



UNIVERSITÀ DEGLI STUDI DI UDINE

Dottorato di Ricerca in Scienze e Biotecnologie Agrarie
Ciclo XXV
Coordinatore: prof. Mauro Spanghero

TESI DI DOTTORATO DI RICERCA

**STUDI FINALIZZATI AD AUMENTARE
L'ESPRESSIONE DI TRANSGENI IN
ENDOSPERMA DI RISO MEDIANTE
INGEGNERIZZAZIONE DI PROMOTORI
CHIMERICI SEME-SPECIFICI**

DOTTORANDO
dott. Piero Cristin

SUPERVISORE
prof. Stefano Marchetti

ANNO ACCADEMICO 2012/2013

SUMMARY

The biopharmaceutical industry is seeking novel, more performing platforms for the production of recombinant proteins. In the last decade plants have been emerging as a good alternative to animal cells in the manufacturing of pharmaceutical proteins, especially for injectable preparations. Several studies have shown that the seed-specific expression of a transgene is more convenient than the constitutive expression. Specifically, the endosperm of cereals can be a suitable tissue to accumulate and durably store high amounts of heterologous proteins. *Oryza sativa* is a promising species for biotechnological application, but up to this time, the yield of recombinant protein is still not enough to push industries toward this innovative manufacturing platform. One of the more effective ways to improve the yield of a recombinant protein is to deal with the promoter. The rice endosperm-specific expression can be easily achieved by the use of a variety of promoters of the rice seed storage proteins (RSSPs). These promoters have been widely studied and some of the *cis*- and *trans*-acting factors responsible of the specific expression have been identified. Several authors have described methods for promoters sequence analysis and manipulation. However, no studies had been published about the engineering of promoters for rice endosperm expression.

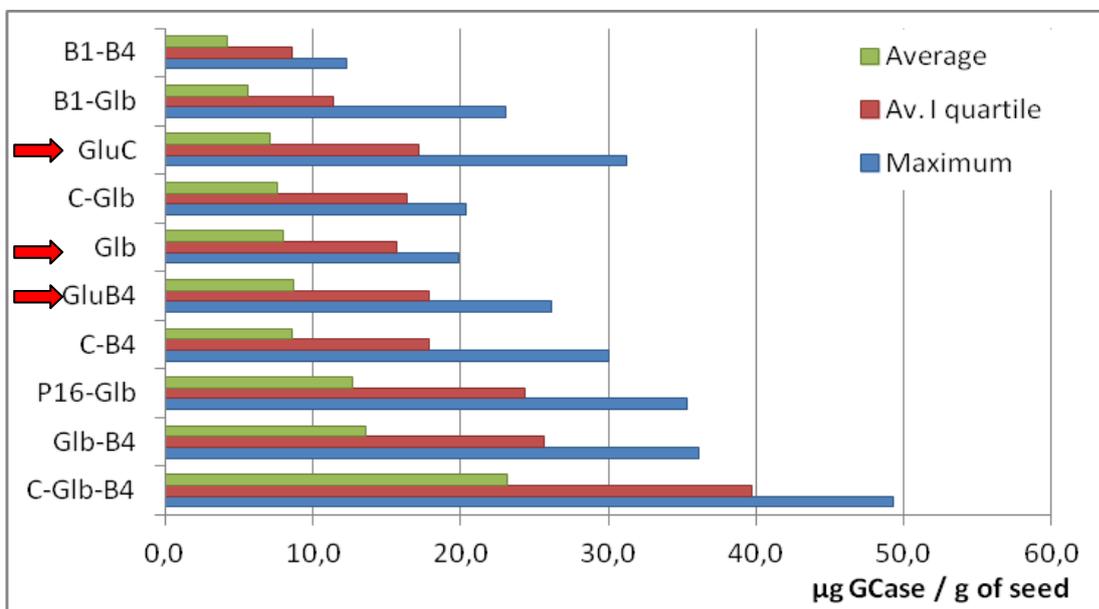
The aim of this PhD has been the application of those manipulation methods to the RSSP promoters. An early goal has been the definition of methods for sequence analysis and manipulation in order to increase the expression level of the RSSP promoters. The ultimate purpose of this PhD has been the construction of a synthetic promoter with activity higher than the natural homologues.

