes Cultivées, avec un séquenceur 16 capillaires ABI 3130XL POP7) et les séquençages des gènes d'*A. thaliana* sur la plateforme de génotypage à haut débit dans l'équipe de Dominique Brunel (INRA, Evry, avec un séquenceur ABI 377XXL 96 puits, Applied Biosystems).

### **III.2.3. Amplification et séquençage des locus indépendants "non-eIF4E"**

Pour distinguer les effets de la sélection des effets démographiques susceptibles d'affecter les gènes *eIF4E*, plusieurs locus indépendants "non-eIF4E", régulièrement répartis sur le génome de chaque espèce, ont été amplifiés et séquencés chez l'ensemble des accessions de l'étude.

Les profils de polymorphisme des gènes *Ca-eIF4E* ont été comparés aux profils de diversité de 11 locus indépendants : 1 gène chloroplastique codant pour la grande sous-unité de la RuBisCo, 2 gènes de ménage nucléaires codant pour le facteur d'élongation eF1 $\alpha$  et l'histone H2B et 8 locus répartis sur plusieurs chromosomes du piment (Figure III.1, Yi *et al.*, 2006). Il s'agit d'ADNg obtenus à partir de séquences EST codant pour des protéines de fonctions diverses. Les amorces utilisées pour amplifier et séquencer ces 11 locus sont répertoriées dans le tableau III.2.

Huit locus indépendants chez la tomate, utilisés précédemment par Rose *et al.* (2007) pour comparer le polymorphisme du gène d'intérêt *Pto* à la diversité globale de l'espèce *Lycopersicon* spp., ont été amplifiés et séquencés. La localisation chromosomique et les protéines codées par ces gènes (Roselius *et al.*, 2005) sont indiquées dans le tableau III.3.A. Les conditions d'amplification de ces 8 locus indépendants sont disponibles sur le site <http://www.zi.biologie.uni-muenchen.de/evol/Download/TomatoPCR.html>. Les amorces utilisées sont répertoriées dans le tableau III.2.

Les profils de polymorphisme des gènes *At-eIF4E* ont été comparés aux profils de diversité des 10 locus utilisés pour générer la core collection de McKhann *et al.* (2004). Il s'agit de fragments de gènes de fonctions diverses (Tableau III.3.B), comprenant des régions exoniques et/ou introniques, distribués sur l'ensemble du génome d'*A. thaliana* (McKhann *et*

*al.*, 2004). Les données de séquence de ces gènes pour les 35 accessions d'*A. thaliana* étudiées ici ont été récupérées sur le site de l'URGI (<http://urgi.versailles.inra.fr/>).

### **III.2.4. Analyse des données de séquence**

#### ***III.2.4.1. Alignement et détection de polymorphisme***

Les séquences sens et antisens ont été alignées et corrigées manuellement avec les logiciels Genalys (<http://software.cng.fr>) et BioEdit (Hall, 1999). Les sites nucléotidiques polymorphes ont été détectés par Genalys. La séquence peptidique a été prédite par BioEdit. Pour chaque position polymorphe, les fréquences de chaque variant nucléotidique et les estimations du polymorphisme ( $\pi$ ,  $\theta_w$ ) ont été obtenues avec le logiciel DnaSP version 4 (Rozas *et al.*, 2003).

#### ***III.2.4.2. Tests statistiques et détection de signatures de sélection***

Les D de Tajima sont des valeurs statistiques basées sur la différence entre 2 estimateurs de la "quantité de polymorphisme" à l'équilibre ( $\theta = 4N_e\mu$ ). Le test de neutralité du D de Tajima a été réalisé avec le logiciel DnaSP (Rozas *et al.*, 2003). Les probabilités (P-values) d'obtenir les D de Tajima observés ont été estimées sur la base de 10 000 simulations de coalescence, sous l'hypothèse d'absence d'évènement de recombinaison et dans une population de taille constante.

Les phylogénies ont été réalisées avec le logiciel MEGA3 (Kumar *et al.*, 2004). Les arbres ont été construits selon la méthode Neighbour-joining qui prend en compte la divergence moyenne de chacune des séquences avec les autres (Saitou and Nei, 1987), en intégrant les valeurs de bootstrap qui permettent d'évaluer la signification statistique de la position des branches dans un arbre phylogénétique.

Les profils de divergence de séquence aux locus *eIF4E* ont été analysés en utilisant des modèles de substitution des codons (Goldman and Yang, 1994). La recherche de sites

sous sélection positive a été effectuée avec le logiciel codeml du package PAML version 3.14 (Yang *et al.*, 2000). Ce programme permet de faire l'analyse phylogénétique à partir des codons selon le principe de maximum de vraisemblance (c'est-à-dire la construction de l'arbre phylogénétique le plus plausible). Le programme analyse la distribution du paramètre  $\omega$ , qui représente le ratio dN/dS (taux de substitutions nonsynonymes par site nonsynonyme sur taux de substitutions synonymes par site synonyme) par site et par branche. Seule de la sélection positive, qui reflète un grand nombre de changements en acides aminés adaptatifs, engendre un  $\omega > 1$ . Plusieurs modèles d'estimation de la distribution de  $\omega$  sont proposés et se différencient par l'hypothèse de départ émise (neutralité ou sélections) (Yang *et al.*, 2000). Le modèle M0 sert de modèle nul pour tester l'hétérogénéité de  $\omega$  le long de la séquence. Les tests par LRT (likelihood ratio test) de M2a contre M1a et de M8 contre M7 (modèles emboîtés) sont les 2 moyens utilisés pour tester statistiquement la présence de sites dont le  $\omega$  dépasse 1. Après l'estimation des paramètres des modèles, le logiciel permet de calculer la probabilité *a posteriori* pour chaque site d'appartenir à la classe  $\omega > 1$ , et ainsi de détecter individuellement les sites sous sélection positive (Nielsen and Yang, 1998).

### **III.2.5. Test de résistance aux potyvirus**

Les souches des potyvirus WMV-Fr (H. Lecoq, INRA, Montfavet, France), TuMV-CDN1 (J. Walsh, Warwick University, UK) et CIYVV (I. Uyeda, Hokkaido University, Sapporo, Japon) ont été utilisées pour les tests phénotypiques chez certaines accessions d'*A. thaliana*. Les accumulations virales ont été détectées 30 jours post-inoculation (jpi) par DAS-ELISA (Clark and Adams, 1977) et/ou RT-PCR (cf annexes 4 et 7).

### **III.3. Polymorphisme des gènes *eIF4E* chez *Capsicum* spp.**

#### **III.3.1. Contexte et objectif**

Trois gènes *eIF4E* (*Ca-eIF4E*) ont été identifiés chez le piment à partir des EST de tomate disponibles dans les bases de données ([http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/est\\_ann.pl?gudb=tomato](http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/est_ann.pl?gudb=tomato)). Cependant, la cartographie génétique des marqueurs RFLP obtenus à partir de sondes EST des gènes *eIF4E1* et *eIF(iso)4E* suggère qu'il existerait au moins 2 copies supplémentaires de ces gènes chez *Capsicum* spp. Un criblage de la banque BAC de piment est actuellement en cours pour isoler l'ensemble des gènes *Ca-eIF4E* (collaboration H. Bergès, Centre National de Ressources Genomiques Végétales, INRA, Toulouse). Les premiers résultats semblent confirmer que 3 gènes *eIF4E* et 2 gènes *eIF(iso)4E* seraient présents chez le piment.

D'un point de vue fonctionnel, peu de choses sont connues sur le rôle respectif des gènes *eIF4E* du piment. L'alignement des séquences en acides aminés des 3 *Ca-eIF4E* avec les protéines eIF4E d'*A. thaliana* (*At-eIF4E*) montre que tous les acides aminés nécessaires pour la fixation à la coiffe des ARNm et les acides aminés essentiels pour la stabilité des protéines sont bien conservés, suggérant que les 3 protéines *Ca-eIF4E* sont potentiellement capables d'initier la traduction. Il a été observé que l'absence de la protéine *Ca-eIF(iso)4E* chez l'accession *C. annuum* Perennial n'affecte pas la fitness de la plante. Ce résultat indique que (i) soit la fonction de *Ca-eIF(iso)4E* n'est pas indispensable, (ii) soit les autres formes *Ca-eIF4E* sont capables de compenser l'absence de *Ca-eIF(iso)4E*. Chez le tabac et *A. thaliana*, la suppression de l'expression du gène *eIF(iso)4E* entraîne effectivement une augmentation compensatrice du niveau de protéines eIF4E (Combe *et al.*, 2005; Duprat *et al.*, 2002), suggérant que les différents paralogues pourraient avoir une certaine redondance fonctionnelle.

Le gène *Ca-eIF4E1* est impliqué dans la résistance au PVY et au TEV chez plusieurs accessions des espèces *C. annuum* et *C. chinense*. Neuf allèles de résistance au locus *Ca-eIF4E2-pvr2* ont été identifiés et se distinguent de l'allèle de sensibilité par différentes combinaisons de substitutions (§ II). Le gène *Ca-eIF(iso)4E* correspond au gène récessif *pvr6* de résistance à un autre potyvirus, le PVMV, uniquement chez l'accession *C. annuum*

Perennial. Ce gène présente une délétion de 82 nucléotides dans sa séquence codante qui entraîne l'apparition d'un codon stop conduisant à la synthèse d'une protéine tronquée non fonctionnelle. A ce jour, aucun locus de résistance aux potyvirus correspondant au gène Ca-eIF4E2 n'a été identifié. Cependant, les résultats présentés au § II.3, montrant que la protéine Ca-eIF4E2 est capable d'interagir avec la VPg de certaines souches de PVY et de TEV, suggèrent que ce facteur pourrait également avoir un rôle dans les interactions plantes/potyvirus.

Nous disposons à l'INRA de Montfavet d'une importante collection de ressources génétiques du genre *Capsicum* spp., regroupant plus de 1150 accessions internationales et locales (Dijan-Caporalino *et al.*, 2006). Des données archéologiques montrent que le piment a subi des événements de domestication indépendants, entre 5 000 et 7 000 ans avant JC, du Nord au Sud du continent américain (Perry *et al.*, 2007). Cinq espèces domestiquées ont été caractérisées parmi les 25 qui composent aujourd'hui le genre : *C. annuum*, *C. frutescens*, *C. chinense*, 3 espèces qui forment le "complexe *C. annuum*", *C. baccatum* et *C. pubescens*. A partir du centre primaire situé en Amérique latine, les piments ont été diffusés à travers le monde vers l'Europe, l'Afrique et l'Asie (Figure III.2). A ce jour, les *C. annuum* sont répartis dans le monde entier, avec des milliers de cultivars résultant de cette diversification secondaire. *C. frutescens* et *C. chinense* sont les 2 autres espèces les plus largement répandues. Ces 3 espèces formant le complexe *C. annuum* peuvent se croiser entre elles et sont probablement issues d'un ancêtre commun récent (Jensen *et al.*, 1979; Pickersgill *et al.*, 1979; Tam *et al.*, 2005); la sous-espèce *C. annuum glabriusculum* apparaît être l'ancêtre le plus probable du complexe des *C. annuum* domestiqués (Dijan-Caporalino *et al.*, 2006). *C. chacoense* et *C. galapagoense* sont des espèces sauvages proches du complexe *C. annuum*. *C. baccatum* et *C. pubescens* sont clairement différenciées des *C. annuum* et restent confinées en Amérique du Sud. *C. microcarpum* et *C. praetermissum* sont les espèces sauvages apparentées aux *C. baccatum* et *C. cardenasii* et *C. eximium* celles apparentées aux *C. pubescens*.

L'objectif de cette partie est de compléter la description de la diversité du gène *Ca-eIF4E1* chez le piment *via* la mise en œuvre de méthodes d'analyse du polymorphisme. Ces analyses seront étendues aux autres gènes *eIF4E* du piment afin de déterminer quel type de pression de sélection (s'il y en a) s'exerce sur ces gènes.

### III.3.2. Recherche des pressions de sélection s'exerçant sur le gène *Ca-eIF4E1* par l'analyse de son polymorphisme nucléotidique

#### III.3.2.1. Polymorphisme intraspécifique du gène *Ca-eIF4E1*

Afin de compléter la description de la diversité du gène *Ca-eIF4E1* décrite précédemment, des analyses portant sur la variabilité intraspécifique de la séquence codante et des introns 2, 3 et 4 de ce gène chez 25 accessions de *C. annuum* ont été réalisées. L'intron 1 (~ 3500 pb) n'a pas été séquencé du fait de la présence de nombreuses séquences répétées (Ruffel *et al.*, 2005).

Les gènes *Ca-eIF4E1* séquencés chez l'ensemble des accessions contiennent 5 exons dont la longueur et la position sont parfaitement conservées. Aucun événement d'insertion/déletion (indel) n'a été détecté ni dans les exons ni dans les introns 2, 3 et 4. Très peu de sites nucléotidiques polymorphes sont détectés et la diversité nucléotidique est plus importante dans les exons que dans les introns ( $\pi_{\text{coding}} = 0,00332$ ;  $\pi_{\text{non coding}} = 0,00116$ ; Tableau III.4). Les analyses par fenêtres glissantes utilisées pour caractériser le profil de diversité le long de la séquence génomique permettent de visualiser que la diversité nucléotidique pour le gène *Ca-eIF4E1* est plus élevée dans les exons, et plus particulièrement dans les exons 1 et 2 (Figure III.3).

Afin de déterminer si la faible diversité nucléotidique de *Ca-eIF4E1* est représentative de la diversité globale du génome de *C. annuum*, elle a été comparée aux valeurs de diversité nucléotidique pour les 11 locus indépendants. La valeur  $\pi$  moyenne pour ces locus est de 0,00283 (de 0,00052 à 0,01273;  $\sigma = \pm 0,0045$ ) (Tableau III.4). Ainsi, le gène *Ca-eIF4E1* présente une diversité nucléotidique faible mais qui ne se différencie pas de la diversité globale du génome. Ces résultats sont en accord avec les analyses de polymorphisme intraspécifiques réalisées à l'aide de marqueurs moléculaires de type SSAP (Sequence-Specific Amplification Polymorphisms; Tam *et al.*, 2005). L'existence d'un goulot d'étranglement étroit dû à la sélection moderne serait à l'origine de cette faible variabilité au sein des *C. annuum* (Dijan-Caporalino *et al.*, 2006).

Des valeurs de  $\pi$  inférieures aux valeurs de  $\theta_w$  (taux de mutation de Watterson qui ne prend pas en compte la fréquence des allèles) et une valeur de D de Tajima négative marquent un déficit en polymorphismes équilibrés, c'est-à-dire un excès de variants alléliques rares (Tableau III.4; Figure III.4). L'hypothèse de sélection balancée pour expliquer le maintien de ce grand nombre d'allèles est donc avancée. Cependant, dans le cas de sélection balancée, une augmentation de la variation nucléotidique aux sites liés au locus qui subit cette pression de sélection est attendue. Or aucun excès de polymorphisme (0 substitutions synonymes) n'est observé (Tableau III.4; Figure III.5). L'hypothèse de sélection balancée est donc rejetée au profit de l'hypothèse d'une pression de sélection positive qui s'exercerait au niveau de certains sites de Ca-eIF4E1 et qui conduit à l'augmentation de la fréquence de plusieurs allèles.

### ***III.3.2.2. Structure haplotypique de Ca-eIF4E1***

A partir des différentes combinaisons de substitutions en acides aminés identifiées, les 25 accessions de *C. annuum* ont pu être regroupées en 10 variants alléliques (Figure III.6). pvr2<sup>+</sup> est l'haplotype majoritaire regroupant 11 accessions alors que les autres haplotypes ne sont constitués que de 3 accessions au maximum. De plus, l'enracinement de l'arbre (avec une espèce sauvage éloignée des *C. annuum*, telle que *C. cardenasii*) se fait au niveau de pvr2<sup>+</sup>, suggérant que cet haplotype majoritaire représente l'ancêtre commun. A partir de pvr2<sup>+</sup>, on distingue 3 chemins évolutifs :

(i) le chemin majoritaire, emprunté par 10 des 14 accessions, passe par la substitution V67E. Une seule accession appartenant à l'haplotype pvr2<sup>4</sup> ne possède que cette substitution (PI322719). A partir de pvr2<sup>4</sup>, 3 chemins sont à nouveau possibles, conduisant aux haplotypes pvr2<sup>5</sup>, pvr2<sup>3</sup> et pvr2<sup>1</sup>. Les branches conduisant aux haplotypes pvr2<sup>2</sup> et pvr2<sup>7</sup> partent de pvr2<sup>1</sup>.

(ii) 3 substitutions en acides aminés conduisent à l'haplotype pvr2<sup>9</sup>, constitué d'une seule accession (Chile de Arbol). A partir de cet haplotype, 1 substitution conduit à l'haplotype pvr2<sup>6</sup> regroupant 2 accessions (Maroc 1 et LP1).

(iii) le dernier chemin conduit à l'haplotype pvr2<sup>8</sup> représenté par une seule accession (PI195301).

### III.3.2.3. Recherche de sites sous sélection positive

La recherche des sites sous sélection positive a été réalisée à partir des séquences ADNc de *Ca-eIF4E1* chez les 25 accessions de *C. annuum*, d'une longueur de 687 nucléotides, soit 229 codons. Deux analyses ont été conduites : la première considère la variation de  $\omega$  entre branches de l'arbre (car un évènement d'adaptation peut avoir lieu dans une branche spécifique) et la seconde considère la variation de  $\omega$  entre sites (car les substitutions adaptatives sont susceptibles de se concentrer à certaines positions dans la séquence). L'analyse des valeurs de LRT (Likelihood Ratio Test) obtenues à partir des modèles d'hétérogénéité de  $\omega$  en fonction des branches de l'arbre nous indique que le modèle 0, qui correspond à un  $\omega$  unique entre branches de l'arbre, est le plus probable. Aucun évènement d'adaptation spécifique ne semble donc avoir eu lieu dans une des branches de l'arbre.

Pour l'identification de la variation du rapport  $\omega$  entre sites, l'estimation des modèles d'hétérogénéité entre codons, et les tests permettant de les comparer, sont reportés dans le tableau III.5. Seuls les résultats obtenus pour les modèles emboîtés M7 et M8, dont la comparaison permet de tester si une catégorie de sites évolue sous sélection positive, sont présentés, car statistiquement plus significatifs que les résultats obtenus avec les modèles M1a et M2a. L'estimation des paramètres des modèles (valeurs de LRT) montre que le modèle M8, qui prend en compte la sélection positive, est plus approprié que le modèle M7. Les valeurs de  $\omega$  pour chaque site ont donc été prédites à partir de ce modèle.

Neuf sites présentent une valeur de  $\omega$  supérieure à 1 (66P, 67V, 68A, 73A, 74A, 79L, 107G, 109D et 205D), dont 5 avec une probabilité statistique supérieure à 95 % (Tableau III.6; Figure III.7). Les 5 acides aminés sous sélection positive sont situés exclusivement dans la région I de la protéine Ca-eIF4E1.

Afin de comparer les valeurs de  $\omega$  prédites en intra-spécifique avec les données inter-spécifique, la recherche de sites sous sélection positive a été effectuée à partir des données de séquences ADNc du gène *Ca-eIF4E1* chez toutes les espèces à l'exception de *C. annuum* (26 accessions de 11 espèces). En inter-spécifique, aucun site n'est prédit pour être sous sélection positive avec une probabilité supérieure à 95 %, ce qui signifie que seules les protéines Ca-eIF4E1 des *C. annuum* présentent des sites sous sélection positive.

### III.3.3. Recherche des pressions de sélection qui s'exercent sur les gènes *Ca-eIF4E2* et *Ca-eIF(iso)4E* par l'analyse de leur diversité nucléotidique

#### III.3.3.1. Polymorphisme intraspécifique des gènes *Ca-eIF4E2* et *Ca-eIF(iso)4E*

La variabilité intraspécifique des séquences codantes des gènes *Ca-eIF4E2* et *Ca-eIF(iso)4E* a été évaluée chez les 25 accessions de *C. annuum*. Les données de variabilité nucléotidique des introns 2, 3 et 4 du gène *Ca-eIF(iso)4E* ont également été obtenues. L'ADNg de *Ca-eIF4E2* n'a pas pu être amplifié à cause de problèmes d'amorçage des amorces lors des réactions de PCR. Tous les gènes séquencés chez l'ensemble des accessions contiennent 5 exons dont la longueur et la position sont parfaitement conservées, mis à part la délétion de 82 nucléotides dans la séquence du gène *Ca-eIF(iso)4E* de l'accèsion Perennial, qui avait déjà été identifiée lors de précédents travaux (Ruffel *et al.*, 2006). Aucun autre évènement d'insertion/délétion (indel) n'a été détecté dans les exons. Dans les introns, seulement trois indels ont été identifiés dans l'intron 3 du gène *Ca-eIF(iso)4E*.

Très peu de sites nucléotidiques polymorphes ont été détectés ( $\pi$  compris entre 0,00012 et 0,00293 dans les exons; Tableau III.4). Les analyses par fenêtres glissantes utilisées pour caractériser le profil de diversité le long de la séquence génomique de *Ca-eIF(iso)4E* (Figure III.3) montrent que la diversité nucléotidique pour ce gène est plus élevée dans les introns que dans les exons, contrairement à *Ca-eIF4E1*. Les gènes *Ca-eIF4E2* et *Ca-eIF(iso)4E* présentent une diversité nucléotidique faible mais qui ne se différencie pas de la diversité globale du génome des *C. annuum*, en comparaison avec les valeurs obtenues pour les 11 locus indépendants.

La diversité nucléotidique du gène *Ca-eIF4E2* est inférieure d'un facteur 10 par rapport à la diversité des 2 autres gènes *Ca-eIF4E* et des locus indépendants (Tableau III.4). Une seule substitution nucléotidique a été détectée chez une seule accession de *C. annuum* (PI187331). Il s'agit d'une substitution non-synonyme, correspondant à la mutation P132L. Seulement 2 variants alléliques ont donc été identifiés en fréquence totalement déséquilibrée à partir de l'identification d'un seul polymorphisme (Figure III.6). L'ensemble de ces résultats

suggère qu'une pression de sélection négative très forte contre les variants alléliques s'exerce sur le gène *Ca-eIF4E2*.

En ce qui concerne *Ca-eIF(iso)4E*, des valeurs de  $\pi$  supérieures aux valeurs de  $\theta_w$  et une valeur de D de Tajima significativement très positive marquent un excès de variants en fréquence équilibrée (Tableau III.4, Figure III.4). En effet, les accessions *C. annuum* se divisent en 2 groupes distincts différenciés par une seule substitution T90A (Figure III.6). Un excès significatif de polymorphismes synonymes (par site synonyme) par rapport aux polymorphismes nonsynonymes (par site nonsynonyme) ( $\pi_a / \pi_s < 1$ , Tableau III.4, Figure III.5), indique que la majorité des substitutions non synonymes est éliminée par la sélection. Ces valeurs appuient l'hypothèse qu'une sélection négative contre les variants alléliques s'exerce sur le gène *Ca-eIF(iso)4E*.

### **III.3.3.2. Recherche de sites sous sélection positive**

Les mêmes analyses de recherche de sites sous sélection positive ont été effectuées sur les ADNc de *Ca-eIF4E2* et *Ca-eIF(iso)4E*, d'une longueur de 672 et 609 nucléotides, soit 272 et 203 codons respectivement. Les valeurs de  $\omega$  pour chaque site ont été prédites avec le modèle M8 (Tableau III.5). Pour *Ca-eIF4E2*, seul le site 132P a un  $\omega$  supérieur à 1, mais avec une probabilité de 78 %. Pour *Ca-eIF(iso)4E*, le site 90A présente un  $\omega > 1$  mais avec une probabilité de 77 % donc peu significative (Tableau III.6). Tous les acides aminés des protéines *Ca-eIF4E2* et *Ca-eIF(iso)4E* présentent donc des valeurs de  $\omega$  prédites inférieures à 1, suggérant que tous les sites sont sous sélection négative.

### **III.3.4. Polymorphisme interspécifique des 3 gènes *Ca-eIF4E* chez 51 accessions de *Capsicum* spp.**

Les alignements des séquences ADNc de chaque gène *Ca-eIF4E* pour les 51 accessions de *Capsicum* sont représentés sur le tableau III.7. Les substitutions nucléotidiques ont été réparties en 4 catégories : espèce-spécifiques (en jaune), ponctuelles (isolées chez une ou quelques accessions d'une même espèce, en bleu), partagées entre accessions de différentes

espèces (en vert) ou localisées dans les régions impliquées dans la résistance aux potyvirus (en rouge). Une observation importante ressort de ces alignements : les gènes *Ca-eIF4E* de la sous-espèce *C. annuum glabriusculum* partagent plusieurs polymorphismes communs avec les autres espèces de *Capsicum*, dont une mutation homoplasique (en position 656) avec *C. cardenasii* et *C. pubescens* pour *Ca-eIF4E1*, et 2 mutations impliquées dans la résistance de certaines accessions de *C. annuum* détectées chez *Ca-eIF4E1*.

Les phylogénies des gènes *Ca-eIF4E* ont été réalisées afin de déterminer si ces locus sont des marqueurs représentatifs de la diversité au sein des *Capsicum*. Les arbres phylogénétiques, construits par la méthode de Neighbour-joining et basés sur la séquence codante de ces gènes, sont comparés avec les arbres obtenus à partir des séquences des locus indépendants non-eIF4E (Figure III.8).

Comme indiqué précédemment par les alignements, les espèces *C. cardenasii* et *C. pubescens* sont les plus distantes. Ces résultats sont en accord avec de précédentes études basées sur des marqueurs SSAP, sur des données de séquence des gènes *apt-rbcL* et *waxy* et sur les données morphologiques (Tam *et al.*, 2005). Les phylogénies sont également respectées pour les espèces *C. chacoense*, *C. baccatum*, *C. microcarpum*, *C. eximium*, *C. praetermissum* et *C. galapogense*. En revanche, sur la base des données de séquences *Ca-eIF4E*, la phylogénie des espèces *C. chinense*, *C. frutescens*, *C. annuum* et *C. annuum glabriusculum* est moins discriminante *i.e.*, les espèces ne sont pas clairement séparées en groupes distincts. Ces résultats sont cependant en accord avec de précédents travaux décrivant la forte proximité de ces espèces, regroupés dans un "complexe *C. annuum*" (Dijan-Caporalino *et al.*, 2006) et résultent probablement du faible polymorphisme des gènes *eIF4E*.

*C. chinense*, *C. frutescens* et *C. annuum* sont des espèces domestiquées tandis que la sous-espèce *C. annuum glabriusculum* est une espèce sauvage. Les arbres phylogénétiques basés sur les séquences ADNc des locus indépendants et des gènes *Ca-eIF4E* confirment que cette sous-espèce serait l'ancêtre commun du complexe "*C. annuum*". La diversité intraspécifique des *C. annuum*, mesurée à partir des séquences des locus indépendants, est réduite d'un facteur 3 par rapport à la diversité intraspécifique des *C. annuum glabrisuculum* ( $\pi_{\text{glabriusculum}} = 0,00294$  et  $\pi_{\text{annuum}} = 0,00097$ ). Cette réduction de diversité est moins marquée dans le cas de tous les gènes *Ca-eIF4E* ( $\pi_{\text{glabriusculum}} = 0,00315$  et  $\pi_{\text{annuum}} = 0,00199$ ). Ces résultats, couplés à la structure des arbres phylogénétiques, confirment que les *C. annuum*

sont issus des *C. annuum glabriusculum* et qu'un goulot d'étranglement récent est à l'origine de la réduction de diversité observée. Le partage de plusieurs polymorphismes communs chez Ca-eIF4E entre les 2 sous-espèces peut être relié à l'existence de flux de gènes (Tableau III.7). Ces processus peuvent se produire entre des plantes cultivées et des espèces sauvages apparentées sexuellement compatibles, et permettre ainsi l'échange d'allèles des plantes cultivées vers leur parent sauvage ou inversement. De façon très intéressante, 2 polymorphismes impliqués dans la résistance aux potyvirus chez Ca-eIF4E1 de *C. annuum annuum* (V67E et L79R) sont partagés avec les accessions El Tajin et Managua de *C. annuum glabriusculum*. Ces résultats suggèrent donc que certains allèles *Ca-eIF4E* de résistance aux potyvirus étaient présents chez *Capsicum* avant la domestication de cette espèce.

#### III.4.4. Conclusions

Les résultats obtenus *via* les analyses fonctionnelles suggèrent fortement que la signature d'évolution positive identifiée chez *Ca-eIF4E1* au niveau de certains sites pourrait résulter d'une pression de sélection exercée par les potyvirus, même si le polymorphisme global de ce gène ne se différencie pas du polymorphisme des autres locus étudiés.

L'ensemble des résultats argumentent très fortement en faveur de l'existence d'une pression de sélection diversifiante exercée par les potyvirus sur les gènes *Ca-eIF4E1* au niveau de certains acides aminés, et plus particulièrement au niveau des acides aminés de la région I de la protéine Ca-eIF4E1 impliqués dans la résistance au PVY. Les acides aminés de la région II n'apparaissent pas sous sélection positive ( $\omega > 1$  mais non significatifs). Les résultats du chapitre 2 montrant le rôle prépondérant de la région I dans la résistance au PVY suggèrent que, dans le cas des *C. annuum*, la pression exercée par le PVY est plus importante que celle exercée par le TEV. Une récente analyse du polymorphisme intraspécifique du gène *eIF4E1* chez *C. chinense* (Jahn *et al.*, données non publiées) indique que les acides aminés en position 21, 49, 54, 69, 74, 84, 107, 109, 110 et 209 sont sous pression de sélection positive. Ces sites ne sont pourtant pas détectés au cours de nos analyses interspécifiques, mises à part les acides aminés en position 74 et 109. Le fait que notre échantillonnage interspécifique ne comprenne que 2 accessions de *C. chinense* pourrait expliquer ces différences de résultats. De nouvelles analyses ont donc été réalisées en intégrant 5 accessions supplémentaires de *C.*

*chinense* (Chi7, Diego, Baili 3, Notto et Chi8). Aucun des sites ne ressort sous pression de sélection positive (résultats non présentés). Il peut s'agir soit d'un biais de notre échantillonnage qui contient trop peu d'accessions possédant des mutations dans cette région, soit d'une réalité biologique indiquant que la pression de sélection exercée par le TEV est moins importante que celle exercée par le PVY. Cette seconde hypothèse est renforcée par le fait que, si l'on considère la variabilité naturelle du genre *Capsicum*, la fréquence de résistance au TEV est faible par rapport au PVY (près de 40% des accessions résistantes, Palloix & Daubèze, communication personnelle).

D'après la structure haplotypique établie pour Ca-eIF4E1, il apparaît que les 11 accessions de *C. annuum* partageant l'haplotype majoritaire (pvr2<sup>+</sup>) sont sensibles aux PVY et TEV tandis que toutes les accessions des autres haplotypes sont résistantes à l'un et/ou l'autre de ces potyvirus, renforçant le lien entre la résistance aux potyvirus et les chemins évolutifs. Parmi les substitutions en acides aminés seule, la D205G n'est pas localisée dans les régions I et II, et n'est pas retrouvée dans les protéines eIF4E de résistance chez les autres espèces. Cette substitution apparaît indépendamment dans 2 chemins évolutifs distincts (pour passer des haplotypes pvr2<sup>+</sup> à pvr2<sup>9</sup>, de pvr2<sup>4</sup> à pvr2<sup>3</sup> et de pvr2<sup>1</sup> à pvr2<sup>7</sup>), suggérant que l'apparition de cette mutation n'est pas due au hasard. De plus, cette substitution n'est jamais retrouvée seule; elle est toujours en combinaison avec 1 ou 2 mutation(s) sous sélection positive impliquée(s) dans la résistance aux potyvirus. D'après ces résultats, 2 hypothèses peuvent être formulées quant au rôle de la substitution D205G : soit elle a un rôle direct dans l'interaction avec d'autres virus à ARN non testés ici, soit elle a un rôle indirect lié à la biochimie de la protéine eIF4E. Les mutations V67E, A68E et A73D entraînent des modifications de la charge électrostatique globale de la protéine, avec l'apparition d'acides aminés hydrophiles acides à la place d'acides aminés hydrophobes. En revanche, la mutation D205G correspond au passage d'un acide aminé acide à un acide aminé neutre. La substitution D205G pourrait donc être une mutation compensatrice permettant de rééquilibrer la charge globale de la protéine eIF4E. Par ailleurs, aucune accession ne possède seulement les mutations A68E ou A73D. Trois hypothèses sont avancées pour expliquer l'absence de ces intermédiaires : (i) la taille de l'échantillonnage est trop réduite; le criblage d'un plus grand nombre d'accessions de *C. annuum* permettrait de tester cette hypothèse, (ii) ces mutations imposent un coût de fitness trop important pour la plante et sont donc compensées rapidement par d'autres mutations, (iii) ces mutations seules ne suffisent pas à la résistance, il est nécessaire d'accumuler plusieurs mutations pour induire la résistance.

Les acides aminés essentiels pour la stabilité de la structure des protéines eIF4E ou pour sa liaison avec la coiffe des ARNm présentent tous des  $\omega < 1$ , indiquant que ces acides aminés sont contraints. Ces résultats appuient en faveur de l'existence d'un compromis évolutif pour le maintien de la double fonction du gène *Ca-eIF4E1* : les acides aminés impliqués dans la fonction de traduction des ARNm sont contraints tandis que l'apparition de nouvelles mutations peut avoir lieu au niveau de certains sites spécifiques impliqués dans l'interaction plante- potyvirus.

Le profil de polymorphisme de *Ca-eIF4E2* et *Ca-eIF(iso)4E* indique que ces gènes sont contraints, que les mutations sont éliminées par la sélection, contrairement à *Ca-eIF4E1* où un relâchement de la contrainte au niveau de certains sites a été identifié. Les substitutions en acides aminés détectées chez ces 2 formes ne semblent pas liées à la résistance aux potyvirus : la substitution P132L a été identifiée dans la séquence de *Ca-eIF4E2* chez *C. annuum* PI187331, qui est une accession sensible aux PVY et TEV et la substitution A90T dans la séquence de *Ca-eIF(iso)4E* chez des accessions sensibles et résistantes, ne se localise pas dans les régions impliquées dans la résistance aux potyvirus.

Deux hypothèses peuvent être avancées pour expliquer la contrainte plus importante observée chez *Ca-eIF4E2* et *Ca-eIF(iso)4E* que chez *Ca-eIF4E*. La première est que les protéines *Ca-eIF4E2* et *Ca-eIF(iso)4E* possèdent d'autres fonctions essentielles dans la cellule de la plante et que les quelques substitutions nécessaires à la résistance induiraient une perte de cette (ces) fonction(s). Cette hypothèse pose cependant question dans le cas de la délétion du gène *Ca-eIF(iso)4E* observée une seule fois dans la variabilité naturelle. Même si les plantes porteuses de cet allèle ne présentent pas de réduction de fitness apparente, il est possible qu'une des fonction normalement remplie par eIF(iso)4E soit altérée. La seconde hypothèse est que seule la protéine *Ca-eIF4E1* peut être utilisée par les potyvirus. Les résultats présentés au chapitre II montrant que certains potyvirus sont capables d'interagir avec la forme *Ca-eIF4E2* et que le PVMV est capable d'interagir avec *Ca-eIF(iso)4E* (Ruffel *et al.*, 206) vont à l'encontre de cette hypothèse.

### **III.4. Polymorphisme des gènes *eIF4E* chez *Lycopersicon* spp.**

#### **III.4.1. Contexte et objectif**

Suite aux résultats obtenus pour le gène *Ca-eIF4E1* chez le piment, une étude du polymorphisme des 3 gènes *Ly-eIF4E* a été initiée chez la tomate, une autre Solanacée proche du piment et également infectée par les potyvirus mais pour laquelle une plus faible fréquence de gènes récessifs de résistance aux potyvirus a été observée dans la variabilité naturelle. L'objectif est de déterminer si l'on retrouve des traces de sélection positive similaires à *Ca-eIF4E1* et une contrainte forte sur les autres gènes *eIF4E*.

Comme chez le piment, à partir des EST de tomate disponibles dans les bases de données, 3 gènes *eIF4E* ont été identifiés : *Ly-eIF4E1*, *Ly-eIF4E2* et *Ly-eIF(iso)4E*. Les profils d'hybridation de sondes RFLP sur l'ADN génomique de cette espèce et leur cartographie montrent qu'il existe, comme chez le piment, certainement 2 paralogues supplémentaires (Ruffel, 2004).

Le gène récessif *pot1* de résistance au PVY et au TEV chez la tomate code pour la protéine eIF4E1 (Ruffel *et al.*, 2005) et représente donc l'orthologue du gène *eIF4E1-pvr2* chez le piment. Ce gène de résistance a été identifié chez l'espèce sauvage *L. hirsutum*, contrairement aux gènes de résistance aux potyvirus chez le piment identifiés chez les espèces cultivées *C. annuum* et *C. chinense*. Un autre type de résistance aux potyvirus, cette fois à spectre étroit (vis-à-vis d'une seule souche PVY) a été identifié chez une accession de *L. hirsutum* (LA1777) et 2 accessions de *L. esculentum* (Cervil et Microtom). Les gènes de résistance correspondants n'ont pas été identifiés mais se localisent au niveau du chromosome 3 dans la même région génomique que *Ly-eIF4E1*.

Le genre *Lycopersicon* comprend 9 espèces, toutes originaires de la région andine du Pérou et de l'Equateur, à l'exception de *L. cheesmanii* que l'on trouve exclusivement dans l'archipel des Galápagos (Rick *et al.*, 1990). La tomate cultivée *L. esculentum* a été introduite en Europe au XVI<sup>ème</sup> siècle. Nous disposons à l'INRA d'Avignon d'une collection de ressources génétiques regroupant plus de 1000 variétés de tomates provenant du monde entier, dont un grand nombre de variétés traditionnelles françaises et des pays méditerranéens ainsi

que des accessions des 8 espèces sauvages apparentées à l'espèce cultivée *L. esculentum*. Les espèces de *Lycopersicon* sont divisées en 2 groupes selon leur compatibilité de reproduction avec l'espèce cultivée *L. esculentum* var. *esculentum* : le complexe *Esculentum* et le complexe *Peruvianum*. Parmi les 9 espèces, 7 appartiennent au complexe *Esculentum* (*L. esculentum*, *L. pimpinellifolium*, *L. hirsutum*, *L. pennellii*, *L. cheesmanii*, *L. chmielewskii* et *L. parviflorum*) et 2 au complexe *Peruvianum* (*L. chilense* et *L. peruvianum*). L'espèce sauvage *L. pimpinellifolium* est l'ancêtre direct des *L. esculentum* (Rick, 1979; Nesbitt and Tanksley 2002). *L. peruvianum* est d'un point de vue génétique et morphologique l'espèce la plus distante de toutes les espèces du genre *Lycopersicon*.

### III.4.2. Analyse du polymorphisme intraspécifique des gènes *Ly-eIF4E*

Les ADNc des gènes *Ly-eIF4E1*, *Ly-eIF4E2* et *Ly-eIF(iso)4E* ont été amplifiés chez les 20 accessions de *L. esculentum*. Seules les données de séquence des introns 2, 3 et 4 de *Ly-eIF4E1* et *Ly-eIF(iso)4E* ont été obtenues. L'ADNg de *Ly-eIF4E2* n'a pas pu être amplifié à cause de problèmes d'ancrage des amorces PCR. La longueur des ADNc varie de 603 à 669 nucléotides (Tableau III.8). La position et la longueur des 5 exons sont conservées chez l'ensemble des accessions. Les introns 2, 3 et 4, totalisant de 811 à 1737 nucléotides pour *Ly-eIF4E1* et *Ly-eIF(iso)4E* respectivement, ont également une structure parfaitement conservée entre accessions.

La diversité nucléotidique des gènes *Ly-eIF4E* est très faible, aussi bien dans les régions codantes que non codantes ( $\pi$  compris entre 0,00029 et 0,00083;  $\sigma = \pm 0,00027$  dans les exons;  $\pi$  compris entre 0,00026 et 0,00086;  $\sigma = \pm 0,00042$  dans les introns) (Tableau III.8). De plus, en comparant cette diversité aux valeurs obtenues pour les 8 locus indépendants, il apparaît que tous les gènes *Ly-eIF4E* présentent une diversité nucléotidique similaire à la diversité globale du génome des *L. esculentum*. Ces résultats confirment des études réalisées à partir de marqueurs moléculaires qui montrent que la diversité intra-*L. esculentum* est très faible (Miller and Tanksley, 1990; Spooner *et al.*, 2005; Tam *et al.*, 2007). Cette réduction de diversité est la conséquence des 3 goulots d'étranglements survenus au cours de l'histoire des cultivars modernes : la domestication initiale, le transfert des variétés vers l'Europe au XVI<sup>ème</sup> siècle et les efforts de sélection moderne (Nesbitt and Tanksley,

2002). La diversité des gènes *Ly-eIF4E* est donc encore plus faible que la diversité des gènes *Ca-eIF4E* chez *C. annuum*, ce qui est cohérent avec la diversité globalement inférieure du génome de *L. esculentum* par rapport à la diversité des *C. annuum* (Tam *et al.*, 2005; Tam *et al.*, 2007).

Les tests d'écart à la neutralité pour les gènes *Ly-eIF4E1* et *Ly-eIF4E2* ne sont pas significatifs à cause de la quasi-absence de polymorphisme de ces gènes (une seule substitution nucléotidique ; Tableau III.9). Pour le gène *Ly-eIF(iso)4E*, une valeur de  $\pi$  inférieure à  $\theta_w$  et un D de Tajima négatif indiquent la présence d'un excès de variants alléliques rares (Tableau III.8). En effet, une seule accession (Mospomorist) présente tous les SNPs détectés (Tableau III.9). Un excès significatif de polymorphismes synonymes (par site synonyme) par rapport aux polymorphismes non-synonymes (par site non-synonyme), indique que la majorité des substitutions non-synonymes est éliminée par la sélection ; les 3 gènes *Ly-eIF(iso)4E*, *Ly-eIF4E1* et *Ly-eIF4E2* sont sous sélection purificatrice.

### III.4.3. Phylogénie des gènes *Ly-eIF4E*

Les 50 accessions appartenant à 9 espèces de *Lycopersion* spp. ont été utilisées pour caractériser le polymorphisme inter-spécifique des gènes *Ly-eIF4E*. Les arbres phylogénétiques basés sur la séquence codante des gènes *Ly-eIF4E* et sur la séquence ADNc des locus indépendants, construits par la méthode Neighbour-Joining, ont été comparés (Figure III.9).

Pour *Ly-eIF4E1*, *Ly-eIF4E2* et les locus indépendants, les phylogénies sont cohérentes par rapport à la taxonomie des espèces de *Lycopersion* spp. (Spooner *at al.*, 2005). Ces phylogénies correspondent également aux phylogénies obtenues à partir de la distribution et du polymorphisme de rétrotransposons chez la tomate (Tam *et al.*, 2007). Pour *Ly-eIF(iso)4E*, la phylogénie ne respecte pas la taxonomie des *Lycopersicon* spp. Le nombre limité de polymorphismes inter-spécifiques et les valeurs de bootstrap très faibles indiquant que les groupes établis n'ont aucune valeur taxonomique, pourraient expliquer l'incohérence de cette phylogénie.

Pour l'ensemble des gènes *Ly-eIF4E*, une très forte proximité entre les accessions de l'espèce domestiquée *L. esculentum* et de l'espèce sauvage *L. pimpinellifolium* est en accord avec le statut de *L. pimpinellifolium* d'ancêtre de l'espèce cultivée.

#### **III.4.4. Polymorphisme interspécifique et résistance aux potyvirus chez *Lycopersicon* spp.**

Les substitutions détectées entre les protéines Ly-eIF4E chez les 50 accessions de *Lycopersicon* spp. sont répertoriées sur le tableau III.10. Trois observations majeures ressortent de ce tableau :

(i) les espèces les plus polymorphes sont *L. hirsutum*, *L. chilense* et *L. peruvianum*, confirmant les résultats de phylogénie obtenus à partir des séquences codantes des gènes *Ly-eIF4E* (§ III.4.3).

(ii) la séquence en acides aminés Ly-eIF4E1 de *L. esculentum* Cervil et Microtom, accessions résistantes au PVY-LYE90, diffère de la séquence des génotypes sensibles de la même espèce par la substitution d'un seul acide aminé (D103E), suggérant que cette substitution pourrait être impliquée dans la résistance à spectre étroit. Deux autres accessions possèdent une protéine Ly-eIF4E avec la substitution D103E (Sweetie et Gardener's Delight) mais le phénotype de résistance/sensibilité au PVY-LYE90V de ces génotypes n'est pas connu. Il est à noter que le remplacement d'un acide aspartique en acide glutamique n'induit pas de changement majeur sur les propriétés biochimiques de la protéine eIF4E.

(iii) *L. hirsutum* PI24 présentent des substitutions en acides aminés dans les régions I et II impliquées dans la résistance aux potyvirus (gène *pot1*). Les substitutions E26D et S79F ne sont détectées que dans la séquence Ly-eIF4E2 de *L. hirsutum* LA1777, accession résistante à la souche PVY-LYE90V. Par alignement avec la séquences des protéines eIF4E impliquées dans la résistance, Ly-eIF4E1-pot1 chez la tomate, Ca-eIF4E1-pvr2 chez le piment, Ls-eIF4E-mo1 chez la laitue ou encore Ps-eIF4E-sbm1 chez le pois, il apparaît que la substitution S79F se localise au niveau de la région I impliquée dans la résistance. De plus, cette substitution induit des changements de propriétés biochimiques importants *via* le remplacement d'un acide aminé hydrophile en acide aminé hydrophobe possédant un cycle aromatique. L'importance de cette substitution dans la résistance de LA1777 au PVY devra donc être confirmée par mutagenèse dirigée et validée par expression transitoire.

### III.4.5. Recherche de sites sous sélection positive

Pour chaque *Ly-eIF4E*, les séquences codantes de toutes les accessions de *Lycopersicon* spp. ont été utilisées pour rechercher des sites sous sélection positive, représentant de 201 à 222 codons. L'estimation des modèles d'hétérogénéité entre codons, et les tests permettant de les comparer sont indiqués sur le tableau III.11. Les résultats des modèles 0 et 1 d'hétérogénéité entre branches de l'arbre montrent que le modèle le plus simple (M0), où toutes les branches évoluent selon un  $\omega$  identique, semble le plus approprié aux données au vu des valeurs de LRT obtenues. Les LRT pour les modèles M7 et M8, utilisés afin d'identifier de la variation de  $\omega$  entre sites, montrent que le modèle M8, qui prend en compte la sélection positive, est plus adapté que le modèle M7 pour *Ly-eIF4E1* et *Ly-eIF(iso)4E*. En revanche, pour le gène *Ly-eIF4E2*, le modèle M8 est moins approprié que le modèle M7, indiquant qu'aucun site de *Ly-eIF4E2* n'est sous sélection positive. Les valeurs de  $\omega$  pour chaque site de *Ly-eIF4E1* et *Ly-eIF(iso)4E* ont été prédites à partir du modèle M8 (Tableau III.12). Pour *Ly-eIF4E1*, 3 sites présentent une valeur de  $\omega$  supérieure à 1 (49L, 60P et 212L), dont un seul avec une probabilité statistique supérieure à 95 %. Cette position correspond à la substitution L49V présente chez 4 accessions de *L. pimpinellifolium*. La valine et la leucine sont 2 acides aminés très proches d'un point de vue biochimique présentant les mêmes polarité et encombrement stérique. Pour *Ly-eIF(iso)4E*, le seul site prédit pour avoir un  $\omega > 1$  correspond à la position 77V ( $P_{(\omega>1)} = 0,997$ ). Ce site correspond à la substitution V77L présente chez 19 accessions d'espèces différentes. Ces résultats montrent que seulement 2 sites sont détectés sous sélection positive chez *Ly-eIF4E1* et *Ly-eIF(iso)4E* et donc que la plupart des acides aminés des protéines *Ly-eIF4E* sont contraints.

### III.4.6. Conclusions

Le gène *Ly-eIF4E1* ne présente qu'un seul site sous sélection positive ( $\omega \approx 5$ ), contrairement au gène *Ca-eIF4E1* chez le piment où 5 sites avec un  $\omega$  supérieur à 9 ont été identifiés. Le site sous sélection positive chez *Ly-eIF4E1* n'a pas de "lien" évident avec la résistance aux potyvirus même si le phénotype de résistance des 4 accessions de *L. pimpinellifolium* possédant cette substitution reste à déterminer. Le gène *Ly-eIF4E2* présente le

même profil évolutif que le gène *Ca-eIF4E2*, avec l'absence de sites sous sélection positive et le profil de polymorphisme du gène *Ly-eIF(iso)4E* est très similaire au profil de diversité du gène *Ca-eIF(iso)4E*. L'ensemble des résultats obtenus indiquent que les gènes *Ly-eIF4E* chez la tomate ne sont pas soumis à une pression de sélection positive exercée par les potyvirus. L'absence de pression évolutive exercée par ces virus sur les gènes *Ly-eIF4E* signifie que les mutations pour la résistance ne sont pas sélectionnées, qu'elles n'apportent pas d'avantage sélectif significatif. Cette observation, couplée au fait que la fréquence des résistances identifiées soit beaucoup plus faible chez la tomate que chez le piment, indiquent que la pression exercée par les potyvirus n'est pas similaire chez ces 2 Solanacées. Ce point sera abordé dans la discussion générale.

### **III.5. Polymorphisme des gènes *eIF4E* chez *A. thaliana***

#### **III.5.1. Contexte**

Malgré les nombreuses tentatives de clonage des différents membres et/ou la recherche d'EST de la famille multigénique eIF4E chez plusieurs espèces, *A. thaliana* est la seule dicotylédone pour laquelle nous disposons, grâce au séquençage complet de son génome, des séquences de l'ensemble des gènes codant pour des protéines de liaison à la coiffe : les 3 gènes codant pour des protéines eIF4E (*At-eIF4E1*, *At-eIF4E2* et *At-eIF4E3*), le gène codant pour l'isoforme eIF(iso)4E spécifique des plantes supérieures, et enfin le gène codant pour une autre protéine de liaison à la coiffe, la nCBP (new Cap Binding Protein). Contrairement aux études décrites précédemment chez les Solanacées, le profil de polymorphisme de tous les membres de la famille multigénique eIF4E pourra donc être caractérisé.

Le TuMV est le seul potyvirus inféodé à la famille des Brassicacées dont fait partie *A. thaliana*. D'autres potyvirus inféodés à d'autres familles végétales sont pourtant capables d'infecter la plante modèle en conditions de laboratoire. Il s'agit du CIYVV, du TEV, du PPV, du LMV et du WMV. Curieusement, et contrairement aux Solanacées, aucune résistance récessive aux virus identifiée dans la variabilité naturelle d'*A. thaliana* n'apparaît correspondre au facteur eIF4E (Decroocq *et al.*, 2006; Revers *et al.*, 2003). La caractérisation du

polymorphisme de ces gènes dans la variabilité naturelle décrite dans ce chapitre contribuera donc à confirmer, ou pas, cette observation.

Même si au niveau international les travaux sur les gènes du complexe d'initiation de la traduction portent principalement sur des aspects fonctionnels, notamment *via* la caractérisation de mutants d'*A. thaliana*, peu de choses sont encore connues sur le rôle respectif de chacune des protéines eIF4E dans la traduction des ARNm ou dans d'autres fonctions cellulaires clés. Ainsi, en plus de l'exploration de la diversité d'*A. thaliana* pour les gènes *eIF4E* et de la recherche d'allèles naturels de résistance aux potyvirus, cette étude a pour objectif de contribuer à la caractérisation du rôle de la redondance de ces gènes, à la fois pour les fonctions cellulaires de la plante hôte et pour l'interaction potentielle avec les potyvirus.

### III.5.2. Polymorphisme intraspécifique des gènes *At-eIF4E*

Les ADNg des gènes *At-eIF4E1*, *At-eIF4E3*, *At-eIF(iso)4E* et *At-nCBP* ont été amplifiés chez les 48 accessions de la core collection d'*A. thaliana* de McKhann (2004) et chez les accessions Col-0 et WS. En revanche, l'ADNg de *At-eIF4E2* n'a pu être amplifié que chez 33 des accessions de la core collection, plus Col-0 et WS. Toutes les données présentées ci-après présentent donc les résultats de polymorphisme obtenus sur un échantillonnage réduit à 35 accessions au lieu des 50 initialement prévues.

La longueur des ADNg séquencés pour les 5 *At-eIF4E* varie de 1420 à 1645 nucléotides (Tableau III.13). Les structures exons/introns sont conservées chez l'ensemble des accessions : 5 exons et 4 introns pour *At-eIF4E1*, *At-eIF4E2*, *At-eIF4E3* et *At-eIF(iso)4E* et 6 exons et 5 introns pour *At-nCBP*.

Très peu d'indels et de sites nucléotidiques polymorphes ont été détectés et les SNP se trouvent majoritairement dans les introns (Tableau III.13). La diversité nucléotidique moyenne  $\pi$  varie, en fonction des gènes, de 0,00047 à 0,02357;  $\sigma = \pm 0,00101$  dans les exons et de 0,00148 à 0,07095;  $\sigma = \pm 0,00299$  dans les introns. Les analyses par fenêtres glissantes montrent clairement que la diversité nucléotidique, pour tous les gènes *At-eIF4E*, est faible bien que plus élevée dans les introns que dans les exons (Figure III.10).

Comme précédemment et afin de déterminer si la diversité nucléotidique observée est représentative de la diversité globale du génome d'*A. thaliana*, ces données de polymorphisme ont été comparées aux valeurs obtenues pour les 10 locus indépendants (McKhann *et al.*, 2004). La diversité nucléotidique  $\pi$  moyenne de ces locus est de 0,0047 (de 0,0010 à 0,0106;  $\sigma = \pm 0,0038$ ) (Tableau III.13). Ainsi, à l'exception de *At-eIF4E2*, tous les gènes *At-eIF4E* présentent une diversité nucléotidique relativement faible mais qui ne se différencie pas de la diversité observée aux 10 locus indépendants.

### **III.5.2.1. Pression de sélection purificatrice sur les gènes *At-eIF4E1*, *At-eIF4E3*, *At-eIF(iso)4E* et *At-nCBP***

Les valeurs de  $\pi$  inférieures aux valeurs de  $\theta_w$  dans le cas des gènes *At-eIF4E1*, *At-eIF4E3*, *At-eIF(iso)4E* et *At-nCBP* (Tableau III.13) indiquent un manque de polymorphismes équilibrés, c'est-à-dire un excès de variants alléliques rares. Ces résultats concordent avec la distribution des fréquences alléliques estimée chez 90 accessions d'*A. thaliana* par Nordborg *et al.* (2005) (Figure III.11.A). Le calcul de l'indice D de Tajima par DnaSP permet en théorie de tester l'hypothèse nulle de neutralité sélective ( $D=0$ ). Les travaux de Nordborg *et al.* (2005) décrivant la variabilité du génome d'*A. thaliana* ont montré que la neutralité chez cette espèce ne correspondait pas à  $D=0$  mais qu'elle était déviée négativement (Figure III.11.B). D'après le test de sélection standard effectué sous DnaSP, on ne peut donc rien conclure des valeurs de D négatives obtenues pour les gènes *At-eIF4E1*, *At-eIF4E3*, *At-eIF(iso)4E* et *At-nCBP* (Tableau III.13). Les valeurs de Nordborg *et al.* (2005) n'ont pas été calculées à partir du même échantillon d'accessions d'*A. thaliana* que celui utilisé au cours de notre étude, rendant impossible la comparaison de nos données. L'ensemble de ces résultats montre tout de même que le polymorphisme des gènes *At-eIF4E1*, *At-eIF4E3*, *At-eIF(iso)4E* et *At-nCBP* correspond à la diversité globale observée chez *A. thaliana*. Les excès d'allèles rares en fréquence extrême et de singletons seraient liés à l'expansion démographique récente de la population d'*A. thaliana* (Nordborg *et al.*, 2005).

Pour les gènes *At-eIF4E3*, *At-eIF(iso)4E* et *At-nCBP*, un excès significatif de polymorphismes synonymes (par site synonyme) par rapport aux polymorphismes nonsynonymes (par site nonsynonyme) ( $\pi_a / \pi_s < 1$ , Tableau III.13, Figure III.12) est

observé, indiquant que les quelques substitutions détectées dans les séquences nucléotidiques sont essentiellement synonymes. Ces valeurs indiquent que ces gènes sont soumis à une sélection purificatrice.

En revanche, le rapport entre le taux de substitutions nonsynonymes et le taux de substitutions synonymes est supérieur à 1 pour le gène *At-eIF4E1* (Tableau III.13, Figure III.12). Ce résultat est cependant modulé par des taux de polymorphismes extrêmement faibles : 2 substitution synonymes contre 3 non-synonymes, suggérant qu'une pression de sélection purificatrice s'exerce sur la protéine At-eIF4E1.

### **III.5.2.2. Pression de sélection balancée sur le gène At-eIF4E2**

Le gène *At-eIF4E2* se distingue des autres membres de la famille multigénique eIF4E. Son niveau de polymorphisme nucléotidique est plus élevé, aussi bien dans les exons que dans les introns. Les valeurs de  $\pi$  inférieures aux valeurs de  $\theta_w$  ainsi qu'une valeur de D de Tajima très positive indiquent un excès de variants en fréquence équilibrée (Tableau III.13). L'excès significatif de polymorphismes synonymes par rapport aux polymorphismes non-synonymes indique que la protéine At-eIF4E2 est contrainte (Tableau III.13, Figure III.12).

Plusieurs hypothèses sont avancées pour expliquer cet excès de polymorphismes synonymes :

(i) *At-eIF4E2* correspond à un pseudogène peu exprimé ou "non fonctionnel" dans les cellules de la plante en comparaison avec les autres gènes *At-eIF4E*. Cette hypothèse est rejetée par la valeur de  $\pi_a / \pi_s$  inférieure à 1 qui indique que la protéine est contrainte par la sélection, donc probablement importante.

(ii) ce gène est un point chaud de recombinaison caractérisé par un taux de mutation local élevé. La valeur de D de Tajima positive et le taux de GC<sub>3</sub> (pourcentage de nucléotides G et C à la troisième position des codons) semblable aux autres *At-eIF4E* vont à l'encontre de cette hypothèse.

(iii) de la sélection balancée agit sur le gène *At-eIF4E2* comme suggéré par la valeur de D de Tajima très positive. La sélection balancée peut en effet maintenir un excès d'allèles à des fréquences intermédiaires sur de très longues périodes de temps. Chez *A. thaliana*, les analyses de génétique des populations montrent que la sélection balancée est la force

prédominante qui agit pour maintenir à la fois des allèles de résistance et de sensibilité aux agents pathogènes. Ce maintien des polymorphismes observé serait dû à de la sélection fréquence dépendante : dans le cas des interactions hôtes-agents pathogènes, la fréquence de l'allèle de résistance dépend de la fréquence de l'allèle d'avirulence, donc de la fréquence de certains génotypes dans chacune des 2 populations (Ellis *et al.*, 1999; Stahl *et al.*, 1999; Tian *et al.*, 2003). Dans le cas du gène *At-eIF4E2*, la nécessité du maintien de ce grand nombre d'allèles en fréquence équilibrée demeure inconnue, car rien à ce jour n'est connu sur la spécialisation fonctionnelle de ce facteur chez les plantes ou sur son implication potentielle dans la résistance aux virus.

### **III.5.3. Recherche de sites sous sélection positive**

Pour chaque *At-eIF4E*, les séquences codantes de toutes les accessions d'*Arabidopsis* (*A. thaliana* et *A. lyrata*) ont été utilisées, représentant de 206 à 241 codons. L'estimation des modèles d'hétérogénéité entre codons, et les tests permettant de les comparer sont indiqués au tableau III.14. Les valeurs de LRT obtenues montrent que le modèle M8, qui prend en compte la sélection positive, est moins approprié que le modèle M7. Le résultat des modèles 0 et 1 d'hétérogénéité entre branches de l'arbre montrent que le modèle le plus simple (M0), où toutes les branches évoluent selon un  $\omega$  identique, semble le plus adapté aux données pour les gènes *At-eIF4E* au vu des LRT obtenus. Cette analyse met en évidence une absence de sites sous sélection positive.

### **III.5.4. Substitutions en acides aminés identifiées dans les 5 protéines At-eIF4E**

Les substitutions en acides aminés détectées dans les protéines At-eIF4E1 (A40V et N211K) ne correspondent ni aux acides aminés essentiels pour la stabilité des protéines eIF4E ou pour la reconnaissance de la coiffe des ARNm, ni aux sites montrés comme étant impliqués dans la résistance aux virus à ARN chez les plantes (Figure III.13).

Les mutations dans les protéines At-eIF4E3, At-eIF(iso)4E et At-nCBP se trouvent à l'extrémité N-terminale, partie connue pour être la plus variable des protéines eIF4E, toutes espèces confondues (Figure III.13; Joshi *et al.*, 2005). Ces résultats suggèrent que les substitutions en acides aminés détectées n'ont aucun impact sur la fonction de traduction des ARNm ou une implication quelconque dans la résistance aux potyvirus.

Seize substitutions en acides aminés ont été détectées chez At-eIF4E2. Trois d'entre elles touchent des acides aminés importants pour la stabilité de la structure de la protéine (F75, D170 et I175). Le fait que des mutations soient retrouvées au niveau des acides aminés essentiels pour la fonction de traduction des ARNm permet d'émettre l'hypothèse que At-eIF4E2 pourrait être impliquée dans d'autres fonctions cellulaires chez les plantes. Par alignement avec la séquence de At-eIF4E1, on peut voir que 3 de ces substitutions se trouvent dans les régions I et II identifiées chez la forme eIF4E1 impliquées dans la résistance aux potyvirus chez plusieurs espèces végétales (F75, I86 et S120; Figure III.13). Ces résultats indiquent que la protéine At-eIF4E2 pourrait potentiellement être impliquée dans la résistance aux virus.