The research project had been structured according to an iterative process and two experimental cycles had been planned. Fourteen sequences of natural RSSP promoters have been analysed, with the purpose of identifying the *cis*-elements. Six of these promoters have been selected on the base of sequence analysis and previously published data. The selected promoters drive the expression of the following RSSPs: B1, B4 and C glutelins, 10 kDa and 16 kDa prolamins and 26 kDa globulin (with acronym:

GluB1, GluB4, GluC, Prol10, Prol16 and Glb, respectively). These promoters have been cloned following PCR isolation from *Oryza sativa* (var. CR W3) genome. A number of chimera sequences have been designed by combining in various ways portions of these promoters. Seven of these chimera sequences have been selected for *in vivo* testing.

In order to estimate the *in vivo* performances, the synthetic promoters have been inserted in an expression cassette with a reporter gene coding for a phytotoxic protein, namely the human acid β -glucosidase (GCCase). The only variable of the experiment has been the sequence upstream the TATA-box. The synthetic promoters have been tested in transformed plants of *Oryza sativa* (var. CR W3) together with three of the strongest natural RSSP promoters, i.e. GluB4, Glb and GluC. A population of at least 20 individuals has been produced for each construct. The amount of the reporter protein in the seed of each plant has been determined and each promoter has been evaluated on the base of three parameters: maximum expression level, average expression level calculated taking account the whole population or the first quartile.

The results of the study have been resumed in the following rank.



The three natural promoters are pointed out by a red arrow. GluB4 has proven to be the strongest natural RSSP promoter, according to the results previously published by other authors. Two of the synthetic

promoters have shown significantly higher activity in endosperm in comparison with GluB4 (+50%). One of the synthetic promoter has determined a two-fold expression level as compared to GluB4 (+113%). Since the reporter protein is phytotoxic, a loss in spatial or temporal expression specificity is supposed to be easily revealed. There has been no evidence of a massive loss of endosperm-specificity, since almost all the rice plants have achieved full maturation and the seed productivity has been enough for analysing a statistically significant number of individuals. Nevertheless, the seed productivity has not been even among the groups. The strongest synthetic promoter has also shown the lowest average seed yield and a high percentage of floral abortion. In order to investigate whether there is a correlation between the incidence of floral abortion and a little loss of spatial or temporal tissue-specificity, the expression level of the reporter gene has been determined in various tissues. A slightly higher content of GCase has been detected in the leaves of the plants transformed with the best synthetic promoter, but the difference with natural promoters has not been statistically significant, also due to the high variability among the individuals of single populations. Among the populations, the average expression level in leaf has been generally proportional to that in endosperm. However, within single populations, there is no correlation either between expression level in leaf and in endosperm or between expression level in leaf and seed yield. No significant difference has been found in the expression of the reporter gene in tissues other than leaf between natural and synthetic promoters. The distribution of the reporter protein between endosperm and aleuronic layer has been almost the same for all natural and synthetic promoters. All these data taken together show that the endosperm specificity had been preserved.

The main target of the study has been achieved, since three synthetic promoters have increased the yield of recombinant protein compared to natural RSSP promoters. The iterative method has allowed getting that result without constructing and testing *in vivo* a large number of synthetic promoters. Furthermore, the methodological approach followed in this work has allowed the identification, within the RSSP promoters, of the regions which are more likely involved in increasing the expression level of the