### **III.5.5. Recherche de combinaisons d'allèles *At-eIF4E* de résistance aux potyvirus**

Dans l'objectif de déterminer si la combinaison de substitutions en acides aminés dans les différentes protéines At-eIF4E pourrait avoir un rôle dans la résistance aux potyvirus, les 35 accessions d'*A. thaliana* ont été regroupées en 13 "polyhaplotypes protéiques". Ces 13 variants alléliques ont été constitués à partir des différentes combinaisons de substitutions en acides aminés détectées pour les 5 protéines At-eIF4E (Tableau III.15). Les accessions appartenant à un polyhaplotype donné codent donc pour les mêmes 5 protéines de liaison à la coiffe. Il est à noter que 7 polyhaplotypes ne sont constitués que d'une seule accession d'*A. thaliana*. Deux accessions représentatives de chacun des 13 polyhaplotypes (quand c'est possible car 7 polyhaplotypes sont constitués d'une seule accession) ont été testées pour leur résistance à 3 potyvirus : CIYVV, TuMV et WMV. Les résultats obtenus par DAS-ELISA et/ou RT-PCR montrent qu'aucune accession ne montre une résistance de haut niveau au

TuMV et au CIYVV. En revanche, 3 accessions (190AV, 21AV et 166AV) sont totalement résistantes au WMV (Figure III.14).

L'accession 190AV appartient au même polyhaplotype que 236AV, qui n'est résistant à aucun des 3 potyvirus testés (Tableau III.15; Figure III.14). Dans ce cas, nous pouvons donc conclure que les combinaisons de substitutions en acides aminés détectées dans les protéines At-eIF4E de 190AV ne sont pas impliquées dans le phénotype de résistance au WMV.

La séquence en acides aminés de At-eIF4E2 de l'accession 21AV (qui constitue à elle seule un polyhaplotype) est 100 % identique à la séquence protéique At-eIF4E2 de l'accession 190AV (Tableau III.15). D'après les conclusions ci-dessus sur l'absence de corrélation entre les substitutions dans les protéines At-eIF4E2 et le phénotype de résistance au WMV de 190AV, on peut affirmer que les mutations dans la protéine At-eIF4E2 ne sont pas responsables du phénotype de résistance au WMV de l'accession 21AV. Cette accession est la seule présentant des substitutions en acides aminés au niveau de la protéine At-eIF4E1, mais ces mutations ne se localisent pas dans les régions impliquées dans la résistance. Il est donc fortement probable que la résistance observée ne soit pas liée au polymorphisme de séquence observée chez At-eIF4E1.

L'accession 166AV, également unique accession représentative de son polyhaplotype, code pour la même protéine At-eIF4E2 que les accessions 200AV et 91AV, qui ne sont pas résistantes au WMV (Tableau III.15; Figure III.14). En revanche, la substitution P49Q dans la séquence en acides aminés de At-eIF4E3 n'est présente que chez 166AV, rendant cette position possiblement impliquée dans la résistance. L'accession 166AV correspond à Cvi-0 qui a été montrée comme étant l'accession la plus divergente par rapport aux autres accessions d'*A. thaliana* (Schmid *et al.*, 2003). Cela pourrait expliquer pourquoi elle possède ce polyhaplotype unique différent de toutes les autres accessions d'*A. thaliana*. Le polymorphisme observé correspondrait donc plutôt à des signatures de type "espèce-spécifique" qu'à des substitutions liées à la résistance aux potyvirus.

Ces analyses ne permettent donc pas de relier clairement la variation allélique observée au sein de la famille multigénique At-eIF4E avec une résistance à l'un des 3 potyvirus testés.

### III.5.6. Phylogénie et polymorphisme interspécifique des gènes *At-eIF4E*

Les 35 accessions d'*A. thaliana* ont été utilisées pour caractériser le polymorphisme intraspécifique et *A. lyrata ssp lyrata* comme groupe externe afin d'estimer la divergence interspécifique des gènes *eIF4E*. La théorie neutraliste moléculaire prédit que les locus les plus variables au sein d'une espèce sont ceux qui montrent le plus de variation entre espèces (la diversité est générée par la mutation uniquement et le taux de mutation neutre est constant au cours du temps). Les résultats du test HKA, utilisé pour comparer les niveaux de variation intra- et inter-spécifique, sont indiqués au tableau III.16. Le modèle neutraliste s'ajuste très étroitement à nos données : les locus *At-eIF4E1*, *At-eIF4E3*, *At-eIF(iso)4E* et *At-nCBP*, peu variables au sein des *A. thaliana*, sont également peu divergents entre *A. thaliana* et *A. lyrata* et le locus *At-eIF4E2* est à la fois le plus variable au sein de l'espèce *A. thaliana* et entre les espèces *A. thaliana* et *A. lyrata*. La comparaison des arbres phylogénétiques basés sur la séquence codante des gènes *At-eIF4E* et ceux basés sur les ADNc *eIF4E* amplifiés chez *A. lyrata*, construits par la méthode de Neighbour-joining, montrent que toutes les accessions *A. thaliana* se regroupent dans un même phylum bien distinct de l'espèce *A. lyrata*.

Les arbres phylogénétiques basés sur la séquence codante des 5 gènes *At-eIF4E* chez les 35 accessions d'*A. thaliana* et sur la séquence codante des locus indépendants non-*eIF4E*, construits par la méthode de Neighbour-joining, ont été comparés (Figure III.15). Les phylogénies obtenues à partir des données *At-eIF4E* et les locus non-*eIF4E* sont très différentes puisqu'aucun groupe d'accessions commun entre les 2 arbres ne peut être identifié. L'incompatibilité peut être due à différents facteurs, tels que des événements "d'allele sorting" qui se caractérisent par des polymorphismes partagés entre accessions très proches ou des flux de gènes. Les représentations radiales permettent de vérifier la cohérence de chaque phylogénie indépendamment. Une phylogénie logique apparaît sous la forme d'une étoile : un point central correspond à l'ancêtre commun à partir duquel divergent les espèces filles. Cette représentation est respectée pour les gènes *At-eIF4E* et les locus non-*eIF4E* (Figure III.15).

L'incohérence des phylogénies intra-spécifiques peut s'expliquer par le très faible taux de polymorphisme détecté entre les différentes accessions d'*A. thaliana*. Sur la base des séquences *At-eIF4E*, il apparaît clairement que toutes les accessions *A. thaliana* utilisées dans cette étude se regroupent dans un phylum bien distinct de l'espèce *A. lyrata*.

### III.5.7. Conclusions

Ce travail a permis la caractérisation des profils d'évolution moléculaire de tous les membres de la famille multigénique eIF4E chez *A. thaliana*. L'analyse du polymorphisme des 5 gènes *At-eIF4E* chez les 35 accessions issues d'une core collection représentant le maximum de la diversité existante chez cette espèce a montré que les gènes *At-eIF4E1*, *At-eIF4E3*, *At-eIF(iso)4E* et *At-nCBP* étaient sous pression de sélection purificatrice (niveau de polymorphisme très réduit et substitutions nonsynonymes très inférieures aux substitutions synonymes). En revanche, le gène *At-eIF4E2* présente des taux de polymorphismes plus élevés et une valeur de D de Tajima très positive, suggérant qu'une pression de sélection balancée s'exercerait sur *At-eIF4E2* (excès de polymorphismes en fréquence équilibrée). Ces résultats indiquent donc qu'au moins 4 des 5 gènes *At-eIF4E* sont très conservés, la sélection négative empêchant les mutations de se fixer, appuyant l'hypothèse que chacun des membres de cette famille multigénique serait maintenu par la sélection. Pour *At-eIF4E2*, l'apparition de mutations touchant les acides aminés impliqués dans la stabilité de la structure des protéines eIF4E suggère que ce gène se spécialiserait / interviendrait dans d'autres processus que la traduction des ARNm. Il est à noter que *At-eIF4E2* et *At-eIF4E3*, issus d'une duplication récente et séparés de seulement 7 Kb, présentent des profils d'évolution très différents. L'acquisition d'une spécialisation fonctionnelle pour *At-eIF4E2* semble donc postérieure à l'évènement de duplication. Dans ce cas, l'évolution au sein de la famille multigénique *At-eIF4E* aurait donc favorisé la différenciation entre les différentes copies eIF4E.

La recherche d'allèles naturels de résistance aux potyvirus correspondant à *At-eIF4E* s'est avérée infructueuse. En effet, les quelques substitutions en acides aminés détectées ne semblent pas être impliquées dans la résistance d'*A. thaliana* au WMV. Ces résultats confirment donc que les résistances récessives chez *A. thaliana* ne sont pas liées à du polymorphisme au niveau des gènes *At-eIF4E*, même si le cas de l'accession 166AV reste à élucider.

## **Chapitre IV**

# **V**ers la caractérisation du rôle fonctionnel des différents facteurs eIF4E

## **IV.1. Introduction**

Toutes les protéines eIF4E chez *A. thaliana* (At-eIF4E1, les protéines At-eIF4E2 et At-eIF4E3 codées par des gènes en tandem séparés de 7 Kb, At-eIF(iso)4E et la nCBP), chez la tomate (Ly-eIF4E1, Ly-eIF4E2 et Ly-eIF(iso)4E) et chez le piment (Ca-eIF4E1, Ca-eIF4E2 et Ca-eIF(iso)4E) apparaissent très contraintes, sous forte sélection purificatrice. Le maintien au cours de l'évolution de l'ensemble de ces paralogues au sein de chaque famille végétale suggère que chacun aurait acquis une spécialisation fonctionnelle dans les cellules de la plante.

Chez les mammifères, eIF4E est un oncogène surexprimé dans de nombreux cancers (Clemens *et al.*, 1999); de nombreuses études se sont donc focalisées sur ce facteur. En plus de son rôle dans l'initiation de la traduction pour le recrutement des ARNm, plusieurs travaux ont montré qu'une fraction significative de protéines eIF4E (environ 68 %) se trouvait dans le noyau afin de réguler l'export d'ARNm spécifiques du noyau vers le cytoplasme (Culjkovic *et al.*, 2005). eIF4E pourrait également jouer un rôle dans la séquestration des ARNm et la stabilité des corps P cytoplasmiques (Ferraiuolo *et al.*, 2005). Une récente étude suggère que eIF4E gouvernerait la progression du cycle cellulaire et la prolifération cellulaire en coordonnant l'expression de plusieurs gènes au niveau post-transcriptionnel (Culjkovic *et al.*, 2007). eIF4E apparaît donc être un facteur pléiotropique impliqué dans des processus cellulaires clés chez les mammifères.

Les études fonctionnelles concernant ce facteur chez les plantes se sont restreintes aux gènes et aux protéines eIF4E1 et eIF(iso)4E (du blé principalement). eIF(iso)4E est une protéine spécifique des plantes supérieures. eIF4E1 et eIF(iso)4E se fixent préférentiellement aux protéines eIF4G et eIF(iso)4G respectivement pour former les complexes eIF4F et eIF(iso)4F impliqués dans l'initiation de la traduction des ARNm (Browning *et al.*, 1987; Browning *et al.*, 1992; Lax *et al.*, 1985). La conservation de ces 2 complexes chez les plantes suggère l'acquisition d'une spécialisation fonctionnelle pour chacun d'entre eux. En effet, chez le blé, eIF4F permet préférentiellement la traduction d'ARNm présentant des structures secondaires en 5' et des ARNm non coiffés et/ou dicistroniques alors que eIF(iso)4F traduirait préférentiellement des ARNm avec des structures linéaires et des coiffes hyperméthylées (Gallie and Browning, 2001). Des expériences *in vitro* et *in vivo* chez le maïs montrent que

eIF(iso)4F se localise au niveau des microtubules. L'association de la protéine eIF(iso)4E aux microtubules pourrait simultanément permettre de relier les composants de la machinerie de la traduction au cytosquelette et de réguler la disposition des microtubules dans les cellules de la plante (Bokros *et al.*, 1995). Les gènes *eIF4E2* et *eIF4E3* ont été identifiés *via* le séquençage complet du génome d'*A. thaliana*. Ils ne sont séparés que de 7 kb sur le chromosome 1 et seraient donc issus d'une duplication récente. Aucune information concernant leur expression transcriptionnelle et traductionnelle n'a été publiée à ce jour; leur rôle biologique dans les

□□□□□□□□□□ú□□□□□□□□□□ú□□□□□□□□□□ú□□□□□□□□□□  
□ú□□□□□□□□□□ú□□□□□□□□□□ú□□□□□□□□□□ú□te famille multigénique car elle est capable d'interagir *in vitro* avec la coiffe des ARNm et eIF(iso)4G pour initier la traduction d'ARNm coiffés (Ruud *et al.*, 1998). En revanche, comme pour eIF4E2 et eIF4E3, son rôle *in vivo* dans la traduction n'a jamais été démontré. Ruud (1998) suggère que la nCBP pourrait avoir un rôle de séquestration des ARNm afin de ralentir ou d'empêcher leur traduction. Le rôle des différentes formes eIF4E chez les plantes reste encore largement incompris mais le maintien de ces protéines homologues suggère fortement qu'il existe une fonction "forme-spécifique" (Duprat *et al.*, 2002; Rodriguez *et al.*, 1998). L'acquisition d'une spécialisation fonctionnelle pour chaque forme eIF4E pourrait permettre, comme chez les mammifères, un contrôle différentiel de la traduction, de l'export nucléaire ou de la séquestration des ARNm. Elle correspondrait à l'acquisition d'un niveau supplémentaire de régulation de l'expression des gènes, avantageux pour des organismes dont le développement est complexe.

Pour aller plus loin dans la compréhension de la fonction de chacun de ces facteurs, une étude moléculaire préliminaire a été initiée visant à caractériser les profils d'accumulation des 5 transcrits *eIF4E* chez *A. thaliana*. Cette première étape dans la compréhension de la régulation des membres de la famille eIF4E a pour objectif de déterminer si les 5 gènes *eIF4E* ont une expression ubiquitaire ou plutôt tissu-spécifique chez *A. thaliana*, et s'il existe des différences d'accumulation entre les différents ARNm. Dans cette étude, les niveaux d'accumulation des transcrits *eIF4E* de l'accession Col-0 d'*A. thaliana* sont donc comparés entre eux et entre plusieurs organes végétatifs (feuilles jeunes ou âgées, racines et tiges) et reproducteurs (fleurs) par RT-PCR semi-quantitative puis par RT-PCR quantitative en temps réel pour affiner les résultats obtenus par la première méthode.

En complément de cette analyse transcriptionnelle, la capacité des différentes protéines eIF4E du piment et d'*A. thaliana* à restaurer la croissance de levures délétées de leur

gène *eIF4E* et donc leur capacité à initier la traduction des ARNm, rôle "premier" de ces facteurs dans les cellules, a été testée par des expériences *in vivo* de complémentation fonctionnelle dans une levure délétée de son gène *eIF4E*. Jusqu'à présent, cette capacité n'avait été vérifiée chez les plantes que pour les protéines eIF4E1 et eIF(iso)4E d'*A. thaliana* (Rodriguez *et al.*, 1998) et avait permis de montrer que 5 des 8 formes eIF4E identifiées chez la drosophile étaient capables d'initier la traduction suggérant une redondance fonctionnelle de ces différentes copies *in vivo* (Hernández *et al.*, 2005).

## **IV. 2. Matériel et méthodes**

### **IV.2.1. RT-PCR semi-quantitative**

Une transcription inverse (RT) est réalisée à partir d'ARN de feuilles en croissance, de feuilles vieilles, de tiges et de racines de l'accession Columbia Col-0 d'*A. thaliana*. La qualité de ces ARN a été vérifiée sur gel dénaturant et leur concentration ajustée. Les conditions de RT et de PCR sont décrites dans l'annexe 7. Seul le nombre de cycles d'amplification varie : 20, 25, 30, 35 et 40. Les séquences des amorces utilisées sont données dans le tableau IV.1. Le témoin PCR est un gène codant pour le facteur d'élongation eF1 $\alpha$ , qui s'exprime dans tous les tissus de la plante et qui montre une bonne stabilité d'expression (Nicot *et al.*, 2005). Toutes les PCR ont été réalisées dans le même thermocycleur (Eppendorf, Hamburg, Germany). Les produits PCR obtenus sont visualisés sur gel d'agarose 1% dans du tampon TAE 1X (Tris 40 mM; Acétate de sodium anhydre 8 mM; EDTA 1 mM) sous ultraviolets après un bain de bromure d'éthidium (BET).

### **IV.2.2. RT-PCR quantitative**

La technique de RT-PCR quantitative permet la quantification relative du niveau d'expression d'un gène cible par rapport à celui d'un gène de référence. Le principe de cette technique est basé sur la détection de l'amplifiat néo-synthétisé après chaque cycle de PCR. Les témoins utilisés dans toutes les expériences sont les gènes codant pour la protéine histone

1 et pour la GAPDHA (Glyceraldéhyde-3-Phosphate Déshydrogénase A). Ces gènes de ménage montrent une bonne stabilité d'expression dans tous les organes de la plante.

Une transcription inverse (RT) est réalisée à partir d'ARN de feuilles en croissance, de feuilles vieilles, de tiges, de racines et de fleurs de l'accession Columbia Col-0 d'*A. thaliana*. La qualité et la quantité des ARN ont été estimées sur gel dénaturant et au spectrophotomètre à 260 et 280 nm. Les ARN sont traités à la RQ1 RNase-free DNase (Promega, Madison, WI, USA) selon le protocole du fournisseur pour éliminer l'ADN résiduel après extraction. Les concentrations en ARN sont ajustées après traitement à 500 ng/μL. L'étape de transcription inverse suit le protocole décrit dans l'annexe 7. Une gamme de dilution à partir d'un mélange équimolaire de l'ensemble des ADNc est alors réalisée (25 μL de chaque ADNc, dilutions aux 1/2, 1/10, 1/20 et 1/100). Les PCR quantitatives ont été réalisées sur l'appareil qPCR Mx3005 (Stratagene, Amsterdam, The Netherlands) avec le kit SYBR® Premix Ex Taq™ Perfect Real Time (Takara, Otsu, Shiga, Japan). Les conditions de PCR sont issues du protocole adapté aux appareils Mx3000P du fournisseur Takara®. Pour une réaction de PCR avec un volume final de 20 μL : 10 μL de SYBR® Premix Ex Taq™ (2X); 0,4 μL de chacune des amorces (10 μM) et 2 μL d'ADNc sont mis en mélange. Chaque plaque PCR 96 puits contient les gènes témoins, la gamme de dilution en duplicat et les échantillons d'ADNc à amplifier en triplicat. Les amorces utilisées sont répertoriées dans le tableau IV.1. Après 30 s à 95°C pour activer l'enzyme, l'amplification est effectuée pendant 40 cycles successifs comprenant les étapes de dénaturation (10 s à 95°C), d'hybridation (20 s, de 50°C à 57°C en fonction des gènes amplifiés) et d'élongation (20 s à 72°C). La fluorescence est détectée après chaque cycle. Suite à la réaction d'amplification, une montée en température, de 55 à 95°C, à raison d'une augmentation de 1°C par 30 s, est réalisée afin de tracer la courbe de fusion qui permet de vérifier l'unicité du produit amplifié. Les résultats sont analysés avec les logiciels MxPro (Stratagene, Amsterdam, The Netherlands) et Rest© de (Pfaffl *et al.*, 2002). Ces expériences ont été réalisées indépendamment sur deux réplicats biologiques distincts. Les valeurs brutes de Ct (nombre de cycles PCR nécessaires pour qu'un signal significatif de fluorescence associé à une amplification soit détecté dans l'échantillon) des gènes témoins par rapport aux gènes d'intérêt *eIF4E* seront comparées pour quantifier le niveau relatif d'expression des gènes *eIF4E*.

### **IV.2.3. Test de complémentation fonctionnelle dans la levure**

Le protocole est le même que celui décrit dans l'article 1. L'ADNc codant pour les 5 gènes *eIF4E* d'*A. thaliana* Col-0 et les 3 gènes *eIF4E* de *C. annuum* Yolo Wonder a été cloné dans le vecteur glucose-dépendant p424GBP/TRP1 (cf annexe 8). Les levures JO55 (cf annexe 5) sont ensuite transformées par choc thermique (15 min à 42°C). Après transformation, les levures sont étalées sur milieu solide sélectif contenant 2 % de glucose. Les boîtes sont incubées 10 jours à 30°C. Si les levures poussent, on pourra en conclure que la protéine eIF4E exogène introduite complète et est donc fonctionnelle.

## **IV. 3. Résultats**

### **IV.3.1. Etude de l'accumulation des transcrits *eIF4E* chez *A. thaliana* par RT-PCR semi-quantitative**

Les analyses d'accumulation des transcrits correspondant aux gènes *At-eIF4E1*, *At-eIF4E2*, *At-eIF4E3*, *At-eIF(iso)4E* et *At-nCBP* ont été effectuées par RT-PCR semi-quantitative en utilisant comme matrice des ARN totaux issus de feuilles jeunes, feuilles vieilles, racines, tiges et de fleurs. Le témoin positif utilisé dans ces expériences est le gène codant pour le facteur eF1 $\alpha$  exprimé constitutivement dans la plante : les ADNc de *eF1 $\alpha$*  sont détectés à partir du 20<sup>ème</sup> cycle de PCR dans les feuilles jeunes, les racines, les tiges et les fleurs et à partir du 25<sup>ème</sup> cycle dans les feuilles vieilles (Figure IV.1). Tous les transcrits *At-eIF4E* sont présents dans l'ensemble des organes : (i) dans les feuilles jeunes, les racines, les tiges et les fleurs, *At-eIF4E1*, *At-eIF4E2* et *At-eIF4E3* sont détectés à partir du 30<sup>ème</sup> cycle de PCR alors que dans les feuilles vieilles, ils ne sont visibles qu'à partir du 35<sup>ème</sup> cycle, (ii) *At-eIF(iso)4E* et *At-nCBP* sont amplifiés dès le 25<sup>ème</sup> cycle dans les feuilles jeunes et seulement à partir du 30<sup>ème</sup> cycle dans tous les autres organes.

Parmi les organes testés, il semble donc que les transcrits *At-eIF4E* s'accumulent majoritairement dans les feuilles jeunes et minoritairement dans les feuilles vieilles. Si on compare visuellement les niveaux d'accumulation des transcrits *At-eIF4E* entre eux, *At-*

*eIF(iso)4E* et *At-nCBP* présentent des niveaux d'accumulation plus élevés que *At-eIF4E1*, *At-eIF4E2* et *At-eIF4E3*.

### **IV.3.2. Etude de l'accumulation des transcrits *eIF4E* chez *A. thaliana* par RT-PCR quantitative**

#### ***IV.3.2.1. Gènes témoins et contrôles de PCR***

Afin de mieux caractériser l'expression des gènes *At-eIF4E* en fonction des organes de la plante, une étude par RT-PCR quantitative en temps réel a été réalisée sur les mêmes organes que ceux décrits précédemment pour les expériences de RT-PCR semi-quantitative. Afin d'homogénéiser les températures d'hybridation entre les différents gènes amplifiés, le gène témoin *eFl $\alpha$*  a été remplacé par 2 autres gènes, l'un codant pour la protéine histone1 et l'autre pour la GAPDHA. Les valeurs brutes des Ct (Cycle threshold) détectées pour ces 2 gènes chez les 5 organes pour l'ensemble des expériences réalisées sont représentées sur la figure IV.2. Ces données montrent une bonne stabilité d'expression des gènes *histone 1* et *GAPDHA* dans tous les organes de la plante au cours de 6 expériences successives sur 2 réplicats biologiques distincts. Avant toute analyse, l'efficacité et la répétabilité de nos résultats ont été vérifiées. Toutes les données obtenues au cours des expériences de RT-PCR quantitative sont exploitables car toutes les valeurs d'efficacité sont comprises entre 90 et 107 % et les coefficients de corrélation (indiquant si les points de la gamme de dilution, réalisée à partir d'un mélange équimolaire de tous les ADNc, sont bien linéaires) sont tous supérieurs à 0,985. Les courbes de fusion permettent de vérifier l'unicité du produit amplifié (Figure IV.3). Les résultats obtenus sont ensuite analysés avec le logiciel Rest (Pfaffl *et al.*, 2002). Un test de randomisation, qui prend en compte à la fois les valeurs de Ct brutes et les efficacités de PCR des gènes témoins et des gènes d'intérêt, permet de quantifier les différences de niveaux d'expression des différents transcrits.

#### ***IV.3.2.1. Comparaison des niveaux d'accumulation des transcrits eIF4E entre organes***

Les transcrits des 5 gènes *At-eIF4E* sont de 4 à 570 fois plus exprimés dans les racines que dans les autres organes testés (Figure IV.4). Les niveaux d'accumulation des transcrits *At-eIF4E* les plus faibles se trouvent dans les feuilles âgées. Dans les tiges, les feuilles en croissance et les fleurs, les transcrits *At-eIF4E* sont retrouvés en proportion équivalente. Les gènes *At-eIF4E* sont donc fortement exprimés dans les organes ayant une activité métabolique intense comme les racines, et beaucoup moins dans les tissus âgés. Ces analyses confirment donc les résultats préliminaires qui avaient été obtenus par RT-PCR semi-quantitative en ce qui concerne la plus faible expression des transcrits *At-eIF4E* dans les feuilles vieilles.

#### ***IV.3.2.2. Comparaison des niveaux d'accumulation des 5 transcrits eIF4E***

Comme toutes les PCR quantitatives réalisées à partir d'un même échantillon biologique ont des efficacités très similaires, il a été possible de comparer le niveau d'expression des 5 gènes *At-eIF4E* (Figure IV.5).

Dans tous les organes, les 2 transcrits majoritaires sont *At-eIF4E1* et *At-eIF(iso)4E* : *At-eIF4E1* est de 1,25 à 203 fois plus exprimé que les autres gènes *At-eIF4E* tout organe confondu; *At-eIF(iso)4E* n'est sous-exprimé que de 1,5 à 2 fois par rapport à *At-eIF4E1*. Ce facteur de différence d'expression très faible entre *At-eIF4E1* et *At-eIF(iso)4E* ne permet pas de conclure sur l'existence d'une réelle différence d'accumulation entre ces 2 transcrits.

Les transcrits minoritaires sont *At-eIF4E2* et *At-eIF4E3* : *At-eIF4E2* apparaît le moins exprimé, avec un niveau d'accumulation de 1,3 à 203 fois inférieur aux autres *At-eIF4E*, et ce dans tous les organes de la plante; *At-eIF4E3* a des niveaux d'accumulation identiques au transcrit *At-eIF4E2* dans les racines, les fleurs et les feuilles âgées et supérieurs d'un facteur 10 dans les feuilles jeunes et les tiges.

Le niveau d'accumulation du transcrit *At-nCBP* est intermédiaire, soit environ de 3 à 10 fois moins exprimé que *At-eIF4E1* dans tous les organes.

Contrairement aux résultats obtenus par RT-PCR semi-quantitative, *At-eIF4E1* apparaît tout autant voir plus exprimé que *At-eIF(iso)4E* et le transcrit *At-nCBP* semble moins exprimé que *At-eIF(iso)4E*. En revanche, les résultats de RT-PCR quantitative confirment que *At-eIF4E2* et *At-eIF4E3* représentent les ADNc les moins exprimés chez *A. thaliana*.

### IV.3.3. Complémentation fonctionnelle dans la levure JO55

Afin de tester la fonctionnalité *in vivo* des 5 protéines eIF4E chez *A. thaliana* et les 3 protéines eIF4E chez *Capsicum* spp., des expériences de complémentation fonctionnelle dans une levure délétée de son gène *eIF4E* (JO55) ont été réalisées.

Les transcrits *At-eIF4E1* et *Ca-eIF4E1* restaurent la croissance des levures JO55. En revanche, en ce qui concerne eIF(iso)4E, seule l'isoforme *At-eIF(iso)4E* est fonctionnelle *in vivo*. (Figure IV.6). Toutes les autres copies, *At-eIF4E2*, *At-eIF4E3*, *nCBP* et *Ca-eIF4E2* permettent la croissance des levures délétées de leur gène *eIF4E*, indiquant que tous ces paralogues sont capables d'initier la traduction des ARNm. Ces résultats appuient donc en faveur de l'existence d'une redondance fonctionnelle *in vivo* entre les différents facteurs eIF4E (Figure IV.6).

## IV.4. Discussion

Ces analyses visant à caractériser les profils d'expression des 5 ADNc *eIF4E* chez *A. thaliana* montrent que tous les membres de la famille multigénique sont transcrits, dans les organes feuilles jeunes, feuilles vieilles, racines, tiges et fleurs. Les profils d'expression de *eIF4E1* et *eIF(iso)4E* dans différents organes chez *A. thaliana* avaient été précédemment analysés par Rodriguez *et al.* (1998). Des expériences de Northern blot complétées par des hybridations *in situ* avaient montré une différence d'accumulation de ces transcrits entre différents tissus : les ARNm *eIF4E1* étaient exprimés dans tous les tissus à l'exception des cellules de la zone spécialisée des racines tandis que les ARNm *eIF4E(iso)4E* étaient abondants dans les organes floraux et les tissus en développement. Globalement, ces travaux montraient une association entre un haut niveau d'ARNm et une prolifération des cellules de la plante, suggérant des fonctions distinctes pour les deux isoformes, mais toujours reliées aux processus de division et prolifération cellulaires, comme cela avait été déjà décrit pour la levure et les cellules animales. Les profils d'expression des gènes *eIF4E* et *eIF(iso)4E* au niveau de différents organes de la laitue ont également été étudiés par hybridation *in situ* (Nicaise, 2006). Aucune différence de niveau d'expression ou de localisation des ARNm de chacune des isoformes d'*eIF4E* selon les organes n'a pu être mise en évidence. Les transcrits

*eIF4E1* et *eIF(iso)4E* ont également été détectés chez le piment et la tomate. Aussi bien les analyses par Northern blot que les expériences de RT-PCR semi-quantitative montraient qu'il n'y avait pas de différence d'expression des gènes *eIF4E* et *eIF(iso)4E* en fonction des organes (Charron, 2004).

Dans ce travail, l'analyse des niveaux d'accumulation des transcrits *eIF4E* et *eIF(iso)4E* entre les organes feuilles jeunes, feuilles âgées, racines, tiges et fleurs par PCR quantitative révèle que ces gènes sont tout autant transcrits dans les racines que dans les autres organes, contrairement à ce qui avait été montré par Rodriguez *et al.*, 1998 par *hybridation in situ*. Ces résultats se rapprochent donc plutôt des profils d'expression observés chez la laitue (Nicaise, 2006), le piment et la tomate (Charron, 2004). Le fait que tous les ARNm *eIF4E* soient plus abondants dans les tissus où se déroulent de nombreux processus métaboliques et en croissance plutôt que dans les feuilles âgées est en adéquation avec la fonction de traduction des protéines eIF4E.

Les transcrits *eIF4E2* et *eIF4E3* apparaissent moins abondants que les autres *eIF4E* dans tous les organes testés. Plusieurs hypothèses pourraient expliquer ce résultat :

- ils ont une fonction moins importante dans les cellules et sont donc transcrits en quantité moins élevée. Cette hypothèse soulève la question de la nécessité de leur conservation au cours de l'évolution. En effet, les gènes "inutiles" devraient être éliminés par la dérive ou la sélection.

- leur transcription n'est sur-exprimée que dans certains cas, en réponse à des stress abiotiques (affectant par exemple les voies de la croissance cellulaire où les eIF4E semblent avoir un rôle clé chez les mammifères) et/ou biotiques (notamment en réponse aux infections virales). Il a été montré que la suppression du facteur *eIF(iso)4E* chez *A. thaliana* résultait en une augmentation compensatrice du niveau de la protéine eIF4E1 (Duprat *et al.*, 2002). Il serait intéressant de regarder le profil d'expression des gènes *eIF4E2* et *eIF4E3* chez des plantes mutées pour les autres gènes de la famille eIF4E afin de mettre en évidence une possible augmentation du niveau d'accumulation des transcrits *eIF4E2* et/ou *eIF4E3*.

- le niveau d'expression de ces ARNm ne correspond pas au niveau d'expression des protéines correspondantes *via* l'existence de régulations post-transcriptionnelles.

Le transcrit *nCBP*, initialement identifié *via* l'analyse des séquences EST disponibles chez *A. thaliana* (Ruud *et al.*, 1998) est exprimé dans tous les organes. Cependant,

l'accumulation des ARNm du transcrit *nCBP* est plus faible que l'accumulation des transcrits *eIF4E1* et *eIF(iso)4E*. De façon similaire, de précédents travaux chez *A. thaliana* montraient que la quantité de protéines nCBP était 10 fois moins importante que la quantité des protéines *eIF4E1* et *eIF(iso)4E*, suggérant une plus faible accumulation de nCBP (Ruud *et al.*, 1998). De plus, la protéine nCBP a une meilleure affinité pour la fixation à la coiffe des ARNm que pour la fixation à la protéine *eIF(iso)4G*, suggérant une plus faible activité de nCBP dans la traduction que les autres formes *eIF4E*. Ainsi, ces résultats argumentent en faveur d'un rôle de la nCBP dans la séquestration des ARNm (Ruud *et al.*, 1998). Cette hypothèse est à nuancer puisque les résultats obtenus ici montrent que la protéine nCBP est capable d'initier la traduction dans la levure. Afin de mieux comprendre le rôle biologique de nCBP dans la traduction et/ou la régulation de l'expression des gènes, il serait intéressant d'identifier ses partenaires protéiques et préciser la localisation subcellulaire de cette protéine spécifique des eucaryotes supérieurs.

Les expériences de complémentation fonctionnelle dans la levure montrent que seul le gène *Ca-eIF(iso)4E* du piment ne restaure pas la croissance des levures, suggérant qu'il n'est pas capable d'initier la traduction des ARNm. Pourtant, au niveau de la séquence en acides aminés, la protéine *Ca-eIF(iso)4E* n'est pas plus différente de *eIF4E* de la levure que ne l'est *eIF(iso)4E* d'*A. thaliana* (Figure IV.7). Cependant, de précédentes expériences montraient que la croissance des levures délétées de leur gène *eIF4E* complémentées avec *At-eIF(iso)4E* était ralentie par rapport à la croissance des levures complémentées avec *At-eIF4E1* (Rodriguez *et al.*, 1998), ce qui n'a pas été observé dans nos expériences menées au cours de ma thèse. Je reviendrai sur ce point dans la discussion générale.

## **Chapitre V**

# **D**iscussion générale et perspectives

Ce travail avait pour objectif de comparer la variabilité naturelle entre les gènes *eIF4E* chez différentes espèces végétales afin de mettre en relation la diversité observée dans les interactions plante-virus et plus particulièrement avec la résistance aux potyvirus. La stratégie développée pour atteindre cet objectif a été de coupler des analyses fonctionnelles avec des études d'évolution moléculaire des gènes *eIF4E* chez 2 Solanacées (piment et tomate) et chez *A. thaliana*. Les Solanacées sont naturellement infectées par les potyvirus et possèdent plusieurs gènes récessifs de résistance correspondant aux facteurs eIF4E tandis que les données disponibles chez *A. thaliana*, espèce végétale dont tous les gènes *eIF4E* sont connus, indiquent qu'aucune des résistances récessives identifiées à ce jour dans la variabilité naturelle ne correspond à eIF4E.

A partir des résultats obtenus et plus largement, des données récentes montrant l'implication des facteurs eIF4E à la fois dans de nombreux processus cellulaires et dans les interactions plantes-virus à ARN, nous aborderons les points de discussion et les perspectives suivants :

- quel est le rôle de la redondance des gènes *eIF4E* dans les fonctions cellulaires de la plante et dans l'interaction plante-poyvirus ?
- quels sont les rôles de eIF4E dans l'interaction plante-potyvirus ?
- comment évoluent les gènes *eIF4E* chez différentes familles végétales ?
- quel est l'impact de l'étude de la diversité des gènes *eIF4E* sur la gestion des résistances aux potyvirus ? Cette question en lien direct avec l'amélioration des plantes abordera les aspects concernant l'efficacité, la durabilité et le coût de la résistance.

## **V.1. Rôle de la redondance des gènes *eIF4E* dans les fonctions cellulaires de la plante et dans l'interaction plante-potyvirus**

### **V.1.1. A la quête de l'ensemble des membres de la famille multigénique eIF4E**

Le séquençage des génomes et le grand nombre de séquences EST disponibles ont permis d'identifier chez les eucaryotes plus de 400 séquences nucléotidiques codant pour des protéines qui partagent des similarités de séquence avec les protéines de liaison à la coiffe

(Joshi *et al.*, 2005). Chez les eucaryotes supérieurs, les gènes *eIF4E* identifiés appartiennent à des familles multigéniques constituées de 2 à 7 copies : 2 chez la grenouille (Wakiyama *et al.*, 2001), 3 chez l'homme (Joshi *et al.*, 2004), 5 chez *C. elegans* (Keiper *et al.*, 2000), 6 chez le poisson zèbre (Robalino *et al.*, 2004), 7 chez la drosophile (Hernández *et al.*, 2005).

Chez les plantes, le nombre exact de gènes codant pour des protéines de liaison à la coiffe n'est connu que chez les espèces dont le génome a été entièrement séquencé *i.e.*, *A. thaliana*, le riz *O. sativa* et le peuplier *Populus trichocarpa* (The Arabidopsis Genome Initiative, 2000; International Rice Genome Sequencing Project, 2005; Tuskan *et al.*, 2006). Trois gènes *eIF4E*, 1 gène *eIF(iso)4E* et 1 gène *nCBP* ont été identifiés chez *A. thaliana* et 1 gène *eIF4E*, 2 gènes *eIF(iso)4E* et 1 gène *nCBP* chez *P. trichocarpa*, la seconde dicotylédone dont le génome a été entièrement séquencé. En revanche, chez l'espèce monocotylédone *O. sativa*, on retrouve bien ces 3 gènes mais en une seule copie. Ainsi, les 3 formes *eIF4E*, *eIF(iso)4E* et *nCBP* sont systématiquement retrouvées chez les plantes supérieures, suggérant une spécialisation fonctionnelle.

A partir de l'ensemble des séquences *eIF4E* répertoriées à ce jour dans les banques d'EST, la différence du nombre de copies de chaque forme *eIF4E* entre dicotylédones et monocotylédones apparaît confirmée : plusieurs copies de ces gènes sont présentes chez différentes dicotylédones comme le tabac, le tournesol (2 *eIF(iso)4E*; Combe *et al.*, 2005) et la pomme de terre (2 *eIF4E*; Joshi *et al.*, 2005) alors que le blé et l'orge ne comptent qu'une copie *eIF4E* et une copie *eIF(iso)4E*. Une copie du gène *nCBP* est retrouvée chez toutes les plantes supérieures. D'un point de vue évolutif, l'hypothèse la plus probable est que la duplication des gènes *eIF4E* a eu lieu après la divergence entre les monocotylédones et les dicotylédones. On ne peut cependant pas exclure que les monocotylédones aient perdu certaines copies au cours de l'évolution après la séparation monocotylédones/dicotylédones.

Ces données confortent les résultats de cartographie qui suggéraient l'existence de plus de 3 copies des gènes *eIF4E* chez les Solanacées, et indiquent également que les ARNm des copies additionnelles sont certainement peu exprimés puisque non retrouvés dans les banques d'EST de tomate pourtant importantes (plus de 213 000 EST disponibles à ce jour; [http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/est\\_ann.pl?gudb=tomato](http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/est_ann.pl?gudb=tomato)). La collaboration mise en place avec le Centre National de Ressources Génomiques Végétales de l'INRA de Toulouse (H. Bergès) permettra, *via* le criblage de la banque BAC de piment, d'isoler

l'ensemble des gènes chez cette espèce. Les orthologues des gènes identifiés pourront ensuite être recherchés chez la tomate.

Dans un premier temps, les nouveaux gènes identifiés seront analysés *in silico*. Les alignements de séquences et la recherche de similarités avec les protéines eIF4E fourniront les premières pistes sur leur rôle fonctionnel. En effet, les alignements de séquences entre 4 protéines eIF4E chez *C. elegans* - IFE-1, IFE-2, IFE-3 et IFE-5 - ont permis de montrer que les copies IFE-1, IFE-2 et IFE-5 présentaient un pourcentage d'identité plus élevé entre elles qu'avec la copie IFE-3, suggérant une divergence fonctionnelle de IFE-3 (Joshi *et al.*, 2005; Keiper *et al.*, 2000). Cette prédiction a été validée par la suite *via* des expériences d'interaction *in vitro* qui montrent que seule la protéine IFE-3 n'interagit pas avec la coiffe triméthylée présente à l'extrémité 5' de 70 % des ARNm de *C. elegans* (Lall *et al.*, 2004). Sur la base de ces alignements de séquences, il sera également possible de déterminer chez le piment et la tomate à quelle forme eIF4E correspondent les nouvelles copies eIF4E identifiées. En effet, des acides aminés spécifiques de chaque forme ont pu être identifiés (Monzingo *et al.*, 2007). *In silico*, il sera également possible de comparer la structure 3D prédite des protéines eIF4E avec la structure de la protéine eIF4E cristallisée chez le blé (Monzingo *et al.*, 2007). Cependant, aucune structure 3D des formes eIF(iso)4E et nCBP n'est disponible. Ces analyses permettront d'avancer des hypothèses quant à l'impact de telles ou telles mutations sur la charge globale de la protéine ou sur l'encombrement stérique qu'elles induisent.

Dans un second temps, des expériences *in vitro* et *in planta* devront être mises en place pour caractériser le rôle fonctionnel de ces formes eIF4E. Des expériences de complémentation fonctionnelle dans la levure et des tests de fixation à la coiffe des ARNm permettront de vérifier la capacité d'initiation de la traduction de ces protéines. Des analyses du niveau d'accumulation des transcrits et des protéines seront nécessaires pour s'assurer de leur expression *in planta*. Ces études pourront être complétées par des expériences de localisation subcellulaire des protéines eIF4E.

## V.1.2. Redondance *versus* spécialisation des protéines eIF4E

### V.1.2.1. Arguments en faveur de la redondance fonctionnelle des protéines eIF4E

Les analyses de séquence couplées aux expériences de complémentation fonctionnelle dans la levure réalisées au cours de ma thèse montrent que (i) les 3 formes eIF4E possèdent les acides aminés importants pour la stabilité de la structure des protéines eIF4E et les acides aminés impliqués dans la fixation à la coiffe des ARNm et à eIF4G et que (ii) tous les facteurs eIF4E, à l'exception de eIF(iso)4E chez le piment, sont capables de restaurer la croissance de levures délétées de leur gène *eIF4E*. De précédentes expériences montraient que la croissance des levures délétées de leur gène *eIF4E* complétement avec le transcrit *eIF(iso)4E* d'*A. thaliana* était ralentie par rapport à la croissance des levures complétement avec *eIF4E1* d'*A. thaliana* (Rodriguez *et al.*, 1998), ce qui n'a pas été observé dans mes expériences. At-eIF(iso)4E présenterait une homologie fonctionnelle partielle avec le facteur eIF4E de levure, qui pourrait s'expliquer par la formation d'un complexe At-eIF(iso)4E/eIF4G de levure moins efficace que le complexe At-eIF4E1/eIF4G de levure et ainsi se traduire dans certaines conditions par un ralentissement de croissance. De façon similaire, il est possible que la protéine eIF(iso)4E de *C. annuum* n'interagisse pas de façon optimale avec la protéine eIF4G de levure. Des études complémentaires d'interaction *in vivo*, comme des expériences d'affinité à la coiffe des ARNm, devront être réalisées pour affiner ces résultats concernant le maintien de la redondance fonctionnelle des différentes protéines eIF4E. Par ailleurs, le fait que At-nCBP restaure la croissance des levures délétées de leur gène *eIF4E* indique que la nCBP pourrait être impliquée dans l'initiation de la traduction et ne se limiterait pas à un rôle dans la séquestration des ARNm comme le proposait Ruud *et al.* (1998).

Des analyses des simples mutants obtenus pour chacun des gènes *eIF4E* chez *A. thaliana* montrent que les plantes ne présentent pas d'altération de leur phénotype. De même, des plantes de piment présentant une délétion naturelle de la protéine eIF(iso)4E (plantes *pvr6*) ne semblent pas être altérées dans leur développement et leur reproduction. Ces résultats suggèrent que l'absence de l'un ou l'autre des facteurs eIF4E peut être compensée par les autres membres de la famille multigénique. D'autres expériences ont montré que la suppression de l'expression du gène *eIF(iso)4E* chez *A. thaliana* et le tabac résultait en une augmentation compensatrice du niveau de protéine eIF4E1 (Duprat *et al.*, 2002; Combe *et al.*,

2005). Il serait intéressant de mesurer le niveau d'accumulation des autres protéines de liaison à la coiffe (eIF4E2, eIF4E3 et nCBP) chez ce mutant *eIF(iso)4E* afin de déterminer si ce phénomène de compensation lié au maintien de la redondance fonctionnelle concerne tous les facteurs eIF4E. A l'inverse, la suppression de l'expression du gène *eIF4E1* chez *A. thaliana* ne résulte pas en une augmentation compensatrice du niveau de protéine eIF4(iso)4E (Duprat *et al.*, 2002). Ces résultats suggèrent que eIF(iso)4E ne pourrait pas remplacer eIF4E1 *in vivo*, indiquant l'acquisition d'une spécialisation de chacun des eIF4E chez les plantes.

#### ***V.1.2.2. Arguments en faveur de la spécialisation des protéines eIF4E***

Une hypothèse pour expliquer pourquoi tous les gènes *eIF4E* sont maintenus au cours de l'évolution est que chacun des membres de la famille multigénique aurait acquis une spécialisation fonctionnelle. La conservation des 3 formes eIF4E chez les plantes supérieures, ainsi que la pression de sélection purificatrice détectée sur l'ensemble des membres de la famille multigénique eIF4E chez toutes les espèces étudiées au cours de ce travail argumentent en faveur de cette hypothèse. Une spécialisation fonctionnelle correspondrait à l'acquisition d'un niveau supplémentaire de régulation de l'expression des gènes, avantageux pour des organismes complexes comme les plantes supérieures. Plusieurs résultats argumentent en faveur de cette hypothèse.

L'initiation de la traduction des ARNm est l'étape la plus limitante et la plus complexe de la régulation de la synthèse protéique chez les eucaryotes. Dans ce processus, eIF4E participe au recrutement de l'ARNm *via* une interaction forte et spécifique avec la protéine eIF4G. Mais des études de plus en plus nombreuses montrent l'implication des différents eIF4E dans tout un ensemble de processus liés au métabolisme des ARNm. En effet, ce facteur intervient dans l'export nucléaire des ARNm, dans la séquestration et dans la stabilisation des ARNm dans le cytosol (Richter and Sonenberg, 2005; Culjkovic *et al.*, 2007). Ainsi, le rôle des protéines eIF4E est considérable puisqu'elles peuvent affecter les niveaux d'accumulation des autres protéines de la cellule par l'un ou l'autre de ces mécanismes.

La fonction de chaque membre de la famille eIF4E chez plusieurs organismes commence à être élucidée. Certaines études se sont intéressées à la localisation des transcrits

eIF4E au niveau des organes, des tissus ou des cellules. Chez le nématode *C. elegans*, l'identification de la forme eIF4E4 uniquement dans les neurones et les tissus musculaires indiquent que cette forme tissu-spécifique est spécialisée dans le recrutement d'ARNm présents dans ces tissus (Dinkova *et al.*, 2005). Chez *A. thaliana*, les expériences de RT-PCR quantitative réalisées au cours de ma thèse ne révèlent pas d'expression organe-spécifique d'un des transcrits *eIF4E* chez *A. thaliana*. Dans ce cas, aucune information sur l'acquisition d'une spécialisation liée à la localisation n'a donc été mise en évidence mais des expériences plus précises de localisation devront être mises en place pour affiner ces résultats et détecter des expressions tissu-spécifique et/ou cellule-spécifique.

Plusieurs études ont montré que certains processus cellulaires stade-spécifiques faisaient intervenir une forme particulière de eIF4E. IFE-1 est requis pour la spermatogenèse chez *C. elegans* (Amiri *et al.*, 2001) et eIF4E-1 est impliqué dans le développement ovarien et embryonnaire chez la Drosophile (Hernández *et al.*, 2004). Chez *Saccharomyces pombe*, eIF4E-1 est impliqué dans la traduction coiffe-dépendante constitutive tandis que eIF4E-2 intervient dans la traduction seulement en condition de stress (Ptushkina *et al.*, 1996; Ptushkina *et al.*, 2004).

Les propriétés de fixation à la coiffe des ARNm de chaque protéine eIF4E sont étudiées pour valider l'implication des eIF4E dans l'initiation de la traduction. Les structures primaire et secondaire de la coiffe des ARNm ont une influence sur l'affinité avec les différents eIF4E. Ainsi, il a été montré que eIF(iso)4F chez le blé préfère se fixer aux ARNm présentant une structure linéaire à leur extrémité 5' tandis que eIF4F préfère interagir avec les structures en épingle à cheveux (Carberry *et al.*, 1991). La nCBP d'*A. thaliana* se fixe avec une affinité 5 à 20 fois plus importante à la coiffe des ARNm que l'isoforme eIF(iso)4E (Ruud *et al.*, 1998). Chez *C. elegans*, seule la protéine IFE-3 n'interagit pas avec la coiffe triméthylée présente à l'extrémité 5' de 70 % des ARNm chez cette espèce. Toujours chez *C. elegans*, certains ARNm impliqués dans la ponte sont traduits spécifiquement par la forme IFE-4. Ces travaux mettent donc en avant la spécificité de fixation de chacune des formes eIF4E en fonction de la structure des ARNm.

Chez de nombreux organismes (levure, drosophile, homme), plus de 60 % des protéines eIF4E sont localisées dans le noyau des cellules où elles sont associées aux corps nucléaires, suggérant que certains facteurs eIF4E auraient acquis une spécialisation

fonctionnelle dans le noyau (Iborra *et al.*, 2001). En effet, plusieurs études ont montré que les protéines eIF4E nucléaires permettent l'export du noyau vers le cytosol de certains ARNm chez les mammifères, *via* la fixation avec la coiffe de ces ARNm (Rousseau *et al.*, 1996; Lai and Borden, 2000; Topisirovic *et al.*, 2002; Culjkovic *et al.*, 2006). eIF4E ne permet l'export que des ARNm qui présentent une séquence d'environ 100 pb dans leur région 3'UTR appelée 4E-sensitivity element (4E-SE; Culjkovic *et al.*, 2006). Certains ARNm sont à la fois exportés et traduits par ces eIF4E nucléaires, signifiant que eIF4E peut agir sur l'expression des protéines qu'il exporte (Culjkovic *et al.*, 2007) Ces études mettent donc en évidence que eIF4E peut se lier de façon discriminante à certains ARNm et qu'il est impliqué dans leur régulation post-transcriptionnelle. Culjkovic *et al.* (2006) estiment qu'une centaine d'ARNm seraient régulés par ce processus.

Certaines formes *eIF4E* peuvent également avoir un rôle dans la séquestration des ARNm. Chez la drosophile, eIF4E-8 séquestre les ARNm dans un complexe inactif avec lequel la forme eIF4E-1 fonctionnelle (seule forme parmi les 8 identifiées capable d'avoir une activité de fixation à la coiffe) et les ribosomes ne peuvent pas interagir (Cho *et al.*, 2005; Hernández *et al.*, 2005). Chez *A. thaliana*, la protéine nCBP présente une meilleure affinité pour la coiffe que pour la protéine eIF(iso)4G, suggérant que la nCBP serait préférentiellement impliquée dans la séquestration des ARNm plutôt que dans leur traduction (Ruud *et al.*, 1998). Ainsi, afin de mieux identifier les spécificités fonctionnelles de chaque membre eIF4E de la famille multigénique, il s'avère nécessaire d'identifier les partenaires protéiques de chaque eIF4E. A ce jour, plus de 15 protéines ont été montrées comme interagissant avec les protéines eIF4E toutes espèces confondues *via* des méthodes classiques d'interactions protéine-protéine (GST-pull-down et double hybride dans la levure; pour revue voir Rhoads *et al.*, 2007). Des méthodes d'analyse à grande échelle des interactions protéine-protéine sont désormais disponibles et des réseaux d'interactions appelés "interactomes" commencent à émerger. Ces bases de données ne sont pour l'instant disponibles que pour une poignée d'organismes modèles (*S. cerevisiae*, *C. elegans* et *D. melanogaster*) et sont en cours de développement chez *A. thaliana* (<http://www.associomics.org/>). Deux méthodes permettent de générer ces "cartes d'interactions" : des criblages par le système de double hybride dans la levure (High-Throughput Yeast two-Hybrid analysis, HT-Y2H) et des analyses de complexes protéiques par purification d'affinité et spectrométrie de masse (AP-MS). Le système HT-Y2H permet de ne détecter que des interactions directes entre 2 protéines tandis que le système AP-MS permet d'identifier tous les composants d'un complexe protéique stable

(Rhoads *et al.*, 2007). Ces méthodes apporteraient donc des informations considérables dans le cas des protéines eIF4E qui participent probablement à de nombreux complexes d'interactions fonctionnels.

### V.1.3. Utilisation sélective des différentes protéines eIF4E par les potyvirus

Les expériences d'interactions entre les 3 protéines eIF4E de *C. annuum* et VPg de plusieurs souches de PVY et de TEV ont montré une utilisation sélective des facteurs eIF4E non seulement par les différents potyvirus mais également par différentes souches d'un même potyvirus, mettant en évidence un niveau supplémentaire de complexité dans les interactions plantes-potyvirus. Le rôle de chaque protéine eIF4E de *C. annuum* dans le contournement des résistances reste ambigu. La VPg des différentes souches de potyvirus testées varient dans leur capacité à interagir ou non avec les différentes protéines eIF4E. La VPg de la souche PVY-SON41 qui contourne les allèles *eIF4E1-pvr2<sup>1</sup>*, *eIF4E1-pvr2<sup>2</sup>* et *eIF4E1-pvr2<sup>3</sup>* interagit avec les protéines eIF4E1 codées par les allèles de résistance mais pas avec les protéines eIF4E2 ou eIF(iso)4E, suggérant que le contournement résulte d'une restauration de l'interaction entre la VPg et les protéines eIF4E1 mutées. En revanche, la VPg de la souche PVY-SON41p, capable de contourner les allèles de résistance *eIF4E1-pvr2<sup>1</sup>* et *eIF4E1-pvr2<sup>2</sup>*, n'interagit pas avec les protéines eIF4E1 codées par les allèles de résistance mais interagit avec la protéine eIF4E2. Dans ce cas, le contournement de la résistance pourrait s'expliquer non pas par la restauration d'une interaction avec les protéines eIF4E1 mutées mais par l'utilisation d'une autre protéine eIF4E.

L'utilisation sélective des différentes protéines eIF4E par les potyvirus est confortée par des analyses réalisées chez des mutants d'*A. thaliana* pour les gènes *eIF4E*. Des mutants T-DNA et/ou EMS pour chacun des 5 gènes sont disponibles (Sato *et al.*, 2005; Duprat *et al.*, 2002; Browning *et al.*, données non publiées; Caranta *et al.*, données non publiées). Des expériences de double hybride dans la levure et des tests ELISA d'interaction *in vitro* montrent que la protéine VPg du TuMV interagit seulement avec la protéine eIF(iso)4E tandis que la VPg du CIYVV interagit spécifiquement avec eIF4E1, eIF4E2 et eIF4E3 (Labay, 2007). De plus, seul le mutant dont le gène *eIF(iso)4E* est éteint est résistant au TuMV et seul le mutant dont le gène *eIF4E1* est éteint est résistant au CIYVV (Lellis *et al.*, 2002; Duprat *et*

*al.*, 2002; Sato *et al.*, 2005). Ces résultats mettent en avant la spécificité d'interaction des VPg avec les différentes protéines eIF4E en fonction du couple plante-potyvirus considéré.

Le domaine protéase de la NIa (forme précurseur de la VPg) du *Tomato ringspot virus* (ToRSV, genre *Nepovirus*) interagit *in vitro* avec la protéine eIF(iso)4E d'*A. thaliana* (Léonard *et al.*, 2002). Des tests par le système de double hybride dans la levure indiquent que la NIa d'un autre nepovirus, l'*Arabidopsis mosaic virus* (ArMV), est capable d'interagir avec les 5 protéines eIF4E d'*A. thaliana* (C. Charron, collaboration T. Wetzel, RLP Agrosiences, Neustadt, Allemagne), suggérant que les nepovirus sont capables d'utiliser toutes les protéines eIF4E chez *A. thaliana*. Ces expériences d'interaction entre les protéines eIF4E et les protéines d'autres genres viraux montrent donc que le mode d'utilisation sélective des facteurs eIF4E par les potyvirus n'est pas généralisable à l'ensemble des interactions plantes-virus à ARN.

## **V.2. Vers l'élucidation du/des rôle(s) de eIF4E dans l'interaction plante-potyvirus**

### **V.2.1. Impact de l'infection virale sur l'expression de eIF4E**

Plusieurs études se sont intéressées aux interactions entre les virus de plantes et la machinerie d'initiation de la traduction de la cellule hôte ainsi qu'à leurs conséquences sur la synthèse des protéines de la plante.

L'infection de *Brassica perviridis* par le TuMV résulte en une augmentation de l'expression de la protéine eIF4E1 (Léonard *et al.*, 2004). Léonard *et al.* (2004) ont proposé que l'infection par le TuMV chez *B. perviridis* conduit à l'inactivation de eIF(iso)4E ou à son recrutement exclusif par le génome viral et que la surexpression de eIF4E1 observée correspond à la réponse de la plante pour compenser la non fonctionnalité de eIF(iso)4E dans la traduction des ARNm cellulaires. D'autres expériences, cohérentes avec ces résultats, ont montré que le knock-out du gène *eIF(iso)4E* chez *A. thaliana* confère une résistance à l'infection par le TuMV (Duprat *et al.*, 2002; Lellis *et al.*, 2002). Ces travaux mettent en évidence un contrôle post-transcriptionnel de l'expression des gènes *eIF4E* chez *B. perviridis*

au cours de l'infection virale. Le contrôle post-transcriptionnel est connu pour être adopté quand une cellule doit répondre rapidement à une situation particulièrement stressante sans faire intervenir les voies nucléaires pour la synthèse de nouveaux ARNm. Ce phénomène de compensation entre 2 formes eIF4E rappelle les résultats obtenus chez le mutant *eIF(iso)4E* d'*A. thaliana* où une surexpression de la protéine eIF4E1 était observée pour faire face à l'absence de la protéine eIF4E1 (Duprat *et al.*, 2002) et ainsi assurer le maintien d'un niveau de synthèse protéique dans les cellules nécessaire au bon développement de la plante. Une étude comparative des niveaux d'accumulation de chaque transcrite et de chaque protéine eIF4E chez des plantes d'*A. thaliana* et de piment, infectées ou non par les potyvirus, est en cours de développement au laboratoire afin de détecter et de préciser, le cas échéant, une régulation transcriptionnelle et/ou traductionnelle liée à l'infection virale.

L'augmentation du niveau d'accumulation d'autres protéines de la plante hôte au cours de l'infection par le PSbMV chez le pois ou par le CMV chez *Cucurbita pepo* a également été mise en évidence (Aranda *et al.*, 1996; Havelda and Maule, 2000). Même si le mécanisme et l'intérêt de cette induction demeurent inconnus, il a été proposé que cette surexpression permettrait de préparer la cellule à la demande plus forte d'activité traductionnelle nécessaire à la réplication du génome viral (Maule *et al.*, 2002).

### **V.2.2. Vers l'identification des partenaires de eIF4E au cours de l'infection virale**

Afin de mieux comprendre les mécanismes moléculaires qui contrôlent les interactions plantes-potyvirus et mieux caractériser l'impact des interactions eIF4E-VPg dans le cycle viral et dans les fonctions cellulaires, il s'avère nécessaire (*i*) d'identifier l'ensemble des partenaires protéiques de eIF4E dans l'interaction avec les potyvirus et (*ii*) de caractériser la localisation subcellulaire des interactions protéines virales-protéines de l'hôte.

Les méthodes de test d'interaction utilisées au cours de ma thèse ne permettent de détecter que les interactions entre 2 partenaires. Or les interactions au cours de l'infection virale dépendent de la formation de complexes protéiques faisant intervenir d'autres protéines virales et/ou des protéines de l'hôte. Des études biochimiques de l'interaction eIF4E-VPg

viennent appuyer cette hypothèse. Le facteur eIF4G, un autre facteur essentiel dans la machinerie d'initiation de la traduction, est capable de former un complexe trimérique avec les protéines eIF4E et VPg. La fixation de eIF4G à eIF4E augmente son affinité pour la VPg, ce qui favoriserait le recrutement de eIF4E par le virus, aux dépens des ARNm cellulaires (Grzela *et al.*, 2006; Michon *et al.*, 2006; Miyoshi *et al.*, 2005). De plus, une étude récente montre que les facteurs eIF4E et eIF4G sont recrutés par les potyvirus de façon coordonnée, suggérant que la réussite de l'infection requiert l'ensemble du complexe eIF4F (Nicaise *et al.*, 2007). Ce sont des approches gènes candidats qui ont permis de mettre en évidence le rôle central des facteurs eIF4E et eIF4G dans l'infection virale; cette approche est donc restreinte par le nombre de gènes candidats connus. Ainsi, la caractérisation de partenaires cellulaires sans *a priori* s'avère nécessaire afin de mieux comprendre les mécanismes mis en jeu au cours de l'infection virale. Des expériences de GST pull-down permettraient notamment de détecter des hétérocomplexes formés *in vivo* (pour revue voir Rhoads *et al.*, 2007).

Peu de données sont disponibles concernant la localisation subcellulaire du virus au cours de l'infection et donc sur la localisation des complexes protéine(s) de l'hôte / protéine(s) virale(s). Plusieurs approches biochimiques pourraient être utilisées pour étudier ces interactions protéine-protéine intracellulaires, telles que le FRET (Fluorescence Resonance Energy Transfert; Gordon *et al.*, 1998) ou le BiFC (Bimolecular Fluorescence Complementation; Walter *et al.*, 2004; Bracha-Drori *et al.*, 2004). Ces approches, suivies d'une expérience de purification des complexes protéiques à partir de noyaux isolés, permettraient également d'identifier d'éventuelles interactions protéine(s) de l'hôte-protéine(s) virale(s) au niveau du noyau des cellules infectées. En effet, plusieurs études ont mis en évidence qu'aussi bien les protéines CP, CI, NIa et NIB des potyvirus que les protéines eIF4E, eIF1A, eIF2, eIF3, eIF4G, eIF5A et PABP impliquées dans l'initiation de la traduction peuvent se localiser dans le noyau des cellules (Baunoch *et al.*, 1991; Strudwick and Borden, 2002). De plus, des études réalisées chez les *Picornavirus* montrent que ces virus dont le cycle infectieux est pourtant cytoplasmique utilisent des composants nucléaires de la cellule hôte au profit de leur réplication et leur traduction (Hiscox, 2003; Weidman *et al.*, 2003). Pour renforcer l'importance de la localisation subcellulaire des interactions, de récents travaux ont montré que les associations entre eIF(iso)4E d'*A. thaliana* et les différentes formes précurseurs de la VPg du TuMV pouvaient être impliquées dans différentes étapes du cycle infectieux. L'interaction eIF(iso)4E/VPg-Pro qui a lieu dans le noyau perturberait des fonctions cellulaires de la plante hôte tandis que l'interaction eIF(iso)4E/6K-VPg-Pro qui a

lieu dans les vésicules endoplasmiques pourrait jouer un rôle dans la réplication et/ou la traduction de l'ARN viral (Beauchemin *et al.*, 2006).

### **V.3. Evolution des gènes *eIF4E* et résistance aux virus à ARN**

#### **V.3.1. Co-évolution entre le facteur eIF4E1 du piment et la VPg des potyvirus**

La combinaison d'approches associant séquençage d'allèles, tests phénotypiques et complémentation génétique a permis de caractériser 9 allèles récessifs de résistance aux PVY et TEV au locus *eIF4E1-pvr2* chez *C. annuum*, espèce diploïde préférentiellement autogame (même si, dans certaines conditions, le taux d'allogamie peut être élevé). Différentes combinaisons de substitutions nucléotidiques exclusivement non-synonymes permettent de distinguer ces 9 variants alléliques. Nous avons pu montrer par le système de double hybride dans la levure que ces substitutions abolissaient l'interaction avec la protéine VPg de souches avirulentes de PVY et de TEV et que cette interaction était rétablie avec les VPg de souches virulentes. Des expériences de complémentation fonctionnelle dans une levure délétée de son gène *eIF4E* ont montré que la protéine eIF4E1 de *C. annuum* peut accumuler les mutations qui affectent la fixation de la VPg des potyvirus sans avoir de répercussion négative sur la fonction de traduction des ARNm dans la levure. De plus, la résistance à large spectre *i.e.*, à plusieurs souches de PVY et de TEV, résulterait de l'effet complémentaire, plutôt qu'individuel, de substitutions en acides aminés dans 2 régions distinctes à la surface de la protéine eIF4E1 et à proximité de la poche de fixation à la coiffe des ARNm. L'analyse du polymorphisme des 3 gènes *eIF4E* chez *C. annuum* a révélé que seul le gène *eIF4E1* présente des traces de sélection positive; les acides aminés sous pression de sélection positive correspondent aux résidus impliqués dans la sensibilité *versus* résistance au PVY, suggérant une pression de sélection positive exercée par les potyvirus. Côté virus, des études montrent que les acides aminés de la protéine virale VPg du PVY sous sélection positive correspondent aux résidus impliqués dans le contournement des résistances médiées par eIF4E (Moury *et al.*, 2004). L'ensemble de ces données mettent en avant l'existence d'une coévolution entre eIF4E1 et la VPg des potyvirus. Ce travail peut être considéré comme pionnier dans le domaine.

Les exemples présentés dans la synthèse bibliographique, et concernant les gènes dominants de résistance aux bactéries et aux champignons, montrent qu'il est difficile d'établir un modèle général d'évolution des gènes de résistance chez les plantes. Récemment, il a été proposé qu'un facteur important régissant l'évolution des gènes *R* était le mode de reconnaissance, direct ou indirect, entre la protéine de résistance et la protéine d'avirulence (Dodds *et al.*, 2006; Mc Dowell et Simon, 2006). Dans le cas des reconnaissances indirectes faisant intervenir une protéine de garde, les gènes *R* présentent une faible diversité allélique maintenue par de la sélection balancée. A l'inverse, dans le cas des reconnaissances directes, une coévolution de type "course aux armements" est retrouvée, avec le maintien d'une forte diversité allélique pour le gène de résistance et le gène d'avirulence. Ainsi, dans le cas d'une interaction directe, l'agent pathogène peut contourner la résistance en modifiant de façon minimale sa protéine Avr pour échapper à la reconnaissance par la protéine de résistance et *vice versa* pour la protéine R. Ce type de reconnaissance est donc caractérisé par de la sélection diversifiante à la fois chez les protéines R et Avr.

Les cas qui se rapprochent le plus de ce que nous observons pour le gène *eIF4E1* du piment sont ceux des gènes de résistance *Pto* de la tomate et *L* du lin, qui présentent un niveau de substitutions non-synonymes élevé et un grand nombre de variants alléliques résultant d'une pression de sélection pour la diversification (Dodds *et al.*, 2006; Rose *et al.*, 2007). L'évolution de *Pto* semble être influencée à la fois par de la sélection purificatrice et de la sélection diversifiante, comme *eIF4E1* chez le piment. Cependant, une pression de sélection balancée est à l'origine du maintien de la variété allélique au locus *Pto*, ce qui ne semble pas être le cas pour *eIF4E1* de *Capsicum* où aucun polymorphisme synonyme adjacent aux polymorphismes non-synonymes n'est détecté. Dans le cas du couple lin-*M. lini*, des analyses portant à la fois sur le polymorphisme des gènes de résistance et de virulence de l'agent pathogène qui interagissent directement ont été réalisées. Des traces de sélection positive sont retrouvées à la fois chez les locus *L* et *Avr*, comme dans le cas du pathosystème piment-potyvirus. Ces interactions directes entre protéines de résistance et protéines d'avirulence caractérisent une évolution rapide de type "course aux armements". Dans le cas de la résistance récessive aux virus à ARN, la résistance est la conséquence de l'incapacité pour le virus d'interagir avec une protéine de l'hôte qui lui est indispensable pour accomplir son cycle infectieux. A l'inverse du mécanisme "actif" des résistances dominantes, le virus doit donc muter pour rétablir l'interaction avec la protéine de l'hôte et non pour échapper à l'interaction. Le parasitisme obligatoire des virus à ARN, comme celui de l'oomycète *M. lini*, implique que

les protéines VPg de virulence coévoluent dans un système clos avec les protéines de résistance eIF4E de la plante hôte, ce qui impose une course évolutive rapide à la fois chez l'hôte et le pathogène. Un pré-requis pour une telle issue évolutive est que les protéines de virulence et de résistance puissent accumuler des mutations qui affectent l'interaction sans imposer un coût de fitness trop important pour la plante ou pour le pathogène. Cette notion de coût sera abordée dans le dernier paragraphe de la discussion générale.

### **V.3.2. Comparaison du profil d'évolution des gènes *eIF4E* chez différentes familles végétales**

Mon travail de thèse s'est focalisé sur l'analyse du polymorphisme des gènes *eIF4E* chez 2 Solanacées et *A. thaliana*. Le caractère générique du rôle des gènes *eIF4E* dans les interactions plante-virus à ARN offre l'opportunité de comparer le polymorphisme et l'évolution moléculaire de ces gènes chez différentes familles végétales et ainsi de vérifier si le modèle de coévolution observé entre le piment et les potyvirus est transposable aux autres couples plante-virus à ARN. Ainsi, ce travail s'inscrit dans un projet financé par le Bureau des Ressources Génétiques (BRG AO 2005-2007) visant à comparer les patrons de diversité des gènes *eIF4E* au sein de 5 familles végétales : Solanacées, Crucifères, Vitacées, Légumineuses et Triticées. L'ensemble des données obtenues dans le cadre de ce projet permettent de comparer le profil d'évolution des gènes *eIF4E* chez différentes familles végétales, possédant ou non des gènes récessifs de résistance aux potyvirus correspondant à eIF4E, et présentant des modes de reproduction et des traits d'histoire de vie différents.

#### ***V.3.2.1. Le profil d'évolution du gène eIF4E1 n'est pas conservé chez les Solanacées***

La tomate est une autre Solanacée proche du piment et également infectée par les potyvirus mais pour laquelle une plus faible fréquence de gènes récessifs de résistance aux potyvirus a été observée dans la variabilité naturelle. Aucun des acides aminés situés dans les régions I et II impliquées dans la résistance aux potyvirus des protéines eIF4E n'apparaît sous sélection positive chez la tomate. Cette absence de pression de sélection pourrait s'expliquer

par le fait que, contrairement au piment, les potyvirus n'ont qu'une incidence faible sur les cultures de tomate. La majorité des souches de potyvirus n'induisent que des symptômes de mosaïque ou de marbrures discrets sur les feuilles de tomate, même si certains isolats apparus dans les années 1980 sont responsables de symptômes plus sévères (croissance des plantes ralentie, production en fruits diminuée; Gebre Selassie *et al.*, 1985). Les potyvirus sont quasiment systématiquement retrouvés en infections mixtes avec d'autres genres viraux, comme le CMV. De plus, les virus prévalant sur tomate ne sont pas les potyvirus mais le *Tomato yellow leaf curl virus* (TYLC, genre *Begomovirus*), le CMV ou encore le *Pepino mosaic virus* (PepMV, genre *Potexvirus*), suggérant que les potyvirus exercent une pression de sélection moins forte sur les gènes *eIF4E* de cette espèce.

Chez la pomme de terre, une des Solanacées les plus touchées par les infections de potyvirus et notamment par le PVY, les résistances récessives ne sont pas facilement détectables à cause de la forte hétérozygotie de leur génome et de la prévalence de la tétraploïdie chez les espèces cultivées de *Solanum* tubéreux. A ce jour, les gènes de résistance sélectionnés pour lutter contre les potyvirus chez la pomme de terre sont de type dominant contrôlant une résistance totale (gènes *R* et *N*; Ross, 1986; Valkonen *et al.*, 1996) et aucune information n'est disponible sur l'implication potentielle des gènes *eIF4E* dans la résistance à ces virus. Une étude préliminaire réalisée au cours de ma thèse sur le polymorphisme de 3 gènes *eIF4E* chez 6 espèces diploïdes de pomme de terre n'a pas permis de détecter des substitutions en acides aminés caractéristiques des allèles de résistance aux potyvirus. Une analyse à plus grande échelle du polymorphisme des 5 gènes *eIF4E* identifiés à partir des EST est actuellement en cours chez une collection d'accessions de l'espèce diploïde *S. phureja* (<http://www.scri.ac.uk/research/pp/plantpathogeninteractions/virusresistance>; L. Torrance, SCRI, Scotland, UK) afin d'identifier d'éventuels variants alléliques *eIF4E* impliqués dans la résistance aux potyvirus.

Même si le piment, la tomate et la pomme de terre ont comme point commun d'être de la famille des Solanacées et d'être infectés par les potyvirus, leurs caractéristiques biologiques conduisent à des profils d'évolution des gènes *eIF4E* différents.

### ***V.3.2.2. A. thaliana ne constitue pas un modèle adapté à l'étude de la coévolution plante-potyvirus***

Chez *A. thaliana*, aucun des 5 gènes *eIF4E* n'est sous pression de sélection positive. Mes résultats confirment de précédentes études qui montrent qu'aucun gène récessif de résistance ne semble correspondre à *eIF4E* chez *A. thaliana*. Ces résultats peuvent s'expliquer par le fait qu'*A. thaliana* ne représente pas un hôte préférentiel pour les potyvirus, à cause de son cycle de vie très court par exemple. Walsh et Jenner (2002) insistent sur l'absence de gènes de résistance au TuMV présents dans la variabilité naturelle d'*A. thaliana*, suggérant que le TuMV n'infecte que très rarement *A. thaliana* et que le maintien de gènes de résistance représente un coût pour la plante. *A. thaliana*, espèce modèle dans de nombreux domaines d'études en biologie, ne constitue pas un modèle pertinent d'étude de la coévolution plante-potyvirus.

### ***V.3.2.3. Le profil d'évolution des gènes eIF4E n'est pas conservé chez les Légumineuses***

L'analyse du polymorphisme des gènes *eIF4E* a également été conduite chez le genre *Medicago* (résultats N. Chantret, J. Ronfort, S. Santoni, UMR DIAPC, Mauguio, France) car de précédentes études mettaient en évidence l'implication de ces gènes dans la résistance aux potyvirus chez d'autres espèces de Légumineuses : (i) chez le pois *Pisum sativum*, plusieurs allèles récessifs de résistance à 3 potyvirus correspondent à des protéines *eIF4E*. Les allèles *sbm1* de résistance au PSbMV, *wlv* de résistance au BYMV-W et *cyy2* de résistance au CIYVV codent tous pour des protéines *eIF4E* présentant des substitutions en acides aminés dans les régions de la protéine impliquées dans la résistance aux potyvirus (Gao *et al.*, 2004b; Bruun-Rasmussen *et al.*, 2007; Andrade *et al.*, données non publiées). L'allèle *mo* de résistance au BYMV code pour l'isoforme *eIF(iso)4E* (Gao *et al.*, 2004a), (ii) chez le haricot commun *Phaseolus vulgaris*, des analyses génétiques indiquent que le locus récessif *desc* de résistance au CIYVV pourrait correspondre à un facteur *eIF4E* (Sato *et al.*, 2003).

Les analyses de séquence des gènes *eIF4E* menées chez des espèces autogames diploïdes du genre *Medicago* ne permettent pas d'identifier de variants alléliques *eIF4E* ou *eIF(iso)4E* potentiellement impliqués dans la résistance aux potyvirus. De plus, que ce soit au

niveau intraspécifique ou interspécifique, aucune pression de sélection positive sur le gène *eIF4E* et le gène *eIF(iso)4E* n'a été identifiée. Aucun des gènes *eIF4E* chez *Medicago* ne semble donc impliqué dans les processus évolutifs liés aux interactions entre les protéines eIF4E et les potyvirus. Le profil de polymorphisme des gènes *eIF4E* diffère donc selon les espèces de Légumineuses étudiées : toute une série de variants alléliques impliqués dans la résistance aux potyvirus identifiées chez le pois, un seul allèle récessif de résistance détecté chez le haricot à ce jour et *a priori* absence de résistance récessive aux potyvirus correspondant aux gènes *eIF4E* chez l'espèce *Medicago*. Des analyses d'évolution des gènes *eIF4E* chez le pois et le haricot permettraient de vérifier si on retrouve les mêmes traces de sélection positive que celles identifiées pour le gène *eIF4E1* chez le piment.

#### ***V.3.2.4. Une pression de sélection positive s'exerce sur le gène eIF4E chez la vigne***

Deux virus du genre *Nepovirus* (*Grapevine fanleaf virus*, GFLV et ArMV) sont responsables de la maladie du Court-Noué chez la vigne, une des maladies les plus importantes chez cette espèce. Le domaine protéase de la VPg-Pro (NIa, forme précurseur de la VPg) du ToRSV (genre *Nepovirus*) interagit *in vitro* avec la protéine eIF(iso)4E d'*A. thaliana*, suggérant un rôle potentiel des facteurs eIF4E dans l'interaction plante-nepovirus (Léonard *et al.*, 2002). Du fait de la forte hétérozygotie du génome de la vigne, seules des résistances dominantes ont été ciblées lors du criblage de ressources génétiques pour la résistance aux nepovirus.

Les analyses de polymorphisme de la séquence codante du gène *eIF4E* (Résultats A.F. Adam-Blondon, URGV Evry, France; S. Santoni, UMR DIAPC Mauguio, France) chez une core collection de 48 accessions de l'espèce cultivée *Vitis vinifera* ssp. *sativa* ont permis de mettre en évidence un excès de substitutions non-synonymes par rapport aux substitutions synonymes ( $\pi_a \gg \pi_s$ ) et un signal très fort de sélection positive : 11 % des sites de la protéine eIF4E présentent un  $\omega$  supérieur à 5, rappelant les traces de sélection positive détectées au niveau du facteur eIF4E1 chez le piment. De plus, plusieurs des substitutions non-synonymes détectées se retrouvent dans les régions de la protéine eIF4E montrées comme étant impliquées dans la résistance aux potyvirus chez d'autres espèces. L'impact des substitutions en acides aminés identifiées sur la résistance aux nepovirus est en cours d'analyse (individus

en cours d'autofécondation pour avoir l'ensemble des variants alléliques à l'état homozygote afin de réaliser des tests de résistance aux nepovirus). Ces résultats pourraient ouvrir la voie à une nouvelle stratégie de lutte contre les nepovirus chez la vigne.

### ***V.3.2.5. Le polymorphisme du gène eIF4E chez le blé se distingue du polymorphisme détecté chez l'orge***

L'orge et le blé, de la tribu des Triticées, sont fréquemment infectés par des virus à ARN du genre *Bymovirus* transmis par le vecteur *Polymyxa graminis* présent dans le sol. Comme chez les potyvirus, une protéine VPg est liée de façon covalente à l'extrémité 5' de l'ARN des bymovirus. Chez l'orge *Hordeum vulgare*, le gène récessif *rym4/5* de résistance à plusieurs souches de bymovirus code pour une protéine eIF4E (Stein *et al.*, 2005). Une première analyse sur 56 accessions d'orge avait permis d'identifier 7 variants alléliques au locus *eIF4E-rym4/5* se différenciant par 9 substitutions nucléotidiques exclusivement non-synonymes. De plus, la substitution S57P se localise dans la région I impliquée dans la résistance aux potyvirus chez les autres espèces et les autres substitutions en acides aminés détectées se trouvent à proximité du domaine de fixation à la coiffe des ARNm (Stein *et al.*, 2005). Une étude plus récente portant sur un échantillon de 127 accessions d'orge cultivé et 4 accessions sauvages a permis de confirmer l'importante variété allélique présente à ce locus et l'identification d'un nombre important de mutations non-synonymes (Stracke *et al.*, 2007). D'autres travaux ont mis en évidence que le déterminant du contournement du gène récessif de résistance *rym4/5* par le BaYMV (genre *Bymovirus*) se situe vraisemblablement dans la zone centrale de la VPg (Kanyuka *et al.*, 2004; Kuhne *et al.*, 2003).

A partir de ces résultats et afin de comparer le profil d'évolution des gènes *eIF4E* au sein de la tribu des *Triticeae*, les analyses de polymorphisme de ces gènes ont été conduites chez des espèces diploïdes de blé (genre *Triticum*; résultats J. David, A. Cenci, UMR DIAPC, Mauguio, France). A partir des séquences codantes des gènes *eIF4E* connues chez le riz et par criblage de la banque BAC du blé dur *T. turgidum*, 1 gène *eIF4E* et 1 gène *eIF(iso)4E* ont été identifiés. Chez les espèces diploïdes *T. monococcum* et *T. urartu*, les gènes *eIF4E* et *eIF(iso)4E* apparaissent sous pression de sélection purificatrice. Cependant, la recherche de sites sous pression de sélection sur un échantillon interspécifique révèle que 4 sites dans la séquence codante de *eIF(iso)4E* présentent un  $\omega$  supérieur à 1, avec notamment le codon en

position 120 dans l'exon 2 qui présente un  $\omega$  supérieur à 4 avec une probabilité de 0,994. Ce site ne varie pas uniquement dans un sous-ensemble d'espèces mais dans toute la population, sans cohérence avec la taxonomie. Cette situation pourrait refléter l'existence, soit d'une évolution convergente, soit de flux des gènes entre espèces, soit d'une sélection balancée. De plus, des analyses de phylogénie des gènes *eIF4E* réalisées sur un échantillon interspécifique de Triticées montrent que l'évolution des gènes *eIF4E* du blé se distingue de l'évolution des gènes *eIF4E* de l'orge, suggérant que d'autres voies pourraient être utilisées chez le blé en réponse aux infections virales. Il est important de rappeler que chez les monocotylédones, l'implication d'un autre facteur du complexe d'initiation de la traduction dans la résistance récessive à un virus du genre *Sobemovirus* a été démontrée : des substitutions dans la séquence codante du gène *eIF(iso)4G* chez le riz sont responsables de la résistance au RYMV (Albar *et al.*, 2006). Ainsi, ce scénario pourrait être retrouvé chez d'autres monocotylédones, comme le blé où les gènes *eIF4E* ne présentent pas de polymorphisme caractéristique des gènes récessifs de résistance aux virus à ARN. Ces résultats mettent en avant l'importance d'une future exploration de la diversité des autres facteurs du complexe d'initiation de la traduction comme eIF4G.

#### ***V.3.2.6. Vers un modèle d'évolution convergente des gènes eIF4E de résistance***

Parmi les genres étudiés à ce jour (*Capsicum* spp., *Lycopersicon* spp., *Arabidopsis*, *Medicago* spp., *Vitis* spp. et *Triticum* spp.), de la sélection positive s'exerçant sur les gènes *eIF4E* n'est retrouvée que dans le cas du gène *eIF4E1* chez le piment et la vigne. Chez le piment, les analyses fonctionnelles montrent que les acides aminés sous sélection positive correspondent à ceux impliqués dans la résistance. Il apparaît clairement que cette trace de sélection n'est pas conservée entre les genres de la famille des Solanacées, genres pourtant tous infectés par les potyvirus et dont certaines possèdent des gènes récessifs de résistance aux potyvirus correspondant à eIF4E. Afin d'étendre ce résultat à d'autres familles végétales, ce type d'étude pourra être étendu à d'autres espèces comme par exemple le pois, où l'on retrouve une grande diversité allélique aux locus de résistance *eIF4E* en comparaison avec d'autres espèces de Légumineuses comme *Medicago* où l'on ne détecte pas de sélection positive.

Malgré cela, la conservation de la position des substitutions en acides aminés impliquées dans la résistance chez des espèces très éloignées - piment, laitue, pois, orge - suggère l'existence d'une évolution convergente des gènes *eIF4E* vers la résistance qui pourrait être conditionnée par une pression de sélection importante exercée par les potyvirus. Une évolution convergente (*i.e.*, lorsque 2 gènes de résistance ne partagent pas d'ancêtre commun et que la fonction de résistance a été acquise indépendamment chez 2 espèces) a été mise en évidence pour les gènes de résistance *Rpg1-b* et *RPM1* chez le soja et *A. thaliana* respectivement (Ashfield *et al.*, 2004). Les protéines R codées par ces gènes permettent la reconnaissance de la même protéine AvrB de *P. syringae* (Innes *et al.*, 1993). Les analyses phylogénétiques basées sur toutes les séquences NBS-LRR du génome d'*A. thaliana* montrent que *Rpg1-b* et *RPM1* appartiennent à 2 clades distincts. Dans chacun de ces clades, des séquences de monocotylédones sont retrouvées indiquant que ces 2 clades ont divergé avant la séparation monocotylédone-dicotylédone, il y a plus de 150 millions d'années. Même si *RPM1* a été maintenu sur une longue période de temps (10 millions d'années; Stahl *et al.*, 1999), cela n'a pas été suffisant pour qu'il soit maintenu chez le soja. Ces résultats suggèrent donc fortement qu'une l'évolution convergente plutôt que le maintien d'une spécificité ancestrale serait à l'origine de l'existence de ces 2 gènes AvrB-spécifiques.

#### **V.4. Diversité des gènes *eIF4E* et gestion des résistances aux potyvirus**

##### **V.4.1. Paramètres influençant l'efficacité et de la durabilité des résistances médiées par *eIF4E***

La substitution D109N, localisée dans la région II de la protéine *eIF4E1* du piment, ne suffit pas à elle seule à contrôler la résistance au TEV. En revanche, la protéine codée par l'allèle *pvr2*<sup>2</sup>, qui présente à la fois la substitution D109N et des substitutions dans la région I, contrôle la résistance au PVY et au TEV. Ainsi, le cumul de substitutions en acides aminés dans les régions I et II de la protéine *eIF4E1* apparaît nécessaire pour abolir l'interaction avec la VPg et induire la résistance au TEV. Une analyse à plus grande échelle, sur un plus grand nombre d'accessions naturelles et sur d'autres couples plantes-potyvirus devra être mise en place pour généraliser l'impact de l'effet cumulé des substitutions en acides aminés dans la protéine *eIF4E1* sur l'élargissement du spectre de résistance. Il serait par exemple intéressant

d'étudier le pois qui présente une fréquence élevée d'allèles récessifs *eIF4E* conférant la résistance à un large spectre de potyvirus (PSbMV, BYMV et CIYVV). A terme, ces études permettraient de générer du matériel végétal résistant à plusieurs potyvirus à partir de l'analyse de la diversité présente dans la variabilité naturelle et/ou la création de nouvelles mutations par mutagenèse dirigée.

Des études réalisées sur les gènes *R* de résistance ont permis de mettre en évidence l'impact de la position et du cumul de mutations générées de façon aléatoire sur le spectre de résistance (Farnham and Baulcombe, 2006). Une approche d'évolution artificielle du domaine LRR de la protéine de résistance Rx, domaine impliqué dans la reconnaissance des protéines de virulence des agents pathogènes, a permis de mettre en évidence l'impact du cumul de mutations obtenues aléatoirement sur l'élargissement du spectre de résistance. A l'origine, la protéine Rx confère la résistance contre seulement quelques souches de PVX. Les mutants artificiels obtenus sont résistants à une souche supplémentaire de PVX et au *Poplar mosaic virus* (PoMV, genre *Carlavirus*). Ainsi, l'évolution *in vitro* des gènes de résistance pourrait être exploitée pour l'amélioration de la résistance aux maladies chez les plantes cultivées.

La durabilité des résistances est aujourd'hui un des critères majeurs dans la sélection des variétés à caractère agronomique. Une résistance est dite durable "lorsqu'elle reste efficace (c'est-à-dire lorsqu'elle n'est pas contournée par des souches virulentes) dans une variété cultivée sur de grandes surfaces, pendant une longue période de temps et dans des conditions favorables à la maladie " (Johnson, 1979; Johnson, 1981).

Des études de la durabilité en condition artificielle des résistances conférées par les allèles *eIF4E1-pvr2<sup>1</sup>*, *eIF4E1-pvr2<sup>2</sup>* et *eIF4E1-pvr2<sup>3</sup>* ont été conduites sur le pathosystème piment-PVY (Ayme, 2005; Ayme *et al.*, 2007). Elles montrent que la résistance conférée par l'allèle *eIF4E1-pvr2<sup>3</sup>* peut être rapidement et efficacement contournée par la souche PVY-SON41p par 5 substitutions dans la zone centrale de la VPg. En revanche, l'allèle *eIF4E1-pvr2<sup>2</sup>* est plus durable que *eIF4E1-pvr2<sup>1</sup>* et *eIF4E1-pvr2<sup>3</sup>*. Ces résultats sont en accord avec ce qui est observé en conditions de culture : l'allèle de résistance *pvr2<sup>2</sup>*, utilisé à grande échelle depuis plus de 20 ans, est particulièrement durable. Les analyses de diversité du gène *eIF4E1* chez le piment montrent que les allèles *pvr2<sup>1</sup>* et *pvr2<sup>3</sup>* présentent 1 ou 2 substitutions en acides aminés dans la région I tandis que l'allèle *pvr2<sup>2</sup>* présente 3 substitutions en acides aminés dans les régions I et II. De même, l'allèle *pvr1* de *C. chinense* codant pour une protéine *eIF4E1* qui

présente des substitutions à la fois dans les régions I et II, confère, en condition naturelle, une résistance durable (Kang *et al.*, 2005). Ces observations permettent de proposer l'hypothèse d'une relation entre la durabilité et le nombre et la position (régions I et II) des substitutions en acides aminés.

Pour cela, il s'avère nécessaire d'étendre l'étude de la durabilité des résistances en conditions de laboratoire à l'ensemble des allèles de résistance identifiés à ce locus (*pvr2*<sup>4</sup> à *pvr2*<sup>9</sup>) chez *C. annuum* et *C. chinense*. Si le cumul des substitutions en acides aminés est lié à une augmentation de la durabilité de la résistance, l'allèle *pvr2*<sup>4</sup>, présentant une seule substitution en acide aminé, devrait conférer une résistance moins durable que les autres allèles cumulant au minimum 2 mutations. Si la localisation des substitutions en acides aminés à la fois dans les régions I et II est liée à l'augmentation de la durabilité, les allèles *pvr2*<sup>2</sup> et *pvr1* devraient conférer une résistance plus durable que tous les autres allèles ne possédant que des substitutions dans la région I. Pour que cette étude soit complète et afin d'identifier quelles sont les substitutions en acides aminés ou les combinaisons de substitutions les moins facilement contournées par les potyvirus, il serait intéressant d'intégrer à ce travail l'analyse de la durabilité d'allèles de résistance obtenus par mutagenèse.

La principale contrainte à l'utilisation des résistances génétiques réside dans la disponibilité, au sein de la variabilité naturelle des espèces cultivées, de gènes de résistance durable. Dans ce contexte, le TILLING (Targeting Induced Local Lesions IN Genomes, Slade and Knauf, 2005) constitue une des stratégies nouvellement adoptée dans le cadre de la lutte anti-virale pour générer artificiellement de nouveaux allèles de résistance. Le TILLING est basé sur la constitution de collections de mutants obtenus par mutagenèse chimique à l'EMS, puis sur la recherche systématique de mutations dans des gènes cibles par détection des mésappariements induits par la mutagenèse (Oleykowski *et al.*, 1998; Colbert *et al.*, 2001). Cette technique permet de générer de la variabilité dans des gènes cibles en évitant de recourir aux techniques de transgénèse (Slade *et al.*, 2005) et présente également l'avantage d'être automatisable à haut débit. Actuellement, cette stratégie a été adoptée au laboratoire pour la recherche d'allèles de résistance aux potyvirus chez la tomate (projet TRANSVIR Génoplante, en collaboration avec l'URGV, INRA-UMR1165, Evry). La recherche de mutations EMS est ciblée sur les 3 gènes *eIF4E* identifiés et sur les domaines fonctionnels des gènes *eIF4G* et *eIF(iso)4G*. La population mutagénisée est constituée de 5000 familles M2 de la tomate d'industrie *L. esculentum* M82. A ce jour, parmi les 5 mutations en acides aminés identifiées

dans la protéine eIF4E1, 1 mutation dans un site d'épissage entre les exons 3 et 4 a été montrée comme impliquée dans la résistance totale au PVY, 1 des 7 virus à ARN testé infectant la tomate (M. Nicolai *et al.*, communication personnelle). L'identification d'une lignée de tomate résistante vis-à-vis d'un potyvirus valide ainsi la stratégie TILLING dans le cadre de l'identification de nouveaux allèles de résistance aux virus à ARN.

#### **V.4.2. Un coût à la résistance contrôlée par les gènes *eIF4E* ?**

On parle de coût de la résistance quand la fitness des plantes résistantes est inférieure à la fitness des plantes sensibles en absence d'invasion par les populations d'agents pathogènes. Le coût de la résistance expliquerait le maintien à la fois des allèles de sensibilité et de résistance observé pour un grand nombre de gènes *R* de résistance, ainsi que la plus faible fréquence d'allèles de résistance dans les populations (Bergelson *et al.*, 2001). Ce coût de la résistance a été démontré dans le cas du gène *RPM1* d'*A. thaliana* qui confère la résistance à *P. syringae* (Tian *et al.*, 2003). Les analyses d'évolution moléculaire des régions non-codantes entourant le locus *RPM1* montrent que les allèles de sensibilité et de résistance coexistent depuis plus de 9 millions d'années (Stahl *et al.*, 1999). L'ancienneté de ce polymorphisme suggère fortement que l'allèle de sensibilité a été maintenu par la sélection naturelle à cause d'un coût de la résistance. Les autres hypothèses expliquant la coexistence de ces allèles ont été éliminées : le mode de reproduction hautement autogame d'*A. thaliana* exclut l'hypothèse de l'avantage aux hétérozygotes et aucune structuration géographique ne peut être reliée à la variation allélique observée. Des études en plein champs de comparaison de rendement en absence de l'agent pathogène entre des plantes transgéniques possédant le gène *RPM1* (*RPM1*<sup>+</sup>) et d'autres dépourvues de ce gène *RPM1* et de son promoteur (*RPM1*<sup>-</sup>) confirment que le gène *R* induit un coût de fitness : les plantes *RPM1*<sup>+</sup> produisent 9 % de graines en moins que les plantes *RPM1*<sup>-</sup>. Ainsi, la sélection qui a lieu au niveau de ce locus est balancée entre l'effet positif sur l'hôte de pouvoir se défendre en cas d'attaque de l'agent pathogène et l'effet négatif sur la fitness de l'hôte en absence de *P. syringae* (Tian *et al.*, 2003). Un autre exemple de coût de la résistance impliqué dans le maintien d'un polymorphisme balancé ancien concerne le gène *RPW8* de résistance au mildiou chez *A. thaliana*. Là encore, une baisse de la production des graines de près de 6 % est observée entre des populations d'*A.*

*thaliana* possédant le gène RPW8 et des populations dépourvues de ce gène (Xiao *et al.*, 2004).

A ce jour, aucune mesure de fitness n'a été réalisée pour estimer le coût de la résistance conférée par les gènes *eIF4E*, que ce soit pour les variants naturels ou les mutants artificiels. Dans le cas du gène récessif de résistance *eIF4E1* du piment, un grand nombre de variants alléliques est observé. Cependant, l'hypothèse de sélection balancée pour expliquer le maintien de cette diversité allélique a été rejetée car aucun polymorphisme synonyme n'a été identifié au locus. La très forte fréquence d'allèles de résistance dans la diversité naturelle de *Capsicum* suggère que la résistance conférée par *eIF4E* n'a pas de coût. L'explication d'un coût de la résistance lié au maintien de l'allèle de sensibilité n'est donc pas aussi évidente que dans les cas énoncés précédemment. Des expériences pour mesurer la fitness des *Capsicum* mutés pour l'un ou l'autre des gènes *eIF4E*, que ce soit pour les plantes possédant les allèles *eIF4E1-pvr2* impliqués dans la résistance aux PVY et TEV ou l'allèle *eIF(iso)4E-pvr6* impliqué dans la résistance au PVMV, permettraient de déterminer si la résistance a un coût. Des expériences avec des mutants d'*A. thaliana* knockout pour les gènes *eIF4E* indiqueraient le coût de la perte d'un des membres de la famille mutigénique pour la fitness de la plante.

Un dispositif expérimental a été mis au point pour estimer le coût de la résistance aux herbicides chez *A. thaliana* (Roux *et al.*, 2006). De nombreuses expériences ne prennent en compte que le poids des graines pour estimer le coût de fitness d'un caractère donné, ce qui induit une estimation biaisée qui ne prend pas en compte tous les paramètres de fitness : le poids des graines ne reflète par exemple pas leur qualité et donc la contribution de chaque plante pour les générations futures dans la population. Pour s'astreindre de ce biais, la méthode utilisée par Roux *et al.* (2006) consiste à mesurer le changement de la fréquence des allèles de résistance en fonction d'un gradient géographique (appelé cline) de plusieurs populations localisées le long d'un environnement variable *i.e.*, entre des zones traitées et non traitées dans le cas de la résistance aux pesticides (Lenormand *et al.*, 1999; Lenormand and Raymond, 2000). L'étude d'un cline permet de distinguer les forces de sélection qui tendent à éliminer les génotypes et les forces de migration qui tendent à homogénéiser la fréquence de l'allèle de résistance le long du cline. Dans leurs expériences, la migration est contrôlée, permettant ainsi de déduire directement le coût de fitness (défini comme correspondant au coefficient de sélection en absence de traitement herbicide) de la modification des fréquences alléliques le long du cline. Ces profils de microévolution ont été analysés sur 3 à 4

générations pour 2 allèles distincts de résistance aux herbicides. Jusqu'à présent, ce type de méthode de mesure de fitness n'avait que rarement été utilisé car cela nécessite le génotypage d'un très grand nombre de génotypes, 10 800 dans le cas des expériences de Roux *et al.* (2006). Le développement actuel de techniques de génotypage à haut débit rend plus facile ce genre d'étude à grande échelle. Ce type d'étude pour estimer le coût de la résistance aux potyvirus contrôlée par les gènes *eIF4E* pourrait être envisagé, avec un cline zones infectées par le potyvirus-zones non infectées. Cependant, un tel dispositif est très difficile à mettre en place et des mesures de fitness plus "simples" pourraient être réalisées afin d'avoir une estimation préliminaire du coût de la résistance contrôlée par les gènes *eIF4E*, avec par exemple des essais conduits en serre pour mesurer des caractères phénotypiques (nombre de graines, nombre de siliques, taille des siliques, taille de la plante...) chez des mutants d'*A. thaliana* knockout pour chaque gène *eIF4E*. Ce type d'expérience pourrait également être mis en place pour des plantes de piment présentant les allèles récessifs de résistance détectés dans la variabilité naturelle.

# **R**éférences bibliographiques

- Albar, L., Bangratz-Reyser, M., Hebrard, E., Ndjiondjop, M.N., Jones, M. and Ghesquiere, A. (2006) Mutations in the eIF(iso)4G translation initiation factor confer high resistance of rice to *Rice yellow mottle virus*. *Plant J.*, **47**, 417-426.
- Altmann, M., Sonenberg, N. and Trachsel, H. (1989) Translation in *Saccharomyces cerevisiae*: initiation factor 4E-dependent cell-free system. *Mol. Cell Biol.*, **9**, 4467-4472.
- Altmann, M. and Trachsel, H. (1997) Translation initiation factor-dependent extracts from yeast *Saccharomyces cerevisiae*. *Methods*, **11**, 343-352.
- Amiri, A., Keiper, B.D., Kawasaki, I., Fan, Y., Kohara, Y., Rhoads, R.E. and Strome, S. (2001) An isoform of eIF4E is a component of germ granules and is required for spermatogenesis in *C. elegans*. *Development*, **128**, 3899-3912.
- Aranda, M.A., Escaler, M., Wang, D. and Maule, A.J. (1996) Induction of HSP70 and polyubiquitin expression associated with plant virus replication. *Proc. Natl. Acad. Sci U S A*, **93**, 15289-15293.
- Ariyaratne, I., Hobbs, H.A., Valverde, R.A., Black, L.L. and Dufresne, P.J. (1996) Resistance of *Capsicum* spp. genotypes to *Tobacco etch potyvirus* isolates from Western hemisphere. *Plant disease*, **80**, 1257-1261.
- Arteaga, M. and Ortega, R.G. (1986) Biological characterization of PVY as isolated from pepper in Spain. In O.R.G. (ed.), *Vlth Eucapria Meeting on Genetics and Breeding on Capsicum and Eggplant*, Zaragoza (Spain), pp. 183-188.
- Ashfield, T., Ong, L.E., Nobuta, K., Schneider, C.M. and Innes, R.W. (2004) Convergent evolution of disease resistance gene specificity in two flowering plant families. *Plant Cell*, **16**, 309-318.
- Ayme, V. (2005) Mécanismes de contournement des résistances et évaluation a priori de leur durabilité dans l'interaction piment (*Capsicum annuum*) - Virus Y de la pomme de terre (PVY). Rapport de thèse, Université Montpellier II, Sciences et techniques du Languedoc, p. 130.
- Ayme, V., Souche, S., Caranta, C., Jacquemond, M., Chadoeuf, J., Palloix, A. and Moury, B. (2006) Different mutations in the genome-linked protein VPg of *Potato virus Y* confer virulence on the pvr2<sup>3</sup> resistance in pepper. *Mol Plant Microbe Interact*, **19**, 557-563.
- Bakker, E.G., Toomajian, C., Kreitman, M. and Bergelson, J. (2006) A genome-wide survey of *R* gene polymorphisms in *Arabidopsis*. *Plant Cell*, **18**, 1803-1818.
- Bamshad, M. and Wooding, S.P. (2003) Signatures of natural selection in the human genome. *Nat Rev Genet*, **4**, 99-111.
- Baulcombe, D.C., Chapman, S. and Santa Cruz, S. (1995) Jellyfish green fluorescent protein as a reporter for virus infections. *Plant J*, **7**, 1045-1053.
- Baunoch, D.A., Das, P., Browning, M.E. and Hari, V. (1991) A temporal study of the expression of the capsid, cytoplasmic inclusion and nuclear inclusion proteins of *Tobacco etch potyvirus* in infected plants. *J. Gen. Virol.*, **72 (Pt 3)**, 487-492.
- Beauchemin, C., Boutet, N. and Laliberte, J.F. (2006) Visualisation of the interaction between the precursors of the viral protein linked to the genome (VPg) of *Turnip mosaic virus* and the translation eukaryotic initiation factor iso4E *in planta*. *J Virol*.
- Bergelson, J., Kreitman, M., Stahl, E.A. and Tian, D. (2001) Evolutionary dynamics of plant *R*-genes. *Science*, **292**, 2281-2285.
- Berger, S., Mitchell-Olds, T. and Stotz, H.U. (2002) Local and differential control of vegetative storage protein expression in response to herbivore damage in *Arabidopsis thaliana*. *Physiol Plant*, **114**, 85-91.

- Bi, X., Ren, J. and Goss, D.J. (2000) Wheat germ translation initiation factor eIF4B affects eIF4A and eIFiso4F helicase activity by increasing the ATP binding affinity of eIF4A. *Biochemistry*, **39**, 5758-5765.
- Boiteux, L.S., Cupertino, F.P., Silva, C., Dusi, A.N., Monte-Neshich, D.C., Vand der Vlugt, R.A.A. and Fonseca, M.E.N. (1996) Resistance to *Potato virus Y* (pathotype 1-2) in *Capsicum annuum* and *Capsicum chinense* is controlled by two independent major genes. *Euphytica*, **87**, 53-58.
- Bokros, C.L., Hugdahl, J.D., Kim, H.H., Hanesworth, V.R., van Heerden, A., Browning, K.S. and Morejohn, L.C. (1995) Function of the p86 subunit of eukaryotic initiation factor (iso)4F as a microtubule-associated protein in plant cells. *Proc. Natl. Acad. Sci. U S A*, **92**, 7120-7124.
- Borgstrom, B. and Johansen, I.E. (2001) Mutations in *Pea seedborne mosaic virus* genome-linked protein VPg alter pathotype-specific virulence in *Pisum sativum*. *Mol. Plant Microbe Interact.*, **14**, 707-714.
- Botella, M.A., Parker, J.E., Frost, L.N., Bittner-Eddy, P.D., Beynon, J.L., Daniels, M.J., Holub, E.B. and Jones, J.D. (1998) Three genes of the *Arabidopsis RPP1* complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. *Plant Cell*, **10**, 1847-1860.
- Bracha-Drori, K., Shichrur, K., Katz, A., Oliva, M., Angelovici, R., Yalovsky, S. and Ohad, N. (2004) Detection of protein-protein interactions in plants using bimolecular fluorescence complementation. *Plant J.*, **40**, 419-427.
- Browning, K.S. (1996) The plant translational apparatus. *Plant Mol Biol*, **32**, 107-144.
- Browning, K.S. (2004) Plant translation initiation factors: it is not easy to be green. *Biochem Soc Trans*, **32**, 589-591.
- Browning, K.S., Lax, S.R. and Ravel, J.M. (1987) Identification of two messenger RNA cap binding proteins in wheat germ. Evidence that the 28-kDa subunit of eIF-4B and the 26-kDa subunit of eIF-4F are antigenically distinct polypeptides. *J. Biol. Chem.*, **262**, 11228-11232.
- Browning, K.S., Webster, C., Roberts, J.K. and Ravel, J.M. (1992) Identification of an isozyme form of protein synthesis initiation factor 4F in plants. *J. Biol. Chem.*, **267**, 10096-10100.
- Brun-Rasmussen, M., Moller, I., Tulinius, G., Hansen, J., Lund, O. and Johansen, E. (2007) The same allele of translation initiation factor 4E mediates resistance against two *Potyvirus* spp. in *Pisum sativum*. *Mol Plant Microbe Interact.*, **20**, 1075-1082.
- Caicedo, A.L., Schaal, B.A. and Kunkel, B.N. (1999) Diversity and molecular evolution of the *RPS2* resistance gene in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A*, **96**, 302-306.
- Caranta, C., Lefebvre, V. and Palloix, A. (1997) Polygenic resistance of pepper to potyviruses consists of a combination of isolate-specific and broad-spectrum quantitative trait loci. *Mol. Plant Microbe Interact.*, **10**, 872-878.
- Caranta, C. and Palloix, A. (1996) Both common and specific genetic factors are involved in polygenic resistance of pepper to several potyviruses. *Theor. Appl. Genet.*, **92**, 15-20.
- Caranta, C., Palloix, A., Gebre-Selassie, K., Lefebvre, V., Moury, B. and Daubeze, A.M. (1996) A complementation of two genes originating from susceptible *Capsicum annuum* lines confers a new and complete resistance to *Pepper veinal mottle virus*. *Phytopathol.*, **86**, 739-743.
- Carberry, S.E., Darzynkiewicz, E. and Goss, D.J. (1991) A comparison of the binding of methylated cap analogues to wheat germ protein synthesis initiation factors 4F and (iso)4F. *Biochemistry*, **30**, 1624-1627.

- Carrington, J. and Freed, D. (1990) Cap-independent enhancement of translation by a plant potyvirus 5' nontranslated region. *J. Virol.*, **64**, 1590-1597.
- Carrington, J., Kaschau, K., Mahajan, S. and Schaad, M. (1996) Cell-to-cell and long distance transport of viruses in plants. *Plant Cell*, **8**, 1669-1681.
- Charlesworth, D., Charlesworth, B. and McVean, G.A. (2001) Genome sequences and evolutionary biology, a two-way interaction. *Trends Ecol Evol*, **16**, 235-242.
- Charron, C. (2004) Rôle du facteur d'initiation de la traduction eIF4E dans la résistance aux potyvirus chez les Solanacées : diversité allélique et études d'expression. Rapport de DEA, Université Aix-Marseille II, Faculté des Sciences de Luminy, p. 19.
- Chiemsombat, P., Sae-Ung, N., Attathom, S., Patarapuwadol, S. and Siriwong, P. (1998) Molecular taxonomy of a new potyvirus isolated from chilli pepper in Thailand. *Arch Virol*, **143**, 1855-1863.
- Cho, P.F., Poulin, F., Cho-Park, Y.A., Cho-Park, I.B., Chicoine, J.D., Lasko, P. and Sonenberg, N. (2005) A new paradigm for translational control: inhibition via 5'-3' mRNA tethering by Bicoid and the eIF4E cognate 4EHP. *Cell*, **121**, 411-423.
- Clark, M. and Adams, A. (1977) Characteristics of the microplate method of enzym-linked immunosorbent assay for the detection of plant virus. *J. Gen. Virol.*, **34**, 475-483.
- Clemens, M.J. and Bommer, U.A. (1999) Translational control: the cancer connection. *Int J Biochem Cell Biol*, **31**, 1-23.
- Colbert, T., Till, B.J., Tompa, R., Reynolds, S., Steine, M.N., Yeung, A.T., McCallum, C.M., Comai, L. and Henikoff, S. (2001) High-throughput screening for induced point mutations. *Plant Physiol.*, **126**, 480-484.
- Combe, J.P., Petracek, M.E., van Eldik, G., Meulewaeter, F. and Twell, D. (2005) Translation initiation factors eIF4E and eIFiso4E are required for polysome formation and regulate plant growth in tobacco. *Plant Mol. Biol.*, **57**, 749-760.
- Couch, B.C., Spangler, R., Ramos, C. and May, G. (2006) Pervasive purifying selection characterizes the evolution of *I2* homologs. *Mol Plant Microbe Interact*, **19**, 288-303.
- Culjkovic, B., Topisirovic, I. and Borden, K.L. (2007) Controlling gene expression through RNA regulons: the role of the eukaryotic translation initiation factor eIF4E. *Cell Cycle*, **6**, 65-69.
- Culjkovic, B., Topisirovic, I., Skrabanek, L., Ruiz-Gutierrez, M. and Borden, K.L. (2005) eIF4E promotes nuclear export of cyclin D1 mRNAs via an element in the 3'UTR. *J. Cell Biol.*, **169**, 245-256.
- Dangl, J.L. and Jones, J.D. (2001) Plant pathogens and integrated defence responses to infection. *Nature*, **411**, 826-833.
- Dawkins, R. and Krebs, J.R. (1979) Arms races between and within species. *Proc R Soc Lond B Biol Sci*, **205**, 489-511.
- de Meaux, J. and Mitchell-Olds, T. (2003) Evolution of plant resistance at the molecular level: ecological context of species interactions. *Heredity*, **91**, 345-352.
- Decroocq, V., Sicard, O., Alamillo, J.M., Lansac, M., Eyquard, J.P., Garcia, J.A., Candresse, T., Le Gall, O. and Revers, F. (2006) Multiple resistance traits control *Plum pox virus* infection in *Arabidopsis thaliana*. *Mol Plant Microbe Interact*, **19**, 541-549.
- Delye, C., Straub, C., Michel, S. and Le Corre, V. (2004) Nucleotide variability at the acetyl coenzyme A carboxylase gene and the signature of herbicide selection in the grass weed *Alopecurus myosuroides* (Huds.). *Mol Biol Evol*, **21**, 884-892.
- Deom, C.M. and He, X.Z. (1997) Second-site reversion of a dysfunctional mutation in a conserved region of the *Tobacco mosaic tobamovirus* movement protein. *Virology*, **232**, 13-18.

- Deulvot, C. (2002) La résistance de l'aubergine aux virus ToMV et PVY : déterminisme génétique et cartographie comparée chez les Solanacées. Rapport de DESS, Université Pierre et Marie Curie, Paris VI, p. 52.
- Diaz, J.A., Nieto, C., Moriones, E., Truniger, V. and Aranda, M.A. (2004) Molecular characterization of a *Melon necrotic spot virus* strain that overcomes the resistance in melon and nonhost plants. *Mol. Plant Microbe Interact.*, **17**, 668-675.
- Diaz-Pendon, J., Truninger, V., Nieto, C., Garcia-Mas, J., Bendahmane, A. and Aranda, M. (2004) Advances in understanding recessive resistance to plant viruses. *Mol. Plant Pathol.*, **5**, 223-233.
- Dijan-Caporalino, C., Lefebvre, V., Sage-Palloix and Palloix, A. (2006) Capsicum. In *Genetic resources, chromosome engineering, and crop improvement*.
- Ding, J., Cheng, H., Jin, X., Araki, H., Yang, Y. and Tian, D. (2007) Contrasting patterns of evolution between allelic groups at a single locus in *Arabidopsis*. *Genetica*, **129**, 235-242.
- Dinkova, T.D., Keiper, B.D., Korneeva, N.L., Aamodt, E.J. and Rhoads, R.E. (2005) Translation of a small subset of *Caenorhabditis elegans* mRNAs is dependent on a specific eukaryotic translation initiation factor 4E isoform. *Mol. Cell Biol.*, **25**, 100-113.
- Dixon, M.S., Hatzixanthis, K., Jones, D.A., Harrison, K. and Jones, J.D. (1998) The tomato *Cf-5* disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. *Plant Cell*, **10**, 1915-1925.
- Dodds, P.N., Lawrence, G.J., Catanzariti, A.M., Teh, T., Wang, C.I., Ayliffe, M.A., Kobe, B. and Ellis, J.G. (2006) Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proc Natl Acad Sci U S A*, **103**, 8888-8893.
- Dodds, P.N., Lawrence, G.J., Pryor, A.J. and Ellis, J.G. (2000) In Dickinson, M.a.B., J. (ed.), *Molecular Plant Pathology*. Sheffield Academic, Sheffield, U.K., pp. 88-107.
- Drake, J.W. and Holland, J.J. (1999) Mutation rates among RNA viruses. *Proc. Natl. Acad. Sci. U S A*, **96**, 13910-13913.
- Duprat, A. (2002) Contribution à l'étude des facteurs d'initiation de la traduction eIF4E chez *Arabidopsis thaliana*. Rapport de thèse, Université de la Méditerranée, p. 110.
- Duprat, A., Caranta, C., Revers, F., Menand, B., Browning, K.S. and Robaglia, C. (2002) The *Arabidopsis* eukaryotic initiation factor (iso)4E is dispensable for plant growth but required for susceptibility to potyviruses. *Plant J.*, **32**, 927-934.
- Ellis, J.G., Dodds, P.N. and Lawrence, G.J. (2007) Flax rust resistance gene specificity is based on direct resistance-avirulence protein interactions. *Annu Rev Phytopathol.*
- Ellis, J.G., Lawrence, G.J., Luck, J.E. and Dodds, P.N. (1999) Identification of regions in alleles of the flax rust resistance gene *L* that determine differences in gene-for-gene specificity. *Plant Cell*, **11**, 495-506.
- Farnham, G. and Baulcombe, D.C. (2006) Artificial evolution extends the spectrum of viruses that are targeted by a disease-resistance gene from potato. *Proc. Natl. Acad. Sci U S A*, **103**, 18828-18833.
- Fay, J.C. and Wu, C.I. (2000) Hitchhiking under positive Darwinian selection. *Genetics*, **155**, 1405-1413.
- Ferraiuolo, M.A., Basak, S., Dostie, J., Murray, E.L., Schoenberg, D.R. and Sonenberg, N. (2005) A role for the eIF4E-binding protein 4E-T in P-body formation and mRNA decay. *J Cell Biol*, **170**, 913-924.
- Flor, H.H. (1971) Current status of the gene for gene concept. *Ann. Rev. Phytopathol.*, **9**, 275-296.

- Fraser, R. (1990) The genetics of resistance to plant viruses. *Annu. Rev. Phytopathol.*, **28**, 179-200.
- French, R. and Stenger, D.C. (2003) Evolution of *Wheat streak mosaic virus*: dynamics of population growth within plants may explain limited variation. *Annu Rev Phytopathol*, **41**, 199-214.
- Fribourg, C. (1979) Host plant reactions, some properties, and serology of *Peru tomato virus*. *Phytopathology*, **69**, 441-445.
- Fu, Y.X. and Li, W.H. (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693-709.
- Fulton, T.M., Chunwongse, J. and Tanksley, S. (1995) Microprep protocol for extraction of DNA from tomato and other Herbaceous plants. *Plant Molecular Biology Reporter*, **13**, 207-209.
- Gallie, D. (2002) Protein-protein interactions during translation. *Plant Mol. Biol.*, **50**, 949-970.
- Gallie, D. and Browning, K. (2001) eIF4G functionally differs from eIF(iso)4G in promoting internal initiation, cap-independent translation and translation of structured mRNAs. *J. Biol. Chem.*, **276**, 36951-36960.
- Gallie, D., Le, H., Tanguay, R. and Browning, K. (1998) Translation initiation factors are differentially regulated in cereals during development and following heat shock. *Plant J.*, **14**, 715-722.
- Gao, Z., Evers, S., Thomas, C., Ellis, N. and Maule, A. (2004a) Identification of markers tightly linked to *sbm* recessive genes for resistance to *Pea seed-borne mosaic virus*. *Theor. Appl. Genet.*, **109**, 488-494.
- Gao, Z., Johansen, E., Evers, S., Thomas, C., Ellis, T. and Maule, A. (2004b) The potyvirus recessive resistance gene, *sbm1*, identifies a novel role for translation initiation factor eIF4E in cell-to-cell trafficking. *Plant J.*, **40**, 376-385.
- Gébré Sélassié, K., Marchoux, G., Delecolle, B. and Pochard, E. (1985) Variabilité naturelle des souches de virus Y de la pomme de terre dans les cultures de piment du Sud-Est de la France. Caractérisation et classification en pathotypes. *Agronomie (Paris)*, **5**, 621-630.
- Glais, L., Tribodet, M. and Kerlan, C. (2002) Genomic variability in *Potato potyvirus Y* (PVY): evidence that PVY(N)W and PVY(NTN) variants are single to multiple recombinants between PVY(O) and PVY(N) isolates. *Arch Virol*, **147**, 363-378.
- Goldman, N. and Yang, Z. (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol*, **11**, 725-736.
- Gordon, G.W., Berry, G., Liang, X.H., Levine, B. and Herman, B. (1998) Quantitative fluorescence resonance energy transfer measurements using fluorescence microscopy. *Biophys. J.*, **74**, 2702-2713.
- Green, S.K. and Kim, J.S. (1991) Characteristics and control of viruses infecting peppers: a literature review. *Asian Vegetable and Development Center. Technical Bulletin*, Vol. 18, p. 60.
- Greenleaf, W.H. (1986) Pepper breeding. In Basset, M.J. (ed.), *Breeding vegetable crops*. AVI Publishing Co. Inc., Westport Connecticut, pp. 67-134.
- Gross, J.D., Moerke, N.J., von der Haar, T., Lugovskoy, A.A., Sachs, A.B., McCarthy, J.E. and Wagner, G. (2003) Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. *Cell*, **115**, 739-750.
- Grzela, R., Strokovska, L., Andrieu, J.P., Dublet, B., Zagorski, W. and Chroboczek, J. (2006) Potyvirus terminal protein VPg, effector of host eukaryotic initiation factor eIF4E. *Biochimie*, **88**, 887-896.

- Guerini, M. and Murphy, J. (1999) Resistance of *Capsicum annuum* 'Avelar' to pepper mottle potyvirus and alleviation of this resistance by co-infection with *Cucumber mosaic cucumovirus* are associated with virus movement. *J. Gen. Virol.*, **80**, 2785-2792.
- Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucl. Acids Symp. Ser.*, **41**, 95-98.
- Hamalainen, J.H., Kekarainen, T., Gebhardt, C., Watanabe, K.N. and Valkonen, J.P. (2000) Recessive and dominant genes interfere with the vascular transport of *Potato virus A* in diploid potatoes. *Mol. Plant Microbe Interact.*, **13**, 402-412.
- Havelda, Z. and Maule, A.J. (2000) Complex spatial responses to *Cucumber mosaic virus* infection in susceptible *Cucurbita pepo* cotyledons. *Plant Cell*, **12**, 1975-1986.
- Hebrard, E. (2007) Les VPgs, des protéines toutes désordonnées. 11<sup>èmes</sup> Rencontres de Virologie Végétale, du 28 janvier au 1<sup>er</sup> février 2007, Aussois, France.
- Hebrard, E., Pinel-Galzi, A., Bersoult, A., Sire, C. and Fargette, D. (2006) Emergence of a resistance-breaking isolate of *Rice yellow mottle virus* during serial inoculations is due to a single substitution in the genome-linked viral protein VPg. *J Gen Virol*, **87**, 1369-1373.
- Hernandez, G., Altmann, M., Sierra, J.M., Urlaub, H., Diez Del Corral, R., Schwartz, P. and Rivera-Pomar, R. (2005) Functional analysis of seven genes encoding eight translation initiation factor 4E (eIF4E) isoforms in *Drosophila*. *Mech. Dev.*, **122**, 529-543.
- Herold, J. and Andino, R. (2001) Poliovirus RNA replication requires genome circularization through a protein-protein bridge. *Mol. Cell*, **7**, 581-591.
- Hiscox, J.A. (2003) The interaction of animal cytoplasmic RNA viruses with the nucleus to facilitate replication. *Virus Res.*, **95**, 13-22.
- Holub, E.B. (2001) The arms race is ancient history in *Arabidopsis*, the wildflower. *Nat Rev Genet*, **2**, 516-527.
- Hong, Y., Levay, K., Murphy, J.F., Klein, P.G., Shaw, J.G. and Hunt, A.G. (1995) A potyvirus polymerase interacts with the viral coat protein and VPg in yeast cells. *Virology*, **214**, 159-166.
- Hudson, R.R. and Kaplan, N.L. (1988) The coalescent process in models with selection and recombination. *Genetics*, **120**, 831-840.
- Hudson, R.R., Kreitman, M. and Aguade, M. (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116**, 153-159.
- Hughes, J.M., Ptushkina, M., Karim, M.M., Koloteva, N., von der Haar, T. and McCarthy, J.E. (1999) Translational repression by human 4E-BP1 in yeast specifically requires human eIF4E as target. *J Biol Chem*, **274**, 3261-3264.
- Iborra, F., Jackson, D. and Cook, P. (2001) Coupled transcription and translation within nuclei of mammalian cells. *Science*, **293**, 1139-1142.
- Innes, R.W., Bisgrove, S.R., Smith, N.M., Bent, A.F., Staskawicz, B.J. and Liu, Y.C. (1993) Identification of a disease resistance locus in *Arabidopsis* that is functionally homologous to the *RPG1* locus of soybean. *Plant J.*, **4**, 813-820.
- Jensen, R., McLeod, M., Eshbaugh, W. and Guttman, S. (1979) Numerical taxonomic analyses of allozymic variation in *Capsicum* (Solanaceae). *Taxon*, **28**, 315-327.
- Johansen, I.E., Lund, O.S., Hjulsgaard, C.K. and Laursen, J. (2001) Recessive resistance in *Pisum sativum* and potyvirus pathotype resolved in a gene-for-cistron correspondence between host and virus. *J. Virol.*, **75**, 6609-6614.
- Johnson, R. (1979) The concept of durable resistance. *Phytopathol.*, **69**, 198-199.
- Johnson, R. (1981) Durable resistance: definition of, genetic control and attainment in plant breeding. *Phytopathol.*, **71**, 567-568.
- Jones, J.D. (2001) Putting knowledge of plant disease resistance genes to work. *Curr Opin Plant Biol*, **4**, 281-287.