transgene in rice endosperm. For instance: Glb and GluB4 had been found to be good bases for the design of the next synthetic promoters; the 400bp region upstream of the TATA-box of both promoters is particularly rich of *cis*-elements that drive high transgene expression levels in the endosperm; the Glb portion comprised between the 5' end and the 400bp downstream region, when fused to the 5' end of the GluB4, has been able to double the yield of the recombinant protein. That portion contains important binding motifs for the RISBZ-2 transcription factor. The importance of those *cis*-acting elements had been previously underlined by other authors, even though these elements had been reported not to improve the expression level when fused to other RSSP promoters, unless RISBZ-2 has been simultaneously overexpressed. The outcomes of this study have demonstrated, discording with what previously published, that the RISBZ-2 binding region of Glb can be used to successfully improve the expression level of a natural RSSP promoter; also the Prol16 region comprised between the 5' end and the 500bp downstream region has shown to positively affect the expression level of the transgene, while the results of this PhD advice against the use of the 5' region of GluB1 and GluC in designing synthetic promoters. The outcomes of the sequence analysis have been generally confirmed by the *in vivo* results. That means the analysis method is consistent and can be exploited to predict the presence of *cis*-elements. Shuffling these elements in novel, even stronger synthetic endosperm-specific promoters may be the way toward the optimization of an industrial, rice-based, protein-production system.

1	<u>INTRODUZIONE</u>	12
1.1	PIANTE COME BIOREATTORI VERDI	12
1.1.1	IMPIEGO DI ORGANISMI GENETICAMENTE MODIFICATI PER LA PRODUZIONE DI BIOFARMACI	12
1.1.2	UTILIZZO DI PIANTE COME BIOFABBRICHE	12
1.1.3	PIPELINE	13
1.1.4	IL SEME COME ORGANO DI ACCUMULO	14
1.1.5	VANTAGGI TECNICI DEL SEME DI RISO	15
1.1.6	MECCANISMI CHE INFLUENZANO IL LIVELLO DI ESPRESSIONE	16
1.2	PROMOTORI ENDOSPERMA-SPECIFICI	17
1.2.1	PROMOTORI DI PROTEINE DI RISERVA DEL SEME DI RISO	17
1.2.2	CLASSIFICAZIONI DEI PROMOTORI DI PROTEINE DI RISERVA DEL SEME DI RISO	19
1.2.3	ELEMENTI <i>CIS</i> -REGOLATORI DEI PROMOTORI DI PROTEINE DI RISERVA DEL SEME	20
1.2.4	APPROCCI SPERIMENTALI NELLO STUDIO DEI FATTORI DI TRASCRIZIONE	24
1.2.5	CARATTERIZZAZIONE DI ALCUNI FATTORI DI TRASCRIZIONE	25
1.3	PROMOTORI SINTETICI	27
1.3.1	STRUMENTI PER L'ANALISI DI SEQUENZA E MANIPOLAZIONE DEI PROMOTORI	28
1.3.2	APPROCCIO SINTETICO COMBINATORIO	30
1.3.3	VERIFICA <i>IN VIVO</i> DELL'ATTIVITÀ DEI PROMOTORI SINTETICI	32
1.3.4	STUDI BASATI SUL PROMOTORE DELLA SUBUNITÀ 35S DEL VIRUS DEL MOSAICO DEL CAVOLFORE	33
1.3.5	UTILIZZO DI ELEMENTI <i>CIS</i> -REGOLATORI PER L'ESPRESSIONE SPECIFICA NELL'ENDOSPERMA	34
2	<u>SCOPO DEL DOTTORATO DI RICERCA</u>	39
3	<u>APPROCCIO METODOLOGICO</u>	41
3.1	PIANIFICAZIONE DEL FLUSSO DELLE ATTIVITÀ	41
3.1.1	SCELTA DEL TIPO DI TRASFORMAZIONE E DEL VETTORE, NUMERO DI CICLI SPERIMENTALI E NUMERO DI TRASFORMAZIONE PER CICLO	42
3.1.2	SCELTA DELLE COMPONENTI DELLA CASSETTA DI ESPRESSIONE	44
3.2	PROGETTAZIONE DEI PROMOTORI	46
3.2.1	METODO DI ANALISI DELLE SEQUENZE DEI PROMOTORI DELLE PROTEINE DI RISERVA DEL SEME DI RISO (PRSR)	46
3.2.2	ANALISI BASATA SULL'INTERROGAZIONE DELLE BANCHE DATI	47

3.2.3 SVILUPPO DI UN METODO BASATO SUI DATI REPERITI IN LETTERATURA.....	48
3.2.4 POSSIBILI APPROCCI ALLA PROGETTAZIONE DI PROMOTORI SINTETICI	53
3.2.5 SCELTA DEI PROMOTORI DA COMBINARE	54
4 MATERIALI E METODI.....	57
4.1 SUPPORTI INFORMATICI	57
4.2 MATERIALI E METODI DI BIOLOGIA MOLECOLARE.....	57
4.2.1 ESTRAZIONE DI DNA GENOMICO DA FOGLIA DI RISO.....	57
4.2.2 METODICHE DI BIOLOGIA MOLECOLARE.....	57
4.3 VETTORI UTILIZZATI	61
4.3.1 pUC18.....	61
4.3.2 pGEM-T	62
4.3.3 pCAMBIA.....	62
4.4 PROMOTORI NATURALI DA GENOMA DI RISO CR W3	64
4.4.1 PROMOTORE DELLA GLUTELINA B4 CON LEADER SINTETICO LLTCK	64
4.4.2 SEQUENZE DI RIFERIMENTO PER GLI ALTRI PROMOTORI NATURALI.....	65
4.4.3 PRIMERS UTILIZZATI PER L'AMPLIFICAZIONE E IL SEQUenziAMENTO DEI PROMOTORI NATURALI.....	65
4.4.4 SEQUenziAMENTO DEGLI AMPLIFICATI	66
4.4.5 INVERSIONE DELL'ORIENTAMENTO DI INSERZIONE DEI PROMOTORI DI GLUB4 E GLB	67
4.5 COSTRUZIONE DEL MODULO ACCETTORE	68
4.5.1 TATA-BOX E LEADER	68
4.5.2 INSERIMENTO DEL TERMINATORE	68
4.5.3 INSERIMENTO DI UN SECONDO <i>ECOR</i> I A MONTE DELLA CASSETTA DI ESPRESSIONE	69
4.5.4 INSERIMENTO DEL GENE <i>REPORTER</i>	69
4.6 COSTRUZIONE DI pCAMBIA-ΔGLUC-GAA	70
4.7 COSTRUZIONE DEI PROMOTORI SINTETICI	71
4.7.1 DERIVATI DI GLB	71
4.7.2 DERIVATI DI PROL10.....	71
4.7.3 DERIVATI DI GLUB4.....	72
4.8 ASSEMBLAGGIO DELLA CASSETTA DI ESPRESSIONE E DEL VETTORE DI TRASFORMAZIONE	72
4.8.1 CASSETTA DI ESPRESSIONE	72
4.8.2 VETTORE DI TRASFORMAZIONE	73
4.9 TRASFORMAZIONE DI <i>ORYZA SATIVA</i> MEDIANTE <i>AGROBACTERIUM TUMEFACIENS</i>	73