- Jones, J.D. and Dangl, J.L. (2006) The plant immune system. *Nature*, **444**, 323-329.
- Joshi, B., Lee, K., Maeder, D.L. and Jagus, R. (2005) Phylogenetic analysis of eIF4E-family members. *BMC Evol. Biol.*, **5**, 48.
- Kang, B.C., Yeam, I., Frantz, J.D., Murphy, J.F. and Jahn, M.M. (2005a) The *pvr1* locus in *Capsicum* encodes a translation initiation factor eIF4E that interacts with *Tobacco etch virus* VPg. *Plant J.*, **42**, 392-405.
- Kang, B.C., Yeam, I. and Jahn, M.M. (2005b) Genetics of plant virus resistance. *Annu Rev Phytopathol*, **43**, 581-621.
- Kanyuka, K., McGrann, G., Alhudaib, K., Hariri, D. and Adams, M.J. (2004) Biological and sequence analysis of a novel European isolate of Barley mild mosaic virus that overcomes the barley *rym5* resistance gene. *Arch Virol*, **149**, 1469-1480.
- Keiper, B., Lamphear, B., Deshpande, A., Jankowska-Anyszka, M., Aamodt, E., Blumenthal, T. and Rhoads, R. (2000) Functional characterization of five eIF4E isoforms in *Caenorhabditis elegans*. *J. Biol. Chem.*, **275**, 10590-10596.
- Kimura, M. (1983) Rare variant alleles in the light of the neutral theory. *Mol Biol Evol*, **1**, 84-93.
- Kozak, M. (1989) The scanning model for translation: an update. *J. Cell Biol.*, **108**, 229-241.
- Kühne, T., Shi, N., Proeseler, G., Adams, M.J. and Kanyuka, K. (2003) The ability of a bymovirus to overcome the *rym4*-mediated resistance in barley correlates with a codon change in the VPg coding region on RNA1. *J. Gen. Virol.*, **84**, 2853-2859.
- Kumar, S., Tamura, K. and Nei, M. (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform*, **5**, 150-163.
- Kunkel, B.N. (1996) A useful weed put to work: genetic analysis of disease resistance in *Arabidopsis thaliana*. *Trends Genet*, **12**, 63-69.
- Kushner, D., Lindenbach, B., Grdzlishvili, V., Noueiry, A., Paul, S. and Ahlquist, P. (2003) Systematic, genome-wide identification of host genes affecting replication of a positive-strand RNA virus. *Proc. Natl. Acad. Sci. U S A*, **100**, 15764-15769.
- Kyle, M.M. and Palloix, A. (1997) Proposed revision of nomenclature for potyvirus resistance genes in *Capsicum*. *Euphytica*, **97**, 183-188.
- Labay, A. (2007) Toward the characterisation of the molecular basis for the specificity in the utilization of eIF4E proteins by potyviruses. Rapport de M2 Biologie, géosciences, agronomie et environnement, Université Aix-Marseille II, Faculté des Sciences de Luminy, p. 20.
- Lai, H.K. and Borden, K.L. (2000) The promyelocytic leukemia (PML) protein suppresses cyclin D1 protein production by altering the nuclear cytoplasmic distribution of cyclin D1 mRNA. *Oncogene*, **19**, 1623-1634.
- Lall, S., Friedman, C.C., Jankowska-Anyszka, M., Stepinski, J., Darzynkiewicz, E. and Davis, R.E. (2004) Contribution of trans-splicing, 5' -leader length, cap-poly(A) synergism, and initiation factors to nematode translation in an embryo cell-free system. *J. Biol. Chem.*, **279**, 45573-45585.
- Lavoie, C., Lachance, P., Sonenberg, N. and Lasko, P. (1996) Alternatively spliced transcripts from the *Drosophila eIF4E* gene produce two different cap-binding proteins. *J. Biol. Chem.*, **271**, 16393-16398.
- Lax, S., Fritz, W., Browning, K. and Ravel, J. (1985) Isolation and characterization of factors from wheat germ that exhibit eukaryotic initiation factor 4B activity and overcome 7-methylguanosine 5'-triphosphate inhibition of polypeptide synthesis. *Proc. Natl. Acad. Sci. U S A*, **82**, 330-333.
- Le, H., Tanguay, R., Balasta, M., Wei, C., Browning, K., Metz, A., Goss, D. and Gallie, D. (1997) Translation initiation factors eIF-iso4G and eIF-4B interact with the poly(A)-

- binding protein and increase its RNA binding activity. *J. Biol. Chem.*, **272**, 16247-16255.
- Lellis, A., Kasschau, K., Whitham, S. and Carrington, J. (2002) Loss-of-susceptibility mutants of *Arabidopsis thaliana* reveal an essential role for eIF(iso)4E during potyvirus infection. *Curr. Biol.*, **12**, 1046-1051.
- Lenormand, T., Bourguet, D., Guillemaud, T. and Raymond, M. (1999) Tracking the evolution of insecticide resistance in the mosquito *Culex pipiens*. *Nature*, **400**, 861-864.
- Lenormand, T. and Raymond, M. (2000) Analysis of Clines with Variable Selection and Variable Migration. *Am. Nat.*, **155**, 70-82.
- Leonard, S., Chisholm, J., Laliberte, J.F. and Sanfacon, H. (2002) Interaction in vitro between the proteinase of *Tomato ringspot virus* (genus *Nepovirus*) and the eukaryotic translation initiation factor iso4E from *Arabidopsis thaliana*. *J. Gen. Virol.*, **83**, 2085-2089.
- Leonard, S., Plante, D., Wittmann, S., Daigneault, N., Fortin, M.G. and Laliberte, J.F. (2000) Complex formation between potyvirus VPg and translation eukaryotic initiation factor 4E correlates with virus infectivity. *J Virol*, **74**, 7730-7737.
- Leonard, S., Viel, C., Beauchemin, C., Daigneault, N., Fortin, M.G. and Laliberte, J.F. (2004) Interaction of VPg-Pro of *Turnip mosaic virus* with the translation initiation factor 4E and the poly(A)-binding protein *in planta*. *J Gen Virol*, **85**, 1055-1063.
- Li, J., Hsia, A.P. and Schnable, P.S. (2007) Recent advances in plant recombination. *Curr Opin Plant Biol*, **10**, 131-135.
- Li, W.H., Wu, C.I. and Luo, C.C. (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol Biol Evol*, **2**, 150-174.
- Luck, J.E., Lawrence, G.J., Dodds, P.N., Shepherd, K.W. and Ellis, J.G. (2000) Regions outside of the leucine-rich repeats of flax rust resistance proteins play a role in specificity determination. *Plant Cell*, **12**, 1367-1377.
- Luo, Y. and Goss, D. (2001) Homeostasis in mRNA initiation: wheat germ poly(A)-binding protein lowers the activation energy barrier to initiation complex formation. *J. Biol. Chem.*, **276**, 43083-43086.
- Malpica, J.M., Fraile, A., Moreno, I., Obies, C.I., Drake, J.W. and Garcia-Arenal, F. (2002) The rate and character of spontaneous mutation in an RNA virus. *Genetics*, **162**, 1505-1511.
- Marcotrigiano, J., Gringras, A., Sonenberg, N. and Burley, S. (1997) Co-crystal structure of the messenger RNA 5' cap binding protein (eIF4E) bound to 7-methyl-GDP. *Cell*, **89**, 951-961.
- Matsuo, H., Li, H., McGuire, A.M., Fletcher, C.M., Gringras, A.C., Sonenberg, N. and Wagner, G. (1997) Structure of translation factor eIF4E bound to m7GDP and interaction with 4E-binding protein. *Nat. Struct. Biol.*, **4**, 717-724.
- Maule, A., Caranta, C. and Boulton, M. (2007) Sources of natural resistance to plant viruses: status and prospects. *Molecular Plant Pathology*, **8**, 223-231.
- Maule, A., Leh, V. and Lederer, C. (2002) The dialogue between viruses and hosts in compatible interactions. *Curr. Op. Plant Biol.*, **5**, 279-284.
- Mauricio, R., Stahl, E.A., Korves, T., Tian, D., Kreitman, M. and Bergelson, J. (2003) Natural selection for polymorphism in the disease resistance gene *Rps2* of *Arabidopsis thaliana*. *Genetics*, **163**, 735-746.
- McDonald, J.H. and Kreitman, M. (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*, **351**, 652-654.

- McDowell, J.M., Dhandaydham, M., Long, T.A., Aarts, M.G., Goff, S., Holub, E.B. and Dangl, J.L. (1998) Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the *RPP8* locus of *Arabidopsis*. *Plant Cell*, **10**, 1861-1874.
- McDowell, J.M. and Woffenden, B.J. (2003) Plant disease resistance genes: recent insights and potential applications. *Trends Biotechnol*, **21**, 178-183.
- McKhann, H.I., Camilleri, C., Berard, A., Bataillon, T., David, J.L., Reboud, X., Le Corre, V., Caloustian, C., Gut, I.G. and Brunel, D. (2004) Nested core collections maximizing genetic diversity in *Arabidopsis thaliana*. *Plant J*, **38**, 193-202.
- Menchari, Y., Camilleri, C., Michel, S., Brunel, D., Dessaint, F., Le Corre, V. and Delye, C. (2006) Weed response to herbicides: regional-scale distribution of herbicide resistance alleles in the grass weed *Alopecurus myosuroides*. *New Phytol*, **171**, 861-873.
- Metz, A., Wong, K., Malmstrom, S. and Browning, K. (1999) Eukaryotic initiation factor 4B from wheat and *Arabidopsis thaliana* is a member of a multigene family. *Biochem. Biophys. Res. Commun.*, **266**, 314-321.
- Meyers, B.C., Dickerman, A.W., Michelmore, R.W., Sivaramakrishnan, S., Sobral, B.W. and Young, N.D. (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J*, **20**, 317-332.
- Meyers, B.C., Kozik, A., Griego, A., Kuang, H. and Michelmore, R.W. (2003) Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell*, **15**, 809-834.
- Michon, T., Estevez, Y., Walter, J., German-Retana, S. and Le Gall, O. (2006) The potyviral virus genome-linked protein VPg forms a ternary complex with the eukaryotic initiation factors eIF4E and eIF4G and reduces eIF4E affinity for a mRNA cap analogue. *Febs J*, **273**, 1312-1322.
- Miller, J. and Tanksley, S. (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. *Theor Appl Genet*, **80**, 437-448.
- Miyoshi, H., Suehiro, N., Tomoo, K., Muto, S., Takahashi, T., Tsukamoto, T., Ohmori, T. and Natsuaki, T. (2005) Binding analyses for the interaction between plant virus genome-linked protein (VPg) and plant translational initiation factors. *Biochimie*.
- Monzingo, A.F., Dhaliwal, S., Dutt-Chaudhuri, A., Lyon, A., Sadow, J.H., Hoffman, D.W., Robertus, J.D. and Browning, K.S. (2007) The structure of eukaryotic translation initiation factor-4E from wheat reveals a novel disulfide bond. *Plant Physiol*, **143**, 1504-1518.
- Morel, C. (2000) Recherche des déterminants moléculaires du virus Y de la pomme de terre (PVY) impliqués dans le contournement de résistance de *Lycopersicon hirsutum*. Rapport de thèse, Université Rennes 1, p. 343.
- Moury, B., Desbiez, C., Jacquemond, M. and Lecoq, H. (2006) Genetic diversity of plant virus populations: towards hypothesis testing in molecular epidemiology. *Adv Virus Res*, **67**, 49-87.
- Moury, B., Morel, C., Johansen, E., Guilbaud, L., Souche, S., Ayme, V., Caranta, C., Palloix, A. and Jacquemond, M. (2004) Mutations in *Potato virus Y* genome-linked protein determine virulence toward recessive resistances in *Capsicum annuum* and *Lycopersicon hirsutum*. *Mol. Plant Microbe Interact.*, **17**, 322-329.
- Moury, B., Morel, C., Johansen, E. and Jacquemond, M. (2002) Evidence for diversifying selection in *Potato virus Y* and in the coat protein of other potyviruses. *J Gen Virol*, **83**, 2563-2573.
- Nei, M. and Gojobori, T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol*, **3**, 418-426.

- Nei, M. and Li, W.H. (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A*, **76**, 5269-5273.
- Nelson, M., Wheeler, R. and Zitter, T. (1982) *Pepper mottle virus*. *CMI/AAB Description Plant viruses*, n°253.
- Nesbitt, T.C. and Tanksley, S.D. (2002) Comparative sequencing in the genus *Lycopersicon*. Implications for the evolution of fruit size in the domestication of cultivated tomatoes. *Genetics*, **162**, 365-379.
- Nicaise, V. (2006) Etude des facteurs cellulaires impliqués dans les interactions plante-virus : rôle du facteur d'initiation de la traduction 4E (eIF4E). *Doctorat de Sciences biologiques et médicales*, Université Victor Ségalen, Bordeaux 2, p. 166.
- Nicaise, V., Gallois, J.L., Chafiai, F., Allen, L.M., Schurdi-Levraud, V., Browning, K.S., Candresse, T., Caranta, C., Le Gall, O. and German-Retana, S. (2007) Coordinated and selective recruitment of eIF4E and eIF4G factors for potyvirus infection in *Arabidopsis thaliana*. *FEBS Lett*, **581**, 1041-1046.
- Nicaise, V., German-Retana, S., Sanjuan, R., Dubrana, M.P., Mazier, M., Maisonneuve, B., Candresse, T., Caranta, C. and Le Gall, O. (2003) The eukaryotic translation initiation factor 4E controls lettuce susceptibility to the potyvirus *Lettuce mosaic virus*. *Plant Physiol.*, **132**, 1272-1282.
- Nicolas, O., Dunnington, S.W., Gotow, L.F., Pirone, T.P. and Hellmann, G.M. (1997) Variations in the VPg protein allow a potyvirus to overcome *va* gene resistance in tobacco. *Virology*, **237**, 452-459.
- Nicot, N., Hausman, J.F., Hoffmann, L. and Evers, D. (2005) Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *J Exp Bot*, **56**, 2907-2914.
- Nielsen, R. and Yang, Z. (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, **148**, 929-936.
- Nieto, C., Morales, M., Orjeda, G., Clepet, C., Monfort, A., Sturbois, B., Puigdomenech, P., Pitrat, M., Caboche, M., Dogimont, C., Garcia-Mas, J., Aranda, M.A. and Bendahmane, A. (2006) An *eIF4E* allele confers resistance to an uncapped and non-polyadenylated RNA virus in melon. *Plant J*, **48**, 452-462.
- Noel, L., Moores, T.L., van Der Biezen, E.A., Parniske, M., Daniels, M.J., Parker, J.E. and Jones, J.D. (1999) Pronounced intraspecific haplotype divergence at the *RPP5* complex disease resistance locus of *Arabidopsis*. *Plant Cell*, **11**, 2099-2112.
- Nordborg, M., Hu, T.T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., Bakker, E., Calabrese, P., Gladstone, J., Goyal, R., Jakobsson, M., Kim, S., Morozov, Y., Padhukasahasram, B., Plagnol, V., Rosenberg, N.A., Shah, C., Wall, J.D., Wang, J., Zhao, K., Kalbfleisch, T., Schulz, V., Kreitman, M. and Bergelson, J. (2005) The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.*, **3**, e196.
- Oleykowski, C.A., Bronson Mullins, C.R., Godwin, A.K. and Yeung, A.T. (1998) Mutation detection using a novel plant endonuclease. *Nucleic Acids Res.*, **26**, 4597-4602.
- Pacaly, V. (2006) Analyse de la résistance au *Watermelon mosaic virus* et au *Cucumber mosaic virus* chez *Arabidopsis thaliana*. *Master M1 biologie, géosciences, agronomie, environnement*, Université Montpellier II, France, p. 17.
- Pain, V.M. (1996) Initiation of protein synthesis in eukaryotic cells. *Eur. J. Biochem.*, **236**, 747-771.
- Palloix, A., Abak, K., Daubèze, A.M., Guldur, M. and Gebre-Selassie, K. (1994) Survey of pepper diseases affecting the main production regions of Turkey with special interest in viruses and potyvirus pathotypes. *Capsicum and Eggplant Newsletter*, **13**, 78-81.
- Parrella, G., Ruffel, S., Moretti, A., Morel, C., Palloix, A. and Caranta, C. (2002) Recessive resistance genes against potyviruses are localized in colinear genomic regions of the

- tomato (*Lycopersicon* spp.) and pepper (*Capsicum* spp.) genomes. *Theor. Appl. Genet.*, **105**, 855-861.
- Perry, L., Dickau, R., Zarrillo, S., Holst, I., Pearsall, D.M., Piperno, D.R., Berman, M.J., Cooke, R.G., Rademaker, K., Ranere, A.J., Raymond, J.S., Sandweiss, D.H., Scaramelli, F., Tarble, K. and Zeidler, J.A. (2007) Starch fossils and the domestication and dispersal of chili peppers (*Capsicum* spp. L.) in the Americas. *Science*, **315**, 986-988.
- Pestova, T., Borukhov, S. and Hellen, C. (1998) Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature*, **394**, 854-859.
- Pfaffl, M.W., Horgan, G.W. and Dempfle, L. (2002) Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res*, **30**, e36.
- Pickersgill, B., Heiser, C. and McNeill, J. (1979) Numerical taxonomic studies on variation and domestication in some species of *Capsicum*. In J.G. Hawkes, R.N.L.a.A.D.S. (ed.), *The biology and taxonomy of the Solanaceae*. Academic Press, London, pp. 679-700.
- Provvidenti, R. and Hampton, R.O. (1992) Sources of resistance to viruses in the Potyviridae. *Arch. Virol. Suppl.*, **5**, 189-211.
- Ptushkina, M., Fierro-Monti, I., van den Heuvel, J., Vasilescu, S., Birkenhager, R., Mita, K. and McCarthy, J.E. (1996) *Schizosaccharomyces pombe* has a novel eukaryotic initiation factor 4F complex containing a cap-binding protein with the human eIF4E C-terminal motif KSGST. *J. Biol. Chem.*, **271**, 32818-32824.
- Ptushkina, M., Malys, N. and McCarthy, J.E. (2004) eIF4E isoform 2 in *Schizosaccharomyces pombe* is a novel stress-response factor. *EMBO Rep.*, **5**, 311-316.
- Raccach, B., Gal-on, A. and Eastpo, V. (1985) The role of flying aphid vectors in the transmission of *Cucumber mosaic virus* and *Potato virus Y* to pepper in Israël. *Ann. Appl. Biol.*, **106**, 451-460.
- Rajamäki, M. and Valkonen, J. (2002) Viral genome-linked protein (VPg) controls accumulation and phloem-loading of a potyvirus in inoculated potato leaves. *Mol. Plant Microbe Interact.*, **15**, 138-149.
- Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Susan, M. and Claverie, J.M. (2004) The 1.2-Mb genome sequence of mimivirus. *Science*.
- Rausher, M.D. (2001) Co-evolution and plant resistance to natural enemies. *Nature*, **411**, 857-864.
- Redondo, E., Krause-Sakate, R., Yang, S.J., Lot, H., Le Gall, O. and Candresse, T. (2001) *Lettuce mosaic virus* pathogenicity determinants in susceptible and tolerant lettuce cultivars map to different regions of the viral genome. *Mol. Plant Microbe Interact.*, **14**, 804-810.
- Revers, F., Guiraud, T., Houvenaghel, M.C., Mauduit, T., Le Gall, O. and Candresse, T. (2003) Multiple resistance phenotypes to *Lettuce mosaic virus* among *Arabidopsis thaliana* accessions. *Mol. Plant Microbe Interact.*, **16**, 608-616.
- Revers, F., Le Gall, O., Candresse, T., Le Romancer, M. and Dunez, J. (1996) Frequent occurrence of recombinant potyvirus isolates. *J Gen Virol*, **77** ( Pt 8), 1953-1965.
- Revers, F., Le Gall, O., Candresse, T. and Maule, A. (1999) New advances in understanding the molecular biology of plant / potyvirus interactions. *Mol. Plant Microbe Interact.*, **12**, 367-376.
- Rhoads, R.E., Dinkova, T.D. and Jagus, R. (2007) Approaches for analyzing the differential activities and functions of eIF4E family members. *Methods Enzymol.*, **429**, 261-297.

- Richly, E., Kurth, J. and Leister, D. (2002) Mode of amplification and reorganization of resistance genes during recent *Arabidopsis thaliana* evolution. *Mol Biol Evol*, **19**, 76-84.
- Richman, A.D. and Kohn, J.R. (1999) Self-incompatibility alleles from *Physalis*: implications for historical inference from balanced genetic polymorphisms. *Proc Natl Acad Sci U S A*, **96**, 168-172.
- Richter, J.D. and Sonenberg, N. (2005) Regulation of cap-dependent translation by eIF4E inhibitory proteins. *Nature*, **433**, 477-480.
- Rick, C.M. (1979) Biosystematic studies in *Lycopersicon* and closely related species of *Solanum*. In Hawkes, J.G., Lester, R. N. & Skelding, A. D. (ed.), *The Biology and Taxonomy of Solanaceae*. Academic Press, New York., pp. 667-677.
- Rick, C.M., Laterrot, H. and Philouze, J. (1990) A revised key for the *Lycopersicon* species. *Report Tomato Genet. Coop.*, 31-40.
- Riechmann, J.L., Lain, S. and Garcia, J.A. (1992) Highlights and prospects of potyvirus molecular biology. *J Gen Virol*, **73 ( Pt 1)**, 1-16.
- Riely, B.K. and Martin, G.B. (2001) Ancient origin of pathogen recognition specificity conferred by the tomato disease resistance gene *Pto*. *Proc Natl Acad Sci U S A*, **98**, 2059-2064.
- Robaglia, C. and Caranta, C. (2006) Translation initiation factors: a weak link in plant RNA virus infection. *Trends Plant Sci*, **11**, 40-45.
- Robalino, J., Joshi, B., Fahrenkrug, S.C. and Jagus, R. (2004) Two zebrafish eIF4E family members are differentially expressed and functionally divergent. *J. Biol. Chem.*, **279**, 10532-10541.
- Rodriguez, C., Freire, M., Camilleri, C. and Robaglia, C. (1998) The *Arabidopsis thaliana* cDNAs coding for eIF4E and eIF(iso)4E are not functionally equivalent for yeast complementation and are differentially expressed during plant development. *Plant J.*, **13**, 465-473.
- Rose, L.E., Bittner-Eddy, P.D., Langley, C.H., Holub, E.B., Michelmore, R.W. and Beynon, J.L. (2004) The maintenance of extreme amino acid diversity at the disease resistance gene, *RPP13*, in *Arabidopsis thaliana*. *Genetics*, **166**, 1517-1527.
- Rose, L.E., Michelmore, R.W. and Langley, C.H. (2007) Natural variation in the *Pto* disease resistance gene within species of wild tomato (*Lycopersicon*). II. Population genetics of *Pto*. *Genetics*, **175**, 1307-1319.
- Roselius, K., Stephan, W. and Stadler, T. (2005) The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. *Genetics*, **171**, 753-763.
- Rousseau, D., Gingras, A.C., Pause, A. and Sonenberg, N. (1996) The eIF4E-binding proteins 1 and 2 are negative regulators of cell growth. *Oncogene*, **13**, 2415-2420.
- Roux, F., Giancola, S., Durand, S. and Reboud, X. (2006) Building of an experimental cline with *Arabidopsis thaliana* to estimate herbicide fitness cost. *Genetics*, **173**, 1023-1031.
- Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X. and Rozas, R. (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496-2497.
- Ruffel, S. (2004) Résistances récessives aux potyvirus chez les solanacées et facteurs du complexe d'initiation de la traduction. Rapport de thèse, Université Aix-Marseille II, Faculté des Sciences de Luminy, p. 159.
- Ruffel, S., Caranta, C., Palloix, A., Lefebvre, V., Caboche, M. and Bendahmane, A. (2004) Structural analysis of the eukaryotic initiation factor 4E gene controlling potyvirus resistance in pepper: exploitation of a BAC library. *Gene*, **338**, 209-216.

- Ruffel, S., Dussault, M.H., Palloix, A., Moury, B., Bendahmane, A., Robaglia, C. and Caranta, C. (2002) A natural recessive resistance gene against *Potato virus Y* in pepper corresponds to the eukaryotic initiation factor 4E (eIF4E). *Plant J.*, **32**, 1067-1075.
- Ruffel, S., Gallois, J.L., Lesage, M.L. and Caranta, C. (2005) The recessive potyvirus resistance gene *pot-1* is the tomato orthologue of the pepper *pvr2-eIF4E* gene. *Mol. Genet. Genomics*, 1-8.
- Ruffel, S., Gallois, J.L., Moury, B., Robaglia, C., Palloix, A. and Caranta, C. (2006) Simultaneous mutations in translation initiation factors eIF4E and eIF(iso)4E are required to prevent *Pepper veinal mottle virus* infection of pepper. *J Gen Virol*, **87**, 2089-2098.
- Ruud, K., Kuhlow, C., Goss, D. and Browning, K. (1998) Identification and characterization of a novel cap-binding protein from *Arabidopsis thaliana*. *J. Biol. Chem.*, **273**, 10325-10330.
- Sacristan, S., Malpica, J.M., Fraile, A. and Garcia-Arenal, F. (2003) Estimation of population bottlenecks during systemic movement of *Tobacco mosaic virus* in tobacco plants. *J Virol*, **77**, 9906-9911.
- Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*, **4**, 406-425.
- Sato, M., Nakahara, K., Yoshii, M., Ishikawa, M. and Uyeda, I. (2005) Selective involvement of members of the eukaryotic initiation factor 4E family in the infection of *Arabidopsis thaliana* by potyviruses. *FEBS Lett*, **579**, 1167-1171.
- Schaad, M., Anderberg, R. and Carrington, J. (2000) Strain-specific interaction of the *Tobacco etch virus* NIa protein with the translation initiation factor eIF4E in the yeast two-hybrid system. *Virology*, **273**, 300-306.
- Schaad, M.C. and Carrington, J.C. (1996) Suppression of long-distance movement of *Tobacco etch virus* in a nonsusceptible host. *J. Virol.*, **70**, 2556-2561.
- Schaad, M.C., Lellis, A.D. and Carrington, J.C. (1997) VPg of *Tobacco etch potyvirus* is a host genotype-specific determinant for long-distance movement. *J Virol*, **71**, 8624-8631.
- Scheper, G., Kollenburg, B.v., Hu, J., Luo, Y., Goss, D. and Proud, C. (2002) Phosphorylation of eukaryotic initiation factor 4E markedly reduces its affinity for capped mRNA. *J. Biol. Chem.*, **277**, 3303-3309.
- Schmid, K.J., Ramos-Onsins, S., Ringys-Beckstein, H., Weisshaar, B. and Mitchell-Olds, T. (2005) A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. *Genetics*, **169**, 1601-1615.
- Schmid, K.J., Sorensen, T.R., Stracke, R., Torjek, O., Altmann, T., Mitchell-Olds, T. and Weisshaar, B. (2003) Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana*. *Genome Res*, **13**, 1250-1257.
- Schwartz, D.C. and Parker, R. (2000) mRNA decapping in yeast requires dissociation of the cap binding protein, eukaryotic translation initiation factor 4E. *Mol Cell Biol*, **20**, 7933-7942.
- Scofield, S.R., Tobias, C.M., Rathjen, J.P., Chang, J.H., Lavelle, D.T., Michelmore, R.W. and Staskawicz, B.J. (1996) Molecular basis of gene-for-gene specificity in bacterial speck disease of tomato. *Science*, **274**, 2063-2065.
- Shen, J., Araki, H., Chen, L., Chen, J.Q. and Tian, D. (2006) Unique evolutionary mechanism in *R*-genes under the presence/absence polymorphism in *Arabidopsis thaliana*. *Genetics*, **172**, 1243-1250.

- Shukla, D., Ward, C. and Brunt, A. (1994) *The Potyviridae*. CAB International Wallingford, U.K. 516p.
- Slade, A.J., Fuerstenberg, S.I., Loeffler, D., Steine, M.N. and Facciotti, D. (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. *Nat. Biotechnol.*, **23**, 75-81.
- Slade, A.J. and Knauf, V.C. (2005) TILLING moves beyond functional genomics into crop improvement. *Transgenic Res.*, **14**, 109-115.
- Spooner, D., Peralta, I. and Knapp, S. (2005) Comparison of AFLPs with other markers for phylogenetic inference in wild tomatoes [*Solanum* L. section *Lycopersicon* (Mill.) Wettst.]. *Taxon*, **54**, 43-61.
- Stahl, E.A., Dwyer, G., Mauricio, R., Kreitman, M. and Bergelson, J. (1999) Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature*, **400**, 667-671.
- Stein, N., Perovic, D., Kumlehn, J., Pellio, B., Stracke, S., Streng, S., Ordon, F. and Graner, A. (2005) The eukaryotic translation initiation factor 4E confers multiallelic recessive Bymovirus resistance in *Hordeum vulgare* (L.). *Plant J*, **42**, 912-922.
- Sterck, L., Rombauts, S., Vandepoele, K., Rouze, P. and Van de Peer, Y. (2007) How many genes are there in plants (... and why are they there)? *Curr Opin Plant Biol*, **10**, 199-203.
- Stracke, S., Presterl, T., Stein, N., Perovic, D., Ordon, F. and Graner, A. (2006) Effects of introgression and recombination on haplotype structure and linkage disequilibrium surrounding a locus encoding Bymovirus resistance in barley. *Genetics*.
- Strudwick, S. and Borden, K. (2002) The emerging roles of translation factor eIF4E in the nucleus. *Differentiation*, **70**, 10-22.
- Tajima, F. (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437-460.
- Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585-595.
- Tam, S.M., Causse, M., Garchery, C., Burck, H., Mhiri, C. and Grandbastien, M.A. (2007) The distribution of copia-type retrotransposons and the evolutionary history of tomato and related wild species. *J Evol Biol*, **20**, 1056-1072.
- Tam, S.M., Mhiri, C., Vogelaar, A., Kerkveld, M., Pearce, S.R. and Grandbastien, M.A. (2005) Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR. *Theor Appl Genet*, **110**, 819-831.
- TheArabidopsisGenomeInitiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796-815.
- Thompson, J., Higgins, D. and Gibson, T. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673-4680.
- Tian, D., Araki, H., Stahl, E., Bergelson, J. and Kreitman, M. (2002) Signature of balancing selection in *Arabidopsis*. *Proc Natl Acad Sci U S A*, **99**, 11525-11530.
- Tian, D., Traw, M.B., Chen, J.Q., Kreitman, M. and Bergelson, J. (2003) Fitness costs of *R*-gene-mediated resistance in *Arabidopsis thaliana*. *Nature*, **423**, 74-77.
- Tiffin, P. and Moeller, D.A. (2006) Molecular evolution of plant immune system genes. *Trends Genet*, **22**, 662-670.
- Tomoo, K., Shen, X., Okabe, K., Nozoe, Y., Fukuhara, S., Morino, S., Ishida, T., Taniguchi, T., Hasegawa, H., Terashima, A., Sasaki, M., Katsuya, Y., Kitamura, K., Miyoshi, H., Ishikawa, M. and Miura, K. (2002) Crystal structures of 7-methylguanosine 5'-triphosphate (m(7)GTP)- and P(1)-7-methylguanosine-P(3)-adenosine-5',5'

- triphosphate (m<sup>7</sup>GpppA)-bound human full-length eukaryotic initiation factor 4E: biological importance of the C-terminal flexible region. *Biochem J*, **362**, 539-544.
- Tomoo, K., Shen, X., Okabe, K., Nozoe, Y., Fukuhara, S., Morino, S., Sasaki, M., Taniguchi, T., Miyagawa, H., Kitamura, K., Miura, K. and Ishida, T. (2003) Structural features of human initiation factor 4E, studied by X-ray crystal analyses and molecular dynamics simulations. *J Mol Biol*, **328**, 365-383.
- Topisirovic, I., Capili, A.D. and Borden, K.L. (2002) Gamma interferon and cadmium treatments modulate eukaryotic initiation factor 4E-dependent mRNA transport of cyclin D1 in a PML-dependent manner. *Mol. Cell Biol.*, **22**, 6183-6198.
- Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R.R., Bhalerao, R.P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.L., Cooper, D., Coutinho, P.M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Dejardin, A., Depamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjarvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leple, J.C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D.R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouze, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C.J., Uberbacher, E., Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., Van de Peer, Y. and Rokhsar, D. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, **313**, 1596-1604.
- Ueda, H., Yamaguchi, Y. and Sano, H. (2006) Direct interaction between the tobacco mosaic virus helicase domain and the ATP-bound resistance protein, N factor during the hypersensitive response in tobacco plants. *Plant Mol Biol*, **61**, 31-45.
- Urcuqui-Inchima, S., Haenni, A.L. and Bernardi, F. (2001) Potyvirus proteins: a wealth of functions. *Virus Res*, **74**, 157-175.
- Vasilescu, S., Ptushkina, M., Linz, B., Muller, P.P. and McCarthy, J.E. (1996) Mutants of eukaryotic initiation factor eIF-4E with altered mRNA cap binding specificity reprogram mRNA selection by ribosomes in *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **271**, 7030-7037.
- Volpon, L., Osborne, M.J., Topisirovic, I., Siddiqui, N. and Borden, K.L. (2006) Cap-free structure of eIF4E suggests a basis for conformational regulation by its ligands. *Embo J*, **25**, 5138-5149.
- Wakiyama, M., Suzuki, A., Saigoh, M., Sakai, N., Miyoshi, H., Kojima, S. and Miura, K. (2001) Analysis of the isoform of *Xenopus* eukaryotic translation initiation factor 4E. *Biosci. Biotechnol. Biochem.*, **65**, 232-235.
- Walsh, J. and Jenner, C. (2002) *Turnip mosaic virus* and the quest for durable resistance. *Mol. Plant Pathol.*, **3**, 289-300.
- Walter, M., Chaban, C., Schutze, K., Batistic, O., Weckermann, K., Nake, C., Blazevic, D., Grefen, C., Schumacher, K., Oecking, C., Harter, K. and Kudla, J. (2004) Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. *Plant J.*, **40**, 428-438.

- Wang, X., Ullah, Z. and Grumet, R. (2000) Interaction between *Zucchini mosaic potyvirus* RNA-dependent RNA polymerase and host poly(A) binding protein. *Virology*, **275**, 433-443.
- Watterson, G.A. (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol*, **7**, 256-276.
- Weidman, M.K., Sharma, R., Raychaudhuri, S., Kundu, P., Tsai, W. and Dasgupta, A. (2003) The interaction of cytoplasmic RNA viruses with the nucleus. *Virus Res.*, **95**, 75-85.
- Wittmann, S., Chatel, H., Fortin, M.G. and Laliberte, J.F. (1997) Interaction of the viral protein genome linked of *Turnip mosaic potyvirus* with the translational eukaryotic initiation factor (iso)4E of *Arabidopsis thaliana* using the yeast two-hybrid system. *Virology*, **234**, 84-92.
- Wright, S.I. and Gaut, B.S. (2005) Molecular population genetics and the search for adaptive evolution in plants. *Mol Biol Evol*, **22**, 506-519.
- Yang, Z., Nielsen, R., Goldman, N. and Pedersen, A.M. (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, **155**, 431-449.
- Yeam, I., Cavatorta, J.R., Ripoll, D.R., Kang, B.C. and Jahn, M.M. (2007) Functional dissection of naturally occurring amino acid substitutions in eIF4E that confers recessive potyvirus resistance in plants. *Plant Cell*.
- Yi, G., Lee, J.M., Lee, S., Choi, D. and Kim, B.D. (2006) Exploitation of pepper EST-SSRs and SSR-based linkage map. *Theor Appl Genet*, **114**, 113-130.
- Yoshii, M., Nishikiori, M., Tomita, K., Yoshioka, N., Kozuka, R., Naito, S. and Ishikawa, M. (2004) The *Arabidopsis* cucumovirus multiplication 1 and 2 loci encode translation initiation factors 4E and 4G. *J Virol*, **78**, 6102-6111.
- Young, N.D. (2000) The genetic architecture of resistance. *Curr Opin Plant Biol*, **3**, 285-290.
- Zhou, T., Wang, Y., Chen, J.Q., Araki, H., Jing, Z., Jiang, K., Shen, J. and Tian, D. (2004) Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Mol Genet Genomics*, **271**, 402-415.

# **A**nnexes

## **Annexe 1 : Conditions de culture des plantes**

Les Solanacées sont semées dans du terreau désinfecté et les terrines placées en pépinière sous serre pendant 2 semaines pour la phase de germination. Les semis sont ensuite éclaircis et transférés en serre chauffée (en période hivernale) ou en chambre climatisée à température constante 22°C, 12 h d'éclairage (intensité 4000 lux). Les plantes d'*A. thaliana* sont directement semées en chambre climatisée.

Pour les expériences de PCR semi-quantitative et quantitative, les plantes de l'accession Col-0 d'*A. thaliana* ont été cultivées *in vitro*. Les graines sont préalablement stérilisées dans un mélange 20% Bayroclor, 80% Ethanol absolu, rincées à l'éthanol absolu puis semées sur milieu de germination ATG (4,3 g sels MS micro et macro éléments; 0,5 g MES 0,05 %; 10 g saccharose 1 %; 8 g agar 0,8 %). Les boîtes sont placées à 4°C pendant 48 h pour l'étape de stratification puis en chambre de culture (20°C, 16 h de luminosité). Les tests de résistance sont soit réalisés dans des compartiments de serre insect-proof et chauffés si besoin, soit en chambre climatisée aux conditions d'éclairage énoncées précédemment.

## **Annexe 2 : Conservation des souches virales**

Les souches de potyvirus sont conservées à +4°C selon la technique de Bos (1969), sous forme de tissus virosés déshydratés. Cette méthode de conservation présente l'avantage de fournir un inoculum de même origine pour toutes les répétitions des tests de résistance. Avant la réalisation des tests et afin d'augmenter la concentration en virus, les souches virales sont pré-multipliées sur un hôte de multiplication. Les souches de PVY et de TEV sont multipliées sur *C. annuum* Yolo Wonder et *Datura stramonium*, respectivement. Le PVX est multiplié sur *N. benthamiana*. Les hôtes de multiplication du WMV, TuMV et CIYVV sont la courgette, le navet et *N. benthamiana* respectivement.

## **Annexe 3 : Inoculation manuelle des plantes**

L'inoculum viral est préparé à partir d'un gramme de feuilles fraîches virosées broyées dans 4 mL de tampon d'extraction (0,03 M Na<sub>2</sub>HPO<sub>4</sub> pH7; 0,2 % DIECA) additionné d'un

abrasif (carborundum 400 mesh, 20 mg/mL) et de charbon actif (20 mg/mL). Le diéthylthiocarbonate de sodium (DIECA) inhibe l'action des oxydases endogènes qui pourraient dénaturer le virus lors du broyage et le charbon actif neutralise les inhibiteurs d'infection. Les plantes sont inoculées mécaniquement à partir de cet inoculum viral, par frottement des feuilles avec le doigt protégé de latex, puis rincées à l'eau 5 minutes après l'inoculation.

#### **Annexe 4 : Détection de l'accumulation virale par DAS-ELISA ou RT-PCR**

La multiplication virale dans les feuilles inoculées et/ou dans les feuilles systémiques est mesurée par la méthode immuno-enzymatique DAS-ELISA (Double Antibody Sandwich Enzyme-Linked Immuno Sorbent Assay, Figure A.1, Clark et Adams, 1977). Elle permet de détecter la présence des protéines de capsid des virus grâce à l'utilisation d'anticorps polyclonaux spécifiques conjugués à la phosphatase alcaline (tableau A.1). Une réaction colorimétrique (visualisée au spectrophotomètre à 405 nm) atteste de la présence du virus. Les échantillons sont considérés comme infectés lorsque la valeur d'absorbance à 405 nm représente 3 fois la valeur obtenue pour le témoin sain (feuille de plantes non inoculées). Des feuilles ayant servi à la multiplication du virus sont utilisées comme témoins d'infection.

La détection du CIYVV se fait par RT-PCR (méthode décrite à l'annexe 7) avec des amorces spécifiques du gène codant pour la capsid du virus (sens GAGTGGACAATGATGGATGG et antisens AATTTAWWWGACGGATACTCTA) car il n'existe pas de sérum commercial pour ce virus.

#### **Annexe 5 : Bactéries et levures**

##### **5.1. Bactéries utilisées pour le clonage**

La souche DH5 $\alpha$  (Invitrogen, Subcloning Efficiency™ DH5 $\alpha$ ™ Chemically Competent Cells) utilisée pour le clonage est une bactérie compétente Gram (-) *Escherichia coli*. Sa transformation se fait par choc thermique (bactéries chimio-compétentes). La souche est conservée à -80°C dans du glycérol.

## 5.2. Levures utilisées pour le système de double hybride

Les souches de levures utilisées pour les expériences de double hybride sont AH109 (MAT a) et Y187 (MAT  $\alpha$ ) du kit Matchmaker GAL4 two-hybrid system 3 (Clontech, Palo Alto, CA, USA). Elles possèdent 3 gènes rapporteurs, ADE2, HIS3 et lacZ permettant des tests d'auxotrophie et d'activité  $\beta$ -galactosidase. Des UAS (Upstream activating sequences), éléments de transcription agissant en cis chez la levure, régulées par GAL4 (le domaine de fixation à l'ADN de GAL4 reconnaît les séquences UAS), sont présents chez ces 2 souches de levures. Les souches sont conservées sur boîtes YPD (Tableau A.2) et restriées toutes les 3 semaines.

## 5.3. Levure utilisée pour la complémentation fonctionnelle

La souche JO55 de *S. cerevisiae* (*cdc33*- $\Delta$  LEU2 leu2 ura3 his3 trp1 ade2 [YCp33supex-h4E URA3] donnée par JMX Hugues (Hughes *et al.*, 1999), porte un gène endogène *eIF4E* éteint (*cdc33*). Sa survie dépend de la présence du plasmide YCp33supex-h4E URA3 qui contient une copie de l'ADNc d'un gène *eIF4E* humain. Ce plasmide est sous le contrôle d'un promoteur GAL, galactose dépendant. La souche est conservée sur boîte SD gal/raf -URA (Tableau A.2) et restriée toutes les 3 semaines.

## 5.4. Milieux de culture

Le tableau A.2 regroupe les protocoles de préparation des milieux de culture : (i) des bactéries compétentes pour les expériences de clonage, (ii) des levures utilisées pour le système de double hybride dans la levure et pour la complémentation fonctionnelle dans la levure JO55.

# **Annexe 6 : Extraction des acides nucléiques**

## 6.1. Extraction d'ARN

Les ARN totaux sont extraits à partir de tissus végétaux en utilisant le kit TRI-Reagent® (Sigma-Aldrich, St Louis, USA). Cette technique d'extraction permet d'obtenir, à

partir de quelques centaines de mg de matériel frais, des ARN purs, de bonne qualité et en forte concentration. Le matériel végétal est broyé dans l'azote liquide puis homogénéisé avec 500µL de solution TRI-Reagent®. Après une étape d'extraction au chloroforme, les ARN sont précipités à l'isopropanol froid. Les culots d'ARN obtenus sont lavés dans 1 volume d'éthanol à 70% puis solubilisés dans de l'eau *RNase free*. La qualité des ARN extraits est analysée sur gel d'agarose en conditions dénaturantes (la migration s'effectue dans du tampon MOPS-EDTA 1X, MOPS 40 mM; AcNa 10 mM; EDTA 1 mM, contenant 6% de formaldéhyde et du BET pour visualiser les ARN sous UV). Leur quantité et leur pureté sont estimées par analyse spectrophotométrique à 260 et 280 nm (1 unité de  $DO_{260}$  = 40 µg d'ARN; les ARN sont de bonne qualité si le rapport  $DO_{260} / DO_{280}$  est compris entre 1,6 et 2).

## **6.2. Microextraction d'ADN génomique de plante**

Le protocole utilisé est adapté de la méthode de Fulton and Tanksley (1995). Le broyage des tissus végétaux est réalisé dans un mélange de tampon sorbitol et de tampon de lyse des noyaux contenant du CTAB (tableau A.3) à l'aide d'un broyeur à piston et de sable de Fontainebleau. Après une étape d'extraction dans un mélange chloroforme / alcool isoamylique (24:1), les ADN sont précipités à l'isopropanol froid. Les culots d'ADN obtenus sont lavés dans 1 volume d'éthanol à 70% puis solubilisés dans de l'H<sub>2</sub>O ultra-pure.

## **Annexe 7 : Amplification des séquences nucléotidiques**

### **7.1. La transcription inverse (ou reverse transcription, RT)**

L'ADNc est synthétisé à partir de 2 à 5 µg d'ARN totaux dans un volume réactionnel contenant : 10 unités de l'enzyme AMV RT (Promega, Madison, WI, USA), un oligo dT (21T, 10 µM), des dNTP (4 mM chacun), le tampon de la transcriptase inverse (5X) et de la RNasine (Promega, Madison, WI, USA). La réaction se fait pendant 1 h 30 à 42°C, puis 15 min à 72°C.

## **7.2. La PCR (Polymerase Chain Reaction)**

Les amorces utilisées pour les expériences de PCR sont répertoriées dans le tableau A.4. Les réactions d'amplification sont réalisées avec 1 unité de Taq polymérase (GoTaq Promega à 5 U/ $\mu$ L), dans le tampon du fournisseur additionné de MgCl<sub>2</sub> (1,5 mM), de dNTPs (0,16 mM de chaque dNTP) et d'un couple d'amorces spécifique du fragment d'ADN à amplifier (0,2 mM). Après 3 min de dénaturation à 94°C, l'amplification est effectuée pendant 30 cycles successifs comprenant les étapes de dénaturation (30 s à 94°C), d'hybridation (30 s à une température qui dépend de la nature et donc de la température de demi-dénaturation (T<sub>M</sub>) des amorces) et d'élongation (1 min à 72°C).

## **Annexe 8 : Techniques de clonage**

### **8.1. Vecteurs de clonage**

Les plasmides pGADT7 et pGBK7 sont fournis dans le kit Matchmaker GAL4 two-hybrid system 3 (Clontech, Palo Alto, CA, USA) pour tester les interactions protéine - protéine par le système de double hybride dans la levure. pGADT7 contient le domaine d'activation du facteur transcriptionnel GAL4 (AD), un gène de résistance à l'ampicilline (Ampr) pour la sélection des clones positifs et un épitope HA qui pourra être reconnu par des anticorps anti-HA lors des expériences de Western blot (annexe 10) (Figure A.2a). pGBKT7 possède le domaine de fixation à l'ADN de GAL4 (BD) et un gène de résistance à la kanamycine (Kanr) (Figure A.2b). Tous les fragments PCR sont préalablement clonés dans le vecteur pGEM®-T Easy (Promega, Madison, WI, USA), qui possède le gène Ampr (Figure A.2c). Le plasmide p424GBP/TRP1 est utilisé pour les expériences de complémentation fonctionnelle dans la levure (Figure A.2d). Ce vecteur est glucose-dépendant et possède le gène Ampr.

### **8.2. Ligation des fragments PCR dans les vecteurs de clonage**

Les conditions de ligation des fragments PCR avec le plasmide pGEM®-T Easy sont indiquées par le fournisseur : 2  $\mu$ L de vecteur à 50 ng, 9  $\mu$ L de tampon de ligation 2X, 1  $\mu$ L

de T4 DNA ligase et 3 µL de produit PCR. La réaction s'effectue à température ambiante pendant 1 h. Les vecteurs de clonage pGADT7, pGBKT7 et p424 sont digérés par des enzymes de restriction adaptées pour ensuite être ligués aux fragments PCR digérés par les mêmes enzymes. La proportion vecteur / insert à respecter pour ce type de ligation est la suivante :

$$\text{concentration insert} = 3 \times \frac{\text{concentration vecteur} \times \text{taille insert}}{\text{taille vecteur}}$$

Le mélange réactionnel contient 1U de T4 DNA ligase (Promega) et son tampon (5X). La ligation se fait à température ambiante pendant 3 h.

### **8.3. Transformation bactérienne et purification des ADN plasmidiques**

Les bactéries DH5α sont transformées avec les plasmides ligués aux fragments PCR. La transformation se fait par choc thermique : les bactéries en mélange avec l'ADN plasmidique sont maintenues dans la glace 30 min puis incubées 45 s à 42°C et replacées immédiatement dans la glace. Les cellules sont incubées dans du milieu LB pendant 1 h à 37°C sous agitation puis étalées sur milieu solide sélectif (Ampicilline pour les vecteurs pGEMT-easy, pGADT7 et p424 et Kanamycine pour le vecteur pGKT7). Les boîtes sont incubées une nuit à 37°C. L'ADN des clones positifs sont ensuite purifiés sur colonnes avec le QIAprep Spin Miniprep kit (Qiagen, Courtaboeuf, France), selon les données du fournisseur. Les plasmides sont repris dans 40µL d'H<sub>2</sub>O ultra-pure.

## **Annexe 9 : Système de double hybride dans la levure**

Pour étudier les interactions protéine-protéine, le système de double hybride dans la levure a été mis en œuvre en utilisant le kit Matchmaker GAL4 two-hybrid system 3 (Clontech, Palo Alto, CA, USA). Le principe général de la technique est le suivant : le plasmide pGADT7 possède le domaine d'activation de l'activateur transcriptionnel GAL4 et pGBKT7 contient le domaine de fixation à l'ADN de GAL4. Si les protéines appât et proie,

clonées dans ces vecteurs transformés respectivement dans les levures AH109 et Y187, interagissent, il y a reconstitution de l'activateur transcriptionnel GAL4 et donc transcription d'un gène rapporteur. Il s'agit d'une méthode d'observation des interactions *in vivo* indirecte : les levures poussent sur milieux sélectifs auxotrophes s'il y a interaction entre les deux partenaires protéiques (Figure A.3a). Le détail des différentes étapes est donné par le fournisseur (Clontech Yeast Protocols Handbook, Mountain View, CA, USA; <http://www.clontech.com/images/pt/PT3024-1.pdf>). Les levures AH109 et Y187 sont rendues compétentes par l'acétate de lithium puis transformées 15 min à 42°C par les plasmides pGADT7 et pGBKT7 respectivement, portant les protéines appât (ADNc eIF4E de plante) et proie (protéine virale). Les transformants AH109-pGADT7-eIF4E sont sélectionnés sur milieu SD-L (sans leucine) et les transformants Y187-pGBKT7-protéine virale sur boîte SD-W (sans tryptophane). L'étape qui suit est la conjugaison entre les transformants (AH109 étant MAT $\alpha$  et Y187 MAT $\alpha$ ). Les doubles transformants sont sélectionnés sur milieu SD-LW puis repiqués 3 fois (triplicat) sur les milieux sélectifs auxotrophes SD-LWH (sans histidine) et SD-LWHA (sans histidine et sans adénine) plus ou moins stringents pour détecter des interactions protéine-protéine plus ou moins fortes (Figure A.3b). Des contrôles positif et négatif sont fournis dans le kit Matchmaker GAL4 two-hybrid system 3 : l'antigène T SV40 interagit avec la protéine p53 tandis qu'il n'interagit pas avec la laminine (Figure A.3).

## **Annexe 10 : Western blot**

L'expression des séquences codant pour des protéines eIF4E1 chez les levures est vérifiée par des analyses de western blot. La stabilité des protéines et leur niveau d'expression sont ainsi contrôlés.

### **10.1. Extraction des protéines totales des levures**

Les protéines totales sont extraites des levures AH109 transformées avec les vecteurs pGADT7 possédant les gènes *eIF4E* selon le protocole d'extraction des protéines urée/SDS du fournisseur Clontech (Yeast Protocols Handbook, Mountain View, CA, USA; <http://www.clontech.com/images/pt/PT3024-1.pdf>). Les levures sont mises en culture dans du

milieu YPD liquide jusqu'à atteindre une  $DO_{600nm}$  d'environ 0,5. Après centrifugation, le culot est resuspendu dans une solution contenant 5 mL de "cracking buffer" (urée 8 M; SDS 5 %, Tris-HCl 40 mM pH 6,8; EDTA 0,1 mM; bleu de bromophénol 0,4 mg/mL), 50  $\mu$ L de  $\beta$ -mercaptoéthanol, 100  $\mu$ L de solution inhibitrice des protéases et 250  $\mu$ L de PMSF 100X (phenylmethyl-sulfonyl fluoride). Les échantillons sont ensuite mélangés à des billes de verre et chauffés à 70°C. Les surnageants peuvent être conservés à -80°C.

## 10.2. SDS-PAGE

La technique utilisée est celle du Slab-gel décrite par Laemmli (1970). Les extraits de polypeptides totaux de levure sont déposés sur un gel polyacrylamide SDS-PAGE (Sodium Dodecyl Sulfate Poly-Acrylamide Gel Electrophoresis) à 12 % après ajout de 20  $\mu$ L de cracking buffer et 2  $\mu$ L de  $\beta$ -mercaptoéthanol et chauffage à 95°C 5 min. La préparation du gel SDS-PAGE ainsi que du tampon de migration est décrite sur le tableau A.5. Le marqueur de taille (SDS-PAGE Molecular Weight Standards, Low Range, BioRad, Hercules, CA, USA) est également déposé. La présence de polypeptides dans les échantillons déposés est vérifiée, avant de passer à l'étape du western blot, par coloration au bleu de Coomassie pendant 30 min du gel de polyacrylamide, suivie de lavages à l'acide acétique 7 % pour éliminer le bleu non fixé. La sensibilité de détection de ce colorant est théoriquement de 50-100 ng de protéines.

## 10.3. Immunoblotting

La première étape est le transfert liquide des polypeptides sur membrane de nitrocellulose. Le montage du western blot est décrit sur la figure A.4. Le tampon de transfert est constitué de 48 mM de Tris-base; 0,037 % de SDS; 39 mM de glycine et 20 % de méthanol. Le transfert s'effectue dans une cuve Hoefer SE 245 Dual Gel Caster (San Francisco, CA, USA) pendant 45 min à 80V. La deuxième étape est la coloration au rouge Ponceau. Il s'agit d'une coloration fugace (protéines non fixées) qui permet de vérifier que les polypeptides ont bien été transférés du gel polyacrylamide vers la membrane de nitrocellulose Hybond ECL (Amersham Biosciences, Saclay, France). La membrane est ensuite rincée à l'eau et les sites aspécifiques sont saturés par incubation de la membrane toute une nuit (O.N.) dans du TBS-Lait 5 % (Tris-base 500 mM pH 8,0; NaCl 2,5 mM) à 4°C. La troisième étape est le dépôt de l'anticorps primaire : la membrane est incubée 4 h à température ambiante

dans un mélange TBS-5 % lait contenant l'anticorps monoclonal HA (12CA5, Roche) à la dilution 1/1000. La membrane est ensuite lavée successivement dans du TBS et du TBS-0,1 % Tween. L'anticorps secondaire, serum anti-souris de lapin couplé à la phosphatase alcaline (Sigma-Aldrich, St Louis, USA), est ensuite déposé : la membrane est incubée 2 h à température ambiante dans un mélange TBS-5 % lait contenant l'anticorps secondaire à la dilution 1/2000. La membrane est ensuite rincée successivement dans des bains de TBS, TBS-0,1 % Tween et TBS. La dernière étape est la révélation en déposant 2 mL du substrat NBT-BCIP sur la membrane.

## **Annexe 11 : Bases de données et liens internet**

Pour analyser les séquences, en plus des logiciels cités dans le chapitre III, les sites internet suivants ont été utilisés :

- <http://align.genome.jp/> : site d'alignement multiple de séquences, qui permet également de sortir les résultats sous forme d'arbres parenthésés, nécessaires pour les analyses dans l'exécutable codeml de PAML.

- [http://www.ch.embnet.org/software/BOX\\_form.html](http://www.ch.embnet.org/software/BOX_form.html) : pour sortir des fichiers d'alignements de séquences au format des publications.

- [http://www.bioinformatics.org/sms2/ident\\_sim.html](http://www.bioinformatics.org/sms2/ident_sim.html) : regroupe tout un ensemble de fonctions d'analyses de séquences de base, de conversions de séquences...

- [http://frodo.wi.mit.edu/cgi-bin/primer3/primer3\\_www.cgi](http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) : pour définir des amorces

- [http://bcf.arl.arizona.edu/resources/online\\_tools/clean.php](http://bcf.arl.arizona.edu/resources/online_tools/clean.php) : pour mettre les séquences au format fasta

Plusieurs bases de données sont également consultées.

Bases généralistes :

- <http://www.tigr.org/> : TIGR, The Institute for Genomic Research, qui répertorie toutes les séquences EST (Expressed Sequence Tag) et TC (Tentative Consensus) connus pour un grand nombre d'espèces (végétales et animales).

- <http://www.ncbi.nlm.nih.gov/sites/entrez> : l'incontournable site de NCBI qu'on ne présente plus. Tous les BLAST sont effectués *via* cette base de données.

- <http://www.ebi.ac.uk/> : EMBL-EBI, European Bioinformatics Institute, l'équivalent européen du site Entrez PubMed de NCBI.

- <http://www.arabidopsis.org> : TAIR, The Arabidopsis Information Resource, qui rassemble l'ensemble des données de biologie moléculaire et de génétique de l'espèce *A. thaliana*.

- <http://soldb.cit.cornell.edu/> : SGN, Sol Genomic Network, contient des informations de génétique, génomique et taxonomiques sur les espèces de Solanacées et Rubiaceae (café).

Bases spécifiques des gènes *eIF4E* :

- <http://www.avignon.inra.fr/evolutif/> : EvoluTIF, Evolutionary dynamics of Translation Initiation Factors in plants, ce site, mis en place dans le cadre du projet BRG-2005-2006, donne accès à une base de données regroupant des séquences des gènes du complexe d'initiation de la traduction chez les plantes. Il est l'un des membres d'une famille d'infoservices dédiés aux études du polymorphisme génétique chez les plantes d'intérêt agronomique : Tritipol (Triticeae Polymorphism) et Eagle (Ecological and Association Genetics in LEgumes).

- <http://umbicc3-215.umbi.umd.edu/iisstart.asp> : site développé par B. Joshi et R. Jagus regroupant l'ensemble des connaissances acquises sur les gènes de la famille multigénique eIF4E.

# Liste des tableaux et figures

## Chapitre I

**Tableau I.1.** Caractéristiques des différentes sélections naturelles et leurs effets sur la généalogie des gènes et les fréquences alléliques (d'après Bamshad and Wooding, 2003).

**Tableau I.2.** Indices du polymorphisme moléculaire.

**Tableau I.3.** Tests de neutralité utilisés couramment pour détecter de la sélection à partir des séquences d'ADN génomique (d'après Bamshad and Wooding, 2003).

**Tableau I.4.** Traces de sélection détectées sur les gènes *R* chez les plantes.

**Tableau I.5.** Répartition des membres de la famille multigénique eIF4E répertoriés dans les bases de données (d'après Joshi *et al.*, 2005).

**Tableau I.6.** Pourcentages d'identité en acides aminés des protéines eIF4E1, eIF(iso)4E et nCBP.

**Tableau I.7.** Interactions biochimiques entre les facteurs eIF4E et les protéines virales.

**Tableau I.8.** Gènes récessifs de résistance aux virus chez les plantes correspondant à des facteurs eIF4E.

**Figure I.1.** Caractéristiques structurales des protéines codées par les gènes *R* chez les plantes (d'après Holub, 2001).

**Figure I.2.** Interactions entre les protéines Avr du pathogène et les protéines R de la plante (d'après McDowell and Woffenden, 2003).

**Figure I.3.** Coévolution d'un gène *R* de l'hôte et du gène *Avr* de l'agent pathogène correspondant (d'après Jones and Dangl, 2006).

**Figure I.4.** Evolution des gènes *R* dans une population (d'après Holub, 2001).

**Figure I.5.** Etapes de l'initiation de la traduction chez les eucaryotes.

**Figure I.6.** Alignement des séquences protéiques de eIF4E et eIF(iso)4E.

**Figure I.7.** Rôles biologiques potentiels de l'interaction eIF4E/VPg dans l'infection virale.

**Figure I.8.** Localisation sur la structure tridimensionnelle eIF4E du blé des acides aminés impliqués dans la résistance aux potyvirus (d'après Monzingo *et al.*, 2007).

**Figure I.9.** Comparaison de la position des acides aminés dans la partie centrale de la VPg impliqués dans le contournement des gènes récessifs de résistance.

**Figure I.10.** Exemples de symptômes causés par les potyvirus sur des plantes cultivées.

**Figure I.11.** Organisation génomique, processus de maturation des potyvirus et fonctions attribuées aux protéines.

**Figure I.12.** Facteurs récessifs de résistance aux potyvirus chez le piment : spectre d'action, mécanisme et localisation chromosomique.

**Figure I.13.** Répartition, en fonction de leur degré de polymorphisme, des 48 accessions d'*A. thaliana* de la core collection (d'après McKhann *et al.*, 2004).

## **Chapitre II**

**Table 1.** Amino acid substitutions observed between deduced eIF4E protein sequences from the twenty five *C. annuum* genotypes and their corresponding alleles.

**Table 2.** Resistance phenotypes against PVY and TEV strains of pepper genotypes used in this study.

**Table 3.** Amino acid differences in the central part of the VPg of PVY and TEV strains used in this study.

**Tableau II.1.** Pourcentages d'identité en acides aminés des protéines eIF4E1, eIF4E2 et eIF(iso)4E chez l'accession *C. annuum* Yolo Wonder.

**Figure 1.** Map showing origin of the twenty five *Capsicum annuum* genotypes used in this study.

**Figure 2.** Alignment of amino acid sequences and secondary structure for wheat eIF4E and predicted eIF4E protein from *C. annuum* Yolo Wonder genotype.

**Figure 3.** Genetic analysis to examine the recessive nature of PVY-LYE84 or TEV-HAT resistance and allelism relationships with *pvr2*.

**Figure 4.** Expression of eIF4E proteins in the yeast two-hybrid system.

**Figure 5.** eIF4E protein-VPg protein interaction assay using the yeast two-hybrid system.

**Figure 6.** Complementation of yeast strain JO55 with *Capsicum annuum* eIF4E cDNAs.

**Figure II.1.** Position des substitutions en acides aminés impliquées dans la résistance aux potyvirus sur la structure 3D prédite de la protéine eIF4E1.

**Figure II.2.** Principe de l'expérience de complémentation fonctionnelle par expression transitoire d'ADNc *eIF4E* via le vecteur viral PVX.

**Figure II.3.** Détection de l'accumulation des protéines de capsid du TEV-S103 et du PVY-LYE72 par DAS-ELISA dans les expériences d'expression transitoire de eIF4E.