4.9.1 PREPARAZIONE E SVILUPPO DI CALLI EMBRIOGENICI DA SCUTELLO DI RISO	74
4.9.2 CO-COLTURA DEI CALLI CON <i>A. TUMEFACIENS</i>	75
4.9.3 SELEZIONE DEI CALLI BASATA SU PMI	76
4.9.4 RIGENERAZIONE DI PIANTINE DI RISO DA CALLI TRASFORMATI.....	76
4.9.5 CONDIZIONI DI COLTIVAZIONE IN CELLA CLIMATICA E IN SERRA	79
4.10 VERIFICA DELLA PRESENZA DEL GENE <i>GAA</i> O <i>GCASI</i> NELLE PIANTE RIGENERATE	80
4.10.1 PCR DI CONTROLLO PER SAGGIARE L'AMPLIFICABILITÀ DEL DNA ESTRATTO	80
4.10.2 PCR PER L'AMPLIFICAZIONE DEL GENE <i>GAA</i> E DEL GENE <i>GCASI</i>	80
4.11 ANALISI PROTEICA DEI TRASFORMATI PRIMARI	81
4.11.1 ESTRAZIONE DI PROTEINE TOTALI DA SEMI <i>GAA</i> E <i>GCASI</i>	81
4.11.2 SAGGIO IMMUNOENZIMATICO (DAS-ELISA).....	82
4.11.3 VALIDAZIONE DEL SAGGIO DAS-ELISA	84
4.11.4 DETERMINAZIONE DEL LIVELLO DI ESPRESSIONE DEL GENE <i>REPORTER</i>	84
4.12 ANALISI DI TESSUTO-SPECIFICITÀ E PROFILO D'ESPRESSIONE	84
4.12.1 ESTRAZIONE DI PROTEINE TOTALI DA FOGLIA	85
4.12.2 ESTRAZIONE DI PROTEINE TOTALI DA SEMI ABORTITI E SEMI IMMATURI.....	85
4.12.3 PROVE DI SBIANCATURA	86
4.12.4 ESTRAZIONE DI PROTEINE TOTALI DALLE GLUME	86
4.12.5 ELABORAZIONE DEI DATI OTTENUTI	86
<u>5 RISULTATI</u>	<u>89</u>
5.1 MAPPATURA DEGLI ELEMENTI <i>CIS</i>-REGOLATORI	89
5.1.1 ELEMENTI <i>CIS</i> -REGOLATORI NEI PROMOTORI DELLE PROLAMINE	89
5.1.2 ELEMENTI <i>CIS</i> -REGOLATORI NEI PROMOTORI DELLE GLUTELINE	90
5.1.3 ELEMENTI <i>CIS</i> -REGOLATORI NEL PROMOTORE DELLA GLOBULINA 26KDA	91
5.2 SCELTA E CLONAZIONE DEI PROMOTORI NECESSARI AL LAVORO.....	92
5.2.1 SCELTA DEI CANDIDATI	92
5.2.2 PROMOTORI AMPLIFICATI DA DNA GENOMICO	93
5.2.3 SEQUENZIAMENTO DEGLI AMPLIFICATI.....	93
5.2.4 CLONAZIONE	94
5.3 COSTRUZIONE DEI PROMOTORI SINTETICI E DELLE RELATIVE CASSETTE DI ESPRESSIONE.....	94
5.3.1 COSTRUZIONE DI PCAMBIA- Δ GLUC-GAA	94
5.3.2 COSTRUZIONE DEL MODULO ACCETTORE	95

5.3.3 COSTRUZIONE DEI PROMOTORI SINTETICI	96
5.3.4 ASSEMBLAGGIO DELLE CASSETTE DI ESPRESSIONE E DEL VETTORE DI TRASFORMAZIONE	99
5.4 PRODUZIONE DELLE PIANTE TRANSGENICHE	100
5.5 PRODUZIONE DI SEME E LIVELLO DI ESPRESSIONE DEL GENE <i>REPORTER</i> NEI TRASFORMATI PRIMARI...	101
5.5.1 PRODUZIONE DI SEME	101
5.5.2 LIVELLI DI ESPRESSIONE INDOTTI DAI PROMOTORI SAGGIATI NEL PRIMO CICLO SPERIMENTALE	101
5.5.3 LIVELLI DI ESPRESSIONE INDOTTI DAI PROMOTORI SAGGIATI NEL SECONDO CICLO SPERIMENTALE	103
5.6 TESSUTO-SPECIFICITÀ	106
5.6.1 LIVELLO DI ESPRESSIONE DI GCASI IN FOGLIA	107
5.6.2 LIVELLO DI ESPRESSIONE DI GCASI IN VARI TESSUTI	109
5.6.3 DISTRIBUZIONE DELLA PROTEINA <i>REPORTER</i> ALL'INTERNO DEL SEME	109
5.7 APPENDICE ALLA SEZIONE DEI RISULTATI	110
5.7.1 RISULTATI DEI SEQUENZIAMENTI DEI PROMOTORI NATURALI	110
5.7.2 COSTRUZIONE DEL MODULO ACCETTORE	113
5.7.3 COSTRUZIONE DI PCAMBIA- Δ GLUC-GAA	115
5.7.4 COSTRUZIONE DEI PROMOTORI SINTETICI	117
5.7.5 ASSEMBLAGGIO DELLE CASSETTE DI ESPRESSIONE E DEL VETTORE DI TRASFORMAZIONE	130
5.7.6 TRASFORMAZIONE DI RISO CON <i>AGROBACTERIUM TUMEFACIENS</i>	135
5.7.7 ANALISI MOLECOLARI SULLE PIANTE TRASFORMATE	137
5.7.8 PRODUZIONE DI SEME	138
5.7.9 RISULTATI DEL PRIMO CICLO SPERIMENTALE	139
5.7.10 RISULTATI DELLA SESSIONE DI TRASFORMAZIONE EFFETTUATA A NOVEMBRE 2011	143
5.7.11 LIVELLI DI ESPRESSIONE NEL SEME DELLE PIANTE TRANSGENICHE PRODOTTE A FEBBRAIO 2012	148
5.7.12 ESTRAZIONE DI PROTEINE DA FOGLIE DEI TRASFORMATI PRIMARI	151
5.7.13 ESTRAZIONE DI PROTEINE DA SEME IMMATURO E DA SEME ABORTITO	151
5.7.14 PROVE DI SBIANCATURA	154
5.7.15 ESTRAZIONE DELLA PROTEINA TOTALE DAI PRODOTTI DI SBIANCATURA	155
5.7.16 LIVELLO DI ESPRESSIONE DI GCASI IN FOGLIA	156
5.7.17 LIVELLO DI ESPRESSIONE DI GCASI IN VARI TESSUTI	158
5.7.18 QUANTIFICAZIONE IN ELISA DI GCASI NEI PRODOTTI DI SBIANCATURA	161
6 <u>DISCUSSIONE</u>	163
6.1 STUDI PRELIMINARI ALLA PROGETTAZIONE	166