**Figure II.4.** Alignement de la séquence en acides aminés des protéines eIF4E1, eIF4E2 et eIF(iso)4E de l'accession *C. annuum* Yolo Wonder.

**Figure II.5.** Alignement de la zone centrale des protéines VPg des 5 souches de potyvirus utilisées.

**Figure II.6.** Interactions entre les protéines eIF4E1 ou eIF4E2 et les VPg de souches virulentes et avirulentes de PVY et de TEV détectées par le système de double hybride dans la levure.

### **Chapitre III**

**Tableau III.1.** Matériel végétal utilisé dans cette étude.

**Tableau III.2.** Amorces utilisées pour amplifier et séquencer les gènes *eIF4E* et les locus indépendants.

**Tableau III.3.** Localisation chromosomique et protéines codées par les locus indépendants.

**Tableau III.4.** Diversité nucléotidique des 3 gènes *Ca-eIF4E* et des 11 locus indépendants chez 25 accessions de *C. annuum*.

**Tableau III.5.** Test d'hétérogénéité des modèles M7 et M8 pour l'estimation du paramètre  $\omega$  par site.

**Tableau III.6.** Sites sous sélection positive détectés par le modèle M8.

**Tableau III.7.** Substitutions nucléotidiques détectées entre chaque gène *Ca-eIF4E* chez les 51 accessions de *Capsicum* spp.

**Tableau III.8.** Diversité nucléotidique des gènes *Ly-eIF4E* et des locus indépendants chez 20 accessions de *L. esculentum*.

**Tableau III.9.** Substitutions nucléotidiques dans les séquences codantes des gènes *Ly-eIF4E* entre les 20 accessions de *L. esculentum*.

**Tableau III.10.** Substitutions en acides aminés dans les protéines Ly-eIF4E entre les 50 accessions de *Lycopersicon* spp.

**Tableau III.11.**

**Tableau III.12.** Sites sous sélection positive détectés par le modèle M8.

**Tableau III.13.** Diversité nucléotidique des 5 gènes *At-eIF4E* et des 10 locus indépendants chez 35 accessions d'*A. thaliana*.

**Tableau III.14.** Test d'hétérogénéité des modèles M7 et M8 pour estimer par maximum de vraisemblance le paramètre  $\omega$  par site.

**Tableau III.15.** Substitutions en acides aminés observées entre les 5 *At-eIF4E* chez 35 accessions d'*A. thaliana* et polyhaplotypes correspondants.

**Tableau III.16.** Test HKA.

**Figure III.1.** Répartition des différents locus "non-eIF4E" sur le génome du piment.

**Figure III.2.** Aires d'origine et de domestication présumées des 5 espèces cultivées de *Capsicum* spp.

**Figure III.3.** Structure des 3 gènes *Ca-eIF4E* et analyse par fenêtre glissante de la diversité nucléotidique et des valeurs de D de Tajima.

**Figure III.4.** Valeurs de D de Tajima pour les 3 gènes *Ca-eIF4E* et les 11 locus indépendants chez 25 accessions de *C. annuum*.

**Figure III.5.** Rapports de  $\pi_a$  sur  $\pi_s$  calculés pour les locus indépendants et les 3 gènes *Ca-eIF4E* chez 25 accessions de *C. annuum*.

**Figure III.6.** Arbres généalogiques enracinés représentant la structure haplotypique des gènes *Ca-eIF4E1*, *Ca-eIF4E2* et *Ca-eIF(iso)4E* chez 25 accessions de *C. annuum*.

**Figure III.7.** Valeurs de  $\omega$  prédites pour chaque acide aminé de la protéine *Ca-eIF4E1*.

**Figure III.8.** Représentation radiale des arbres phylogénétiques construits à partir des séquences codantes des gènes *Ca-eIF4E* et des locus indépendants concaténés.

**Figure III.9.** Représentations radiales des arbres phylogénétiques (Neighbour-Joining) construits à partir des séquences codantes des gènes *Ly-eIF4E* et des locus indépendants concaténés.

**Figure III.10.** Structure de l'ADNg des 5 *At-eIF4E* et analyse par fenêtre glissante de la diversité nucléotidique.

**Figure III.11.** Caractéristiques du profil de polymorphisme chez *A. thaliana* estimé par Nordborg *et al.* (2005).

**Figure III.12.** Rapports de  $\pi_a$  sur  $\pi_s$  calculés pour les locus indépendants et les 5 gènes *At-eIF4E* chez 35 accessions d'*A. thaliana*.

**Figure III.13.** Alignement des séquences en acides aminés des 5 protéines *At-eIF4E*.

**Figure III.14.** Pourcentage de plantes infectées par le WMV, le TuMV et le CIYVV chez 19 accessions d'*A. thaliana* représentatives des différents polyhaplotypes protéiques eIF4E.

**Figure III.15.** Comparaison des arbres phylogénétiques obtenus à partir des ADNc *At-eIF4E* concaténés et des ADNc des locus indépendants concaténés.

## Chapitre IV

**Tableau IV.1.** Amorces utilisées pour les expériences de RT-PCR semi-quantitatives et quantitatives.

**Figure IV. 1.** Profils d'expression des gènes *At-eIF4E* chez *A. thaliana* analysés par RT-PCR semi-quantitative.

**Figure IV.2.** Expression des gènes témoins *histone 1* et *GAPDHA* dans les 5 organes chez *A. thaliana*.

**Figure IV.3.** Contrôles nécessaires à l'exploitation des résultats obtenus par RT-PCR quantitative.

**Figure IV.4.** Comparaison des niveaux d'accumulation en quantité relative des transcrits *At-eIF4E* entre les 5 organes de la plante.

**Figure IV.5.** Comparaison du niveau d'accumulation de chacun des 5 transcrits *At-eIF4E* pour chaque organe.

**Figure IV.6.** Test de complémentation fonctionnelle dans la levure JO55 des 5 ADNc *At-eIF4E* et des 3 *Ca-eIF4E*.

**Figure IV.7.** Alignement des séquences en acides aminés des protéines eIF4E d'*A.thaliana* Col-0 et *C. annuum* Yolo Wonder avec la protéine eIF4E de *S. cerevisiae*.

## Annexes

**Tableau A.1.** Sérums utilisés pour la détection d'accumulation virale par la méthode DAS-ELISA.

**Tableau A.2.** Composition des milieux de culture utilisés pour les levures et les bactéries.

**Tableau A.3.** Solutions nécessaires pour la microextraction d'ADN à partir de matériel végétal.

**Tableau A.4.** Amorces PCR répertoriées pour chaque type d'expérience.

**Tableau A.5.** Composition du gel polyacrylamide SDS-PAGE et du tampon de migration.

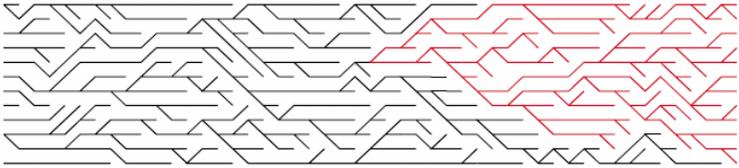
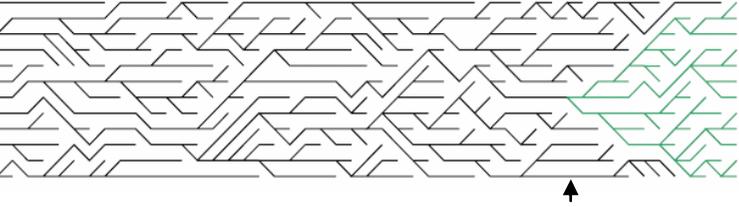
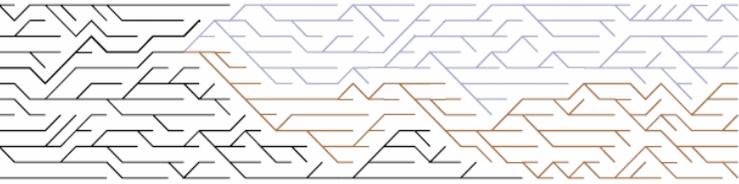
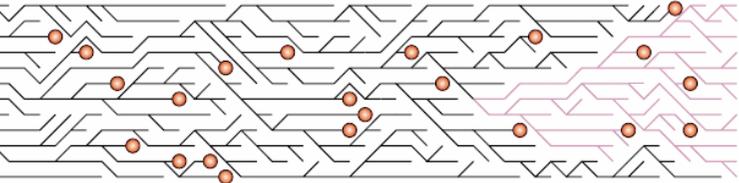
**Figure A.1.** Principe du test DAS-eLISA pour détecter l'accumulation virale dans les feuilles inoculées.

**Figure A.2.** Schémas des vecteurs et de leur site de clonage multiple utilisés lors de ce travail.

**Figure A.3.** Principe et méthode du système de double hybride dans la levure pour détecter *in vivo* des interactions protéine/protéine.

**Figure A.4.** Montage du Western blot pour transférer les polypeptides du gel de polyacrylamide SDS-PAGE vers une membrane de nitrocellulose.

**Tableau I.1. Caractéristiques des différentes sélections naturelles et leurs effets sur la généalogie des gènes et les fréquences alléliques (d'après Bamshad and Wooding, 2003).**

Sélection naturelle	Caractéristiques	Effets de la sélection sur la généalogie des gènes et les fréquences alléliques*
Neutre	Absence de sélection	 <p>Fixation de l'allèle neutre (en rouge)</p>
Positive, adaptative, diversificatrice, darwinienne	Favorise les variants génétiques	 <p>Fixation rapide d'un allèle (en vert) après un évènement de sélection positive (flèche)</p>
Balancée, équilibrante	Maintien le polymorphisme allélique dans une population	 <p>Ni fixation ni extinction de 2 allèles (en bleu et marron)</p>
Négative, purificatrice	Diminue la fréquence ou élimine les variants génétiques	 <p>Fixation d'un allèle (en rose). Chaque rond rouge représente l'élimination d'une mutation délétère (létale)</p>

\*Chaque panel représente la généalogie complète d'une population de 12 individus haploïdes. Chaque ligne retrace la généalogie d'un individu, les lignes en couleur représentent tous les descendants qui ont hérité soit d'un allèle neutre, soit d'un allèle affecté par la sélection naturelle à partir de leur ancêtre commun (coalescence de la généalogie).

**Tableau I.2. Indices du polymorphisme moléculaire.**

Indice	Définition
$N_e$	Nombre d'individus d'une population idéale, aussi appelé taille efficace de la population
$\mu$	Taux de mutation par gène et par génération
$\pi$	Diversité nucléotidique (nombre moyen de différences nucléotidiques par base entre 2 séquences prises au hasard dans un échantillon)
$\theta_w$	Taux populationnel de mutation de Watterson (nombre de sites en ségrégation par base en tenant compte de la taille efficace de la population)

**Tableau I.3. Tests de neutralité utilisés couramment pour détecter de la sélection à partir des séquences d'ADN génomique (d'après Bamshad and Wooding, 2003).**

Tests	Ce qu'ils comparent	Références
Tests basés sur le polymorphisme intraspécifique		
D de Tajima	Le nombre moyen de différences de nucléotides par paire de séquences est comparé au nombre total de sites qui ségrègent (SNPs)	Tajima, 1989
D de Fu and Li	Le nombre de variants nucléotidiques dérivés présents une seule fois est comparé au nombre total de variants dérivés	Fu and Li, 1993
F de Fu and Li	Le nombre de variants nucléotidiques dérivés présents une seule fois est comparé au nombre moyen de différences entre paires de séquences	Fu and Li, 1993
H de Fay and Wu	Le nombre de variants nucléotidiques observés à des fréquences faibles et élevées est comparé au nombre de variants observés à des fréquences intermédiaires	Fay and Wu, 2000
Rapport dN/dS = $\omega$	Compare les taux de substitutions non-synonymes et des substitutions synonymes dans les régions codantes	Li et al., 1985 Nei and Gojobori, 1986
Tests basés sur la comparaison entre la divergence interspécifique et le polymorphisme intraspécifique		
Test Hudson-Kreitman-Aguade (HKA)	Compare le polymorphisme intraspécifique et la divergence entre espèces à au moins 2 locus	Hudson et al., 1987
Tests McDonald and Kreitman	Compare les nombres de substitutions non-synonymes et de substitutions synonymes contribuant au polymorphisme intraspécifique et à la divergence interspécifique	McDonald and Kreitman, 1991

**Tableau I.4. Traces de sélection détectées sur les gènes R chez les plantes.**

Hôte végétal	Pathogène	Locus R de résistance	Pression de sélection exercée sur le gène R	Références
<i>A. thaliana</i>	<i>Peronospora parasitica</i> (oomycète)	<i>Rpp1</i>	Sélection balancée	Botella <i>et al.</i> , 1998
		<i>Rpp5</i>	Sélection balancée	Noel <i>et al.</i> , 1999
		<i>Rpp8</i>	Sélection balancée	McDowell <i>et al.</i> , 1998; Ding <i>et al.</i> , 2007
		<i>Rpp13</i>	Sélection balancée	Rose <i>et al.</i> , 2004; Ding <i>et al.</i> , 2007
	<i>Pseudomonas syringae</i> (bactérie)	<i>Rpm1</i>	Sélection balancée	Stahl <i>et al.</i> , 1999; Shen <i>et al.</i> , 2006
		<i>Rps2</i>	Sélection balancée	Caicedo <i>et al.</i> , 1999; Mauricio <i>et al.</i> , 2003; Ding <i>et al.</i> , 2007
		<i>Rps5</i>	Sélection balancée	Tian <i>et al.</i> , 2002; Shen <i>et al.</i> , 2006
		<i>Rps4</i>	Sélection diversificatrice	Bergelson <i>et al.</i> , 2001
Tomate	<i>Pseudomonas syringae</i> (bactérie)	<i>Pto</i>	Sélection balancée + purificatrice	Rose <i>et al.</i> , 2005; Rose <i>et al.</i> , 2007
	<i>Fusarium oxysporum</i> (champignon)	<i>I2</i>	Sélection purificatrice	Couch <i>et al.</i> , 2006
	<i>Cladosporium fulvum</i> (champignon)	<i>Cf2/5</i>	Sélection faiblement balancée (à l'extrémité 5') et faiblement diversificatrice (à l'extrémité 3')	Dixon <i>et al.</i> , 1998
Lin	<i>Melampsora lini</i> (champignon)	<i>L</i>	Sélection diversificatrice	Dodds <i>et al.</i> , 2000; Dodds <i>et al.</i> , 2006; Ellis <i>et al.</i> , 2007

**Tableau I.5. Répartition des membres de la famille multigénique eIF4E répertoriés dans les bases de données (d'après Joshi *et al.*, 2005).**

Règnes	Nombre de membres de la famille eIF4E identifiés <sup>1</sup>	Nombre d'espèces représentées	Nombre de membres de la famille eIF4E/espèce
Metazoaires	186	89	2,09
Plantes	155	83	1,87
Champignons	42	36	1,17
Protozoaires	28	22	1,27
<b>Total</b>	<b>411</b>	<b>230</b>	<b>1,78</b>

<sup>1</sup> eIF4E dont la séquence nucléotidique codante a été identifiée

**Tableau I.6. Pourcentages d'identité en acides aminés des protéines eIF4E1, eIF(iso)4E et nCBP chez *A. thaliana*, *L. esculentum* (tomate), *Oryza sativa* (riz) et *Triticum aestivum* (blé) et des protéines eIF4E chez *Homo sapiens* (homme) et *Saccharomyces cerevisiae* (levure). Les pourcentages d'identité protéiques entre les orthologues sont surlignés en jaune, ceux des paralogues en blanc.**

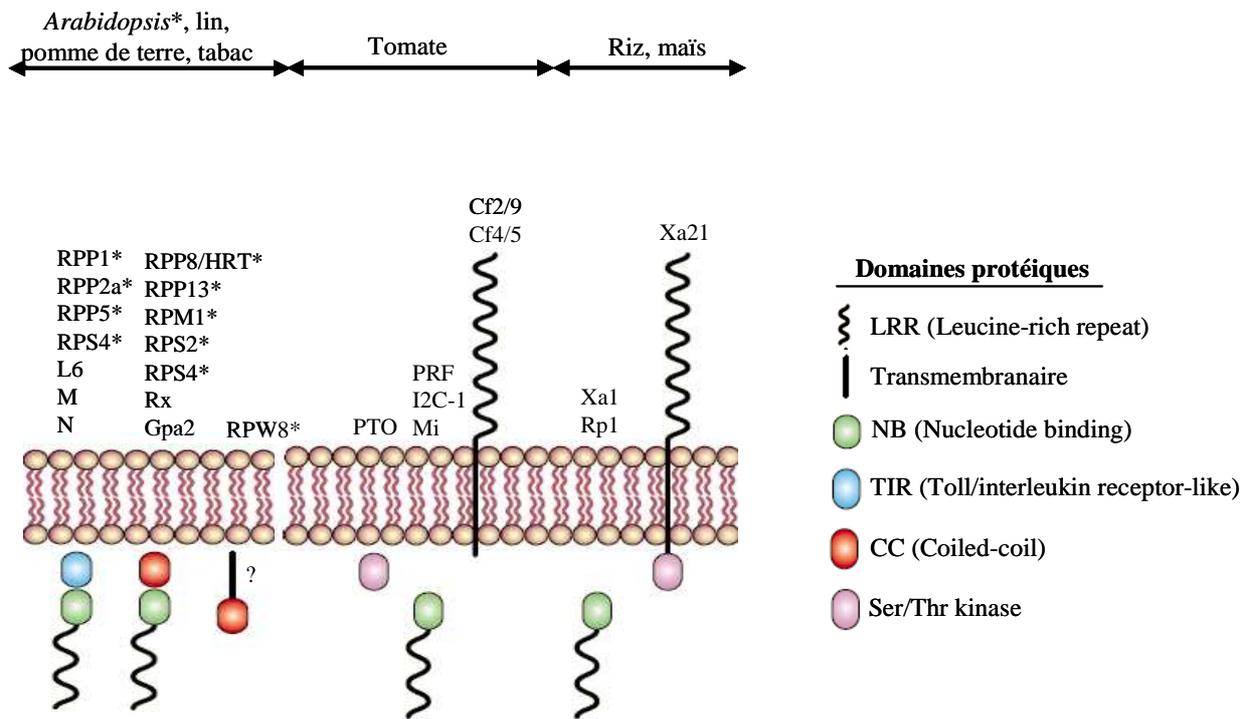
		<i>Arabidopsis thaliana</i> <i>Lycopersicon esculentum</i> <i>Oryza sativa</i> <i>Triticum aestivum</i> <i>Homo sapiens</i>	<i>Arabidopsis thaliana</i> <i>Lycopersicon esculentum</i> <i>Oryza sativa</i> <i>Triticum aestivum</i>	<i>Arabidopsis thaliana</i> <i>Lycopersicon esculentum</i> <i>Oryza sativa</i> <i>Triticum aestivum</i>			
		eIF4E1		eIF(iso)4E	nCBP		
<i>Arabidopsis thaliana</i>	eIF4E1			40	26		
<i>Lycopersicon esculentum</i>		62		41	24		
<i>Oryza sativa</i>		62	64		41	24	
<i>Triticum aestivum</i>		59	62	80		43	24
<i>Homo sapiens</i>		30	31	34	33		
<i>Saccharomyces cerevisiae</i>		27	28	26	27	29	
<i>Arabidopsis thaliana</i>	eIF(iso)4E			40	25		
<i>Lycopersicon esculentum</i>			41	62	25		
<i>Oryza sativa</i>			41	61	69	24	
<i>Triticum aestivum</i>			43	60	67	82	27
<i>Arabidopsis thaliana</i>	nCBP			26	25		
<i>Lycopersicon esculentum</i>			24	25	71		
<i>Oryza sativa</i>			24	24	70	72	
<i>Triticum aestivum</i>			24	27	68	70	84

**Tableau I.7. Interactions biochimiques entre les facteurs eIF4E et les protéines virales.**

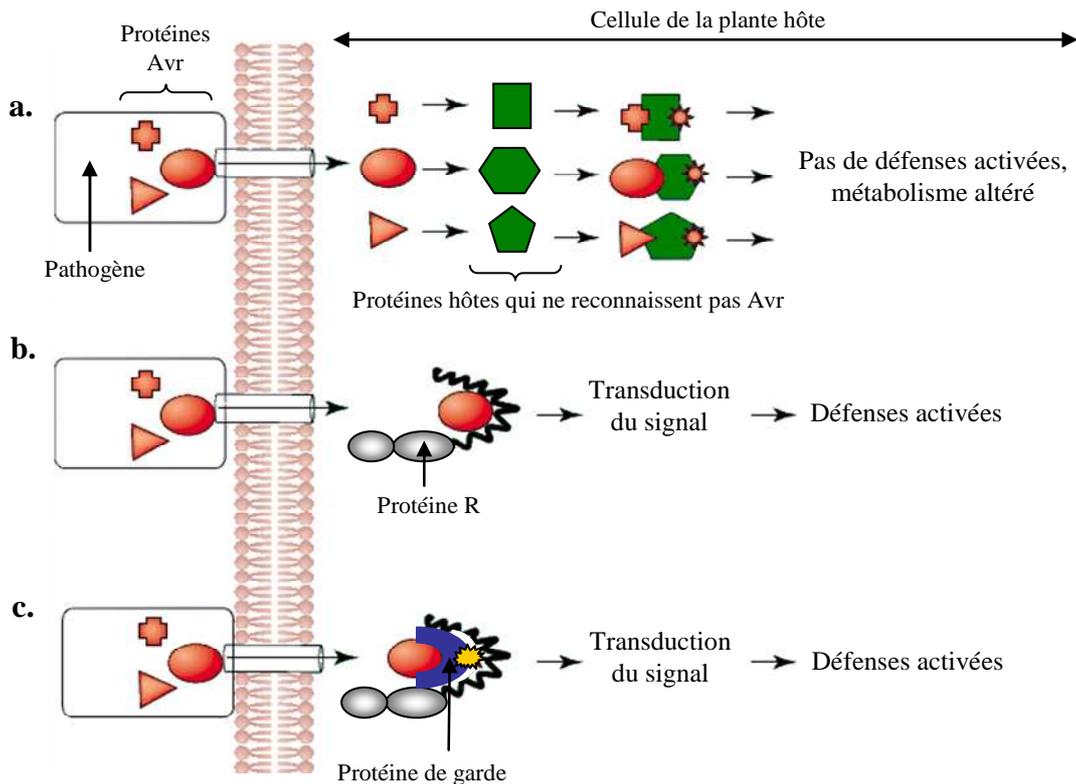
Protéines de la plante	Protéines virales	Tests biochimiques d'interaction	Références
eIF4E1 <i>A. thaliana</i>	VPg-Pro TuMV	Système ELISA ( <i>in vitro</i> )	Léonard <i>et al.</i> , 2000
	VPg PVY	GST pull-down ( <i>in vitro</i> ) + baculovirus recombinants ( <i>in vivo</i> )	Grzela <i>et al.</i> , 2006
eIF(iso)4E <i>A. thaliana</i>	VPg TuMV	Système double hybride dans la levure ( <i>in vivo</i> )	Wittmann <i>et al.</i> , 1997
		Chromatographie d'affinité ( <i>in vitro</i> )	Miyoshi <i>et al.</i> , 2006
	VPg-Pro ToRSV	Système ELISA ( <i>in vitro</i> )	Léonard <i>et al.</i> , 2002
	VPg PVY	GST pull-down ( <i>in vitro</i> ) + baculovirus recombinants ( <i>in vivo</i> )	Grzela <i>et al.</i> , 2006
eIF4E blé	VPg PVY	GST pull-down ( <i>in vitro</i> ) + baculovirus recombinants ( <i>in vivo</i> )	Grzela <i>et al.</i> , 2006
eIF(iso)4E blé	VPg-Pro TuMV	Système ELISA ( <i>in vitro</i> )	Léonard <i>et al.</i> , 2000
	VPg PVY	GST pull-down ( <i>in vitro</i> ) + baculovirus recombinants ( <i>in vivo</i> )	Grzela <i>et al.</i> , 2006
eIF4E1 tomate	VPg-Pro TEV	Système double hybride dans la levure ( <i>in vivo</i> )	Schaad <i>et al.</i> , 2000
eIF4E1 tabac	VPg-Pro TEV	Système double hybride dans la levure ( <i>in vivo</i> )	Schaad <i>et al.</i> , 2000
eIF4E1 laitue	VPg LMV	Spectroscopie	Michon <i>et al.</i> , 2006

**Tableau I.8. Gènes récessifs de résistance aux virus chez les plantes correspondant à des facteurs eIF4E.**

	Gènes	Plantes	Virus	Références
<b>Allèles naturels</b>				
<i>pvr2</i> <sup>1</sup> , <i>pvr1</i>	<i>eIF4E1</i>	Piment ( <i>Capsicum</i> spp.)	PVY, TEV ( <i>Potyvirus</i> )	Ruffel <i>et al.</i> , 2002; Kang <i>et al.</i> , 2005
<i>pvr2</i> + <i>pvr6</i>	<i>eIF(iso)4E</i>	Piment ( <i>Capsicum</i> spp.)	PVMV ( <i>Potyvirus</i> )	Ruffel <i>et al.</i> , 2006
<i>pot1</i>	<i>eIF4E1</i>	Tomate ( <i>Lycopersicon</i> spp.)	PVY, TEV ( <i>Potyvirus</i> )	Ruffel <i>et al.</i> , 2005
<i>sbm1</i>	<i>eIF4E1</i>	Pois ( <i>Pisum sativum</i> )	PSBMV ( <i>Potyvirus</i> )	Gao <i>et al.</i> , 2004
<i>mo1</i> <sup>1</sup> , <i>mo1</i> <sup>2</sup>	<i>eIF4E1</i>	Laitue ( <i>Lactuca sativa</i> )	LMV ( <i>Potyvirus</i> )	Nicaise <i>et al.</i> , 2003
<i>rym4/5</i>	<i>eIF4E1</i>	Orge ( <i>Hordeum vulgare</i> )	BMMV, BYMV, BYMV2 ( <i>Bymovirus</i> )	Stein <i>et al.</i> , 2005; Kanyuka <i>et al.</i> , 2005
<i>nsv</i>	<i>eIF4E1</i>	Melon ( <i>Cucumis melon</i> )	MNSV ( <i>Carmovirus</i> )	Nieto <i>et al.</i> , 2006
<b>Mutants</b>				
<i>At-eIF4E1</i>	<i>eIF4E1</i>	<i>A. thaliana</i>	CIYVV ( <i>Potyvirus</i> )	Sato <i>et al.</i> , 2005
<i>At-eIF4E1(cum1)</i>	<i>eIF4E1</i>	<i>A. thaliana</i>	CMV ( <i>Cucumovirus</i> )	Yoshii <i>et al.</i> , 2004
<i>At-eIF(iso)4E</i>	<i>eIF(iso)4E</i>	<i>A. thaliana</i>	TEV ( <i>Potyvirus</i> )	Lellis <i>et al.</i> , 2002
<i>At-eIF(iso)4E</i>	<i>eIF(iso)4E</i>	<i>A. thaliana</i>	TuMV, LMV ( <i>Potyvirus</i> )	Duprat <i>et al.</i> , 2002
<i>At-eIF(iso)4E</i>	<i>eIF(iso)4E</i>	<i>A. thaliana</i>	PPV ( <i>Potyvirus</i> )	Decroocq <i>et al.</i> , 2006



**Figure I.1. Caractéristiques structurales des protéines codées par les gènes *R* chez les plantes (d'après Holub, 2001).**

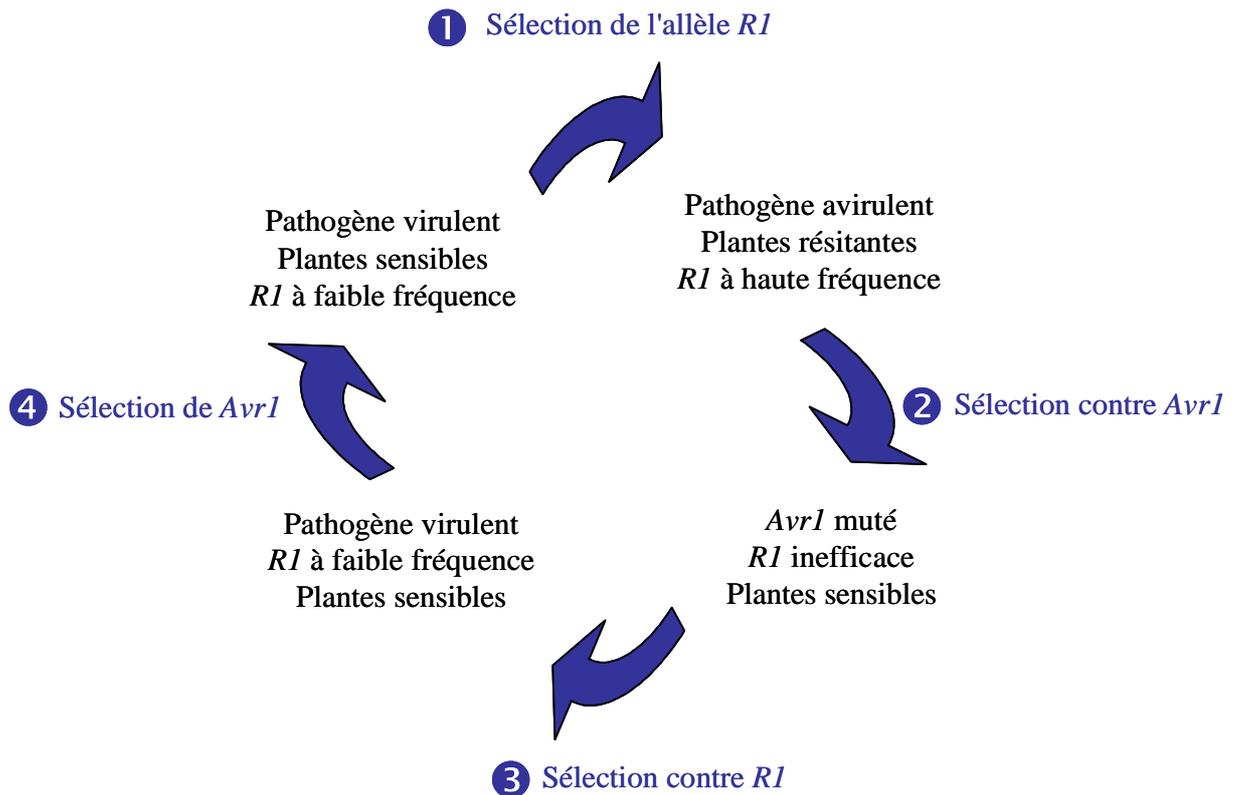


**Figure I.2. Interactions entre les protéines Avr du pathogène et les protéines R de la plante (d'après McDowell and Woffenden, 2003).**

**a.** Absence de protéines R capables de reconnaître les protéines Avr : la plante ne détecte pas le pathogène.

**b.** Modèle d'interaction directe "éliciteur-récepteur" : l'interaction directe entre les protéines Avr et R déclenche une cascade de signalisation conduisant à la réaction de défense.

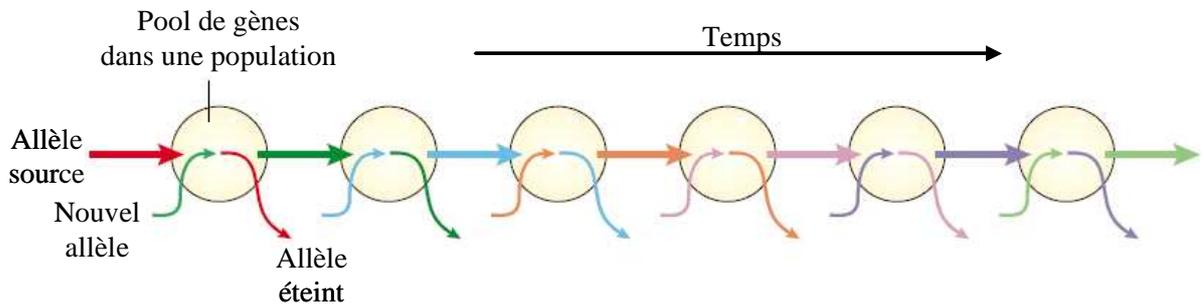
**c.** Modèle de garde : la protéine Avr interagit avec la protéine de garde. Cette interaction induit des modifications dans la protéine de garde qui vont être reconnues par la protéine R; la cascade de signalisation à l'origine de la résistance de la plante est alors déclenchée.



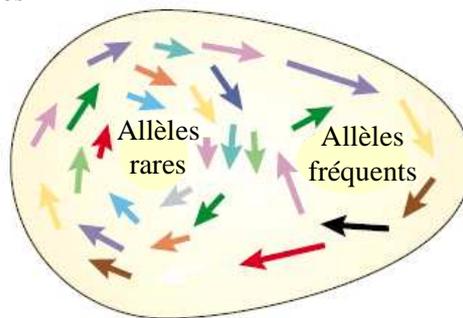
**Figure I.3. Coévolution d'un gène *R* de l'hôte et du gène *Avr* de l'agent pathogène correspondant.**

L'agent pathogène porte un gène d'avirulence (*AvrI*) qui est reconnu par un allèle rare *RI* (1). Ceci entraîne une augmentation de la fréquence de l'allèle *RI* dans la population. Les pathogènes pour lesquels *Avr* est muté sont ensuite sélectionnés car ils peuvent infecter les plantes portant l'allèle *RI* (2). L'efficacité de *RI* diminue et comme un coût de fitness est associé aux gènes *R*, les plantes portant l'allèle *RI* ont une fitness réduite (3), donc la fréquence de *RI* diminue. La population de pathogènes possède toujours *AvrI*. En absence de *RI*, *AvrI* confère une meilleure fitness et sa fréquence dans la population va augmenter (4). Ceci reconduit à une sélection de *RI*. Dans les populations de plantes et de pathogènes, ce cycle est permanent (d'après Jones and Dangl, 2006).

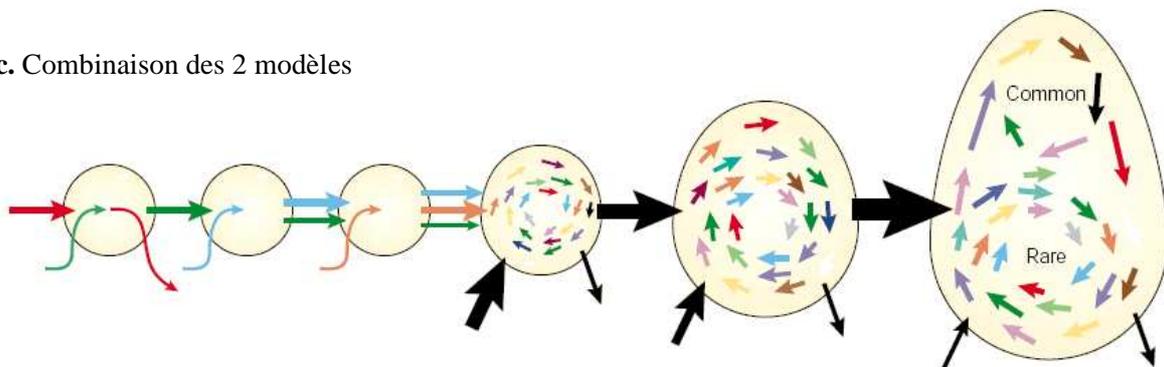
**a. Modèle "course aux armements"**



**b. Modèle "guerre des tranchées"**

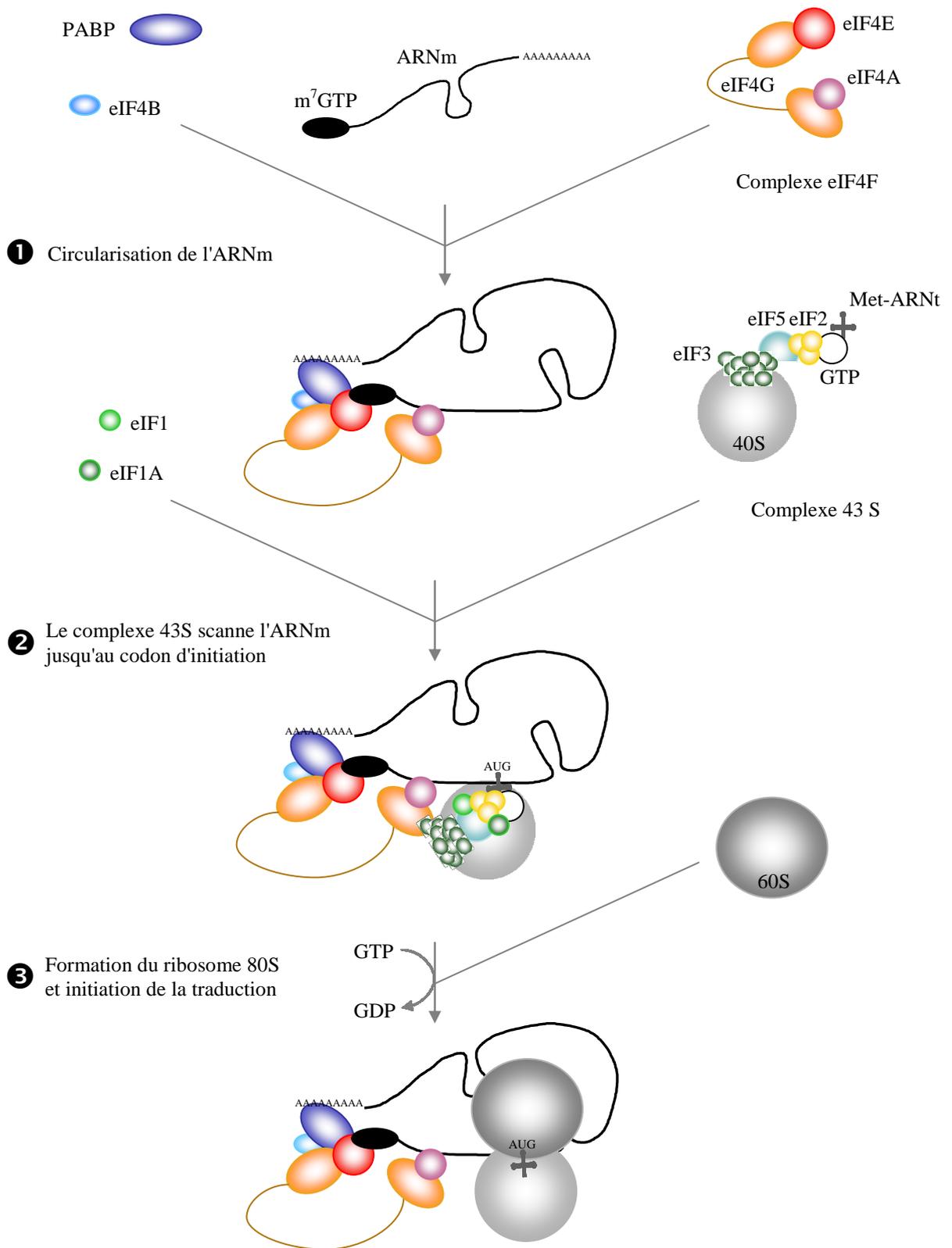


**c. Combinaison des 2 modèles**

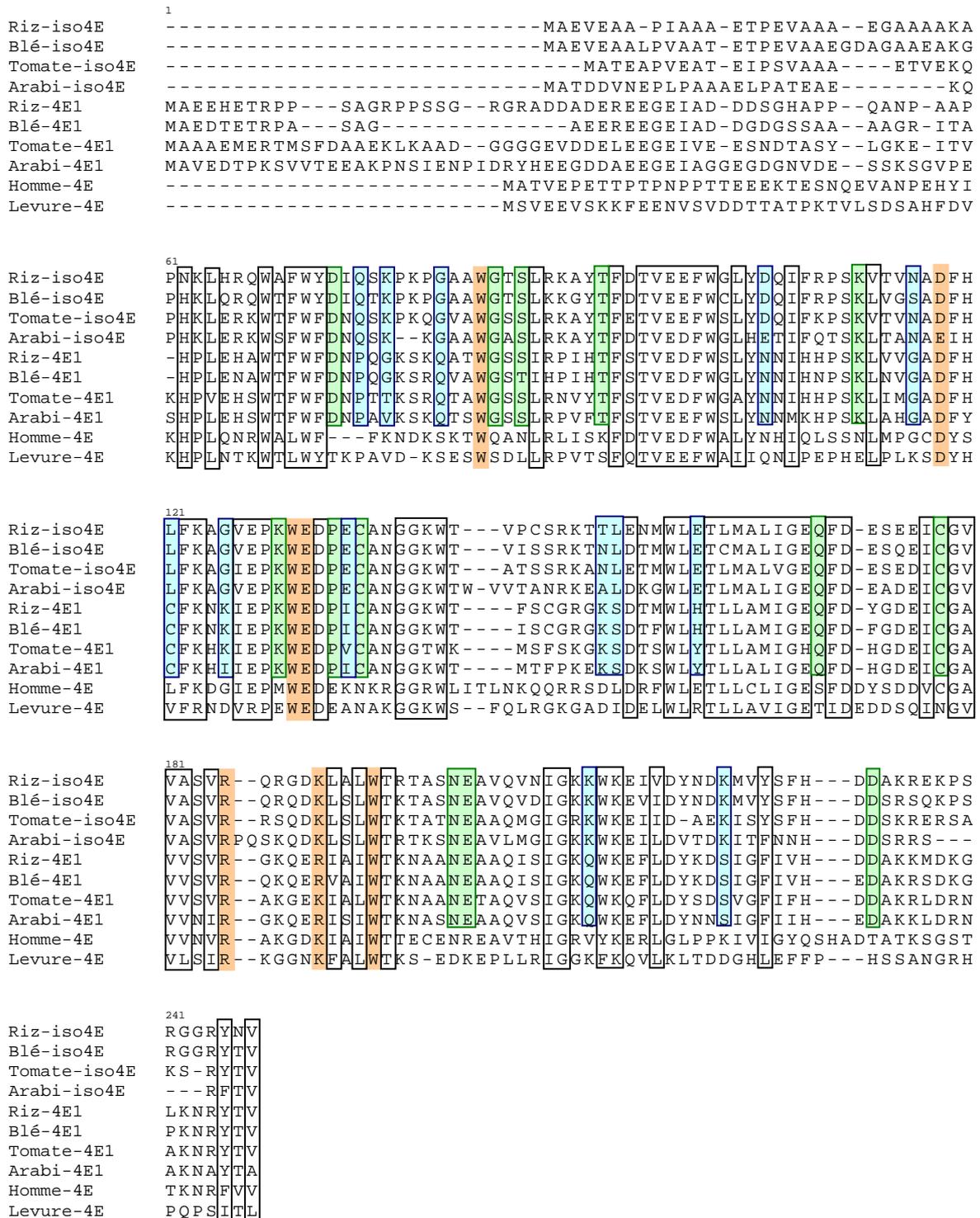


**Figure I.4. Evolution des gènes *R* dans une population (d'après Holub, 2001).**

**a.** Modèle "course aux armements" ou "transitoire" : les anciens allèles *R* sont remplacés à chaque cycle par un nouvel allèle *R* suite à l'émergence de nouvelles formes d'avirulence du pathogène. Le nouvel allèle *R* a soit divergé à partir d'un allèle existant par mutation ou recombinaison soit il vient d'une autre population. **b.** Modèle "guerre des tranchées" : beaucoup d'allèles sont maintenus dans la même population par du polymorphisme balancé. Les allèles sont fréquents quand ils confèrent une résistance efficace à un temps donné vis-à-vis du pathogène avirulent correspondant; Lorsque de nouvelles formes virulentes émergent et contournent la résistance, les allèles *R* rares vont augmenter en fréquence et les autres, incapables de conférer la résistance à ces nouveaux pathogènes, vont diminuer en fréquence. **c.** Combinaison des 2 modèles : la population progresse à partir de cycles transitoires puis il y a un maintien des variations alléliques qui s'accumulent dans la population de l'hôte.



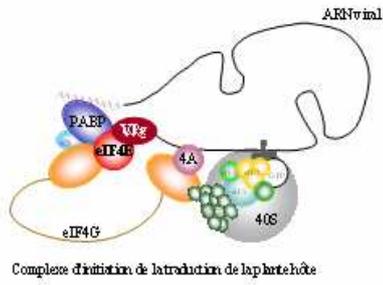
**Figure I.5. Etapes de l'initiation de la traduction chez les eucaryotes.**  
PABP, Poly-A Binding Protein; eIF, eucaryotic Initiation Factors.



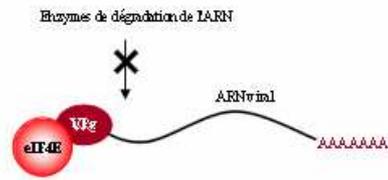
**Figure I.6. Alignement des séquences protéiques de eIF4E1 et eIF(iso)4E de *O. sativa* (riz), *T. aestivum* (blé), *L. esculentum* (tomate) et *A. thaliana* (Arabi) avec les eIF4E de *Homo sapiens* (homme) et *S. cerevisiae* (levure).**

Les acides aminés importants pour la stabilité de la structure de la protéine sont encadrés en noir. Les acides aminés impliqués dans la fixation à la coiffe des ARNm sont surlignés en orange. Les acides aminés "spécifiques des plantes" sont surlignés en vert. Les acides aminés surlignés en bleu sont ceux qui différencient les isoformes eIF4E et eIF(iso)4E (d'après Monzingo et al., 2007).

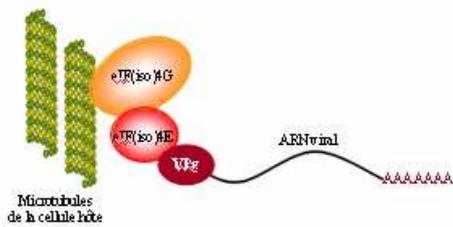
a. Initiation de la traduction de l'ARN viral



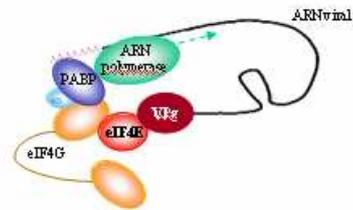
b. Stabilité du génome viral



c. Transport intracellulaire du génome viral

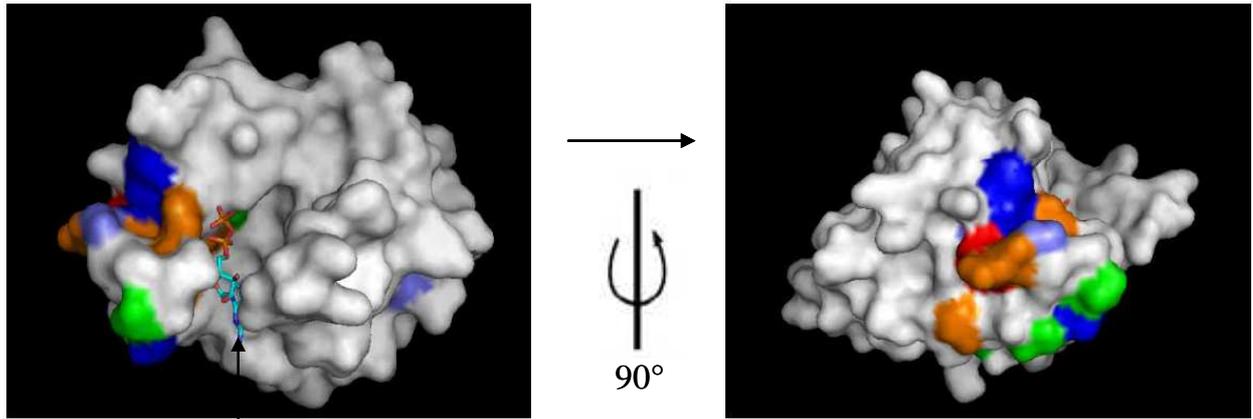


d. Réplication du génome viral



**Figure 1.7. Rôles biologiques potentiels de l'interaction eIF4E/VPg dans l'infection virale.**

- (a) Dans l'initiation de la traduction de l'ARN viral
- (b) Pour la stabilité du génome du potyvirus
- (c) Pour le transport intracellulaire du génome viral
- (d) Pour la réplication de l'ARN des potyvirus

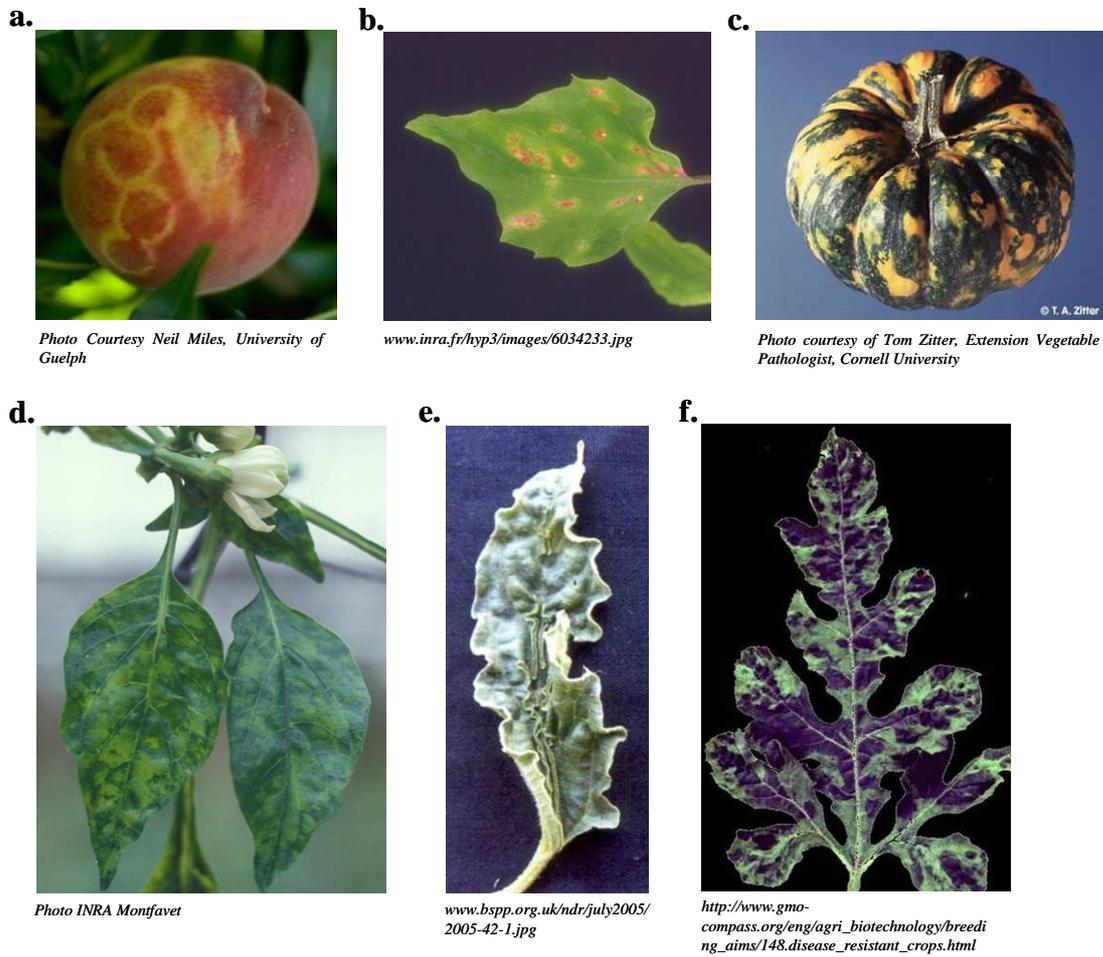


m<sup>7</sup>GDP dans la poche de fixation à la coiffe

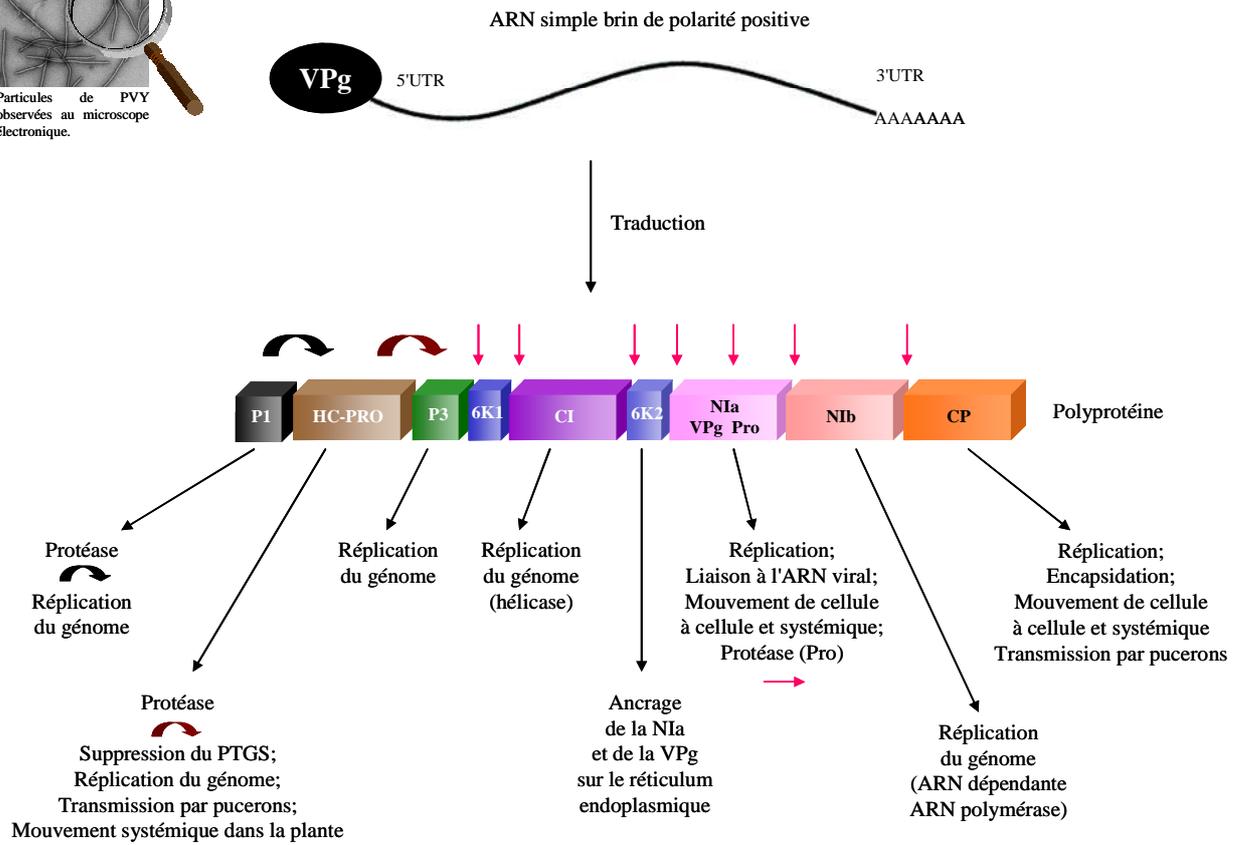
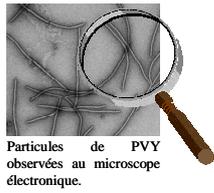
**Figure I.8. Localisation sur la structure tridimensionnelle eIF4E du blé des acides aminés impliqués dans la résistance aux potyvirus.**

La vue de gauche montre le m<sup>7</sup>GDP dans la poche de fixation à la coiffe. La vue de droite est une rotation de 90° à droite par rapport à l'axe vertical. Les acides aminés impliqués dans la résistance aux potyvirus sont en rouge pour la tomate (Ruffel *et al.*, 2005), en bleu pour la laitue (Nicaise *et al.*, 2003), en vert pour le pois (Gao *et al.*, 2004) et en orange pour le piment (Kang *et al.*, 2005; Ruffel *et al.*, 2002). D'après Monzingo *et al.* (2007).

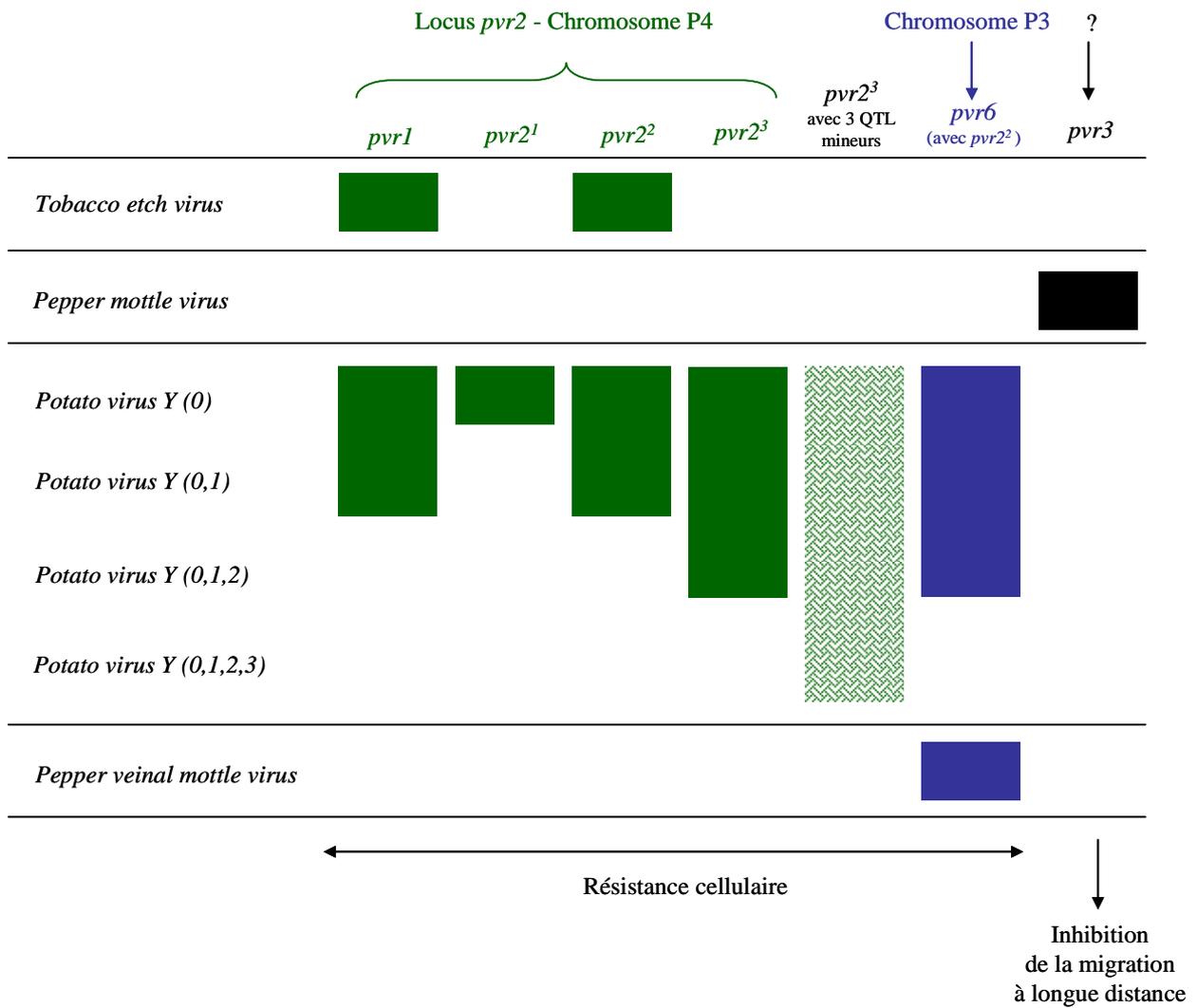




**Figure I.10. Exemples de symptômes causés par les potyvirus sur des plantes cultivées.**  
**a.** *Plum pox virus* (PPV) sur fruit du pêcher, **b.** *Pea seed borne mosaic virus* (PSbMV) sur feuille de pois, **c.** *Watermelon mosaic virus* (WMV) sur potiron, **d.** *Potato virus Y* (PVY) sur feuilles de piment, **e.** *Zucchini yellow mosaic virus* (ZYMV) sur feuille de melon et **f.** *Papaya ringspot virus* (PRV) sur feuille de papaye.

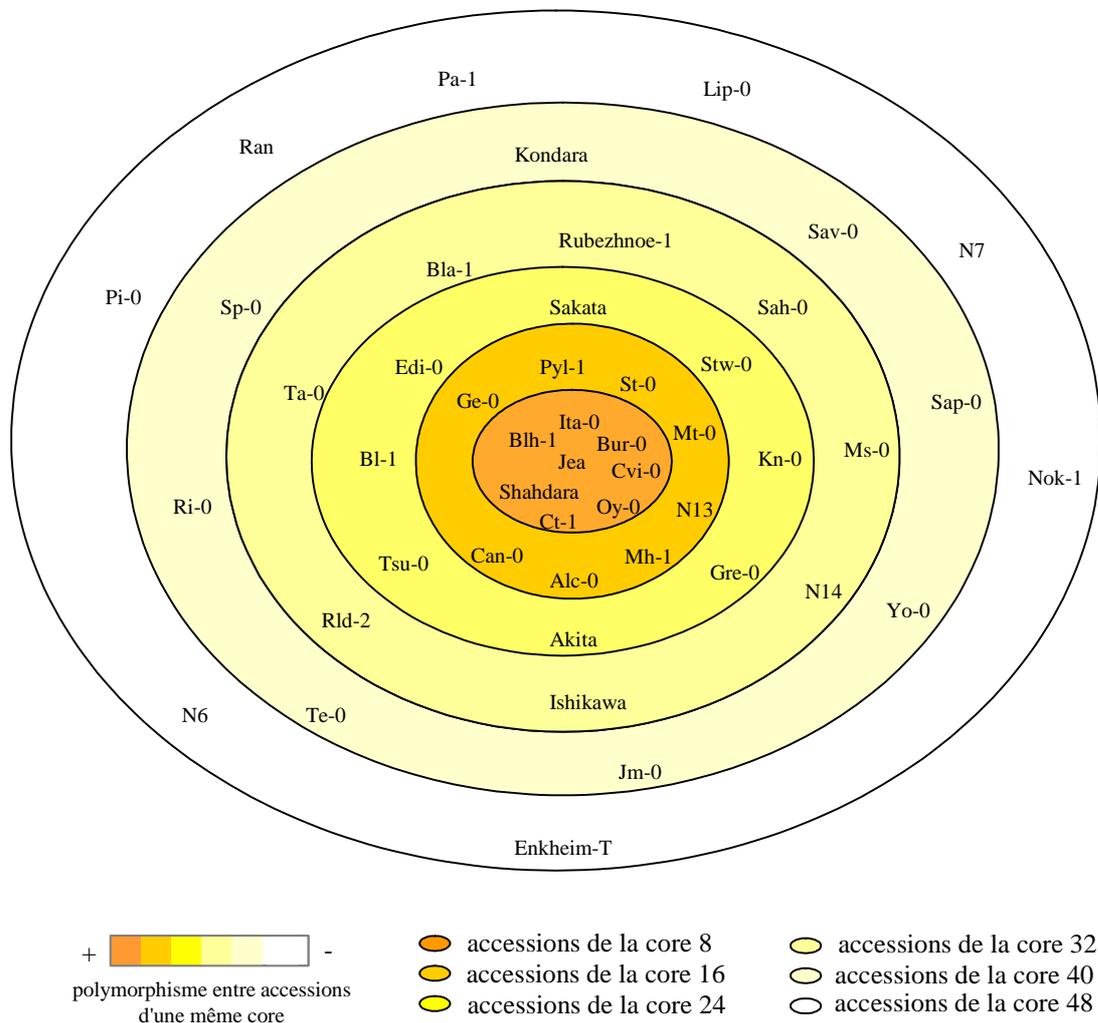


**Figure I.11. Organisation génomique, processus de maturation des potyvirus et fonctions attribuées aux protéines.**

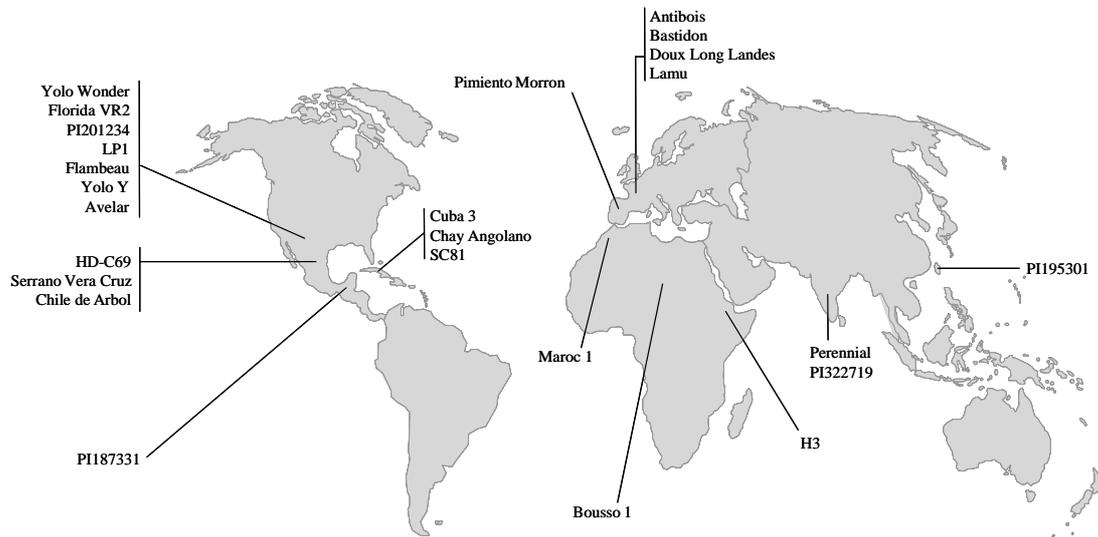


**Figure I.12. Facteurs récessifs de résistance aux potyvirus chez le piment : spectre d'action, mécanisme et localisation chromosomique.**

- Résistances qualitatives
- ▨ Résistance quantitative



**Figure I.13. Répartition, en fonction de leur degré de polymorphisme, des 48 accessions d'*Arabidopsis thaliana* de la core collection (d'après McKhann *et al.*, 2004).**



**Figure 1. Map showing origin of the 25 *C. annuum* genotypes used in this study.**

**Table 1. Amino acid substitutions observed between deduced eIF4E protein sequences from the twenty five *C. annuum* genotypes and their corresponding alleles.**

<i>C. annuum</i> accessions	Alleles	Position of amino acid substitutions								
		66	67	68	73	74	79	107	109	205
Yolo Wonder (YW)	pvr2 <sup>+</sup>	P	V	A	A	A	L	G	D	D
Flambeau	pvr2 <sup>+</sup>	- <sup>b</sup>	-	-	-	-	-	-	-	-
PI187331	pvr2 <sup>+</sup>	-	-	-	-	-	-	-	-	-
Cuba3	pvr2 <sup>+</sup>	-	-	-	-	-	-	-	-	-
Pimiento Morron	pvr2 <sup>+</sup>	-	-	-	-	-	-	-	-	-
Antibois	pvr2 <sup>+</sup>	-	-	-	-	-	-	-	-	-
Bastidon	pvr2 <sup>+</sup>	-	-	-	-	-	-	-	-	-
Doux Long Landes	pvr2 <sup>+</sup>	-	-	-	-	-	-	-	-	-
Lamu	pvr2 <sup>+</sup>	-	-	-	-	-	-	-	-	-
Bouso1	pvr2 <sup>+</sup>	-	-	-	-	-	-	-	-	-
H3	pvr2 <sup>+</sup>	-	-	-	-	-	-	-	-	-
Yolo Y (YY)	pvr2 <sup>1</sup>	-	E	-	-	-	R	-	-	-
Avelar	pvr2 <sup>1</sup>	-	E	-	-	-	R	-	-	-
Florida VR2 (F)	pvr2 <sup>2</sup>	-	E	-	-	-	R/L <sup>c</sup>	-	N	-
Chay Angolano	pvr2 <sup>2</sup>	-	E	-	-	-	R	-	N	-
Perennial	pvr2 <sup>3</sup>	-	E	-	-	-	-	-	-	G
HD-C69 <sup>a</sup>	pvr2 <sup>3</sup>	-	E	-	-	-	-	-	-	G
PI201234	pvr2 <sup>3</sup>	-	E	-	-	-	-	-	-	G
PI322719	pvr2 <sup>4</sup>	-	E	-	-	-	-	-	-	-
SC81	pvr2 <sup>5</sup>	T	E	-	D	-	-	-	-	-
Maroc1	pvr2 <sup>6</sup>	-	-	E	D	D	-	-	-	G
LP1	pvr2 <sup>6</sup>	-	-	E	D	D	-	-	-	G
Serrano Vera Cruz	pvr2 <sup>7</sup>	-	E	-	-	-	R	-	-	G
PI195301	pvr2 <sup>8</sup>	-	-	E	-	-	-	R	-	-
Chile de Arbol	pvr2 <sup>9</sup>	-	-	E	D	-	-	-	-	G

<sup>a</sup> HDC69 is a doubled-haploid line obtained from the F<sub>1</sub> hybrid (Criollo de Morelos 334 X YW) that possesses only a recessive resistance allele (from Criollo de Morelos 334) for resistance to PVY-0.

<sup>b</sup> amino acids identical to those of eIF4E from Yolo Wonder.

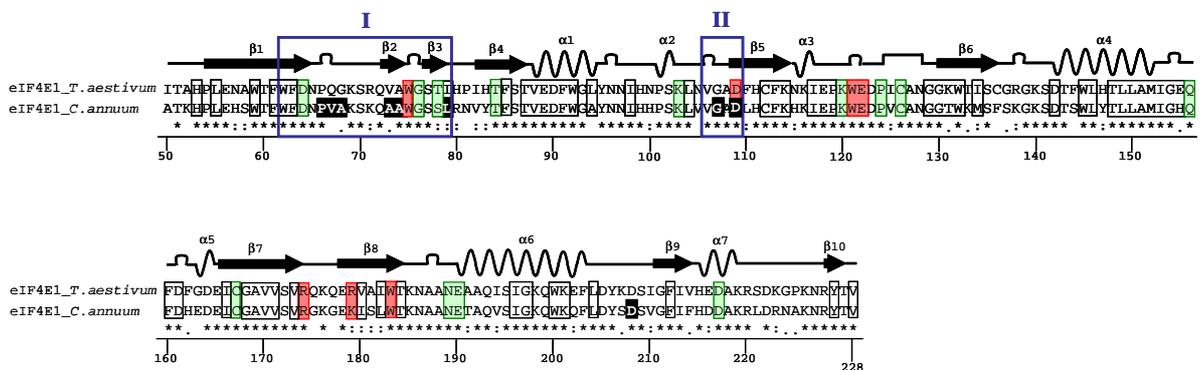
<sup>c</sup> additional sequencing of eIF4E cDNA from distinct Florida VR2 plants coming from various seed batches showed that amino acid at position 79 is either a L or a R.

**Table 2. Resistance phenotypes against PVY and TEV strains of pepper genotypes used in this study.**

Genotypes	eIF4E alleles	PVY- LYE84	PVY- SON41	TEV- HAT	TEV- CAA10
Yolo Wonder (YW)	pvr2 <sup>+</sup>	S <sup>a</sup>	S	S	S
Flambeau	pvr2 <sup>+</sup>	S	S	S	S
PI187331	pvr2 <sup>+</sup>	S	S	S	S
Cuba3	pvr2 <sup>+</sup>	S	S	S	S
Bousso1	pvr2 <sup>+</sup>	S	S	S	S
Pimiento Morron	pvr2 <sup>+</sup>	S	S	S	S
H3	pvr2 <sup>+</sup>	S	S	S	S
Doux Long Landes	pvr2 <sup>+</sup>	S	S	S	S
Lamu	pvr2 <sup>+</sup>	S	S	S	S
Bastidon	pvr2 <sup>+</sup>	S	S	S	S
Antibois	pvr2 <sup>+</sup>	S	S	S	S
Yolo Y (YY)	pvr2 <sup>1</sup>	<b>R</b>	S	S	S
Avelar	pvr2 <sup>1</sup>	<b>R</b>	S	S	S
Florida VR2 (F)	pvr2 <sup>2</sup>	<b>R</b>	S	<b>R</b>	S
Chay Angolano	pvr2 <sup>2</sup>	<b>R</b>	S	<b>R</b>	S
Perennial	pvr2 <sup>3</sup>	<b>R</b>	S	S	S
PI201234	pvr2 <sup>3</sup>	<b>R</b>	S	S	S
HD-C69 <sup>b</sup>	pvr2 <sup>3</sup>	<b>R</b>	S	S	S
PI322719	pvr2 <sup>4</sup>	<b>R</b>	S	S	S
SC81	pvr2 <sup>5</sup>	<b>R</b>	<b>R</b>	S	S
Maroc1	pvr2 <sup>6</sup>	<b>R</b>	<b>R</b>	S	S
LP1	pvr2 <sup>6</sup>	<b>R</b>	<b>R</b>	S	S
Serrano Vera Cruz	pvr2 <sup>7</sup>	<b>R</b>	<b>R</b>	S	S
PI195301	pvr2 <sup>8</sup>	<b>R</b>	<b>R</b>	S	S
Chile de Arbol	pvr2 <sup>9</sup>	<b>R</b>	<b>R</b>	S	S

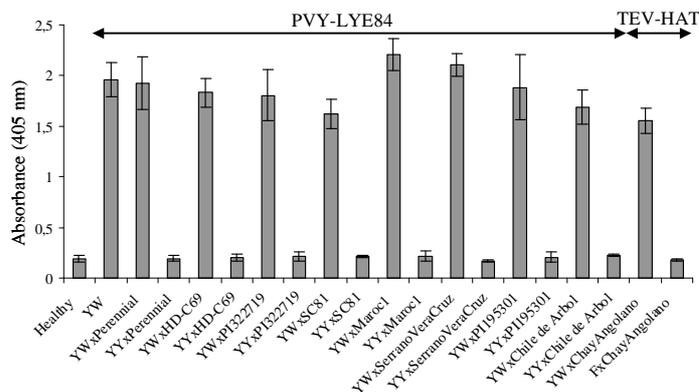
<sup>a</sup> R, resistant and S, susceptible

<sup>b</sup> HD-C69 is a doubled-haploid line obtained from the F<sub>1</sub> hybrid between Criollo de Morelos 334 and Yolo Wonder that possesses only a recessive resistance allele (from Criollo de Morelos 334) for resistance to PVY-0



**Figure 2. Alignment of amino acid sequences and secondary structure for wheat (*Triticum aestivum*) eIF4E and predicted eIF4E protein from *C. annuum* Yolo Wonder genotype.**

Amino acid positions are numbered according to the predicted eIF4E protein sequence from *C. annuum*. Polymorphic amino acids identified in this study are in black boxes. Blue open boxes indicate the two neighbouring regions previously demonstrated to be involved in resistance against potyviruses (Robaglia and Caranta, 2006). Amino acids involved in m<sup>7</sup>GTP binding are boxed in red. Amino acids important for stabilizing the structure of the protein are in black open boxes. "Plant-specific" amino acids (*i.e.*, conserved in more than 95% of plant species and different from those in other eukaryotes) are boxed in green (Monzingo *et al.*, 2007). Beta sheets are represented by arrows, alpha helix by oscillations and turns by open squares.



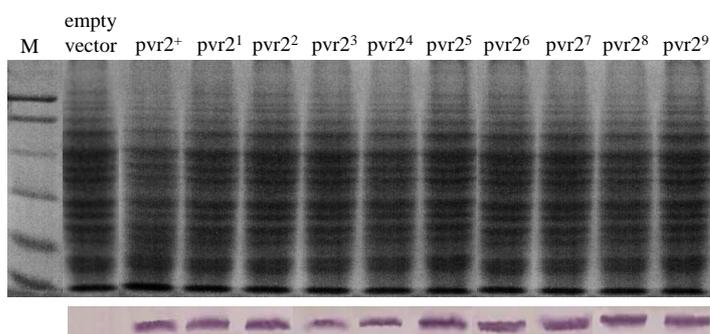
**Figure 3. Genetic analysis to examine the recessive nature of PVY-LYE84 or TEV-HAT resistance (F1 hybrids with YW-*pvr2+*) and allelism relationships with *pvr2* (F1 hybrids with YY-*pvr21* or with F-*pvr22*).**

F1 hybrids were inoculated with PVY-LYE84 or TEV-HAT and assayed for viral coat protein accumulation by DAS-ELISA at 30 dpi in non-inoculated leaves. The assay includes a healthy control (non-inoculated plants of YW) and a susceptible control (YW inoculated plants). DAS-ELISA was considered to be positive when the absorbance value at 405 nm of the sample was at least three times greater than the mean value of the healthy controls. Error bars indicate standard error.

**Table 3. Amino acid differences in the central part of the VPg of PVY [A] (Moury *et al.*, 2004) and TEV [B] strains used in this study. Numbers are amino acid positions in the VPg.**

[A]	Position of amino acid substitutions				
	101	105	115	119	123
PVY-LYE84	S	K	M	R	N
PVY-SON41	G	R	P	Y	S

[B]	Position of amino acid substitutions			
	111	112	113	115
TEV-HAT	I	E	P	S
TEV-CAA10	L	D	H	D

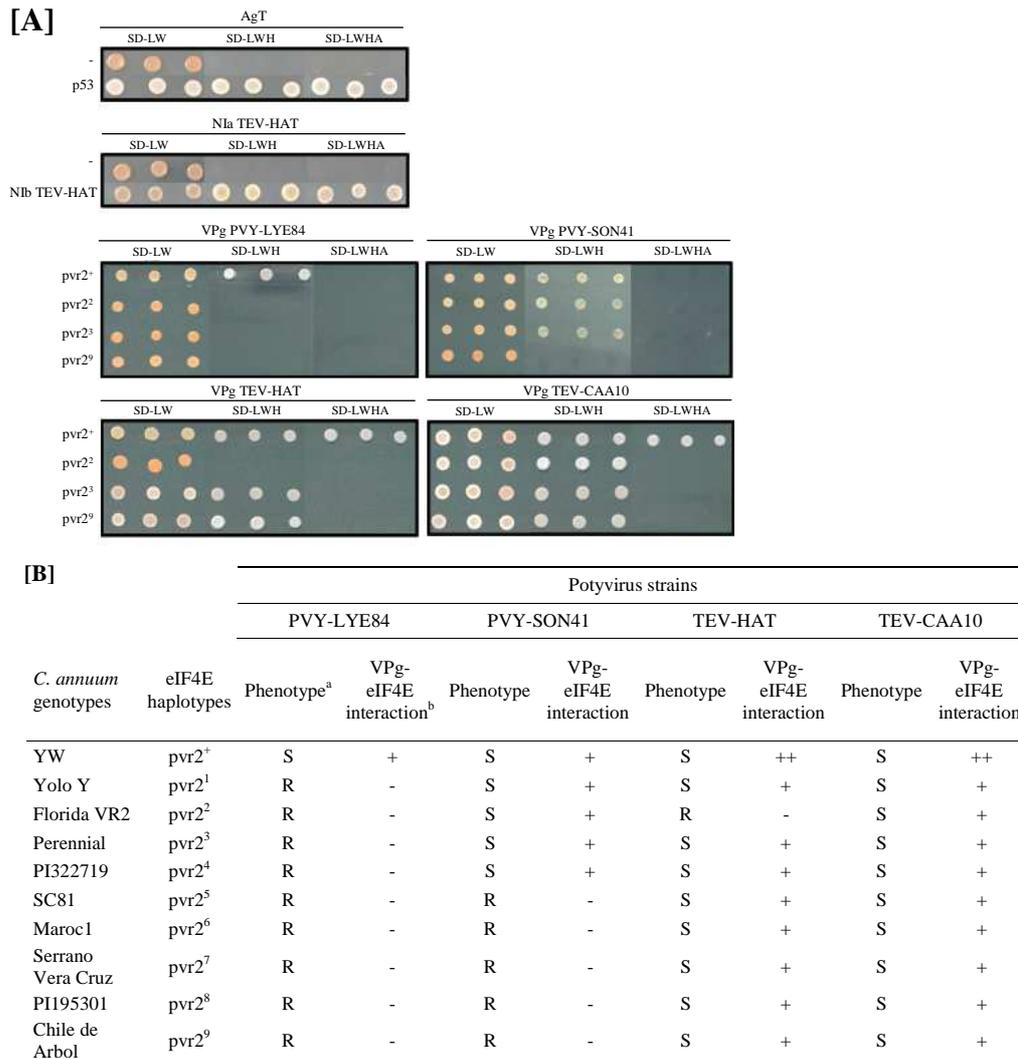


**Figure 4. Expression of eIF4E proteins in the yeast two-hybrid system.**

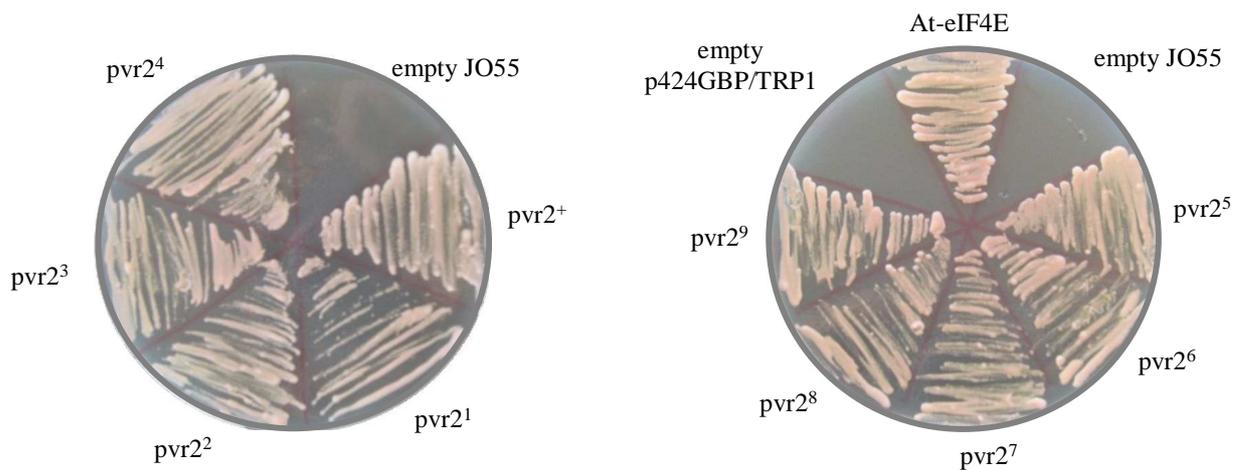
Lane 1: SDS PAGE molecular weight standard (Low Range, Bio-RAD); lane 2: yeast cells with empty vector; lanes 3 to 12: eIF4E fusion gene in pGADT7 vector.

[A] Yeast total protein extracts obtained using the Urea/SDS method.

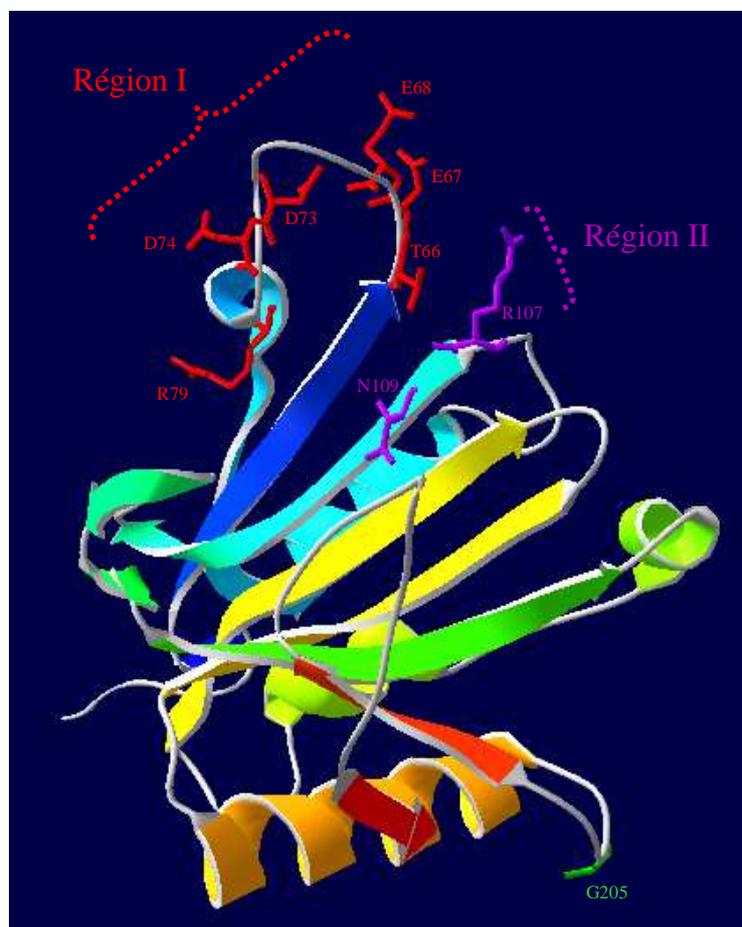
[B] Western blot analysis with the anti-HA antibody.



**Figure 5. eIF4E protein-VPg protein interaction assay using the yeast two-hybrid system.** [A] Examples of eIF4E-VPg protein interactions detected by yeast two-hybrid system. Yeast expressing both “bait” and “prey” recombinant proteins were obtained by first transforming yeast cells with the individual plasmid construction followed by separate conjugation between yeast expressing eIF4E proteins and those expressing VPg proteins. Conjugations were then cultured on control plates (-LW) and on two selective media lacking leucin, tryptophan and histidine (-LWH) and leucin, tryptophan, histidine and adenine (-LWHA). [B] Summary of interactions detected by the yeast two-hybrid system between *C. annuum* eIF4E proteins and VPg proteins together with the resistance phenotype against PVY and TEV strains controlled by the different allelic variants at the *pvr2* locus. <sup>a</sup>S:susceptible; R: resistant; <sup>b</sup> -: no interaction; +: interaction on SD-LWH media; ++: interaction on SD-LWHA media.

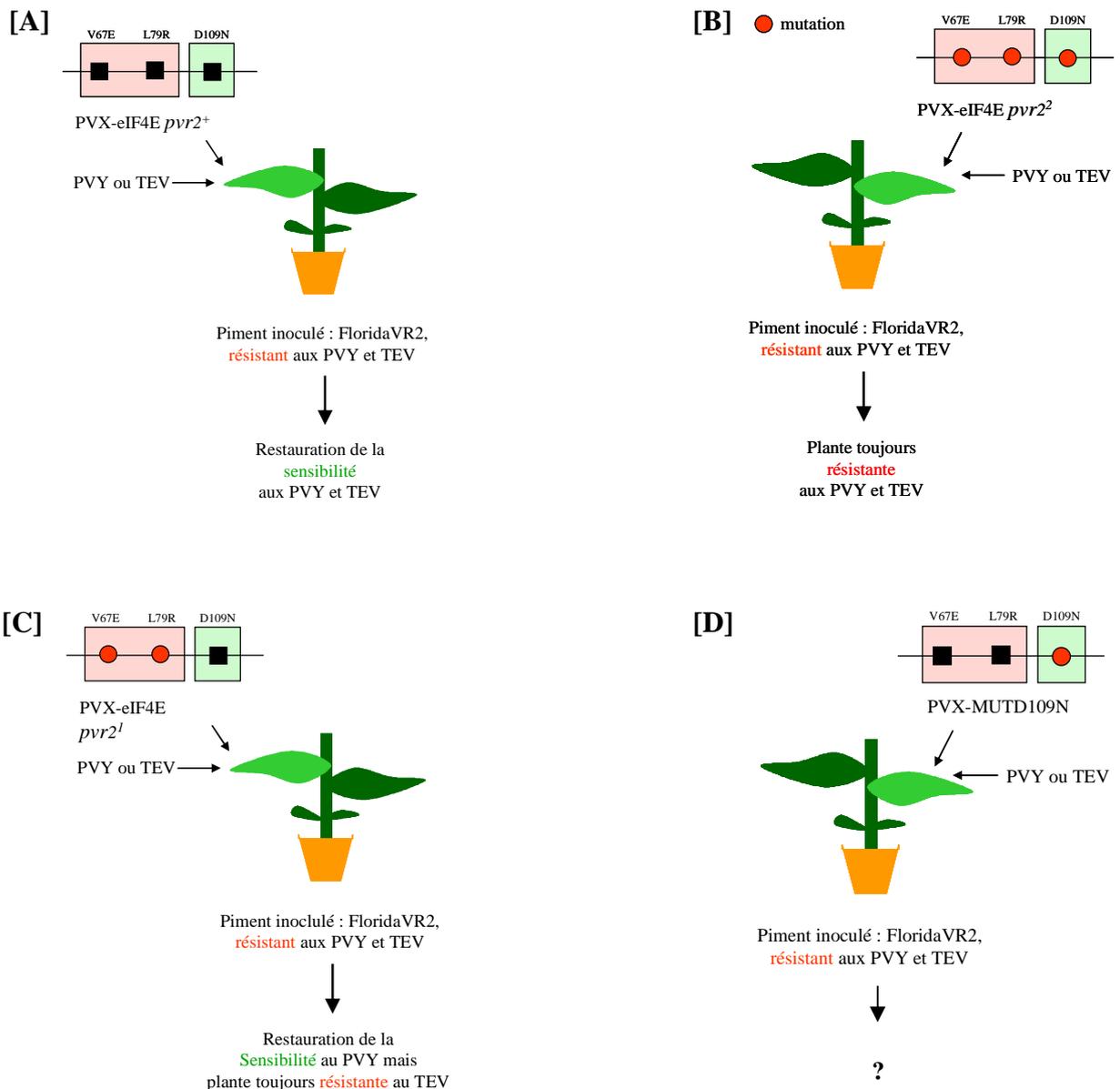


**Figure 6. Complementation of yeast strain JO55 with *Capsicum annuum eIF4E* cDNAs.** The yeast strain JO55 was transformed with the plasmids p424GBP/TRP1:*eIF4E* (pvr2<sup>1</sup> to pvr2<sup>9</sup>). Control transformations were done with no DNA (empty JO55), the empty plasmid p424GBP/TRP1 (negative controls) and with p424GBP/TRP1:*At-eIF4E* (At4g18040, positive control). Equal amounts of transformed yeast cells were plated on appropriate selective nutrient drop-out media containing glucose. Cells were grown for 10 days at 30°C.



**Figure II.1. Position des substitutions en acides aminés identifiées chez eIF4E1-pvr2 sur la structure 3D prédite de la protéine.**

Cette prédiction est basée sur la structure cristallisée de la protéine eIF4E de blé *T. aestivum* (Protein Data Bank accession 2IDR; Monzingo *et al.*, 2007), réalisée *via* SWISS-MODEL (Guex and Peitsch, 1997). Les mutations sont insérées et visualisées en utilisant le logiciel DeepView-SWISS pdbViewer (Guex *et al.*, 1995). Les extrémités N- et C-terminales sont représentées en bleu foncé et en rouge respectivement.



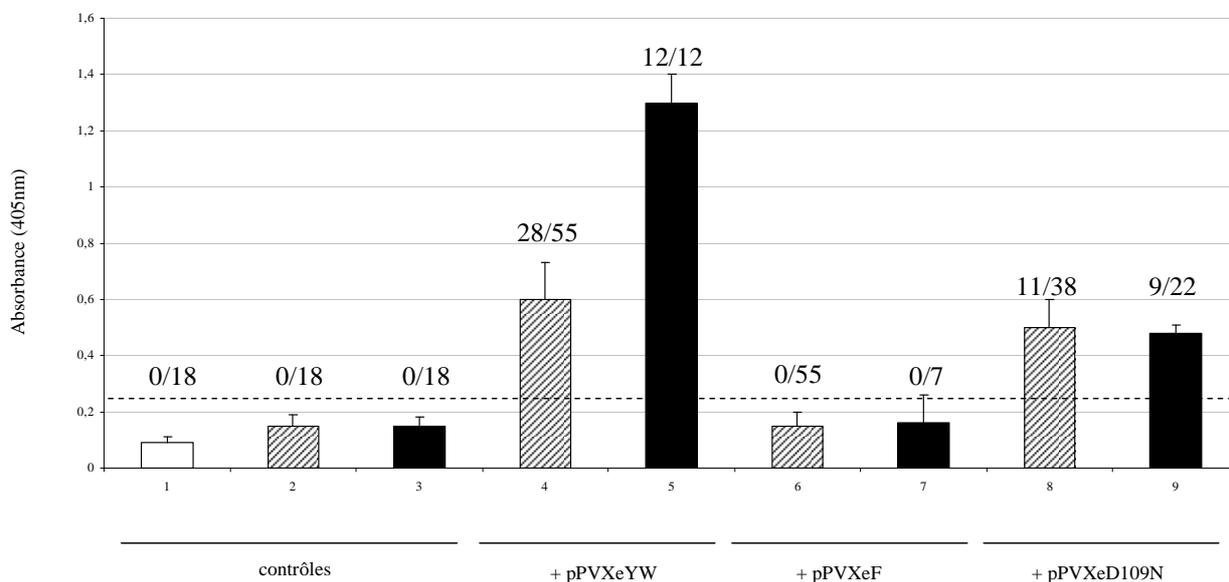
**Figure II.2. Principe de l'expérience de complémentation fonctionnelle par expression transitoire d'ADNc *eIF4E* via le vecteur viral PVX (Baulcombe *et al.*, 1995).**

**[A]** L'inoculation de l'ADNc *eIF4E1-pvr2*<sup>+</sup> sur une feuille de piment résistant au PVY et au TEV restaure la sensibilité à ces 2 potyvirus  $\Rightarrow$  *pvr2*<sup>+</sup> est un allèle de sensibilité aux PVY et TEV.

**[B]** L'inoculation de l'ADNc *eIF4E1-pvr2*<sup>2</sup> sur une feuille de piment résistant au PVY et au TEV ne restaure pas la sensibilité à ces 2 potyvirus  $\Rightarrow$  *pvr2*<sup>2</sup> est un allèle de résistance aux PVY et TEV.

**[C]** L'inoculation de l'ADNc *eIF4E1-pvr2*<sup>1</sup> sur une feuille de piment résistant au PVY et au TEV restaure la sensibilité au PVY mais pas au TEV  $\Rightarrow$  *pvr2*<sup>1</sup> est un allèle de sensibilité au PVY et de résistance au TEV.

**[D]** Expérience d'inoculation de l'ADNc *eIF4E-pvr2*<sup>+</sup> portant la mutation D109N introduite par mutagenèse dirigée pour déterminer si ce mutant est un allèle de sensibilité ou de résistance au PVY et/ou au TEV.



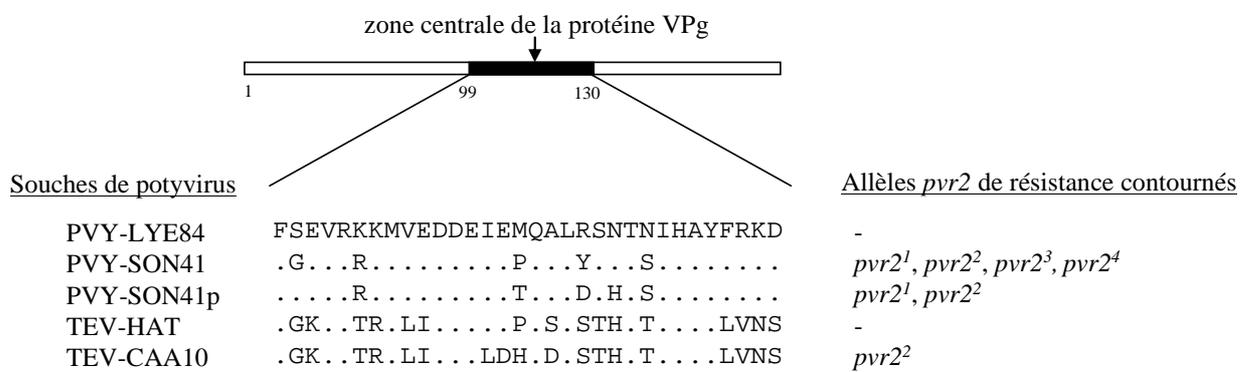
**Figure II.3. Détection de l'accumulation des protéines de capsid du TEV-S103 et du PVY-LYE72 par DAS-ELISA dans les expériences d'expression transitoire de eIF4E.**

Les histogrammes représentent les mesures d'accumulation virale au cours de 3 tests indépendants de complémentation fonctionnelle par expression transitoire d'ADNc *eIF4E*, via le vecteur viral PVX (Baulcombe *et al.*, 1995), chez l'accession Florida VR2 de *C. annuum* (*pvr2<sup>2</sup>*), résistante aux TEV et PVY.

La ligne pointillée désigne la valeur d'absorbance à 405 nm au-dessus de laquelle les échantillons sont considérés comme infectés (3 fois la valeur moyenne des témoins sains). Les barres blanche, hachurées et noires représentent l'accumulation virale dans des plantes respectivement non-inoculées, inoculées avec du TEV-S103 ou avec du PVY-LYE72. Les plantes témoin sont co-inoculées avec les constructions PVX vides (2 et 3).

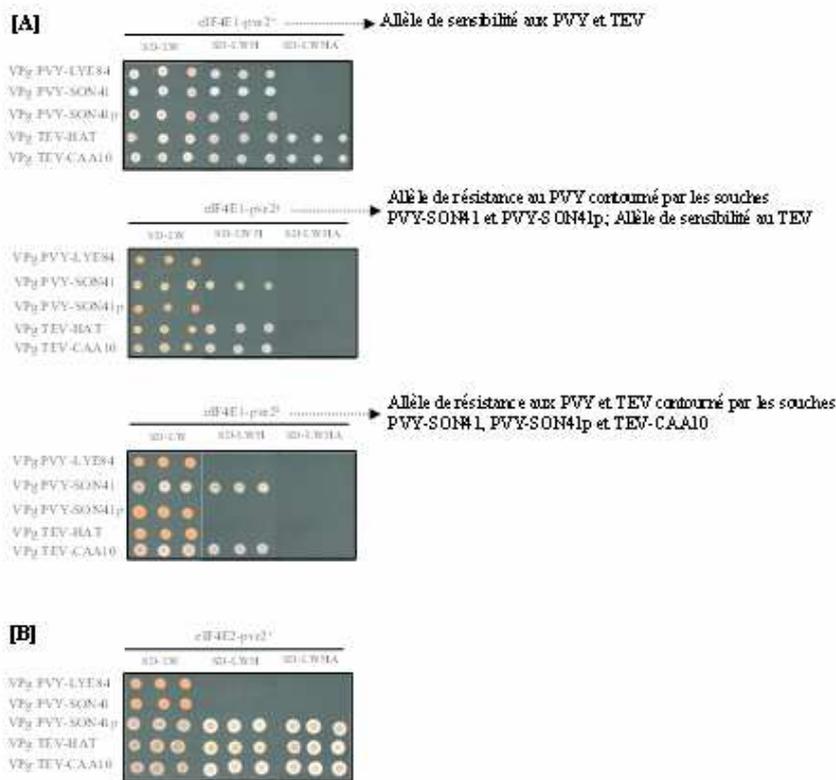
Au-dessus des histogrammes sont indiqués les rapports "nombre de feuilles infectées sur nombre de feuilles inoculées".





**Figure II.5. Alignement de la zone centrale de la protéine VPg des 5 souches de potyvirus utilisées.**

A droite sont indiqués les allèles de résistance contournés.



**Figure II.6. Interactions entre les protéines eIF4E1 ou eIF4E2 et les VPg de souches virulentes et avirulentes de PVY et de TEV détectées par le système de double hybride dans la levure.**

Les interactions fortes sont détectées sur milieu sélectif sans leucine, tryptophane, histidine et adénine (-LWHA) et les interactions plus faibles sur milieu sans leucine, tryptophane et histidine (-LWH).

**[A]** Interactions entre les protéines eIF4E1 codées par les allèles *pvr2<sup>+</sup>*, *pvr2<sup>+</sup>* et *pvr2<sup>+</sup>* et les VPg des souches PVY-LYE84, PVY-SON41, PVY-SON41p, TEV-HAT et TEV-CAA10.

**[B]** Interactions entre la protéine eIF4E2 et les VPg des 5 souches de potyvirus.

**Tableau III.1. Matériel végétal utilisé dans cette étude.**  
**[A] *Capsicum* spp.; [B] *Lycopersicon* spp.; [C] *A. thaliana***

Espèces	Nombre d'accessions	Accessions	Codes	Origines géographiques
<i>C. annuum</i>	25	Yolo Wonder	CanYW	USA
		Antibois	CanAT	France
		Pimiento morron	CanPM	Espagne
		Chite de Arbol	CanCA	Mexique
		Bastidon	CanBA	France
		Yolo Y	CanYY	USA
		Florida VR2	CanAV	USA
		Perennial	CanPE	Inde
		Avelar	CanAV	USA
		Serrano Vera Cruz	CanSVC	Mexique
		Chay Angolano	CanCHA	Cuba
		SC81	CanSC	Cuba
		Maroc1	CanM1	Maroc
		PI322719	Can32	Inde
		PI195301	Can19	Taiwan
		PI201234	Can20	USA
		LP1	CanLP	USA
		Flambeau	CanFL	USA
		CM334	CanCM	Mexique
		PI187331	Can18	Guatemala
		Cuba3	CanCU	Cuba
		Bousso1	CanB1	Tchad
		Doux Long des Landes	CanDLL	France
		Lamu	CanLA	France
		H3	CanH3	Ethiopie
<i>C. annuum var glabriusculum</i>	7	Turrialba (C81)	CagTU	Costa Rica
		Managua CB10	CagMA	Nicaragua
		Antigua (390)	CagAN	Guatemala
		Chihuahua (504bp)	CagCH	Mexique
		El Tajin	CagET	Mexique
		Panama city	CagPA	Panama
		Pitalito	CagPI	Colombie
<i>C. chinense</i>	2	PI260438	Chi38	Bolivie
		Hmong Rouge	ChiHR	Guyane
<i>C. frutescens</i>	3	Capsicum sauvage	CfuCS	Mexique
		Boca vieja	CfuBV	Costa Rica
		Grain de café	CfuGC	Amérique
<i>C. chacoense</i>	2	PI260426	Cha26	Argentine
		PI260437	Cha37	Bolivie
<i>C. baccatum</i>	2	PI230461	Cha61	Bolivie
		PI439367	Cha67	Bresil
<i>C. pubescens</i>	2	Rocoto	CpuRO	Bolivie
		C. pubescens 54P	Cpu54	Pérou
<i>C. eximium</i>	2	Exi2	Cex2	Nicaragua
		Exi3	Cex3	Bolivie
<i>C. cardenasii</i>	3	Ulla Pica	CcaUP	Bolivie
		La Paz	CcaLP	Bolivie
		Quilacollo	CcaQU	Bolivie
<i>C. galapogense</i>	1	C. galapogense	Cga1	Iles Galapagos
<i>C. microcarpum</i>	1	C. microcarpum SA 326	CmiSA	Pérou - Bolivie
<i>C. praetermissum</i>	1	EX1 x PRA 20-201	Cpr20	Bresil

Espèces	Nombre d'accessions	Accessions	Codes	Distribution géographique et écologique <sup>1</sup>			
<i>L. esculentum</i> ( <i>Solanum lycopersicum</i> L.) <sup>2</sup>	20	Momor	LesMOM	Forme exclusivement cultivée présente dans divers habitats écologiques du monde entier.			
		Microtom	LesMIC				
		M82	LesM82				
		Plovdivska Conserva	LesPLO				
		Nagarlan	LesNAG				
		San Marzano	LesMRZ				
		Saint Pierre	LesPIE				
		Marmade	LesMRM				
		Gardener's Delight	LesGAR				
		Poivron des Andes	LesPOI				
		Alisa Craig	LesALI				
		Sweetie	LesSWE				
		Plovdiv XXIVa	LesPOV				
		WVA106	Les106				
		Yellow pear	LesYEL				
		Carabio	LesCAR				
		Levovil	LesLE				
		Ferum	LesFE				
		Mospomorist	LesMOS				
		Cervil	LesCE				
		<i>L. pimpinellifolium</i> ( <i>S. pimpinellifolium</i> L.) <sup>2</sup>	9		LA 1279	Lp179	Originaire de l'Equateur Central jusqu'au Chili Central, jusqu'à 500 m d'altitude.
					LA 0121	Lp121	
					LA 1582	Lp182	
					LA 1478	Lp178	
					63280	Lp180	
66063	Lp163						
LA1589	Lp189						
Wva 700	LpWVA						
Hirsute	LpHIR						
<i>L. hirsutum</i> ( <i>S. habrochaites</i> S.) <sup>2</sup>	10			B	LhB	Dans les forêts de l'Ouest des Andes, de 500 à 2500 m d'altitude.	
		G1290	Lh90				
		87610012	Lh79				
		10496	Lh104				
		PI247087 (PI24)	Lh24				
		PI134417 (PI13)	Lh13				
		LA1777	Lh77				
G1-1560	Lh60						
H2	LhH2						
PI390660 (PI39)	Lh39						
<i>L. pennellii</i> ( <i>S. pennellii</i> ) <sup>2</sup>	2	LA716	Lpe716	Nord du Pérou et Nord du Chili, de 0 à 3000 m d'altitude.			
		Clayberg	LpeCLA				
<i>L. chilense</i> ( <i>S. chilense</i> ) <sup>2</sup>	2	LA1969	Lcl69	Côte Est des Andes, jusqu'à 600 m d'altitude.			
		LA1971	Lcl71				
<i>L. cheesmanii</i> ( <i>S. cheesmanii</i> ) <sup>2</sup>	2	LA1401	Lcm01	Endémique aux Iles Galapagos, jusqu'à 500 m d'altitude.			
		LA1450	Lcm50				
<i>L. peruvianum</i> ( <i>S. peruvianum</i> L.) <sup>2</sup>	2	CMV sel INRA	LprCMV	Pérou Central et Nord du Chili, jusqu'à 600 m d'altitude.			
		PI126435	Lpr126				
<i>L. chmielewskii</i> ( <i>S. chmielewskii</i> ) <sup>2</sup>	2	LA1317	Lcw17	Dans les vallées désertiques des Andes, entre 2300 et 2880 m.			
		LA1318	Lcw18				
<i>L. parviflorum</i> ( <i>S. neorickii</i> ) <sup>2</sup>	1	LA1321	Lpa21	Dans les vallées désertiques des Andes, entre 1950 et 2600 m d'altitude.			

<sup>1</sup>http://sofkb.cit.oregonstate.edu/about/solanum\_nomenclature.pl; <sup>2</sup>Nouvelle nomenclature *Solanum* (Spooner et al., 2005).

**Tableau III.1. suite - Matériel végétal utilisé dans cette étude.**  
**[A] *Capsicum* spp.; [B] *Lycopersicon* spp.; [C] *A. thaliana***

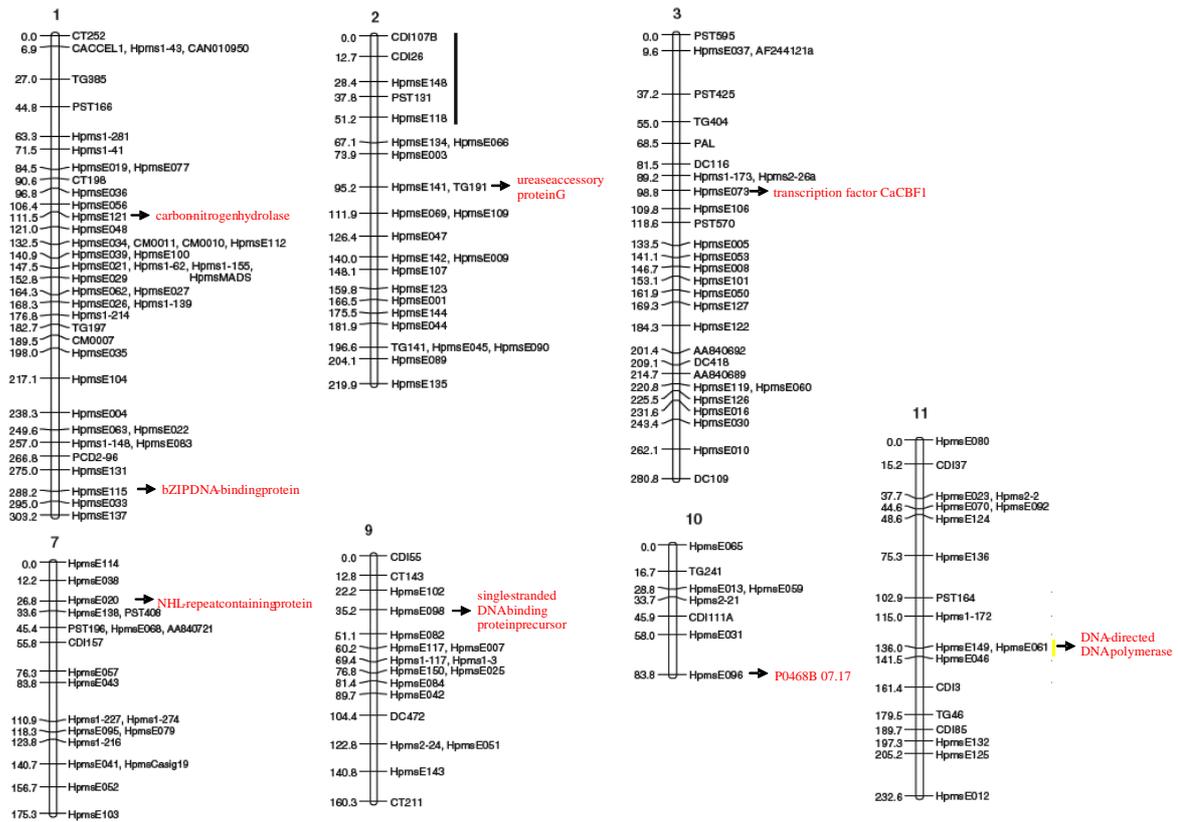
Noms	Numéros d'identification	Cores <sup>1</sup>	Origines géographiques	Longitudes, Latitudes
Ran	21AV	48	Estuaire de la Rance, France	W2°00', N48°3'
Jea	25AV	8	St Jean Cap Ferrat, France	E7°20', N43°41'
Pi-0	40AV	48	Pitztal / Tirol, Autriche	E10°51', N47°04'
Bl-1	42AV	24	Bologna, Italie	E11°20', N44°29'
Pa-1	50AV	48	Palermo, Italie	E13°22', N38°07'
Ta-0	56AV	32	Tabor, République Tchèque	E14°40', N49°25'
St-0	62AV	16	Stockholm, Suède	E18°03', N59°19'
Te-0	68AV	40	Tenela, Finlande	E23°18', N60°04'
Ka-0	70AV	24	Kaunas, Lituanie	E23°54', N54°54'
Bla-1	76AV	32	Blanes / Gerona, Espagne	E2°48', N41°41'
Edi-0	83AV	24	Edinburgh, Royaume-Uni	W3°13', N50°57'
Tsu-0	91AV	24	Tsu, Japon	E29°19', N34°19'
Stw-0	92AV	24	Stobowa / Orel, Russie	E36°04', S2°57'
Ri-0	160AV	40	Richmond (B.C.), Canada	E22°58', N49°1'
Ct-1	162AV	8	Catania, Italie	E15°06', N37°3'
Can-0	163AV	16	Unknow, Espagne	W15°30', N28°00'
Cvi-0	166AV	8	Iles du Cap Vert	W24°00', N16°00'
Bur-0	172AV	8	Burren, Irlande	W9°04', N53°07'
Alc-0	178AV	16	Alcala de Henares, Madrid, Espagne	W3°22', N40°29'
Bih-1	180AV	8	Bulhary, République Tchèque	E16°45', N48°49'
Col-0	186AV	8	Gorzow Wielkopolski, Pologne	E15°15', N52°44'
Kondara	190AV	40	Kondara, Tadjikistan	E68°49', N38°48'
Gre-0	200AV	24	Greenville (MD), USA	W85°15', N43°11'
Oy-0	224AV	8	Oystese, Norvège	E6°13', N60°23'
Rid-2	229AV	32	Rschew, Russie	E34°19', N56°15'
Rubezhnoe-1	231AV	32	Rubezhnoe, Ukraine	E37°19', N48°13'
Sah-0	233AV	32	Sierra Alhambra, Espagne	E3°07', N38°52'
Sap-0	234AV	40	Slapy, République Tchèque	E14°24', N49°49'
Sav-0	235AV	40	Slavice, République Tchèque	E12°56', N49°49'
Shakdara	236AV	24	Shakdara River, Tadjikistan	E71°30', N37°29'
WS	244AV	40	Wassilewskija, Biélorussie	E30°38', N52°13'
Akita	252AV	24	Akita Pref., Japon	E40°06', N39°43'
N6	262AV	48	Karelian region, Russie	E34°00', N62°00'
N7	263AV	48	Pinguba, Russie	E34°34', N61°51'
N14	267AV	32	Sampo Mountain, Russie	E34°34', N61°62'

<sup>1</sup>à partir des données de séquence de 10 locus, des collections embouteillées ont été générées, composée chacune de 8 accessions pour atteindre un total de 48. Les accessions regroupées dans la core 8 représentent le maximum de diversité génétique pour ces 10 locus de l'espèce *A. thaliana*.

**Tableau III.2. Amorces utilisées pour amplifier et séquencer les gènes *eIF4E* et les locus indépendants.**

Genre ou espèce	Matériel génétique	Gène	Amorce sens	Amorce antisens		
<i>Capsicum</i> spp.	ADNc	<i>Ca-eIF4E1</i>	AAAAGCACACAGCACCAACA	TATTCGACATTGCATCAAGAA		
		<i>Ca-eIF4E2</i>	GATAGTTTGTGTGAGGATTCAC	GCAACCGCAAGTGACTGCTGAT		
	ADNg	<i>Ca-eIF(iso4E)</i>	ATGGCCACCGAAGCACCACCACCGG	TCACACGGTGTATCGGCTCTTAGCT		
		<i>Ca-eIF4E1</i>	GAAGATCCTGTATGTGCCAATG	CTGTGTAACGATTCTTTGC		
		<i>Ca-eIF4E2</i>	GAGGATCCTGTATGTGCCAATGGAG	CTATACGGGAATAACGATTCTTG		
		<i>Ca-eIF(iso)4E</i>	GAGTGCCTAATGGTGGCAAGTG	TCACACGGTGTATCGGCTC		
		<i>Ca-HpmsE121</i> <sup>1</sup>	TGAGAGTTGGATCCCTCTTCTCG	CTTGTCGGGCTTGAAAGAGGA		
		<i>Ca-HpmsE115</i> <sup>1</sup>	TCATCTCATAGCCTGCCCCCTA	CCACTTGAAGAAGCCATGACCA		
		<i>Ca-HpmsE141</i> <sup>1</sup>	TCCCAACAACCTCAAATGGCTTC	TGGAGGTGCCCTTCTGGTAAAG		
		<i>Ca-HpmsE073</i> <sup>1</sup>	TTATTCAGGCCCACTTATCGAA	CAGCAGCCAAATCTTGATTTTC		
		<i>Ca-HpmsE020</i> <sup>1</sup>	CCCCCGAGAGGAACAGAATCAT	TTCCATTTTGGTCCAGTACCA		
		<i>Ca-HpmsE098</i> <sup>1</sup>	CCACAAACTCCCTAAACCCACT	CAAACCCCTTCAACTTCTCTCA		
		<i>Ca-HpmsE096</i> <sup>1</sup>	CGGGTCAAACAAAACCGAAGT	GCTTGTGGTGTAGCTCGCTCTT		
		<i>Ca-HpmsE061</i> <sup>1</sup>	CCCAAGAAAGAAGTTGGGAATGG	TTCGACGAGCTTGAAGGTTGAT		
		<i>Ca-eF1 α</i>	ATGGGTAAAGAGAAGATTTC	TCACTTCCCTTCTTCTCG		
		<i>Ca-H2B-1</i>	GGCCCGACAAGAAGAAGAAGA	CTGCCTTGGTTCCTTCAGAAA		
		<i>Ca-rbcL</i>	ATGTCACCACAAACAGAGAC	CTCACAGCAGCAGCTAGT		
		<i>Lycopersicon</i> spp.	ADNc	<i>Ly-eIF4E1</i>	CTGAAATGGAGAGAACGATGT	CACTGCATCAAGAACTATACGG
				<i>Ly-eIF4E2</i>	GGGACGAAAACACCAAAAATG	CCCTGTGTGAACGATAGAAGTA
				<i>Ly-eIF(iso4E)</i>	GACGGGAAAACAATGGCCACCG	ATTGACGAGACAAGCCG
ADNg	<i>Ly-eIF4E1</i> <sup>2</sup>		GAAGATCCTGTATGTGCCAATG	CTATACGGTGTAAACGATTCTTG		
	<i>Ly-eIF4E2</i> <sup>2</sup>		GAGGATCCTGTATGTGCCAATGGAG	CTACACGGTGTAAACGATTTTG		
	<i>Ly-eIF(iso)4E</i> <sup>2</sup>		GAGTGTGCCAATGGTGGCAAGTG	CTACACAGTATATCGGCTCTTAG		
	<i>CT066</i> <sup>3</sup>		CGCTGTCCCTCTTACCACC	AATGTCTGCTCCACTTTCGCTAC		
	<i>CT093</i> <sup>3</sup>		CTCCCTCGGCTACAGCATT	AGCAGCCCTTCAGAACGGACT		
	<i>CT148</i> <sup>3</sup>		TCGTGTCAAGGCTGTGTAGT	GGTTGCTTCCCAGATAGTTGT		
	<i>CT179</i> <sup>3</sup>		CGAATTCATCTCCACACTCA	TAAGACCAGCCAAACTACCAC		
	<i>CT189</i> <sup>3</sup>		CAAGATCCCCTGACCACTG	TGGAGGCAGCAAGAAGTGTGA		
	<i>CT198</i> <sup>3</sup>		TGACAACTACCGAATTACGA	GGTGATTTATTTAGTGCCACA		
	<i>CT208</i> <sup>3</sup>		CTATGGAGTTATATTTTACCACA	ACTTTTGAGAGGACATCAATTT		
	<i>Sucr</i> <sup>3</sup>		GGACCATTGTATCACAAGGGATG	CAAAGGGAAGGATTGAATATTAGC		
	<i>A. thaliana</i>		ADNg	<i>At-eIF4E1</i>	GAGAAAGAGAAGCAGTTCCG GATGAATTGCTATGTTGGTG <sup>4</sup>	GACAGACTGTCAAGTCATG
				<i>At-eIF4E2</i>	GGAGCATAGTGAACGGGAAAA CCGTTGCTTCTTGGATTT <sup>4</sup>	TTTACAGTAAAGATAGAGCAAC
				<i>At-eIF4E3</i>	CGGAGCAAAGTGAACGGGAAAAG GTATCTGTACATAAACCTG <sup>4</sup>	GTAAAGATAGATCAGAGGCAAC
<i>At-eIF(iso)4E</i>		GCCGTCTCAGAAGAAAACCTCACTGCG GGAGGAAAGTGGACTTGGGTTGTTACTGC <sup>4</sup>		CCGACCAACAGTATCAC		
<i>At-nCBP</i>		CTTCTGATTGGAATTTTAGGGC CTAGGTTGAAGGATTTTGGCCTG <sup>4</sup>		GCCTCCGGTACAGTATGGTATTGG		

<sup>1</sup>Yi *et al.*, 2006; <sup>2</sup>sans l'intron 1; <sup>3</sup>Roselius *et al.*, 2005; <sup>4</sup>Amorces internes pour le séquençage



**Figure III.1. Répartition des différents locus "non-eIF4E" sur le génome du piment (d'après Yi *et al.*, 2006).**

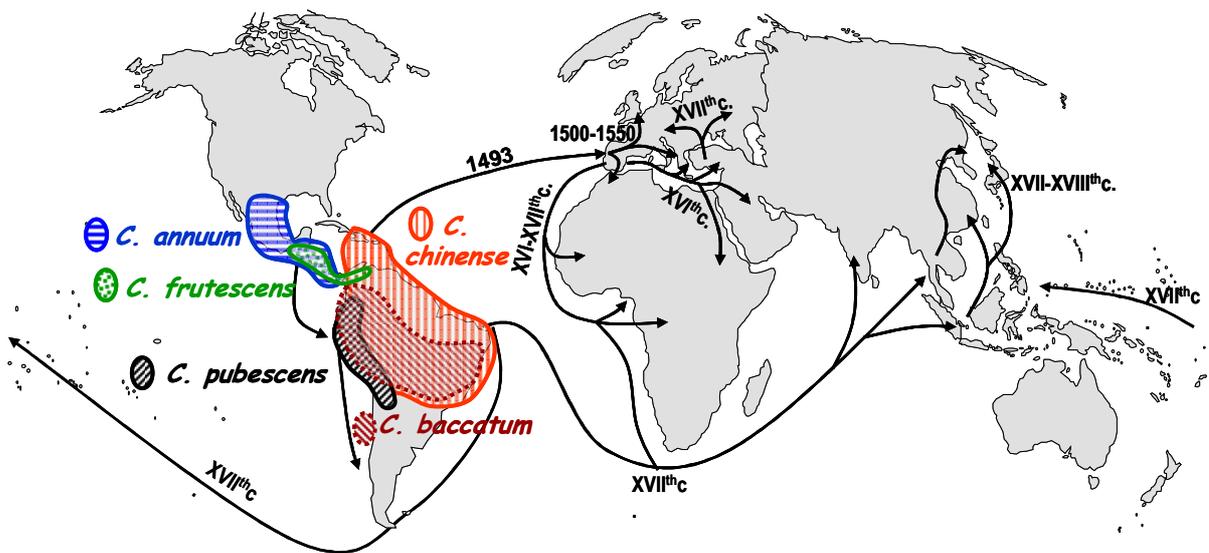
Les protéines codées par les gènes en question sont indiquées en rouge.