6.2	DISCUSSIONE SULL'APPROCCIO METODOLOGICO SEGUITO IN BASE AI RISULTATI OTTENUTI DALL'ANALISI DI SEQUENZA	167
6.3	LIVELLO DI ESPRESSIONE DEL GENE <i>REPORTER</i>	169
6.4	CONSIDERAZIONI SULLA RIPRODUCIBILITÀ DEL SAGGIO <i>IN VIVO</i>	177
6.5	TESSUTO-SPECIFICITÀ.....	179
6.6	IN CONCLUSIONE	185
<u>7</u>	<u>BIBLIOGRAFIA</u>	<u>187</u>
<u>8</u>	<u>SEQUENZE DEI PROMOTORI SINTETICI</u>	<u>199</u>
<u>9</u>	<u>INDICE DEI GRAFICI.....</u>	<u>207</u>
<u>10</u>	<u>INDICE DELLE FIGURE.....</u>	<u>209</u>
<u>11</u>	<u>INDICE DELLE TABELLE</u>	<u>215</u>
<u>12</u>	<u>RINGRAZIAMENTI.....</u>	<u>219</u>

1 INTRODUZIONE

1.1 Piante come bioreattori verdi

In questo capitolo verranno ripercorse brevemente le tappe che hanno portato all'attuale stato di sviluppo della tecnologia del *molecular farming*, tecnologia alla quale si intende applicare il frutto della ricerca svolta durante questo dottorato.

1.1.1 Impiego di organismi geneticamente modificati per la produzione di biofarmaci

Le applicazioni industriali della tecnologia del DNA ricombinante (rDNA) hanno rivoluzionato la produzione di proteine di interesse farmaceutico, e hanno dato vita a un settore industriale il cui fatturato ammonta a circa cento miliardi di euro (Egelkrout *et al.*, 2012). La produzione di proteine ricombinanti che utilizza i batteri come organismi ospite è di certo l'opzione tecnologicamente meno complessa e meno onerosa dal punto di vista finanziario; di contro, organismi di questo genere non si prestano sempre allo scopo per una serie di ragioni, fra cui l'assenza di un ambiente adeguato a favorire il corretto *folding* della proteina, la mancanza di processamento post-traduzionale (Oszvald *et al.*, 2008), la tossicità esercitata dalla proteina ricombinante nei confronti dell'organismo ospite e, nel caso di produzione di biofarmaci iniettabili per via endovenosa, il significativo rischio di contaminazione del prodotto finale da parte di componenti altamente tossiche dell'organismo ospite (Fischer e Emans, 2000; Giddings *et al.*, 2000; Daniell *et al.*, 2001). Per questo motivo, a fianco degli organismi ospiti microbici, sono stati sviluppati sistemi fondati su colture di cellule mammifere (es. CHO) o di insetto (es. SF9), di cellule e tessuti vegetali, nonché sulla produzione di animali e piante transgenici.