**Tableau III.3. Localisation chromosomique et protéines codées par les locus indépendants.**

[A] chez *Lycopersicon* spp. , d'après Roselius *et al.* (2005)

[B] chez *A. thaliana*, d'après McKhann *et al.* (2004)

[A]			[B]		
Locus	Chromosome	Protéine codée	Nom du marqueur	Chromosome	Protéine codée
<i>CT066</i>	10	Arginine decarboxylase	125C F	1	Acetyl CoA synthétase
<i>CT093</i>	5	S-adenosylmethionine decarboxylase proenzyme	67D R	1	Auxin responsive protein
<i>CT148</i>	8	Copper / zinc superoxyde dismutase	100D F	2	Chorismate mutase precursor
<i>CT179</i>	3	Tonoplast intrinsic protein Δ-type	49A R	2	Cycloartenol synthase
<i>CT189</i>	12	40S ribosomal protein S19	DFR	3	Dihydroflavonol reductase
<i>CT198</i>	9	Submergence induced protein 2-like	82A F	3	Enoyl CoA hydratase
<i>CT208</i>	9	alcohol dehydrogenase class III	133C R	4	Geranylgeranyl reductase
<i>Sucr</i>	3	Vacuolar invertase	92B R	4	Long-chain fatty acid CoA ligase
			76D F	5	Methionyl tRNA synthetase like
			115A R	5	Peptidase



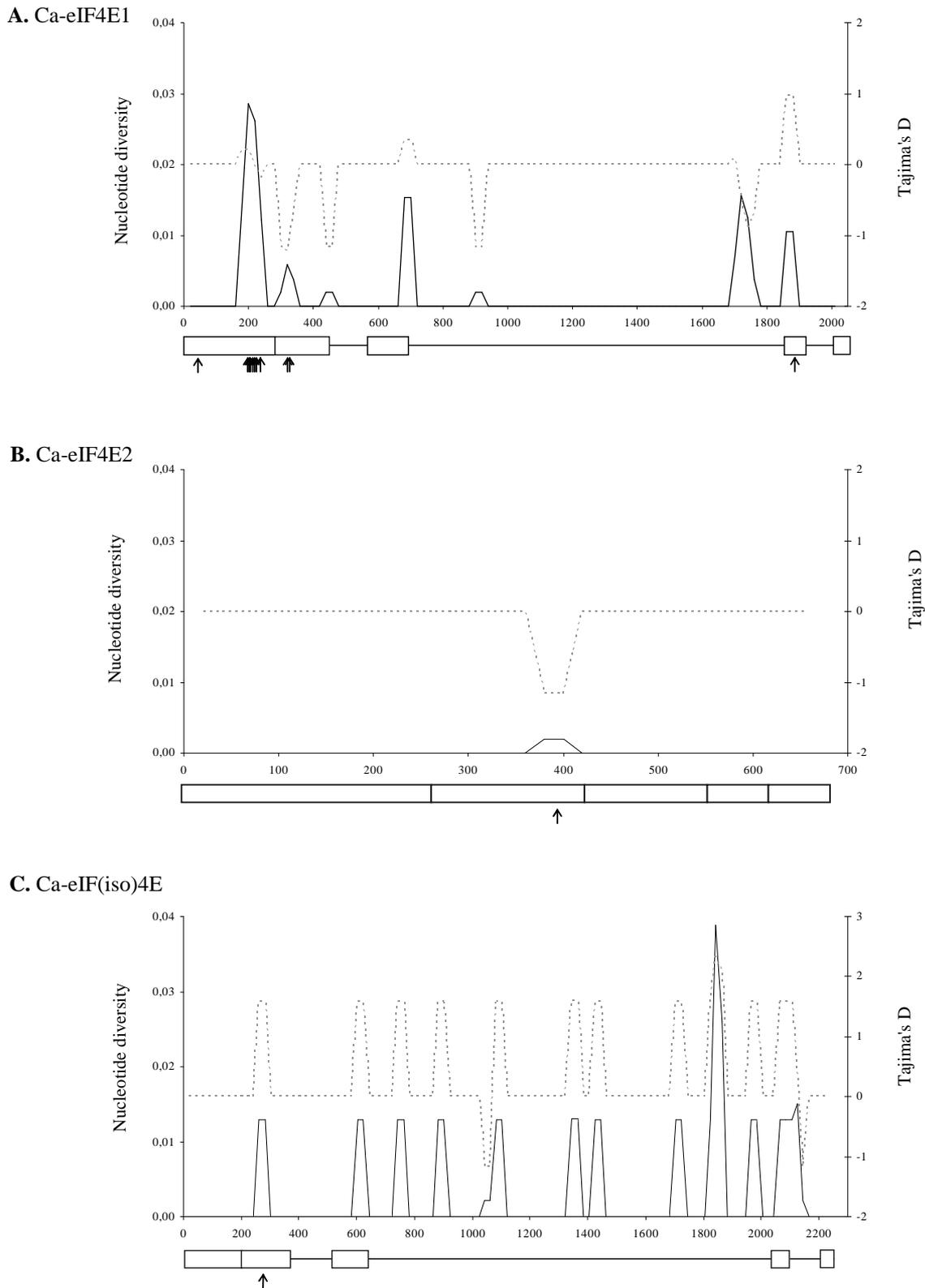
**Figure III.2. Aires d'origine et de domestication présumées des 5 espèces cultivées de *Capsicum* spp.**

Les principales routes de diffusion du piment à travers le monde entre le XV<sup>ème</sup> et le XVIII<sup>ème</sup> siècle sont représentées par des flèches (Dijan-Caporalino *et al.*, 2006).

**Tableau III.4. Diversité nucléotidique des 3 gènes *Ca-eIF4E* et des 11 locus indépendants chez 25 accessions de *C. annuum*.**

Gènes	Régions	Nombre de sites	Nombre de sites en ségrégation	Singletons	$\pi$ par site $\times 10^{-4}$	$\pi_s$ (sites) $\times 10^{-4}$	$\pi_a$ (sites) $\times 10^{-4}$	$\pi_a/\pi_s$	$\theta_w$ par site $\times 10^{-4}$	D de Tajima	Signification statistique du D de Tajima
<i>Ca-eIF4E1</i>	codant	687	9	2	33,2	0,0 (139,93)	41,9 (544,07)	-	34,7	-0,141	>0,1
	noncodant <sup>1</sup>	1335	7	2	11,6				15,9	-0,847	>0,1
	Total	2022	16	4	19,0				22,3	-0,528	>0,1
<i>Ca-eIF4E2</i>	codant	672	1	1	1,2	0,0 (145,46)	1,5 (523,54)	-	3,9	-1,158	>0,1
	noncodant	ND	ND	ND	ND				ND	ND	ND
	Total	ND	ND	ND	ND				ND	ND	ND
<i>Ca-eIF(iso)4E</i>	codant	609	3	0	29,3	93,1 (111,00)	12,5 (411,00)	0,134	15,1	2,351	<0,05*
	noncodant <sup>1</sup>	1633	13	2	36,0				21,4	2,380	<0,05*
	Total	2242	16	2	34,4				19,9	2,609	<0,01**
<i>Elongation factor 1-<math>\alpha</math></i>	codant	1250	0	0	-	0,0000	0,0000	0,000	-	-	-
	noncodant	-	-	-	-	-	-	-	-	-	-
	Total	1250	0	0	-	-	-	-	-	-	-
<i>RuBisCo large subunit</i>	codant	1279	3	0	5,2	0,0 (313,29)	8,2 (964,71)	-	6,2	-0,420	>0,1
	noncodant	-	-	-	-	-	-	-	-	-	-
	Total	1279	3	0	5,2	-	-	-	6,2	-0,420	>0,1
<i>Histone H2B-1</i>	codant	237	4	2	34,9	157,6 (52,49)	0,0 (184,51)	0,000	44,7	-0,593	>0,1
	noncodant	-	-	-	-	-	-	-	-	-	-
	Total	237	4	2	34,9	-	-	-	44,7	-0,593	>0,1
<i>HpmsE121</i>	codant	199	4	3	26,1	48,3 (54,85)	30,7 (143,15)	0,636	53,2	-1,376	>0,1
	noncodant	-	-	-	-	-	-	-	-	-	-
	Total	199	4	3	26,1	-	-	-	53,2	-1,376	>0,1
<i>HpmsE115</i>	codant	-	-	-	-	-	-	-	-	-	-
	noncodant	218	2	0	17,4	-	-	-	24,6	-0,640	>0,1
	Total	218	2	0	17,4	-	-	-	24,6	-0,640	>0,1
<i>HpmsE141</i>	codant	266	2	1	11,5	55,0 (40,00)	4,5 (176,00)	0,082	20,4	-0,941	>0,1
	noncodant	-	-	-	-	-	-	-	-	-	-
	Total	266	2	1	11,5	-	-	-	20,4	-0,941	>0,1
<i>HpmsE073</i>	codant	-	-	-	-	-	-	-	-	-	-
	noncodant	239	3	1	21,8	-	-	-	33,7	-0,879	>0,1
	Total	239	3	1	21,8	-	-	-	33,7	-0,879	>0,1
<i>HpmsE020</i>	codant	-	-	-	-	-	-	-	-	-	-
	noncodant	204	1	0	25,0	-	-	-	13,2	1,473	>0,1
	Total	204	1	0	25,0	-	-	-	13,2	1,473	>0,1
<i>HpmsE098</i>	codant	191	3	2	27,8	33,4 (48,49)	16,6 (137,51)	0,497	41,0	-0,796	>0,1
	noncodant	-	-	-	-	-	-	-	-	-	-
	Total	191	3	2	27,8	-	-	-	41,0	-0,796	>0,1
<i>HpmsE096</i>	codant	226	4	4	14,3	14,8 (54,17)	1,5 (161,83)	0,101	47,5	-1,886	<0,05*
	noncodant	-	-	-	-	-	-	-	-	-	-
	Total	226	4	4	14,3	-	-	-	47,5	-1,886	<0,05*
<i>HpmsE061</i>	codant	276	10	0	12,7	299,3 (61,03)	78,7 (211,97)	0,263	96,3	1,062	>0,1
	noncodant	-	-	-	-	-	-	-	-	-	-
	Total	276	10	0	12,7	-	-	-	96,3	1,062	>0,1

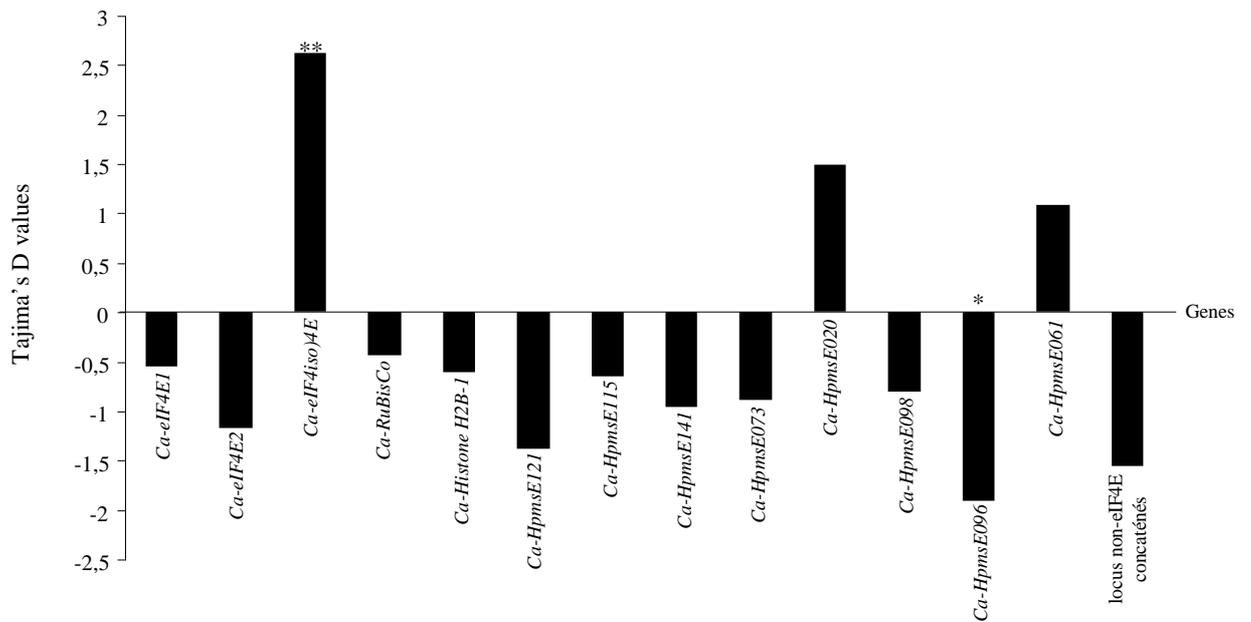
<sup>1</sup>introns 2, 3 et 4



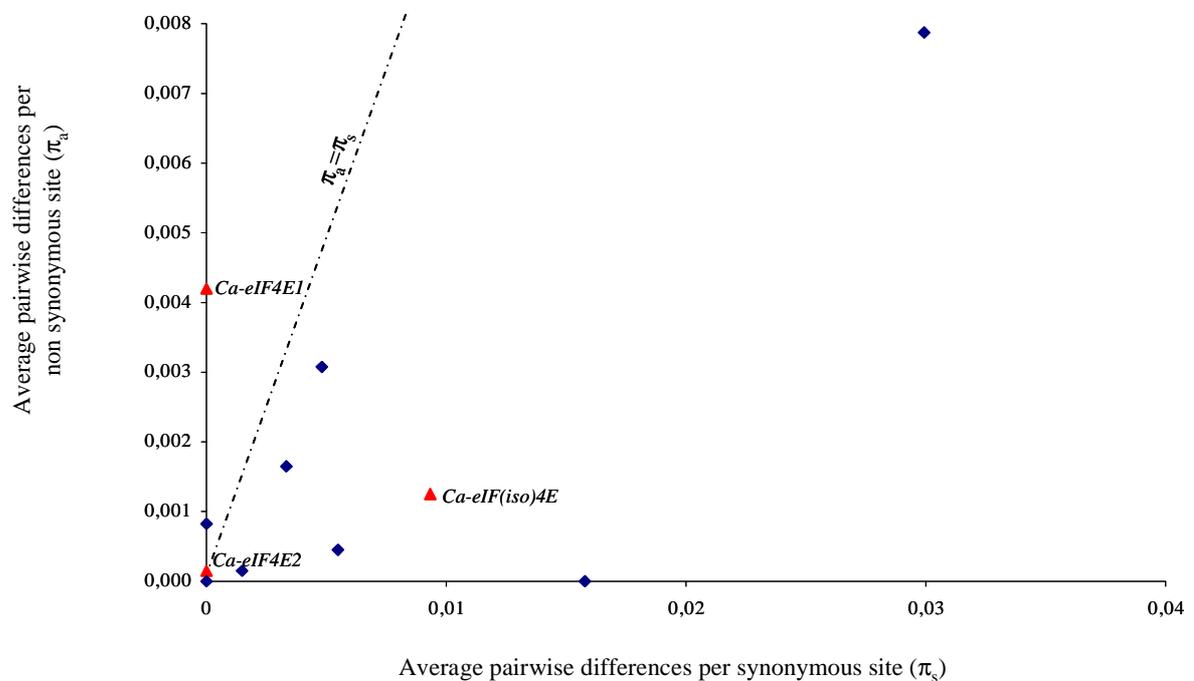
**Figure III.3. Structure des 3 gènes *Ca-eIF4E* et analyse par fenêtre glissante de la diversité nucléotidique (traits pleins) et des valeurs de *D* de Tajima (traits pointillés).**

Les exons sont représentés par les rectangles sous le graphe. Les flèches indiquent la position des substitutions en acides aminés.

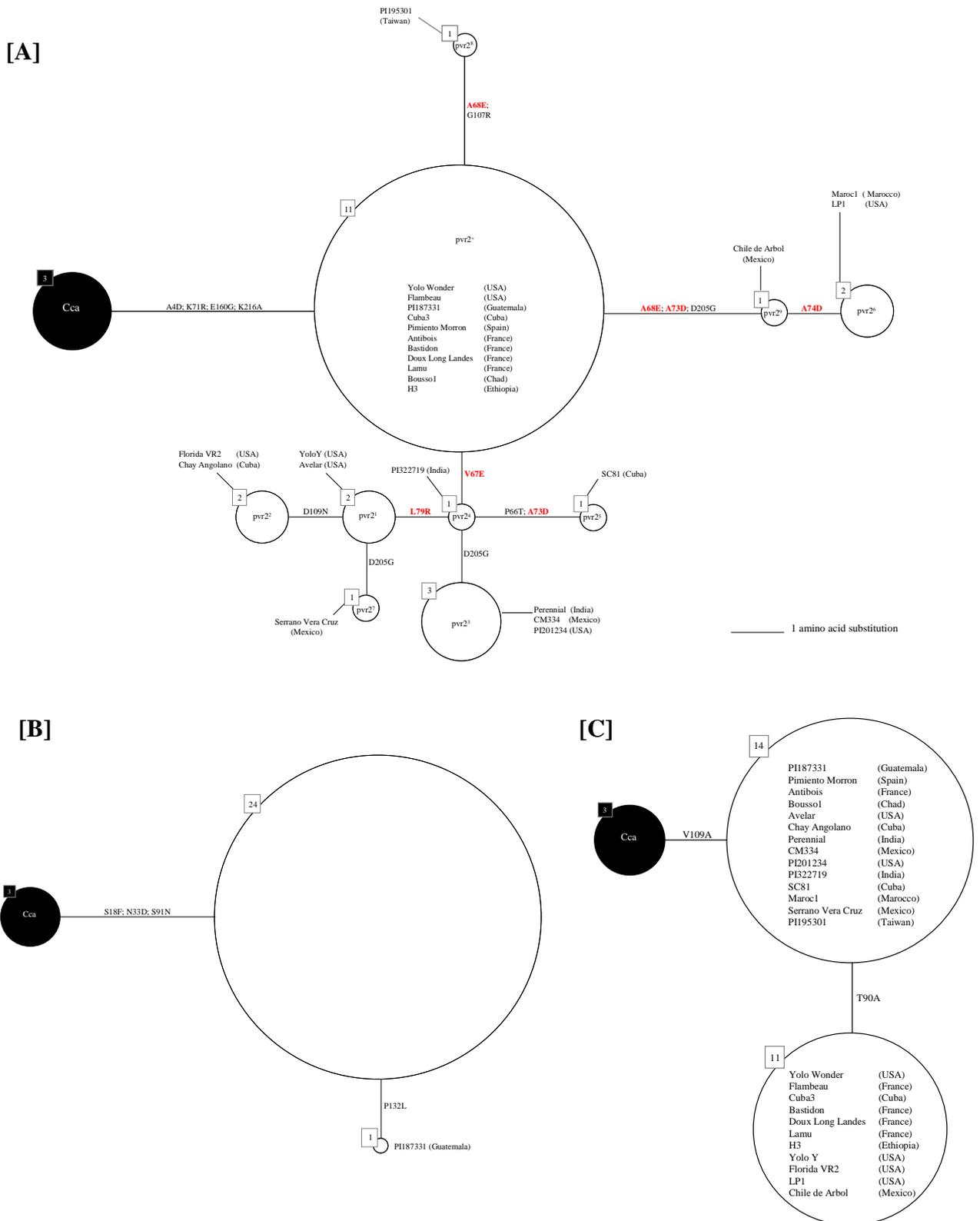
[A] *Ca-eIF4E1*; [B] *Ca-eIF4E2*; [C] *Ca-eIF(iso)4E*.



**Figure III.4. Valeurs de D de Tajima pour les 3 gènes *Ca-eIF4E* et les 11 locus indépendants chez 25 accessions de *C. annuum*.**  
 Les astérisques indiquent le seuil de significativité des D de Tajima obtenus : 5% (\*) and 1% (\*\*).



**Figure III.5. Rapports de  $\pi_a$  sur  $\pi_s$  calculés pour les locus indépendants (losanges bleus) et les 3 gènes *Ca-eIF4E* chez 25 accessions de *C. annuum* (triangles rouges).**



**Figure III.6. Arbres généalogiques enracinés représentant la structure haplotypique des gènes *Ca-eIF4E1* [A], *Ca-eIF4E2* [B] et *Ca-eIF(iso)4E* [C] chez 25 accessions de *C. annuum*.**

Chaque haplotype est représenté par un cercle proportionnel à sa fréquence (indiquée dans le carré adjacent). La racine de l'arbre (cercle noir) est constituée des 3 accessions de l'espèce *C. cardenasii*. Les substitutions en acides aminés pour passer d'un haplotype à l'autre sont indiquées sur les branches. Les acides aminés sous sélection positive sont indiqués en rouge.

**Tableau III.5. Test d'hétérogénéité des modèles M7 et M8 pour l'estimation du paramètre  $\omega$  par site.**

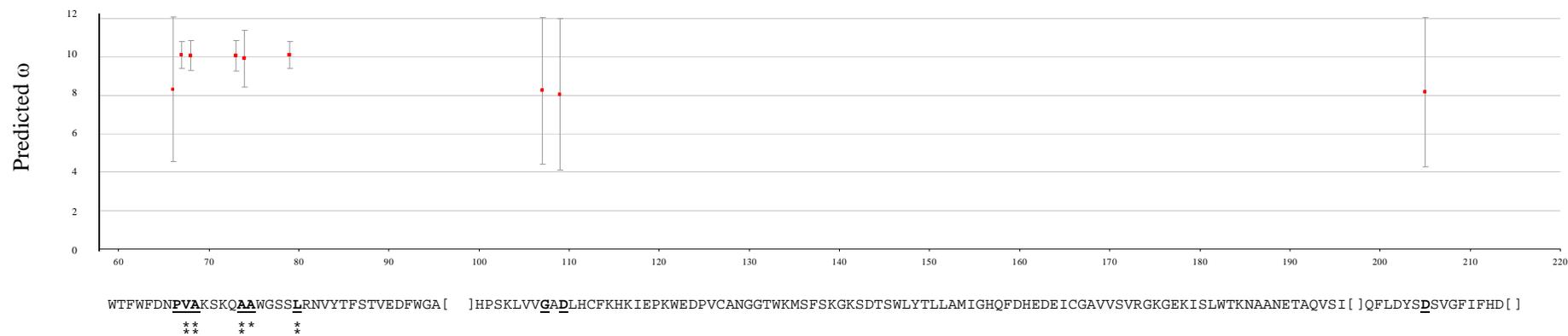
Gènes	Modèles	$p$	Paramètres $\beta$	$\omega_s$	$p_s$	logL	LRT	$\chi^2$
<i>Ca-eIF4E1</i>	M7 ( $\beta$ )	2	0.006, 0.018	NE	NE	-1313.43	53.28	2.69E-12
	M8 ( $\beta + \omega > 1$ )	4	0.016, 0.037	20.45	0.024	-1286.79		
<i>Ca-eIF4E2</i>	M7 ( $\beta$ )	2	0.005, 0.014	NE	NE	-1076.67	15.10	5.26E-04
	M8 ( $\beta + \omega > 1$ )	4	0.006, 0.009	61.71	0.005	-1069.12		
<i>Ca-eIF(iso)4E</i>	M7 ( $\beta$ )	2	0.005, 0.015	NE	NE	-932.00	0.28	8.69E-01
	M8 ( $\beta + \omega > 1$ )	4	0.005, 1.965	1.79	0.159	-931.86		

$p$  : nombre de paramètres du modèle; NE : non estimé dans le modèle;  $\omega_s$  : valeur de  $\omega$  maximum;  $p_s$  : proportion de codons évoluant pour une valeur  $\omega_s$ ; logL : log vraisemblance; LRT : test statistique de rapport de vraisemblance.

**Tableau III.6. Sites sous sélection positive détectés par le modèle M8.**

Protéines	Sites sous sélection positive	$p_{\omega>1}$	$\omega$ estimé	Ecart type
<i>Ca-eIF4E1</i>	66P	0.810	8.30	3.77
	67V	1,000 **	10.10	0.70
	68A	0.999 **	10.10	0.78
	73A	0.998 **	10.10	0.79
	74A	0.980 *	9.91	1.47
	79L	1,000 **	10.10	0.70
	107G	0.804	8.23	3.82
	109D	0.784	8.04	3.95
	205D	0.796	8.16	3.87
<i>Ca-eIF4E2</i>	132P	0.782	5.76	3.48
<i>Ca-eIF(iso)4E</i>	90A	0.775	3.20	2.43

\* $P > 95\%$ ; \*\* $P > 99\%$



**Figure III.7. Valeurs de  $\omega$  prédites pour chaque acide aminé de la protéine Ca-eIF4E1.**

Seules les valeurs de  $\omega > 1$  sont représentées avec l'écart type correspondant. La séquence en acides aminés de Ca-eIF4E1 est représentée sur l'axe des abscisses. Les sites avec un  $\omega > 1$  sont soulignés et les acides aminés impliqués dans la résistance aux potyvirus sont en gras.

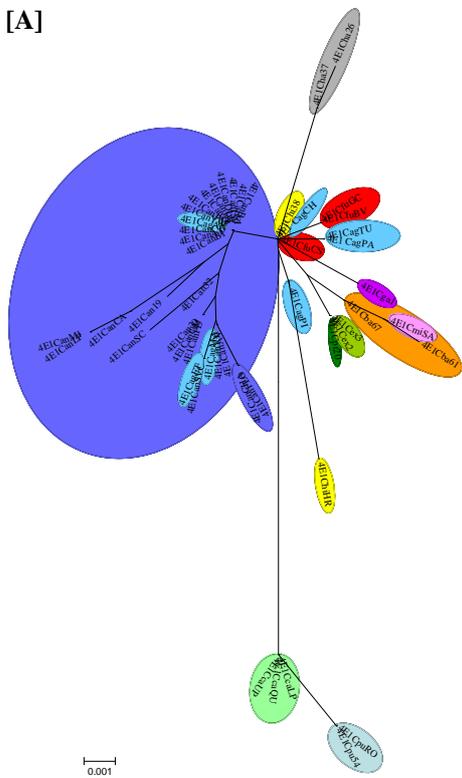
\* $p_{(\omega > 1)} > 95\%$ , \*\* $p_{(\omega > 1)} > 99\%$ .

**Tableau III.7. Substitutions nucléotidiques détectées entre chaque gène *Ca-eIF4E* chez les 51 accessions de *Capsicum* spp.**  
**[A] *Ca-eIF4E1*; [B] *Ca-eIF4E2*; [C] *Ca-eIF(iso)4E***

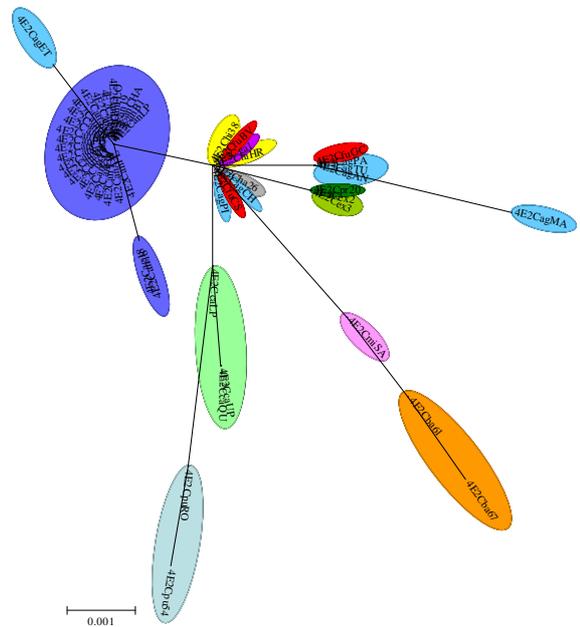
[A]

Accessions	Position des substitutions nucléotidiques																																																		
	11*	30	43*	44*	85*	119*	193*	196	200*	203*	212*	218*	221*	233*	236*	243	319*	324	325*	387	392*	446	449	452	471	479*	480*	511	564	614*	642*	645	646	647*	656*	659															
CagAN	C	G	G	C	G	A	A	C	T	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CagCH	C	G	G	C	G	A	A	C	T	C	G	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CagET	C	G	G	C	G	A	A	C	A	C	A	C	C	C	G	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CagMA	C	G	G	C	G	A	A	C	A	C	A	C	C	C	G	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CagPA	C	G	G	T	G	A	A	C	T	C	G	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CagPI	C	G	G	C	G	A	A	C	T	C	G	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CagTU	C	G	G	T	G	A	A	C	T	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
Can18	C	G	G	C	G	A	A	C	T	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
Can19	C	G	G	C	G	A	A	C	T	C	A	C	C	C	T	C	A	G	A	C	G	T	A	C	A	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
Can20	C	G	G	C	G	A	A	C	A	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
Can32	C	G	G	C	G	A	A	C	A	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanAT	C	G	G	C	G	A	A	C	T	A	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanAV	C	G	G	C	G	A	A	C	A	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanB1	C	G	G	C	G	A	A	C	T	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanBA	C	G	G	C	G	A	A	C	T	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanCA	C	G	G	C	G	A	A	C	T	A	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanCHA	C	G	G	C	G	A	A	C	A	C	A	C	C	C	G	C	G	A	A	C	G	T	A	C	A	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanCM	C	G	G	C	G	A	A	C	A	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanCU	C	G	G	C	G	A	A	C	T	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanDLL	C	G	G	C	G	A	A	C	T	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanFL	C	G	G	C	G	A	A	C	T	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanFLO	C	G	G	C	G	A	A	C	A	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanFLO	C	G	G	C	G	A	A	C	A	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanH3	C	G	G	C	G	A	A	C	T	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanLA	C	G	G	C	G	A	A	C	T	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanLP	C	G	G	C	G	A	A	C	T	A	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanM1	C	G	G	C	G	A	A	C	T	A	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanPE	C	G	G	C	G	A	A	C	A	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanPM	C	G	G	C	G	A	A	C	T	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanSC	C	G	G	C	G	A	A	C	A	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanSVC	C	G	G	C	G	A	A	C	A	C	A	C	C	C	G	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanYW	C	G	G	C	G	A	A	C	T	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CanYY	C	G	G	C	G	A	A	C	T	C	A	C	C	C	G	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
Cba61	C	A	G	C	G	A	A	C	T	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
Cba67	C	A	G	C	G	A	A	C	T	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CcaLP	C	A	G	C	G	A	A	C	T	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CcaQU	C	A	G	C	G	A	A	C	T	C	A	C	C	C	T	C	G	A	G	A	C	G	G	C	G	C	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CcaLP	C	A	G	C	G	A	A	C	T	C	A	C	C	C	T	C	G	A	G	A	C	G	G	C	G	C	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
Cex2	C	G	G	C	A	A	A	C	T	C	G	C	C	C	T	C	G	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A										
Cex3	C	G	G	C	A	A	A	C	T	C	G	C	C	C	T	C	G	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A										
CfuBV	C	G	G	C	G	A	A	C	T	C	G	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CfuCS	C	G	G	C	G	A	A	C	T	C	G	C	C	C	T	C	G	A	G	A	T	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
CfuGC	C	G	G	C	G	A	A	C	T	C	A	C	C	C	T	C	G	A	G	A	T	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
Cga1	C	G	G	C	G	A	A	C	T	C	A	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
Cha26	C	G	A	C	G	A	A	C	T	C	G	C	C	C	T	T	G	A	G	C	C	T	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
Cha37	C	G	A	C	G	A	A	C	T	C	G	C	C	C	T	T	G	A	G	C	C	T	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
Cha38	C	G	A	C	G	A	A	C	T	C	G	C	C	C	T	T	G	A	G	C	C	T	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
ChiHR	C	G	G	C	G	A	A	C	T	C	G	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
ChiHR	C	G	G	C	G	A	A	C	T	C	G	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
ChiHR	C	G	G	C	G	A	A	C	T	C	G	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
ChiHR	C	G	G	C	G	A	A	C	T	C	G	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
ChiHR	C	G	G	C	G	A	A	C	T	C	G	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
ChiHR	C	G	G	C	G	A	A	C	T	C	G	C	C	C	T	C	G	A	G	A	C	G	T	A	C	A	A	A	G	A	T	A	A	G	C	A	T	A	A	G	C	A									
ChiHR	C	G	G	C	G	A	A	C	T	C	G																																								

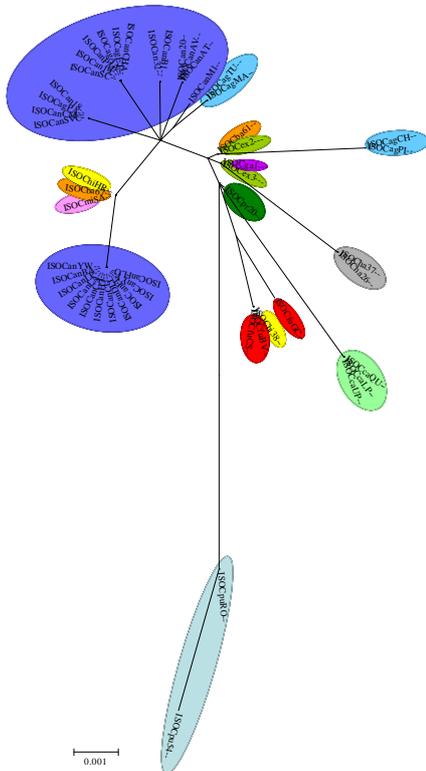
[A]



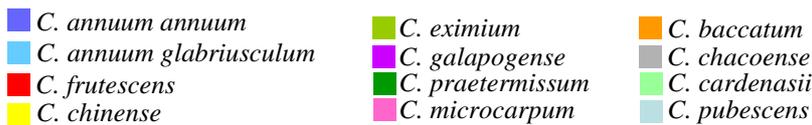
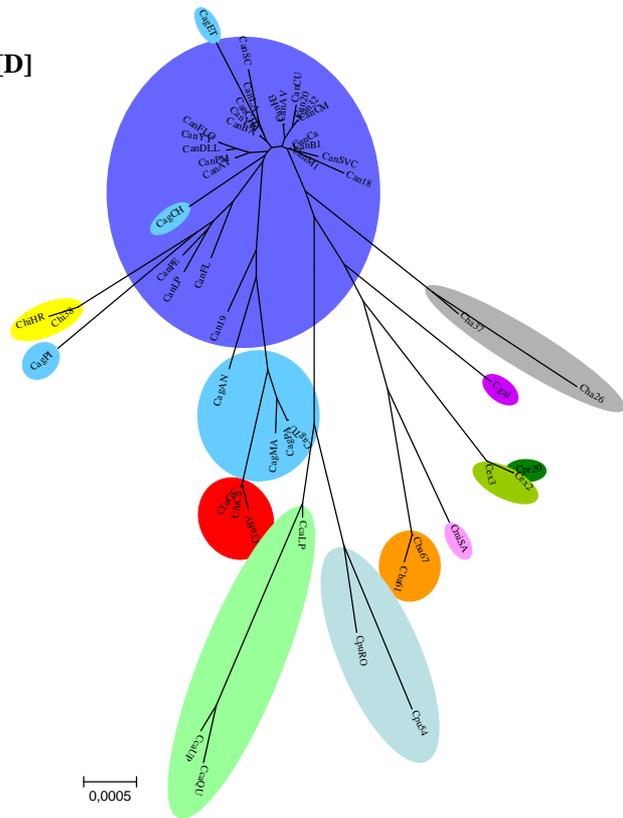
[B]



[C]



[D]



**Figure III.8. Représentation radiale des arbres phylogénétiques (Neighbour-joining) construits à partir des séquences codantes des gènes *Ca-eIF4E* et des locus indépendants concaténés. [A] *Ca-eIF4E1*; [B] *Ca-eIF4E2*; [C] *Ca-eIF(iso)4E*; [D] locus indépendants concaténés.**

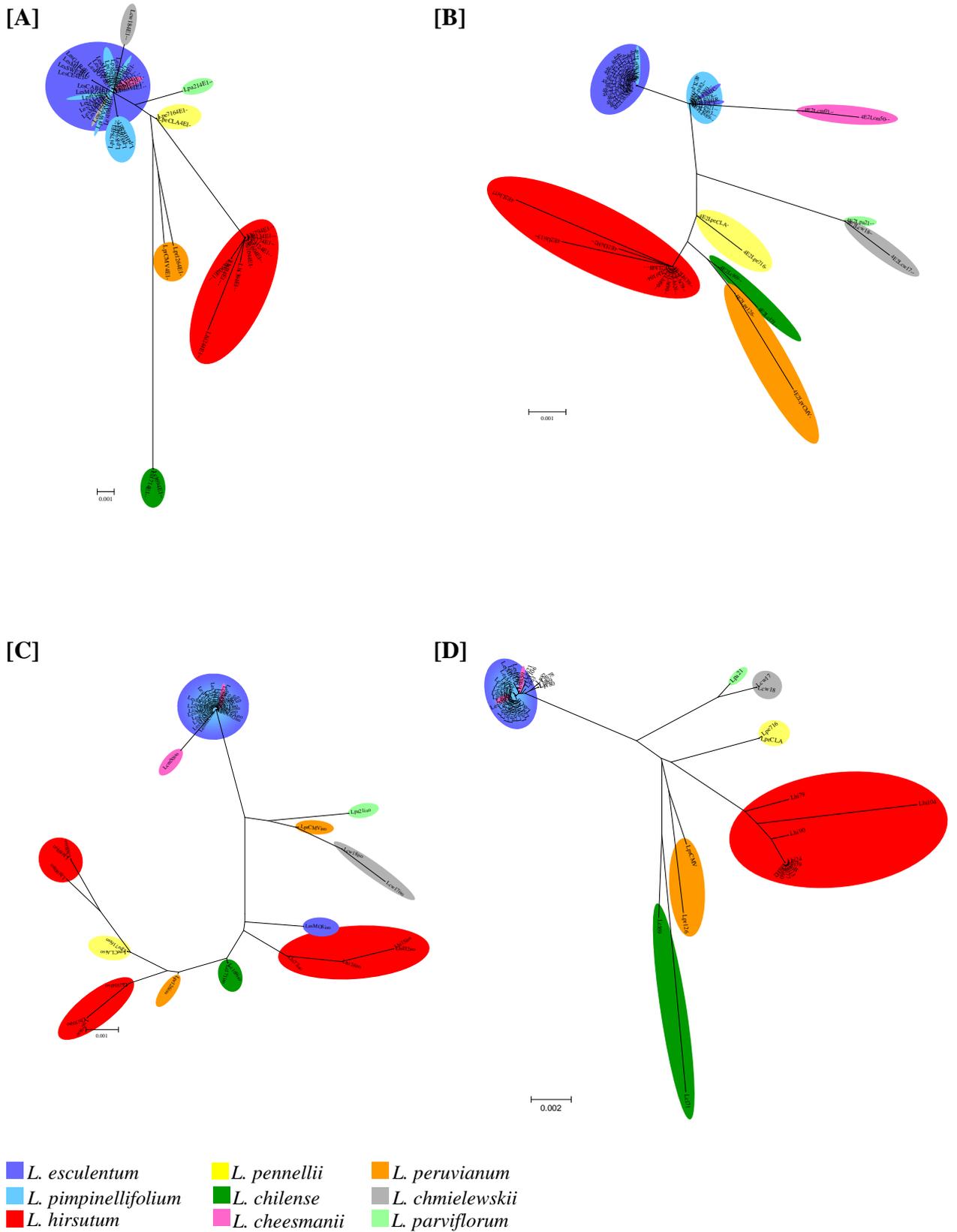
**Tableau III.8. Diversité nucléotidique des gènes *Ly-eIF4E* et des locus indépendants chez 20 accessions de *L. esculentum*.**

Gènes	Régions	Nombre de sites	Sites en ségrégation	Singletons	$\pi$ par site $\times 10^{-4}$	$\pi_s$ (sites) $\times 10^{-4}$	$\pi_a$ (sites) $\times 10^{-4}$	$\pi_a/\pi_s$	$\theta_w$ par site $\times 10^{-4}$	D de Tajima	Signification statistique du D de Tajima
<i>Ly-eIF4E1</i>	codant	669	1	0	5,1	0	6,4	-	4,2	0,35195	>0,1
	noncodant1	811	2	0	8,6	-	-	-	7,2	0,45727	>0,1
	Total	1480	3	0	5,9	0	6,4	-	5,8	0,05543	>0,1
<i>Ly-eIF4E2</i>	codant	666	1	0	2,9	0	3,6	-	4,3	-0,59155	<0,1
	noncodant	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
	Total	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
<i>Ly-eIF(iso)4E</i>	codant	603	5	5	8,3	19,7	5,3	0,27	23,4	-1,97429	<0,05*
	noncodant1	1737	2	0	2,6	-	-	-	3,2	-0,44022	>0,1
	Total	2340	7	5	3,7	19,7	5,3	0,27	7,2	-1,57027	>0,1
<i>CT066<sup>2</sup></i>	codant	900	0	0	-	-	-	-	-	-	-
	noncodant	0	-	-	-	-	-	-	-	-	-
	Total	900	0	0	-	-	-	-	-	-	-
<i>CT093<sup>2</sup></i>	codant	915	2	0	5,9	0	7,7	-	6,2	-0,11187	>0,1
	noncodant	0	-	-	-	-	-	-	-	-	-
	Total	915	2	0	5,9	0	7,7	-	6,2	-0,11187	>0,1
<i>CT148<sup>2</sup></i>	codant	333	2	2	6	12,5	4	0,32	16,9	-1,51284	>0,1
	noncodant	664	3	2	6	-	-	-	13	-1,44071	>0,1
	Total	997	5	4	6	12,5	4	0,32	14,3	-1,78003	0,05<p<0,1
<i>CT179<sup>2</sup></i>	codant	251	0	0	-	-	-	-	-	-	-
	noncodant	676	7	7	11,3	-	-	-	31,7	-2,12144	<0,05*
	Total	927	7	7	8,1	-	-	-	22,9	-2,12144	<0,05*
<i>CT189<sup>2</sup></i>	codant	185	0	0	-	-	-	-	-	-	-
	noncodant	588	0	0	-	-	-	-	-	-	-
	Total	773	0	0	-	-	-	-	-	-	-
<i>CT198<sup>2</sup></i>	codant	297	1	0	9	38,3	0	0	9,5	-0,0861	>0,1
	noncodant	415	3	1	15,5	-	-	-	20,6	-0,65939	>0,1
	Total	712	4	1	12,8	38,3	0	0	15,9	-0,57056	>0,1
<i>CT208<sup>2</sup></i>	codant	923	1	1	1,1	0	1,4	-	3,1	-1,16439	>0,1
	noncodant	0	-	-	-	-	-	-	-	-	-
	Total	923	1	1	1,1	0	1,4	-	3,1	-1,16439	>0,1
<i>Sucr<sup>2</sup></i>	codant	880	2	1	3,7	0	4,9	-	7,2	-1,1407	>0,1
	noncodant	0	-	-	-	-	-	-	-	-	-
	Total	880	2	1	3,7	0	4,9	-	7,2	-1,1407	>0,1

**Tableau III.9. Substitutions nucléotidiques dans les séquences codantes des gènes *Ly-eIF4E* entre les 20 accessions de *L. esculentum*.**

Accessions	Position des substitutions nucléotidiques						
	<i>Ly-eIF4E1</i>	<i>Ly-eIF4E2</i>	<i>Ly-eIF(iso)4E</i>				
	307*	61*	72	137*	229	231*	417
Les106	G	R	G	T	G	T	G
LesALI	G	R	G	T	G	T	G
LesCAR	G	R	G	T	G	T	G
LesCE	A	G	G	T	G	T	G
LesFE	G	R	G	T	G	T	G
LesGAR	A	R	G	T	G	T	G
LesLE	G	R	G	T	G	T	G
LesM82	G	R	G	T	G	T	G
LesMIC	A	G	G	T	G	T	G
LesMOM	G	R	G	T	G	T	G
LesMOS	G	R	A	C	T	G	A
LesMRM	G	R	G	T	G	T	G
LesMRZ	G	R	G	T	G	T	G
LesNAG	G	R	G	T	G	T	G
LesPIE	G	R	G	T	G	T	G
LesPLO	G	R	G	T	G	T	G
LesPOI	G	R	G	T	G	T	G
LesPOV	G	R	G	T	G	T	G
LesSWE	A	R	G	T	G	T	G
LesYEL	G	R	G	T	G	T	G

\*substitutions non-synonymes



**Figure III.9. Représentation radiale des arbres phylogénétiques (Neighbour-Joining) construits à partir des séquences codantes des gènes *Ly-eIF4E* et des locus indépendants concaténés.**

[A] *Ly-eIF4E1*; [B] *Ly-eIF4E2*; [C] *Ly-eIF(iso)4E*; [D] locus indépendants concaténés.

**Tableau III.10. Substitutions en acides aminés dans les protéines Ly-eIF4E entre les 50 accessions de *Lycopersicon* spp.**

Accessions	Position des substitutions en acides aminés																														
	Ly-eIF4E1															Ly-eIF4E2							Ly-eIF(iso)4E								
	19	32	48	49	59	60	67	68	71	76	93	96	97	100	103	114	131	166	211	212	215	11	21	26	27	28	30	79	46	77	133
Les106	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	R	E	I	V	E	S	V	V	E
LesALI	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	R	E	I	V	E	S	V	V	E
LesFE	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	R	E	I	V	E	S	V	V	E
LesLE	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	R	E	I	V	E	S	V	V	E
LesM82	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	R	E	I	V	E	S	V	V	E
LesMOM	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	R	E	I	V	E	S	V	V	E
LesMRM	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	R	E	I	V	E	S	V	V	E
LesMRZ	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	R	E	I	V	E	S	V	V	E
LesNAG	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	R	E	I	V	E	S	V	V	E
LesPIE	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	R	E	I	V	E	S	V	V	E
LesPLO	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	R	E	I	V	E	S	V	V	E
LesPOI	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	R	E	I	V	E	S	V	V	E
LesPOV	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	R	E	I	V	E	S	V	V	E
LesCAR	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	R	E	I	V	E	S	V	V	E
LesYEL	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	R	E	I	V	E	S	V	V	E
LesMOS	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	R	E	I	V	E	S	A	L	E
LesSWE	V	S	L	L	N	P	T	A	S	V	H	S	K	M	E	K	S	A	R	L	N	E	R	E	I	V	E	S	V	V	E
LesGAR	V	S	L	L	N	P	T	A	S	V	H	S	K	M	E	K	S	A	R	L	N	E	R	E	I	V	E	S	V	V	E
LesMIC	V	S	L	L	N	P	T	A	S	V	H	S	K	M	E	K	S	A	R	L	N	E	R	E	I	V	E	S	V	V	E
LesCE	V	S	L	L	N	P	T	A	S	V	H	S	K	M	E	K	S	A	R	L	N	E	G	E	I	V	E	S	V	V	E
Lpi21	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	G	E	I	V	E	S	V	V	E
Lpi82	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	G	E	I	V	E	S	V	V	E
Lpi79	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	G	E	I	V	E	S	V	V	E
Lpi63	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	G	E	I	V	E	S	V	V	E
LpiWVA	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	G	E	I	V	E	S	V	V	E
Lpi78	V	S	L	V	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	G	E	I	V	E	S	V	V	E
Lpi80	V	S	L	V	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	G	E	I	V	E	S	V	V	E
Lpi89	V	S	L	V	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	G	E	I	V	E	S	V	V	E
LpiHIR	V	S	L	V	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	G	E	I	V	E	S	V	V	E
Lpe716	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	G	E	I	V	E	S	A	L	E
LpeCLA	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	G	E	I	V	E	S	A	L	E
Lpr126	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	G	E	I	V	E	S	A	L	E
LprCMV	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	G	E	I	V	D	S	A	L	E
Lcw17	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	G	E	I	V	E	S	A	V	D
Lcw18	V	L	L	L	N	P	T	A	S	V	H	C	K	M	D	K	S	A	R	L	N	E	G	E	I	V	E	S	A	L	E
Lhi24	V	S	F	L	K	S	T	D	S	L	H	S	K	L	D	Q	S	A	R	L	S	E	G	E	I	V	E	S	A	L	E
Lhi60	V	S	L	L	N	S	T	A	S	L	H	S	K	M	D	Q	S	A	R	L	S	E	G	E	I	V	E	S	A	L	E
Lhi104	V	S	L	L	N	S	T	A	S	L	H	S	K	M	D	Q	S	A	R	L	S	E	G	E	I	V	E	S	A	L	E
Lhi39	V	S	L	L	N	S	T	A	S	L	H	S	K	M	D	Q	S	A	R	L	S	E	G	E	I	V	E	S	A	L	E
Lhi13	V	S	L	L	N	S	T	A	S	L	H	S	K	M	D	Q	S	A	R	L	S	E	G	E	I	I	E	S	A	L	E
Lhi77	V	S	L	L	N	S	T	A	S	L	H	S	K	M	D	Q	S	A	R	L	S	E	G	D	I	V	E	F	A	L	E
Lhi79	V	S	L	L	N	S	T	A	S	L	H	S	K	M	D	Q	S	A	R	L	S	E	G	E	I	V	E	S	A	L	E
Lhi90	V	S	L	L	N	S	T	A	S	L	H	S	K	M	D	Q	S	A	R	L	S	E	G	E	I	V	E	S	A	L	E
LhiB	V	S	L	L	N	S	T	A	S	L	H	S	K	M	D	Q	S	A	R	L	S	E	G	E	I	V	E	S	A	L	E
LhiH2	V	S	L	L	N	S	T	A	S	L	H	S	K	M	D	Q	S	A	R	L	S	E	G	E	I	V	E	S	A	L	E
Lcl69	V	S	L	L	N	H	N	A	I	V	R	S	N	M	D	K	P	T	K	D	N	E	G	E	I	V	E	S	A	L	E
Lcl71	V	S	L	L	N	H	N	A	I	V	R	S	N	M	D	K	P	T	K	D	N	E	G	E	I	V	E	S	A	L	E
Lpa21	I	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	E	G	E	I	V	E	S	A	L	E
Lcm50	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	D	G	E	T	V	E	S	V	V	E
Lcm01	V	S	L	L	N	P	T	A	S	V	H	S	K	M	D	K	S	A	R	L	N	D	G	E	I	V	E	S	V	V	E

Jaune : substitutions espèces-spécifiques; Bleu : substitutions ponctuelles; Rouge : substitutions impliquées dans la résistance aux potyvirus; Vert : substitutions partagées entre plusieurs espèces.

**Tableau III.11. Test d'hétérogénéité des modèles M7 et M8 pour estimer par maximum de vraisemblance le paramètre  $\omega$  par site.**

Gènes	Modèles	$p$	Paramètres $\beta$	$\omega_s$	$p_s$	logL	LRT	$\chi^2$
<i>Ly-eIF4E1</i>	M7 ( $\beta$ )	2	0,016; 0,038	NE	NE	1205,60		
	M8 ( $\beta + \omega > 1$ )	4	0,709; 1,950	16,48	0,01	1200,79	<b>9,62</b>	8,15E-03
<i>Ly-eIF4E2</i>	M7 ( $\beta$ )	2	13,77; 99,00	NE	NE	1063,28		
	M8 ( $\beta + \omega > 1$ )	4	13,77; 99,00	1	0	1063,28	0,00	1,00E+00
<i>Ly-eIF(iso)4E</i>	M7 ( $\beta$ )	2	0,013; 0,179	NE	NE	995,45		
	M8 ( $\beta + \omega > 1$ )	4	2,462; 99,00	20,28	0,005	986,84	<b>17,22</b>	1,82E-04

$p$  : nombre de paramètres du modèle; NE : non estimé dans le modèle;  $\omega_s$  : valeur de  $\omega$  maximum;  $p_s$  : proportion de codons évoluant pour une valeur  $\omega_s$ ; logL : log vraisemblance; LRT : test statistique de rapport de vraisemblance, les valeurs significatives au seuil de 5 % sont en gras.

**Tableau III.12. Sites sous sélection positive détectés par le modèle M8.**

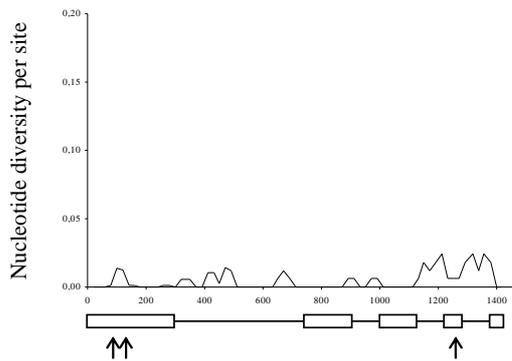
Protéines	Sites sous sélection positive	$P_{\omega>1}$	$\omega$ estimé	Ecart type
<i>Ly-eIF4E1</i>	49L	0,991 **	5,847	2,379
	60P	0,770	4,699	3,013
	212L	0,875	5,330	2,810
<i>Ly-eIF4E2</i>	-			
<i>Ly-eIF(iso)4E</i>	77V	0,997 **	7,042	2,735

\* $P > 95\%$ ; \*\* $P > 99\%$

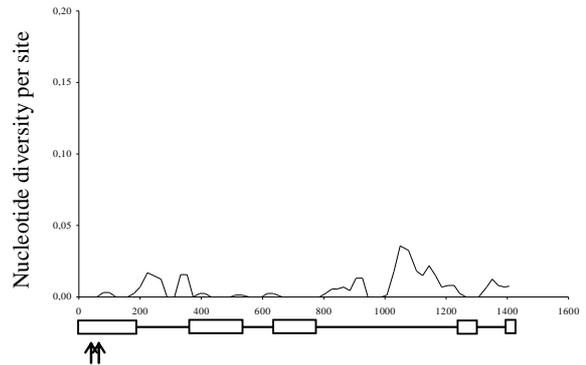
**Tableau III.13. Diversité nucléotidique des 5 gènes *At-eIF4E* et des 10 locus indépendants chez 35 accessions d'*A. thaliana*.**

Gènes	Régions	Nombre de sites	Nombre de sites en ségrégation	Singletons	$\pi$ par site $\times 10^{-4}$	$\pi_s$ (sites) $\times 10^{-4}$	$\pi_a$ (sites) $\times 10^{-4}$	$\pi_a/\pi_s$	$\theta_w$ par site $\times 10^{-4}$	D de Tajima	Signification statistique du D de Tajima
<i>At-eIF4E1</i>	codant	708	5	4	10,2	7,8 (146,02)	10,9 (558,98)	1,40	17,1	-1,063	>0,1
	noncodant	720	24	13	48,8						
	Total	1428	29	17	29,8						
<i>At-eIF4E2</i>	codant	723	34	0	235,7	553,1 (128,59)	150,3 (477,41)	0,27	136,2	2,594	<0,01**
	noncodant	922	105	4	709,5						
	Total	1645	139	4	459,0						
<i>At-eIF4E3</i>	codant	723	5	4	4,7	13,8 (163,50)	1,9 (589,50)	0,14	16,8	-1,891	<0,05*
	noncodant	858	12	9	14,8						
	Total	1581	17	13	10,1						
<i>At-eIF(iso)4E</i>	codant	606	7	5	8,5	36,6 (123,15)	1,2 (467,85)	0,03	28,5	-2,000	<0,05*
	noncodant	814	58	9	90,9						
	Total	1420	65	14	54,4						
<i>At-nCBP</i>	codant	666	7	1	16,4	43,7 (148,60)	8,1 (514,40)	0,19	25,7	-1,035	>0,1
	noncodant	818	12	5	15,8						
	Total	1484	19	6	14,6						
Acetyl CoA synthétase	codant	450	4	2	8,7	0,0 (109,78)	11,5 (340,22)	-	21,6	-1,782	0,1>p>0,05
	noncodant	200	4	1	24,4						
	Total	660	8	3	13,5						
Auxin responsive protein	codant	413	6	3	33,0	87,4 (96,51)	16,4 (314,49)	0,19	35,4	-0,187	>0,1
	noncodant	217	1	1	2,6						
	Total	630	7	4	14,3						
Chorismate mutase precursor	codant	296	3	2	7,6	8,9 (64,32)	7,3 (229,68)	0,82	24,6	-1,559	>0,1
	noncodant	363	3	3	5,0						
	Total	659	6	5	10,4						
Cycloartenol synthase	codant	478	22	4	78,0	153,0 (106,70)	53,2 (343,30)	0,35	120,6	-1,214	>0,1
	noncodant	160	14	1	196,9						
	Total	638	36	5	102,8						
Dihydroflavonol reductase	codant	485	9	5	42,4	116,6 (106,67)	21,5 (358,33)	0,18	46,0	-0,236	>0,1
	noncodant	173	4	1	78,1						
	Total	658	13	6	78,3						
Enoyl CoA hydratase	codant	81	0	0	-	-	-	-	-	-	<0,05*
	noncodant	515	4	4	4,5						
	Total	596	4	4	3,9						
Geranylgeranyl reductase	codant	462	10	1	94,9	306,1 (109,97)	28,9 (352,03)	0,09	52,6	-2,467	<0,05*
	noncodant	200	5	1	89,7						
	Total	662	15	2	101,6						
Long-chain fatty acid CoA ligase	codant	193	1	0	23,0	103,8 (42,72)	0,0 (146,28)	0,00	12,6	1,228	>0,1
	noncodant	459	9	4	26,6						
	Total	652	10	4	25,5						
Methionyl tRNA synthetase like	codant	429	12	3	58,9	86,7 (97,96)	50,7 (331,04)	0,59	67,9	-0,419	>0,1
	noncodant	88	0	0	-						
	Total	517	12	3	49,0						
Peptidase	codant	284	3	0	34,3	79,3 (66,93)	20,6 (215,07)	0,26	25,7	0,763	>0,1
	noncodant	326	11	0	116,3						
	Total	610	14	0	70,7						

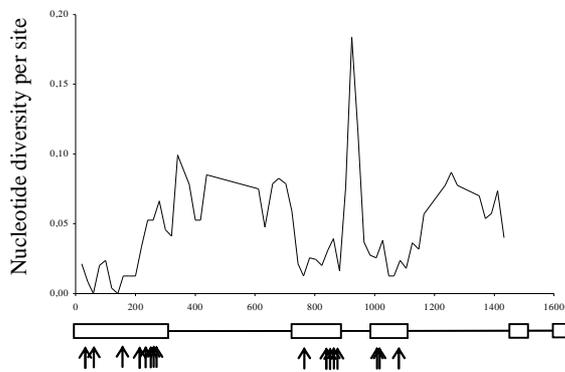
[A] *At-eIF4E1*



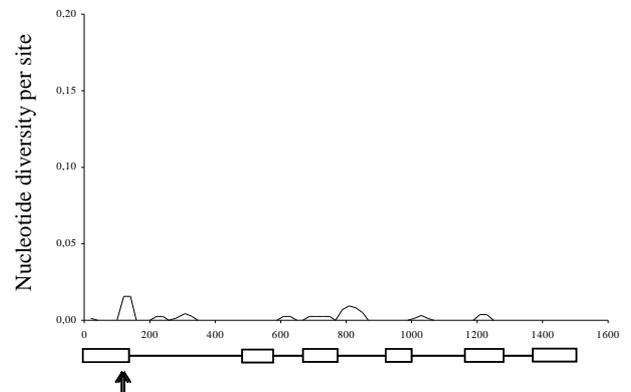
[D] *At-eIF(iso)4E*



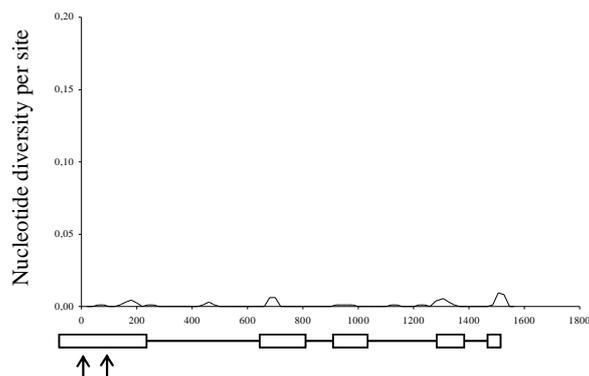
[B] *At-eIF4E2*



[E] *At-nCBP*



[C] *At-eIF4E3*

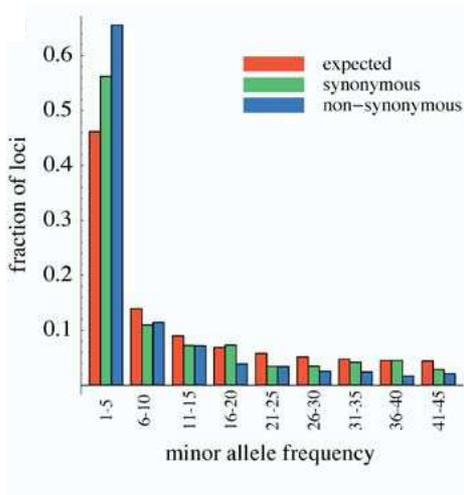


**Figure III.10. Structure des 5 ADNg *At-eIF4E* et analyse par fenêtre glissante de la diversité nucléotidique.**

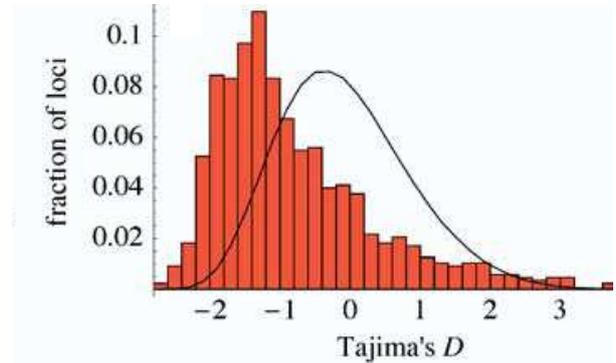
Les exons sont représentés par des rectangles sous le graphe. Les flèches indiquent la position des substitutions en acides aminés.

[A] *At-eIF4E1*; [B] *At-eIF4E2*; [C] *At-eIF4E3*; [D] *At-eIF(iso)4E*; [E] *At-nCBP*

[A]



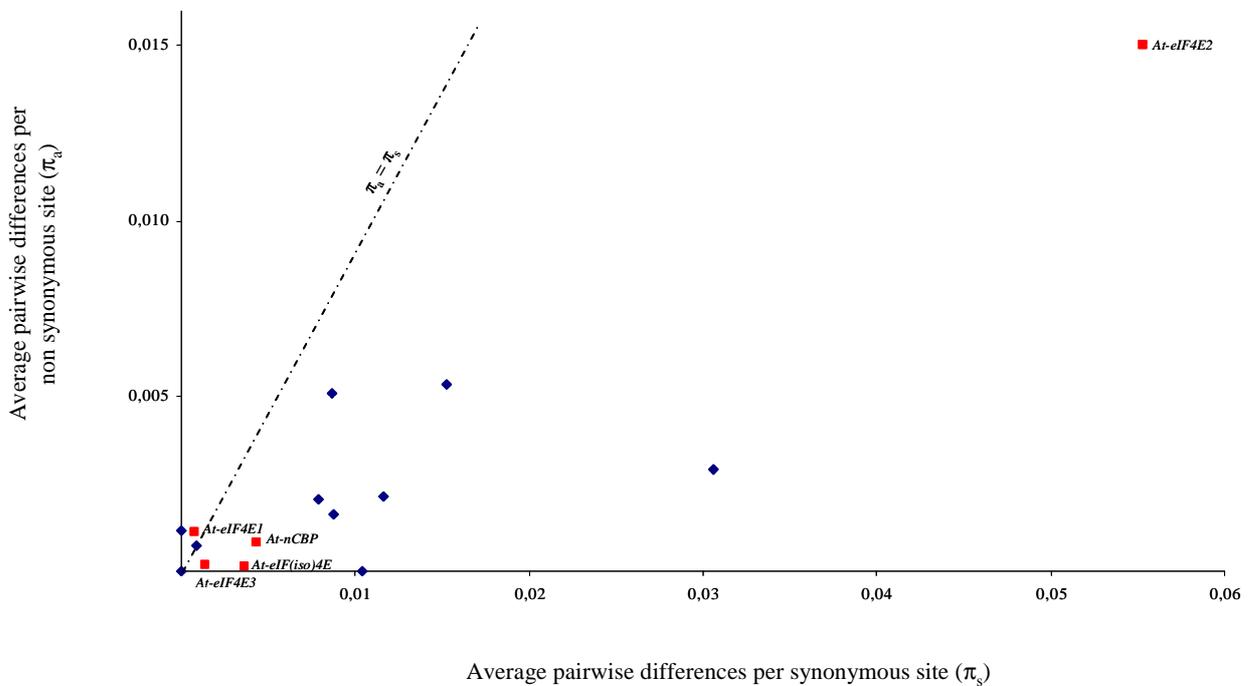
[B]



**Figure III.11. Caractéristiques du profil de polymorphisme chez *A. thaliana* estimé par Nordborg *et al.* (2005).**

[A] Distribution de la fréquence allélique pour les SNPs synonymes et non-synonymes à partir d'un échantillon de 90 individus.

[B] Distribution des valeurs de D de Tajima's D et distribution attendue dans une population à taille constante.

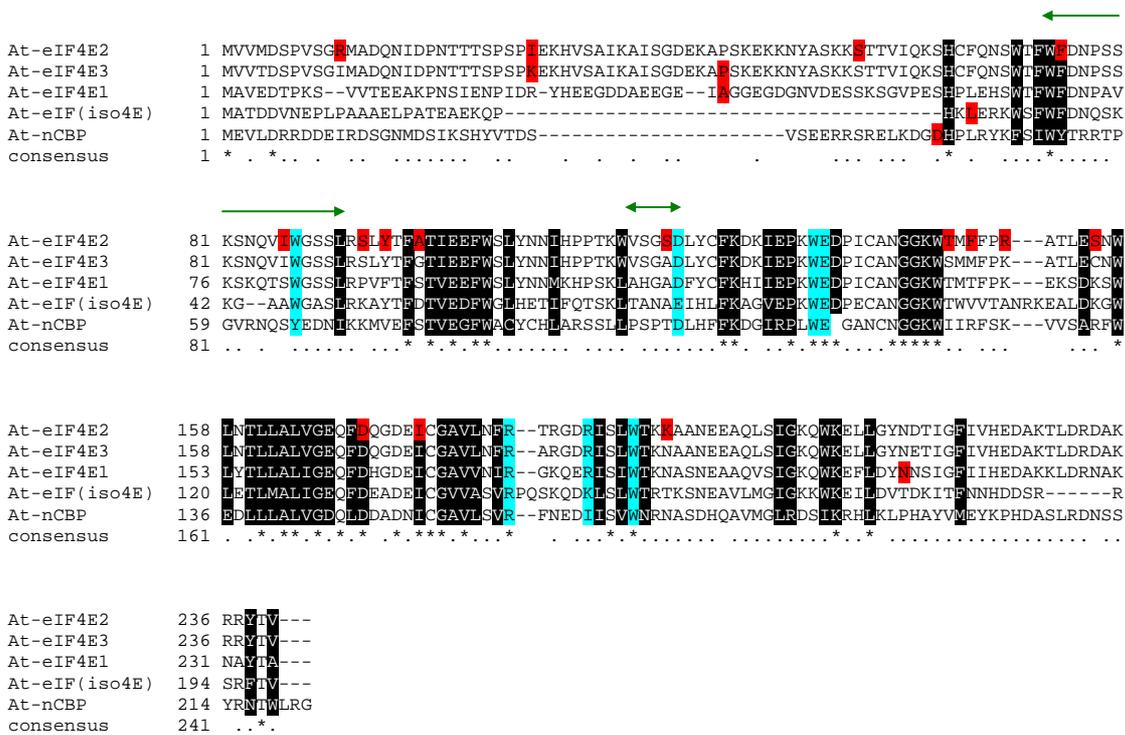


**Figure III.12. Rapports de  $\pi_a$  sur  $\pi_s$  calculés pour les locus indépendants (losanges bleus) et les 5 gènes *At-eIF4E* chez 35 accessions d'*A. thaliana* (carrés rouges).**

**Tableau III.14. Test d'hétérogénéité des modèles M7 et M8 pour l'estimation du paramètre  $\omega$  par site.**

Gène	Modèle	$p$	Paramètres $\beta$	$\omega_s$	$p_s$	logL	LRT	$\chi^2$
<i>At-eIF4E1</i>	M7 ( $\beta$ )	2	91.46, 99	NE	NE	-975.83	0.00	1.00E+00
	M8 ( $\beta + \omega > 1$ )	4	92.57, 99	1	0	-975.83		
<i>At-eIF4E2</i>	M7 ( $\beta$ )	2	0.005, 0.015	NE	NE	-1068.05	0.90	6.38E-01
	M8 ( $\beta + \omega > 1$ )	4	0.005, 2.353	1.71	0.172	-1067.60		
<i>At-eIF4E3</i>	M7 ( $\beta$ )	2	34.85, 99	NE	NE	-1001.74	0.00	1.00E+00
	M8 ( $\beta + \omega > 1$ )	4	10.18, 28.71	1	0	-1001.74		
<i>At-eIF(iso)4E</i>	M7 ( $\beta$ )	2	0.019, 0.336	NE	NE	-846.30	0.14	9.32E-01
	M8 ( $\beta + \omega > 1$ )	4	0.005, 11.5	8.08	0.008	-846.23		
<i>At-nCBP</i>	M7 ( $\beta$ )	2	0.012, 0.17	NE	NE	-957.51	5.78	5.56E-02
	M8 ( $\beta + \omega > 1$ )	4	0.005, 2.95	17.01	0.005	-954.62		

$p$  : nombre de paramètres du modèle; NE : non estimé dans le modèle;  $\omega_s$  : valeur de  $\omega$  maximum;  $p_s$  : proportion de codons évoluant pour une valeur  $\omega_s$ ; logL : log vraisemblance; LRT : test statistique de rapport de vraisemblance.

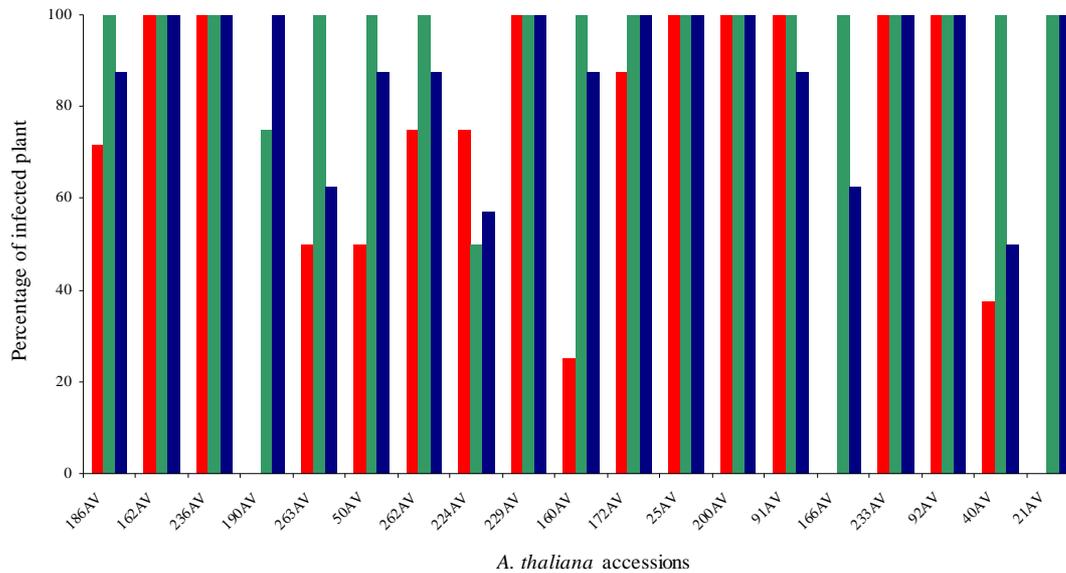


**Figure III.13. Alignement des séquences en acides aminés des 5 protéines At-eIF4E.**

Les substitutions en acides aminés détectées à partir des analyses de polymorphisme sont surlignées en rouge, les acides aminés essentiels pour la stabilité des protéines eIF4E en noir et ceux impliqués dans la fixation de la structure coiffe des ARNm en bleu. Les flèches vertes indiquent les régions impliquées dans la résistance chez les autres espèces.

**Tableau III.15. Substitutions en acides aminés observées entre les 5 At-eIF4E chez 35 accessions d'*A. thaliana* et polyhaplotypes correspondants.**

eIF4E-polyhaplotype	Accessions	Position des substitutions en acides aminés pour chaque protéine At-eIF4E																						
		eIF4E1		eIF4E2															eIF4E3		eIF(iso)4E	nCBP		
		40	211	11	28	57	75	86	93	95	98	120	145	147	150	155	170	175	195	27	49	28	42	
11111	186AV	A	N	R	I	S	F	I	S	Y	A	S	T	F	R	S	D	I	K	K	P	L	D	
14112	172AV, 25AV, 68AV, 76AV	-	-	-	-	-	-	T	-	-	-	-	-	-	-	K	-	-	-	-	-	-	L	
16112	233AV, 92AV	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	L	
15112	200AV, 91AV	-	-	-	-	-	-	-	-	-	-	-	-	-	-	K	-	-	-	-	-	-	L	
12111	236AV, 190AV	-	-	-	K	L	L	T	A	H	-	A	-	-	-	K	-	H	V	N	-	-	-	
13111	262AV	-	-	I	K	-	-	-	-	-	G	-	S	M	K	C	-	-	N	-	-	-	-	
22112	21AV	V	K	-	K	L	L	T	A	H	-	A	-	-	-	K	-	H	V	N	-	-	L	
15312	166AV	-	-	-	-	-	-	-	-	-	-	-	-	-	-	K	-	-	-	-	Q	-	L	
13212	160AV	-	-	I	K	-	-	-	-	-	G	-	S	M	K	C	-	-	N	S	-	-	L	
11112	162AV	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	L	
16122	40AV	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	F	L	
13112	224AV, 229AV, 234AV, 56AV, 62AV	-	-	I	K	-	-	-	-	-	G	-	S	M	K	C	-	-	N	-	-	-	L	
12112	163AV, 178AV, 180AV, 231AV, 235AV, 244AV, 252AV, 263AV, 267AV, 42AV, 50AV, 70AV, 83AV	-	-	-	K	L	L	T	A	H	-	A	-	-	-	K	-	H	V	N	-	-	-	L

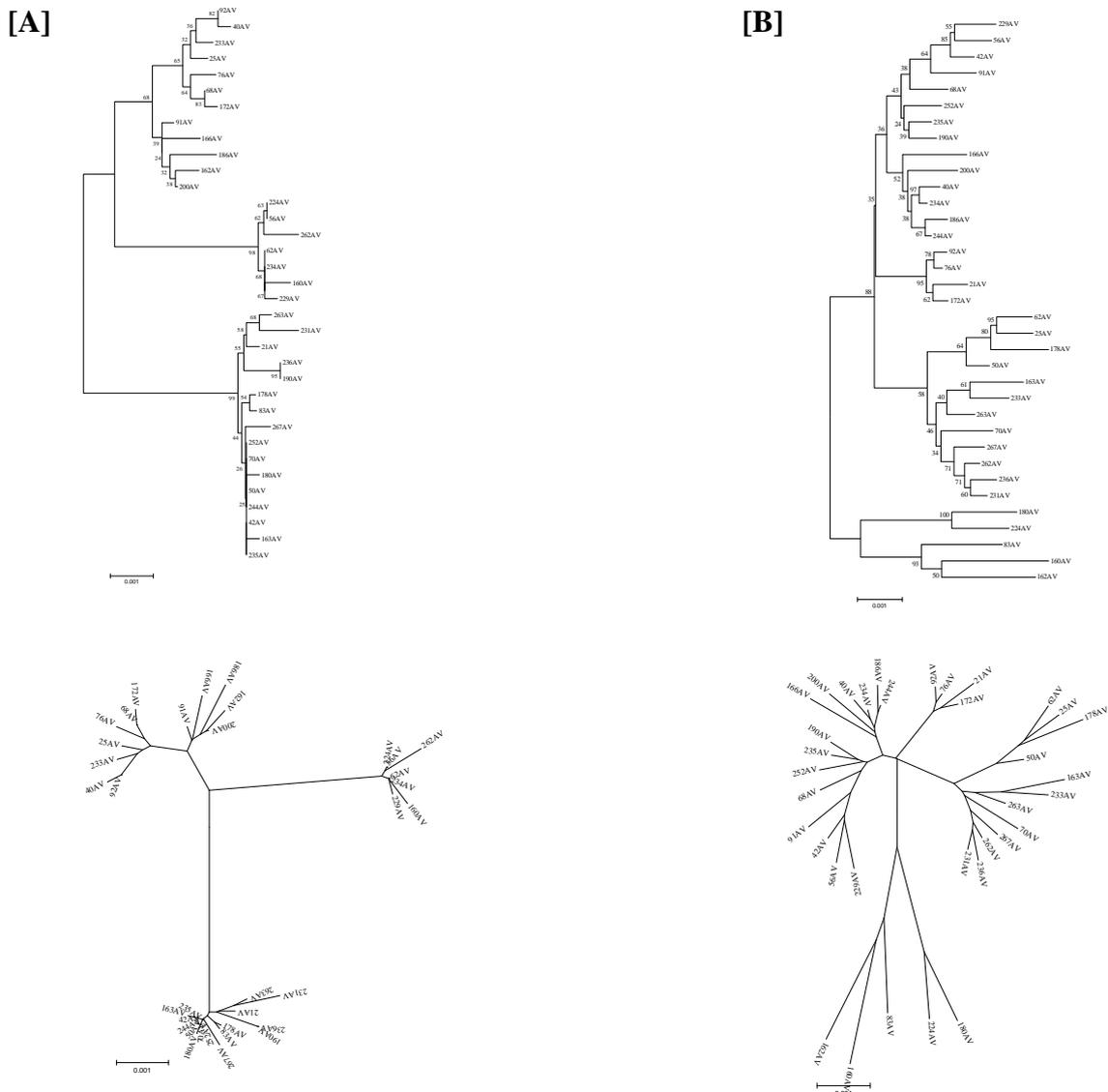


**Figure III.14. Pourcentage de plantes infectées par le WMV (en rouge), le TuMV (en vert) et le CIYVV (en bleu) chez 19 accessions d'*A. thaliana* représentatives des différents polyhaplotypes protéiques eIF4E.**

**Tableau III.16. Test HKA.**

Gène	Nombre de sites polymorphes		Divergence interspécifique	
	Observé	Attendu	Observé	Attendu
<i>At-eIF4E1</i>	5	6.05	100.51	99.47
<i>At-eIF4E2</i>	34	26.11	180.31	185.24
<i>At-eIF4E3</i>	5	6.70	121.17	119.48
<i>At-eIF(iso)4E</i>	7	5.80	77.11	78.31
<i>At-nCBP</i>	7	5.45	70.51	72.06

Les 35 accessions d'*A. thaliana* ont été utilisées pour estimer le polymorphisme intraspécifique et *A. lyrata* a été utilisée comme groupe externe pour estimer la divergence interspécifique.



**Figure III.15. Comparaison des arbres phylogénétiques (Neighbor-joining) obtenus à partir des ADNc *At-eIF4E* concaténés (A) et des ADNc des locus indépendants concaténés (B).**

En haut, représentations traditionnelles et en bas, représentations radiales.

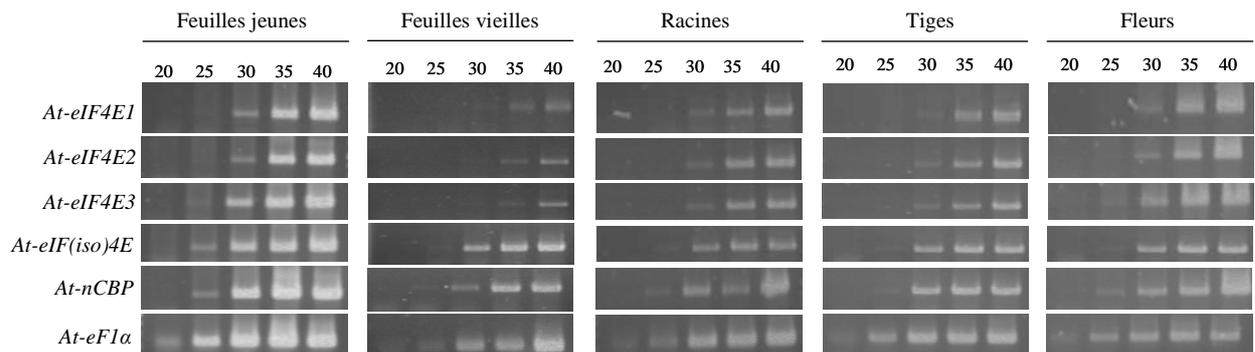
**Tableau IV.1 -Amorces utilisées pour les expériences de RT-PCR semi-quantitative et quantitative.**

[A] pour la PCR semi-quantitative

Gène		Amorce sens (5' -> 3')		Amorce antisens (5' -> 3')	Taille fragment PCR (en pb)
<i>eIF4E1</i>	Z588F	GAGAAAGAGAAGCAGTTCGG	Z588R	GACAGACTGTCAAGTCATG	885
<i>eIF4E2</i>	Z590F	GGAGCATAGTGAACGGGAAAA	Z737R	GGTCATGTTCTAATATGGCTTCCC	957
<i>eIF4E3</i>	Z589F	CGGAGCAAAGTGAACGGAAAAG	Z522R	GTAAGATAGATCAGAGGCAAC	955
<i>eIF(iso)4E</i>	Z591F	GCCGTCTCAGAAGAAAACACTGCG	Z591R	CCGACCAAACAGTATCAC	695
<i>nCBP</i>	Z608F	CTTCCTGATTGGAATTTAGGGC	Z608R	GCCTCCGGTACAGTATGGTATTGG	844
<i>eF1 α</i>	Z975F	ATGAACCACCCTGGTCAG	Z975R	TCAC TTCGACCCTTCTTG	218

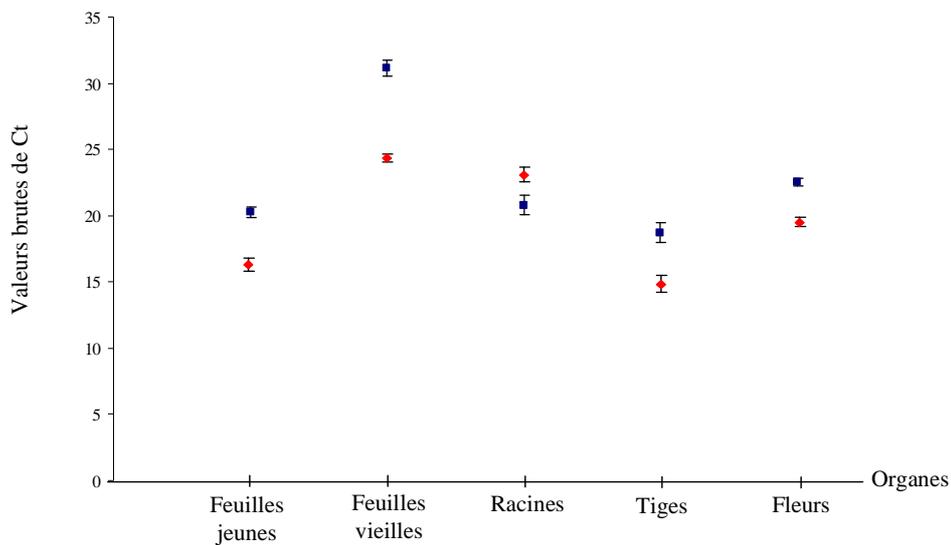
[B] pour la PCR quantitative

Gène		Amorce sens (5' -> 3')		Amorce antisens (5' -> 3')	Taille fragment PCR (en pb)
<i>eIF4E1</i>	Z1056F	TCGAGCAAATCCGGTGTTCCTG	Z1056R	ACAGTAGAAGTCAGCTCCGTG	207
<i>eIF4E2</i>	Z1057F	CTTTGTATACATTCGCTACC	Z1057R	TAACGTGTTAAGCCAATTAG	206
<i>eIF4E3</i>	Z1058F	CCTTGTATACATTCGGTACT	Z1058R	TAACGTGTTAAGCCAATTGC	206
<i>eIF(iso)4E</i>	Z1169F	AGTGGACTTGGGTTGTTACTG	Z1169R	GAACAGCTTCATTAGACTTCG	190
<i>nCBP</i>	Z1170F	TTCTCAAAAGTTGTATCTGCTC	Z1170R	GAGTCTCTCAAACCCATCACT	181
<i>histone 1</i>	Z1197F	AGAAAACAACCTCCTCTCACCCTA	Z1197R	AGGCAGTGATTTATGTTCTCCTC	180
<i>GAPDHA</i>	Z1199F	TTGACTACCAACTGTCTTGCT	Z1199R	AGCTGCTCTGTCTCCTTAGA	185



**Figure IV.1. Profils d'expression des gènes *eIF4E* chez *A. thaliana* analysés par RT-PCR semi-quantitative.**

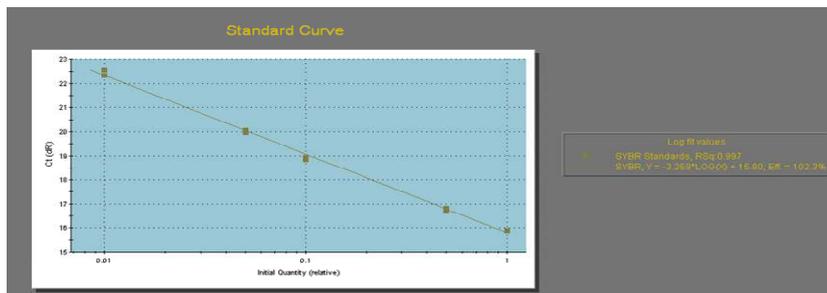
L'accumulation des transcrits est observée dans les organes feuilles jeunes, feuilles vieilles, racines, tiges et fleurs, après 20, 25, 30, 35 et 40 cycles de PCR. L'amplification d'un fragment de l'ADNc de *eF1α* d'*A. thaliana* est utilisée comme contrôle.



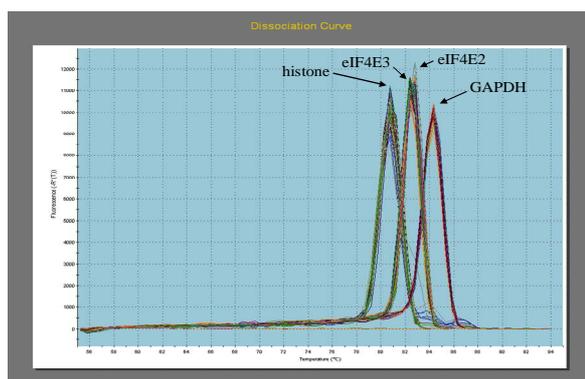
**Figure IV.2. Expression des gènes témoins *histone 1* et *GAPDHA* dans les 5 organes chez *A. thaliana*.**

Les valeurs brutes de Ct indiquées ici sont mesurées pour les gènes *histone 1* (carré bleu) et *GAPDHA* (losange rouge), au cours de 6 expériences de RT-PCR quantitative distinctes et pour 2 réplicats biologiques de chaque organe.

[A]



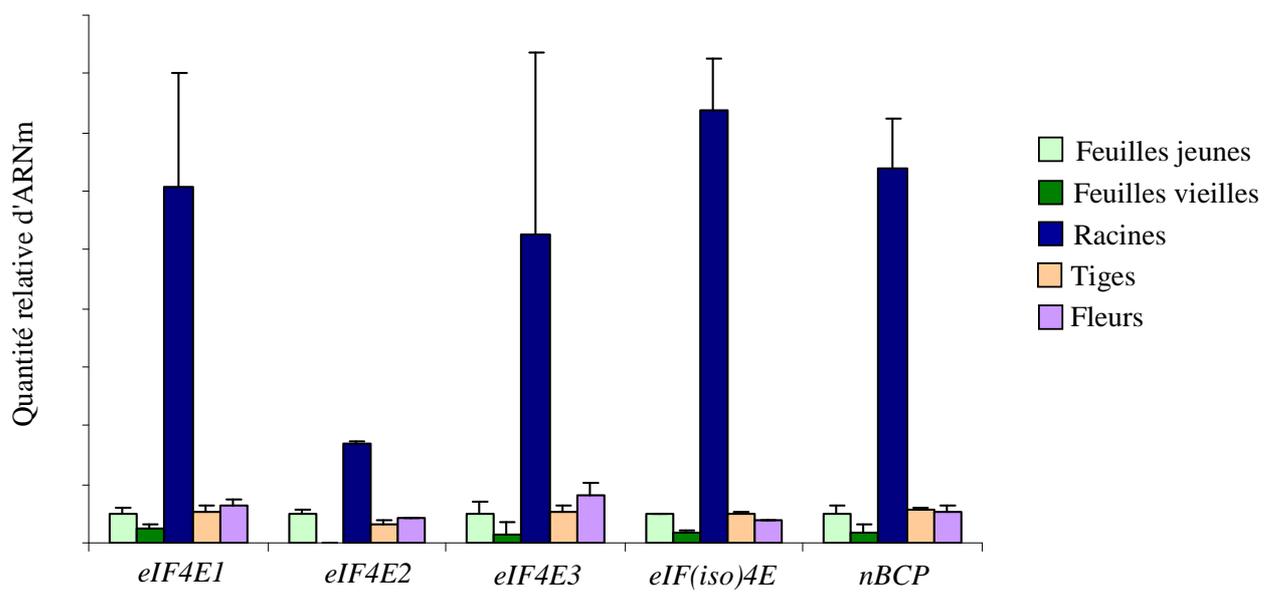
[B]



**Figure IV.3. Contrôles nécessaires à l'exploitation des résultats obtenus par RT-PCR quantitative.**

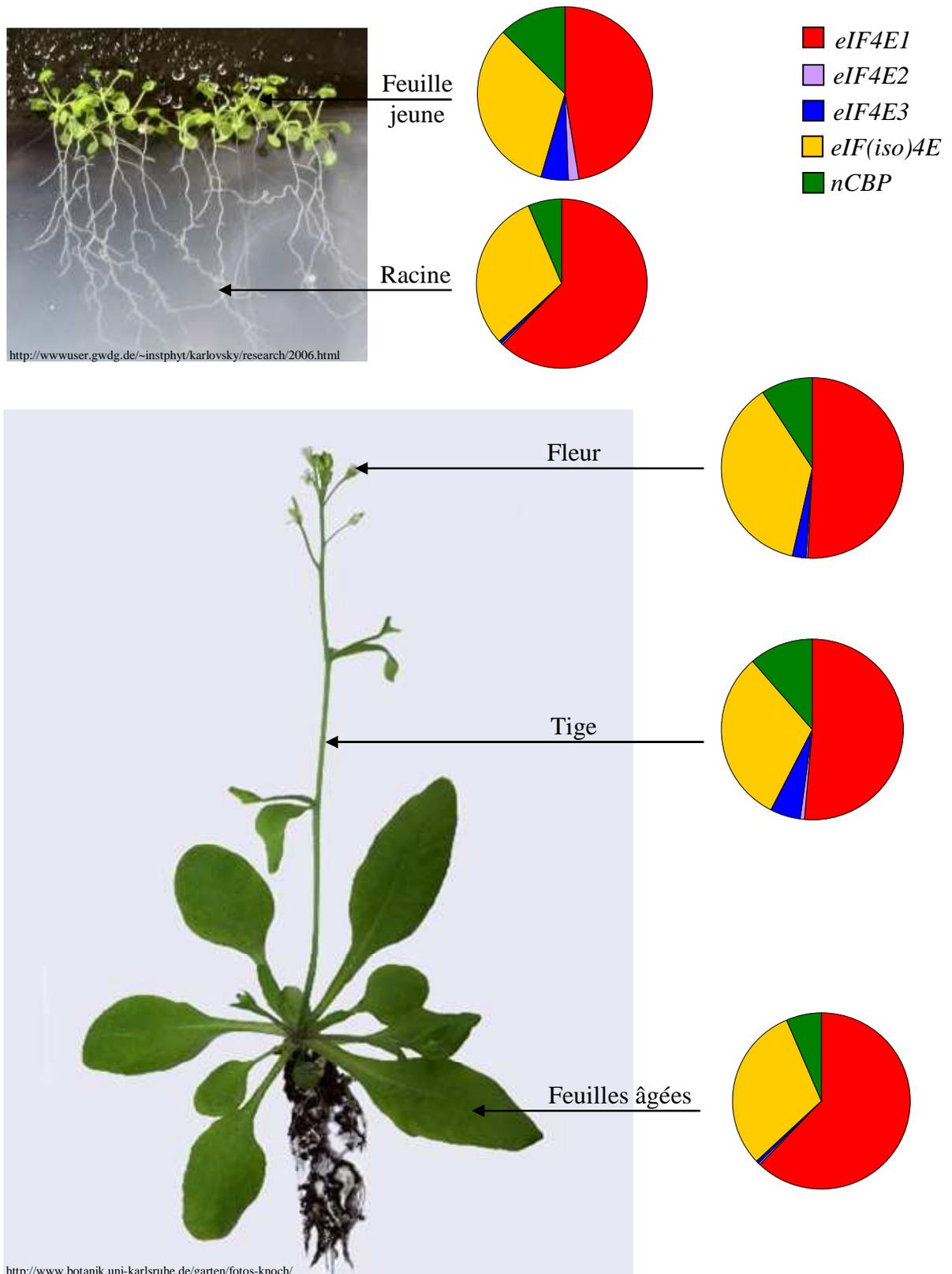
[A] Exemple de courbe standard pour vérifier l'efficacité de la PCR quantitative. L'efficacité est calculée à partir de la pente de la droite obtenue pour une gamme de dilution de 5 points en duplicat réalisée à partir d'un mélange équimolaire de tous les ADNc.

[B] Exemple de courbe de fusion où chaque pic représente un ADNc amplifié.



**Figure IV.4. Comparaison des niveaux d'accumulation des transcrits *eIF4E* entre les 5 organes de la plante.**

Les analyses ont été effectuées par PCR quantitative en temps réel (moyenne de deux réplicats biologiques répliqués 3 fois pour chaque expérience).



**Figure IV.5. Comparaison du niveau d'accumulation de chacun des 5 transcrits *eIF4E* pour chaque organe**

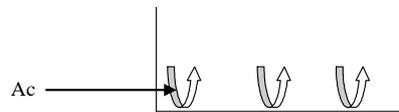
Les résultats obtenus par PCR quantitative en temps réel présentés correspondent à la moyenne de 2 réplicats biologiques répliqués 3 fois pour chaque expérience.



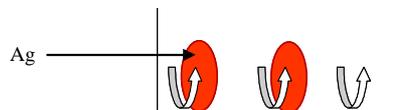
**Tableau A.1. Sérums utilisés pour la détection d'accumulation virale par la méthode DAS-ELISA.**

Virus	Serum	Dilution IgG	Dilution Conjugué
PVY	Piperade II, INRA, Avignon	1/500	1/800
TEV	Estève, INRA, Avignon	1/500	1/500
WMV	Adgen, Scotland, UK	1/750	1/350
TuMV	Adgen, Scotland, UK	1/750	1/750

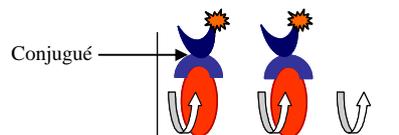
1. Sensibilisation des plaques : dépôt des anticorps dans les puits d'une plaque



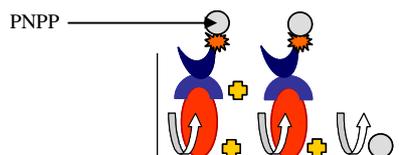
2. Fixation de l'antigène (protéines virales extraites des échantillons de feuilles broyées) aux anticorps



3. Fixation du conjugué (anticorps conjugué à la phosphatase alcaline)



4. Révélation : dépôt du substrat de la phosphatase alcaline (PNPP)



S'il y a réaction antigène-anticorps, la phosphatase alcaline dégrade le substrat, ce qui se traduit par une réaction colorimétrique (incolore ->jaune).

**Figure A.1. Principe du test DAS-ELISA.**

**Tableau A.2. Composition des milieux de culture utilisés pour les levures et les bactéries.**

Nom du milieu	Composants	Quantité pour 1L
<b>Levures</b>		
YPD*	Difco peptone	20 g
	Extrait de levure	10 g
	Glucose	20 g
SD-L*	Dropout Leu Trp	0,64 g
	Nitrogen w/o amino acid	6,7 g
	Dextrose	20 g
	Tryptophane	20 mg
SD-W*	Dropout Leu Trp	0,64 g
	Nitrogen w/o amino acid	6,7 g
	Dextrose	20 g
SD-LW*	Leucine	100 mg
	Dropout Leu Trp	0,64 g
	Nitrogen w/o amino acid	6,7 g
	Dextrose	20 g
SD-LWH*	Dropout Leu Trp	0,64 g
	Nitrogen w/o amino acid	6,7 g
	Dextrose	20 g
	Adénine hémisulfate	20 mg
SD-LWHA*	Dropout Leu Trp Ade His	0,6 g
	Nitrogen w/o amino acid	6,7 g
	Dextrose	20 g
SD gal/raf - URA*	Dropout Leu Trp Ura His	0,6 g
	SD gal/raf	37 g
	Leucine	100 mg
	Tryptophane	20 mg
	Histidine	20 mg
<b>Bactéries</b>		
LB*	Bacto-tryptone	10 g
	Bacto-yeast extract	5 g
	NaCl	5 g

\* Rajouter 20g d'agar pour faire les milieux solides

**Tableau A.3. Solutions nécessaires pour la microextraction d'ADN à partir de matériel végétal.**

Solutions	Composants	Concentration ou volume
Tampon sorbitol	D-sorbitol	0,35 M
	Tris	0,10 M
	EDTA	5,00 mM
	Ajuster le pH à 7,5	
	A conserver à 4°C	
Tampon de lyse des noyaux	Tris	0,20 M
	EDTA	0,05 M
	NaCl	2,00 M
	CTAB	2%
	A conserver à 20°C	
Tampon de microextraction	Tampon sorbitol	25 mL
	Tampon de lyse des noyaux	25 mL
	Sarcosyl 5%	10 mL
	Bisulfite de sodium (Na <sub>2</sub> S <sub>2</sub> O <sub>5</sub> )	0,2 g

**Tableau A.4. Amorces PCR répertoriées pour chaque type d'expérience.**

**[A] Pour amplifier les gènes *eIF4E* en vue de leur séquençage**

Plante	Gène	Amorce sens (5' -> 3')	Amorce antisens (5' -> 3')	Taille fragment PCR (en pb)
<i>Arabidopsis</i>				
ADNg	<i>eIF4E1</i>	Z588F GAGAAAGAGAAGCAGTTCCG	Z588R GACAGACTGTCAAGTCATG	1540
	<i>eIF4E2</i>	Z590F GGAGCATAGTGAAACGGGAAAA	Z523R TTTACAGTAAAGATAGAGCAAC	1720
	<i>eIF4E3</i>	Z589F CGGAGCAAAGTGAACGGAAAAAG	Z522R GTAAAGATAGATCAGAGGCAAC	1802
	<i>eIF(iso)4E</i>	Z591F GCCGTCTCAGAAGAAAACTCAACTGCC	Z591R CCGACCAAACAGTATCAC	1482
	<i>nCBP</i>	Z608F CTTCTGATTGGAATTTAGGGC	Z608R GCCTCCGGTACAGTATGGTATTGG	1662
<i>Capsicum</i> spp.				
ADNc	<i>eIF4E1</i>	Z119F AAAAGCACACAGCACCACA	Z119R TATTCGACATTGCATCAAGAA	729
	<i>eIF4E2</i>	Z401F GATAGTTTGTGTGAGGATTTTAC	Z403R GCAACCGCAAGTGACTGCTGAT	832
	<i>eIF(iso)4E</i>	Z320F ATGGCCACCGAAGCACCACCCGG	Z321R TCACACGGTGTATCGGCTCTTAGCT	613
ADNg	<i>eIF4E1</i>	Z1245F GAAGATCCTGTATGTGCCAATG	Z1245R CTGTGTACGATTTCTTTTC	1600
	<i>eIF(iso)4E</i>	Z1247F GAGTGGCCTAATGGTGGCAAGTG	Z1247R TCACACGGTGTATCGGCTC	2200
<i>Lycopersion</i> spp.				
ADNc	<i>eIF4E1</i>	Z1068F CTGAAATGGAGAGAAGCATGT	Z1070R CACTGCATCAAGAACTATACGG	687
	<i>eIF4E2</i>	Z449F GGGACGAAAAACCAAAAAATG	Z449R CCCTGTTGTAAACGATAGAACTA	703
	<i>eIF(iso)4E</i>	Z382F GACGGGAAAAACAATGGCCACCG	Z340R ATTTGCGAGCAGACAAGCGC	711
ADNg	<i>eIF4E1</i>	Z1245F GAAGATCCTGTATGTGCCAATG	Z1248R CTATACGGTGTAAACGATTTCTG	1000
	<i>eIF(iso)4E</i>	Z1250F GAGTGGCCTAATGGTGGCAAGTG	Z1250R CTACACAGTATATCGGCTCTTAG	2200
<i>Solanum</i> spp.				
ADNc	<i>eIF4E1</i>	Z1165F ATGGCAGCAGCTGAAATGGC	Z1167R GGCATGAATTCAAACTAAACA	840
	<i>eIF4E2</i>	Z449F GGGACGAAAAACCAAAAAATG	Z449R CCCTGTTGTAAACGATAGAACTA	703
	<i>eIF(iso)4E</i>	Z382F GACGGGAAAAACAATGGCCACCG	Z340R ATTTGCGAGCAGACAAGCGC	711

**Tableau A.4. -suite 1- Amorces PCR répertoriées pour chaque type d'expérience.**

**[B] Pour amplifier les gènes *eIF4E* et les gènes viraux en vue de leur clonage**

Plante, virus	Gène, allèle	Amorce sens (5' -> 3')	Amorce antisens (5' -> 3')	Taille fragment PCR (en pb)	Sites enzymatiques introduits
<i>Arabidopsis</i>					
ADNc	<i>eIF4E1</i>	Z441F AGAGGATTCATGGCGGTAGAAGACACTC	Z441R AGACTCGAGTCAAGCGGTGTAAGCGTTC	724	EcoRI / XhoI
	<i>eIF4E2</i>	Z581 AGAGGATCCATATGGTGGTTATGGATTC	Z520R AGACTCGAGTCAACAGTATATCGACG	739	BamHI / XhoI
	<i>eIF4E3</i>	Z517 AGAGGATCCATATGGTGGTTACGGATTC	Z520R AGACTCGAGTCAACAGTATATCGACG	739	BamHI / XhoI
	<i>eIF(iso)4E</i>	Z442F AGAGGATTCATGGCGCAGCATGATGTGA	Z442R AGAGGATCCCTCAACAGTGAATCGGCTTC	622	EcoRI / BamHI
	<i>nCBP</i>	Z587F AGGATCCATATGGAGGTTTTGGATAG	Z587R ACTCGAGCTATCCTCTCAGCCAT	682	BamHI / XhoI
<i>Capsicum</i> spp.					
ADNc	<i>eIF4E1</i>	Z302 AGAGGATTCATGGCAACAGCTGAAATGGAG	Z303 AGAGGATCCCTTACTGTGTAAACGATTCCT	705	EcoRI / BamHI
	<i>eIF(iso)4E, pvr6<sup>+</sup></i>	Z515 AGAGGATCCATATGGCCACCGAAGCACCACCA	Z516R AGACTCGAGTCAACCGGTGTAATCGGCTCT	627	BamHI / XhoI
	<i>eIF(iso)4E, pvr6</i>	Z515 AGAGGATCCATATGGCCACCGAAGCACCACCA	Z597 ACTCGAGTCAACTTGTGGGCTTTGAATA	542	BamHI / XhoI
	<i>eIF4E1-mutD109N</i>	pvx1 CATCGATTTATGSCAACAGCTGAAATGGAGAA	pvx3 TCTCGAGATTTTACTGTGTAAACGATTC	701	ClaI / XhoI
<i>Potato virus Y</i>					
LYE84	<i>VPg</i>	Z304 AGAGGATCCATGGGAAAAATAATCCAAAAGA	Z306 AGAGGATCCCTCATTCATGCTCTACTTCTGTT	582	BamHI / BamHI
	<i>Nla</i>	Z304 AGAGGATCCATGGGAAAAATAATCCAAAAGA	Z308 AGAGGATCCCTCACTGCTCCACACTACATCATG	750	BamHI / BamHI
SON41p	<i>Nlb</i>	Z310 AGAGGATCCCTAGCTAAACACTCTGCATGGATG	Z312 AGAGGATCCCTCACTGGTGTATGACTTCAAGA	1575	BamHI / BamHI
	<i>VPg</i>	Z305 AGAGGATCCATGGGAAAAATAATCCAAAGAGA	Z307 AGAGGATCCCTCATTCATGCTCTACTTCTGTT	582	BamHI / BamHI
SON41g	<i>Nla</i>	Z305 AGAGGATCCATGGGAAAAATAATCCAAAGAGA	Z309 AGAGGATCCCTCACTGCTCCACACCATCATG	750	BamHI / BamHI
	<i>Nlb</i>	Z311 AGAGGATCCCTAGCCAGCACTCTGCATGGATG	Z313 AGAGGATCCCTCACTGATGATGCACCTCATAGA	1575	BamHI / BamHI
SON41g	<i>VPg</i>	Z305 AGAGGATCCATGGGAAAAATAATCCAAAGAGA	Z307 AGAGGATCCCTCATTCATGCTCTACTTCTGTT	582	BamHI / BamHI
	<i>Tobacco etch virus</i>				
HAT	<i>VPg</i>	Z435 AGAGGATTCGGGAAAGAGAATCAGAAG	Z643 AGACTCGAGCTATTCAAACGTCAAGTCTCTC	582	EcoRI / XhoI
	<i>Nla</i>	Z435 AGAGGATTCGGGAAAGAGAATCAGAAG	Z436 AGACTCGAGCTATTGCGAGTACACCAATTC	744	EcoRI / XhoI
	<i>Nlb</i>	Z438 AGAGGATTCGGGAAAGAGAATG	Z439 AGACTCGAGCTACTGAAAATAAAGATTTCTC	1554	EcoRI / XhoI
CAA10	<i>VPg</i>	Z435 AGAGGATTCGGGAAAGAGAATCAGAAG	Z643 AGACTCGAGCTATTCAAACGTCAAGTCTCTC	582	EcoRI / XhoI
	<i>Nla</i>	Z435 AGAGGATTCGGGAAAGAGAATCAGAAG	Z436 AGACTCGAGCTATTGCGAGTACACCAATTC	744	EcoRI / XhoI

**Tableau A.4. -suite 2- Amorces PCR répertoriées pour chaque type d'expérience.**

**[C] Pour amplifier les gènes de ménage et les EST nécessaires aux analyses de polymorphisme des gènes *eIF4E***

Plante	Gène, marqueur	Amorce sens (5' -> 3')	Amorce antisens (5' -> 3')	Taille fragment PCR (en pb)
<i>Capsicum</i> spp.				
ADNg	HpmsE121 <sup>1</sup>	Z1316F TGAGAGTTGGATCCCTCTTCTTCG	Z1316R CTGTTCGGCTTGAAAGAGGA	~ 200
	HpmsE115 <sup>1</sup>	Z1317F TCATCTCATAGCCTGCCCCCTA	Z1317R CCACCTGAAGAAGCCATGACCA	~ 200
	HpmsE141 <sup>1</sup>	Z1318F TCCCAACAACCTCAAATGGCTTC	Z1318R TGGAGTGCCCTTCTGGTAAAG	~ 250
	HpmsE073 <sup>1</sup>	Z1319F TTATTCCAGGCCACTTATCGAA	Z1319R CAGCAGCAAATCTTGATTTC	~ 250
	HpmsE006 <sup>1</sup>	Z1321F GCTGACCGTTTTCGTTTGGG	Z1321R CAAAATCAACCGCACCAACA	~ 200
	HpmsE146 <sup>1</sup>	Z1322F CCCTCTTCTCCTTCCACCATCA	Z1322R AGCGTGAAAGGGTTA TGAGGA	~ 200
	HpmsE015 <sup>1</sup>	Z1323F TTGTGAGGTTTGACACTGGGA	Z1323R CCGAGCTCGATGAGGATGAACCT	~ 150
	HpmsE020 <sup>1</sup>	Z1325F CCCCAGAGAGGAAACAGAATCAT	Z1325R TTCCATTTGGTCCAGTACCA	~ 200
	HpmsE098 <sup>1</sup>	Z1326F CCAGCAAACTCCCTAAACCCACT	Z1326R CAAACCCCTTCAACTTCCTCA	~ 200
	HpmsE096 <sup>1</sup>	Z1327F CGGGTCAAAACAAAACCGAAGT	Z1327R GCTTGTGGTTGAGCTCGCTCTT	~ 200
	HpmsE061 <sup>1</sup>	Z1328F CCCAGAAAGAAAGTTGGGAATGG	Z1328R TTCGACGAGCTTGGAAAGTGAT	~ 300
	HpmsE132 <sup>1</sup>	Z1329F ACATCCACAGCAAAAGGAAAAA	Z1329R GTGGAAGTTTGGTGGATCAGA	~ 200
	<i>eF1 α</i>	Z1089F ATGGGTAAAGAGAAGATTTC	Z1089R TCACCTCCCCTTCTTCTG	1400
	<i>H2B-1</i>	Z1053F GGCCCGACAAGAAGAAGAAGA	Z1053R CTGCTTGGTTCCTTCAGAAA	430
	<i>ndhF5</i>	Z1183F CGGGTCTTATACATGCTGCTAC	Z1075R GATACCTTCTCTACAAT	2100
	<i>rbcL</i>	Z1076F ATGTCCACCAAAACAGAGAC	Z1076R CTCACAAGCAGCAGTAGT	1400
<i>Lycopersicon</i> spp.				
ADNg	CT066 <sup>2</sup>	CT066F CGCTGTCCTCTTACCACCC	CT066R AATTGCTCTGCCACTTTTCCTAC	1346
	CT093 <sup>2</sup>	CT093F CTCCCTCGGCTCAGCATT	CT093R AGCAGCCCTCAGAACGGACT	1415
	CT148 <sup>2</sup>	CT148F TCGTGTCAAGGCTGTGTAGT	CT148R GGTGTCTCCAGATAGTTGT	1497
	CT179 <sup>2</sup>	CT179F CGAATTCATCTCCACTCA	CT179R TAAGACCAGCAAACTACCAC	995
	CT189 <sup>2</sup>	CT189F CAAGATCCCGCTGACCCTG	CT189R TGGAGCAGCAAGAAGTGTGA	1463
	CT198 <sup>2</sup>	CT198F TGACAACTACCGAATTACGA	CT198R GGTGATTTATTAGTGCCACA	779
	CT208 <sup>2</sup>	CT208F CTATGGAGTTATATTTCCACACA	CT208R ACTTTTGAGAGGACATCAATT	1767
	Sucr <sup>2</sup>	SucrF GGACCAATGTATCACAAGGATG	SucrR CAAAGGGAAGGATGAATATTAC	1575
	<i>eF1 α</i>	Z1089F ATGGGTAAAGAGAAGATTTC	Z1089R TCACCTCCCCTTCTTCTG	1400
	<i>H2B-1</i>	Z1053F GGCCCGACAAGAAGAAGAAGA	Z1053R CTGCTTGGTTCCTTCAGAAA	430
	<i>ndhF5</i>	Z1183F CGGGTCTTATACATGCTGCTAC	Z1075R GATACCTTCTCTACAAT	2100
	<i>rbcL</i>	Z1076F ATGTCCACCAAAACAGAGAC	Z1076R CTCACAAGCAGCAGTAGT	1400

<sup>1</sup> d'après Yi et al., 2006; <sup>2</sup> d'après Roselius et al., 2005

**Tableau A.4. -suite 3- Amorces PCR répertoriées pour chaque type d'expérience.**

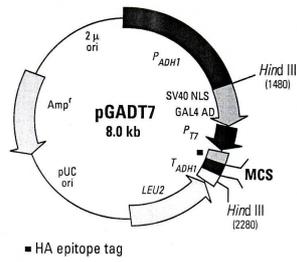
**[D] pour la PCR quantitative**

Plante	Gène, allèle	Amorce sens (5' -> 3')	Amorce antisens (5' -> 3')	Taille fragment PCR (en pb)
<i>A. thaliana</i>				
ADNc	<i>eIF4E1</i>	Z1056F TCGAGCAAATCCGGTGTCTCTG	Z1056R ACAGTAGAAGTCAGCTCCGTG	207
	<i>eIF4E2</i>	Z1057F CTTTGTATACATTGCTACC	Z1057R TAACGTGTTAAGCCAATTAG	206
	<i>eIF4E3</i>	Z1058F CCTTGTATACATTGCGTACT	Z1058R TAACGTGTTAAGCCAATTGC	206
	<i>eIF(iso)4E</i>	Z1169F AGTGGACTTGGGTGTTACTG	Z1169R GAACAGCTTCATTAGACTTCG	190
	<i>nCBP</i>	Z1170F TTCTCAAAGTTGTATCTGCTC	Z1170R GAGTCTCTCAAAACCATCACT	181
	<i>histone 1</i>	Z1197F AGAAAACAACCTTCCCTCTCACCCTA	Z1197R AGGCAGTGATTTATGTTTCTCCTC	180
	<i>GAPDHA</i>	Z1199F TTGCACCTACCACTGTCTTGCT	Z1199R AGCTGCTTTGCTCTCTCTAGA	185

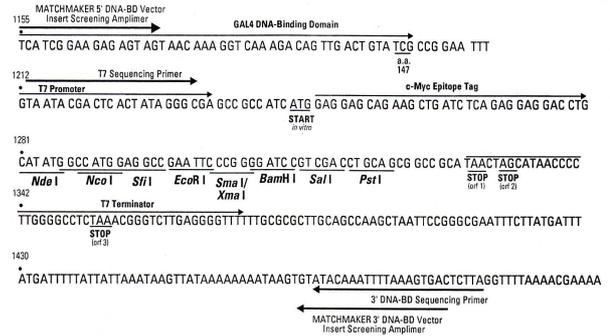
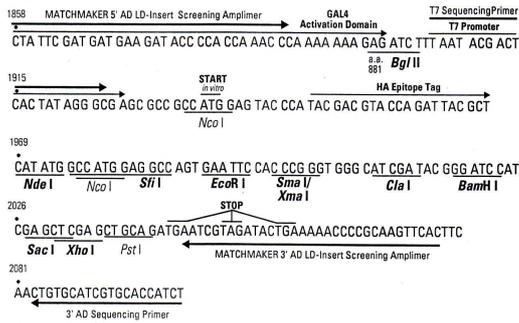
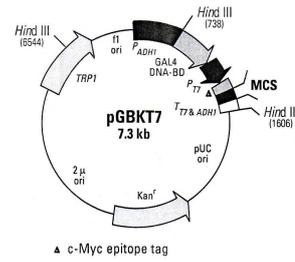
**[E] pour la PCR semi-quantitative**

Plante	Gène, allèle	Amorce sens (5' -> 3')	Amorce antisens (5' -> 3')	Taille fragment PCR (en pb)
<i>A. thaliana</i>				
ADNc	<i>eIF4E1</i>	Z588F GAGAAAGAGAAGCAGTTCGG	Z588R GACAGACTGTCAAGTCATG	885
	<i>eIF4E2</i>	Z590F GGAGCATAGTGAACGGGAAAA	Z737R GGTGATGTTCTAATATGGCTTCCC	957
	<i>eIF4E3</i>	Z589F CGGAGCAAAGTGAACGGGAAAA	Z522R GTAAGATAGATCAGAGGCAAC	955
	<i>eIF(iso)4E</i>	Z591F GCCGCTCAGAAAGAACTCAACTGCG	Z591R CCGACCAACAGTATCAC	695
	<i>nCBP</i>	Z608F CTTCTGATTTGGAATTTAGGGC	Z608R GCCTCCGGTACAGTATGGTATTGG	844
	<i>eF1 α</i>	Z975F ATGAACCCCTGGTTCAG	Z975R TCACCTCCACCCCTTCTTG	218

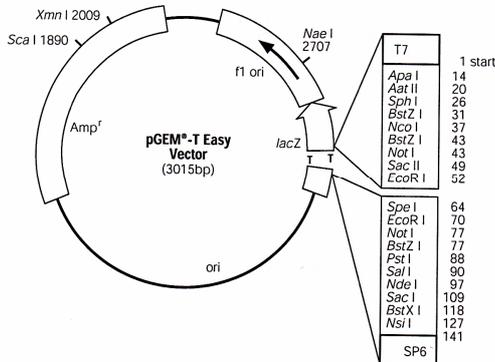
a



b



c



d

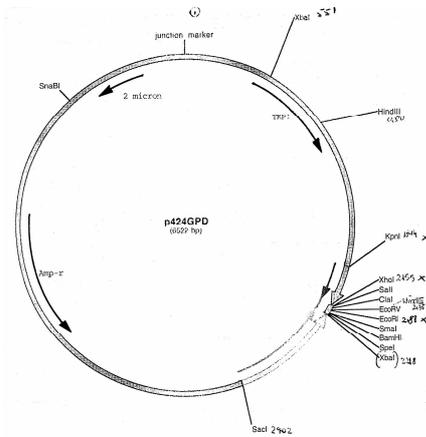
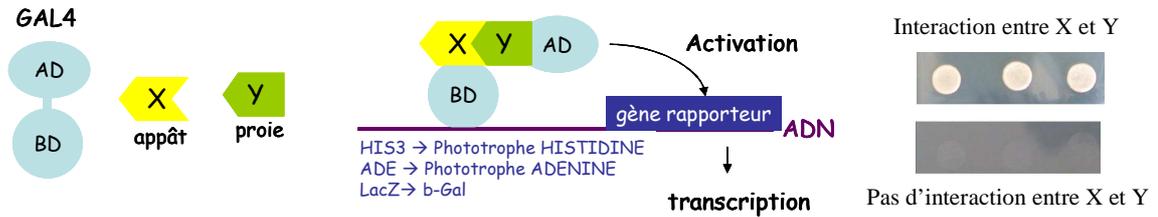
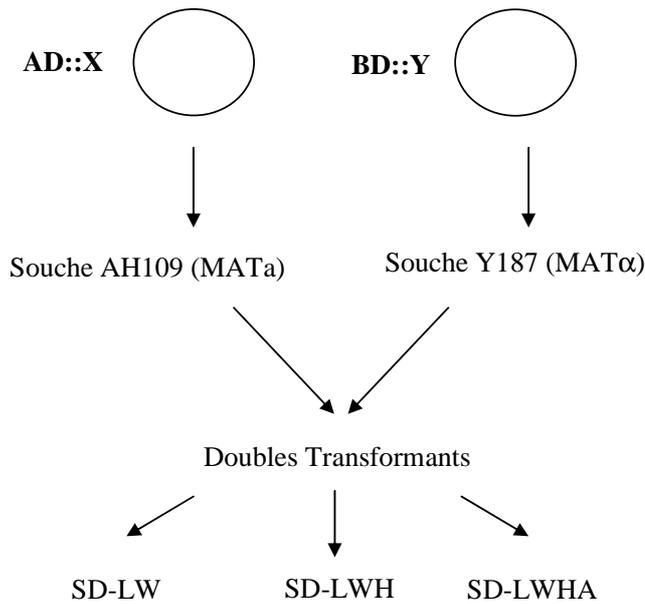


Figure A.2. Schémas des vecteurs et de leur site de clonage multiple utilisés lors de ce travail. pGADT7 (a), pGBKT7 (b), pGEMT-easy (c) et p424GBP/TRP1 (d).

a.



b.



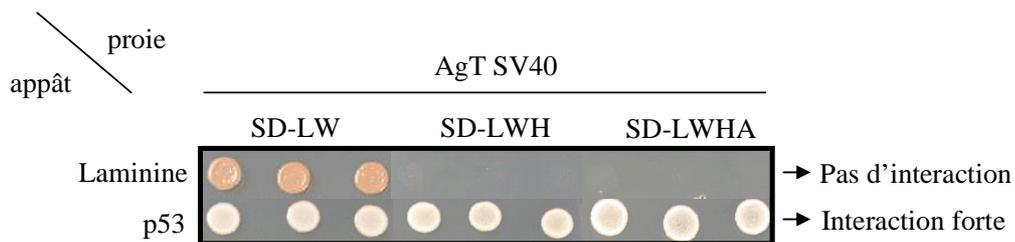
1. Clonage des partenaires en fusion traductionnelle avec le domaine d'activation (AD) du vecteur pGADT7 et le domaine de liaison (BD) du vecteur pGBKT7

2. Transformation des levures AH109 et Y187 avec les plasmides pGADT7 et pGBKT7 portant les gènes proie et appât. Les transformants sont étalés sur milieux SD-L et SD-LW respectivement.

3. Conjugaison. Les doubles transformants sont sélectionnés sur boîte SD-LW.

4. Tests d'auxotrophies par repiquage en triplicat sur milieux sélectifs SD-LWH et SD-LWHA. Le milieu SD-LWH permet de détecter les interactions protéine / protéine faibles.

c.

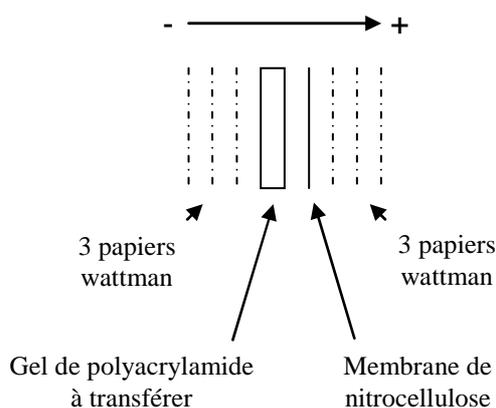


**Figure A.3. Principe (a) et méthode (b) du système de double hybride dans la levure pour détecter *in vivo* des interactions protéine / protéine.**

Les contrôles négatif et positif (c) sont fournis par le kit Matchmaker GAL4 two-hybrid system 3 (Clontech, Mountain View, CA, USA) : l'AgTSV40 n'interagit pas avec la laminine tandis qu'il interagit fortement avec la protéine p53 de souris.

**Tableau A.5. Composition du gel de polyacrylamide SDS-PAGE et du tampon de migration.**

		Composant	Quantité ou volume
Gel acrylamide	Gel de résolution (12%)	acrylamide 40%	2,5 mL
		Tris-HCl 1,5 M pH 8,8 5X	1,5 mL
		SDS 10%	75 µL
		H <sub>2</sub> O <sub>m</sub> Q	3,6 mL
		APS 10%	60 µL
		TEMED	2,5 µL
<i>Laisser polymériser 1 h avant de couler le gel de concentration</i>			
Gel de concentration (4%)		acrylamide 40%	250 µL
		Tris-HCl 1 M pH 6,8 8X	310 µL
		SDS 10%	25 µL
		H <sub>2</sub> O <sub>m</sub> Q	1,8 mL
		APS 10%	25 µL
		TEMED	2,5 µL
<i>Ajouter une pointe de bleu pour visualiser les puits</i>			
<i>Placer le peigne et laisser polymériser 1 h avant de déposer les échantillons</i>			
Tampon de migration		Tampon Tris (0,25 M) - Glycine (1,92 M) pH 8,3	25 mL
		SDS 10%	2,5 mL
		H <sub>2</sub> O <sub>m</sub> Q	qsp 250 mL



**Figure A.4. Montage du Western blot pour transférer les protéines du gel de polyacrylamide SDS-PAGE vers une membrane de nitrocellulose.**

Les feuilles de papier wattman et la membrane de nitrocellulose, découpées aux dimensions du gel, ont été préalablement imbibées de tampon de transfert. Ce montage est placé dans le support amovible de la cuve de transfert, en respectant l'orientation, de l'électrode négative vers l'électrode positive.

**Carine CHARRON**

**Caractérisation fonctionnelle et évolution moléculaire des gènes codant pour les facteurs d'initiation de la traduction eIF4E : des facteurs clés dans la résistance des plantes aux *Potyvirus***

L'objectif de cette étude est de décrire puis analyser et comparer le polymorphisme de séquence des gènes *eIF4E* (*eukaryotic Initiation Factor 4E*) de résistance aux potyvirus, afin de mettre en relation ce polymorphisme avec la double fonction des protéines eIF4E dans les processus cellulaires de l'hôte et dans l'interaction avec les potyvirus. La stratégie repose sur l'association d'analyses fonctionnelles et d'études d'évolution moléculaire des gènes *eIF4E* chez deux Solanacées (piment et tomate) et chez l'espèce modèle *Arabidopsis thaliana*. En ce qui concerne le rôle de eIF4E dans les processus cellulaires, la pression de sélection purificatrice forte identifiée pour l'ensemble des gènes *eIF4E* argumente en faveur de l'acquisition d'une spécialisation fonctionnelle, et ce bien que toutes les protéines eIF4E apparaissent capables de compléter une levure délétée de son gène *eIF4E*. En ce qui concerne le rôle de eIF4E dans les interactions plantes-potyvirus, une importante diversité de variants alléliques a été identifiée au locus *eIF4E1-pvr2* du piment, diversité caractérisée par des substitutions nucléotidiques exclusivement non-synonymes. La plupart de ces substitutions, impliquées dans la résistance aux PVY (*Potato virus Y*) et TEV (*Tobacco etch virus*) via une modification de l'interaction avec la protéine virale VPg (*Viral Protein genome-linked*), sont sous pression de sélection positive. Ces résultats couplés à ceux montrant que des mutations dans la VPg permettent aux potyvirus de s'adapter aux protéines eIF4E1 de résistance, mettent en évidence l'existence d'une coévolution entre eIF4E1 et VPg. Cette coévolution n'est cependant pas retrouvée chez la tomate et *A. thaliana*, indiquant que leurs caractéristiques biologiques d'interaction avec les potyvirus "*in natura*" conduisent à des profils différents d'évolution des gènes *eIF4E*. Plus globalement, ces résultats associés à ceux montrant la conservation de la position des substitutions en acides aminés dans les protéines eIF4E de résistance d'espèces très éloignées suggèrent l'existence d'une évolution convergente.

**Mots clés :** eIF4E, gènes récessifs de résistance, diversité allélique, piment, tomate, *A. thaliana*, *Potyvirus*, coévolution, sélection positive, évolution convergente.