1.1.2 Utilizzo di piante come biofabbriche

L'uso di piante come bioreattori ha innegabili vantaggi quali il basso costo di produzione della biomassa, assenza di patogeni animali (virus e prioni), facilità di combinazione di transgeni diversi mediante incrocio, possibilità di processamento della biomassa in impianti già sviluppati per

altre finalità e rapida scalabilità delle colture. Tuttavia, il problema delle basse rese in proteina ricombinante ha in passato messo in secondo piano questi aspetti positivi. Nell'ultimo decennio però, i bioreattori verdi hanno conosciuto un rinnovato interesse da parte dell'industria, in ragione della possibilità di accumulare proteina ricombinante in specifici tessuti di riserva con rese promettenti (Egelkrout *et al.*, 2012).

A tutt'oggi la produzione di farmaci biotecnologici avviene per la quasi totalità in sistemi animali, per lo più in colture cellulari. Ciò sebbene il mercato richieda, spesso con urgenza, piattaforme che abbattano i costi e garantiscano un approvvigionamento meno condizionato da limiti tecnici insiti negli attuali metodi di produzione.

Uno dei motivi per cui non si è ancora riusciti a portare il sistema vegetale alla grande produzione è costituito dai limiti che questa piattaforma ancora ha in termini di resa sulla biomassa.

1.1.3 Pipeline

Tabella 1 Esempi di sintesi di proteine farmaceutiche in pianta (a esclusione di vaccini, anticorpi, nutraceutici).

Proteina	Pianta ospite	Applicazioni/commenti
Ormone della crescita	tabacco, girasole	prima proteina umana in pianta
Siero albumina	tabacco, patata, riso	cirrosi epatica
Fattore di crescita epidermide	tabacco	riparazione ferite
Interferone α	riso, rapa	epatite B e C
Eritropoietina	tabacco	anemia
GM-CSF	tabacco	neutropenia
IL-2 e IL-4	sosp. cell. tabacco	ematopoiesi
Antitripsina α -1	riso	fibrosi cistica
Fosfatasi alcalina di placenta	tabacco	prodotta per secrezione da foglia
Aprotinina	mais	inibitore della tripsina
Collagene	tabacco	primo polimero strutturale umano

1.1.4 Il seme come organo di accumulo

Il primo e più evidente vantaggio della produzione di proteine ricombinanti confinata nel seme dell'organismo ospite consiste nel fatto che il prodotto viene accumulato ad elevati livelli in un ambiente stabile nel quale può essere conservato per lunghi periodi a temperatura ambiente (Egelkrout *et al.*, 2012).

Proteine ricombinanti prodotte in seme possono essere accumulate senza incorrere in problemi di fitotossicità, a differenza di ciò che è stato frequentemente osservato in sistemi che ricorrono a un'espressione negli organi vegetativi. Pertanto, il sequestro di una proteina eterologa nei tessuti di riserva del seme, persino nel caso di elevati livelli di espressione, previene l'inibizione di crescita o l'abbattimento della resa in biomassa dovute all'azione della proteina stessa sul tessuto vegetativo dell'ospite (Takaiwa *et al.*, 2008).

Il problema, dovuto a livelli di accumulo inferiori al punto percentuale, che fu il principale svantaggio delle piattaforme di produzione basate su pianta, è ora superato dall'uso di promotori seme-specifici (Daniell *et al.*, 2001; Stoger *et al.*, 2005).

Il seme può accumulare quantità più elevate di proteina ricombinante rispetto ad altri tessuti vegetativi, perché è un naturale organo di deposito per le proteine, che servono come fonte di azoto durante la germinazione (Takaiwa *et al.*, 2008).

E' stato recentemente dimostrato che in endosperma è possibile esprimere proteine la cui espressione non avviene o non può avvenire in altri tessuti (Takaiwa *et al.*, 2007); inoltre, prodotti dotati di attività biologica espressi in seme non perdono tale attività neppure dopo essere stati conservati a temperatura ambiente per un anno (Fischer *et al.*, 2004; Stoger *et al.*, 2005).

1.1.5 Vantaggi tecnici del seme di riso

Il riso è il cereale più consumato al mondo, rappresentando l'alimento primario di più di metà della popolazione del pianeta. Sono disponibili diverse prove concernenti la possibilità di utilizzare il riso come piattaforma per la produzione di biofarmaci. Ciò dimostra il potenziale di questo cereale come bioreattore industriale per la produzione di peptidi e proteine di interesse farmaceutico (Takaiwa *et al.*, 2008). Il processamento della biomassa ricombinante è la parte più costosa della produzione, contribuendo a formare circa l'80% del costo totale (Hood *et al.*, 2002).

L'espressione seme-specifica di proteine farmaceutiche in riso ha numerosi vantaggi:

- sono stati raggiunti elevati livelli di espressione con spiccata tessuto-specificità;
- il riso ha un'elevata resa di seme per unità di biomassa (5000–6000 kg per ettaro);
- esistono protocolli consolidati di trasformazione stabile, anche con marcatore selezionabile positivo o senza marcatore selezionabile;
- il genoma di una varietà di riso è stato interamente sequenziato e sono disponibili sequenze parziali di numerose altre varietà;
- il riso è una specie autogama e ciò riduce drasticamente il rischio di contaminazione di varietà alimentari;
- infine i processi di coltivazione, raccolta, lavorazione e conservazione si avvalgono di una secolare esperienza nelle regioni in cui questo cereale ricopre da sempre un ruolo primario nella dieta (Takaiwa *et al.*, 2008).

Sulla base di queste considerazioni *Oryza sativa* L. è ritenuta una delle specie più promettenti per applicazioni biotecnologiche. In questo senso numerosi sforzi sono attualmente rivolti ad aumentare il livello di espressione di transgeni in questa pianta.

1.1.6 Meccanismi che influenzano il livello di espressione

Qualunque sia la specie vegetale o il tessuto considerato, i due principali fattori che influenzano l'accumulo di una proteina sono la velocità di sintesi e la velocità di degradazione. La velocità di sintesi è prevalentemente influenzata dalla velocità di trascrizione, dalla stabilità dell'mRNA e dalla velocità di traduzione. Il ritmo di trascrizione è fortemente determinato dall'efficienza del promotore, ma è altrettanto importante il numero di copie di T-DNA e la posizione genomica in cui si integrano i singoli T-DNA (Takaiwa *et al.*, 2008). La velocità di degradazione dipende dalla stabilità intrinseca della proteina e dalla localizzazione dell'accumulo (Egelkrout *et al.*, 2012). In questo senso, l'espressione seme-specifica è un fattore che abbatta drasticamente la velocità di degradazione. Oltre alla tessuto specificità, un ruolo importante per l'accumulo e soprattutto per la stabilità della proteina va attribuito alla localizzazione subcellulare; infatti, il comparto a cui verrà indirizzata la proteina nascente ne influenzerà il ripiegamento, l'eventuale assemblaggio di varie subunità, il livello e il tipo di glicosilazione (Twyman *et al.*, 2003).

Il complesso sistema formato dal reticolo endoplasmico nel seme in maturazione ha un'elevata efficienza nel processare proteine nascenti grazie all'abbondanza di chaperoni molecolari. Svariati peptidi segnale all'N-terminale con lunghezze che vanno dai 18 ai 30 amminoacidi sono in grado di indirizzare la proteina a cui sono legati al lume del reticolo. L'aggiunta di un peptide segnale è perciò necessaria per l'accumulo in seme. L'assenza o la rimozione del peptide segnale conduce a un'espressione non rilevabile anche quando il gene viene trascritto a livelli elevati (Sharma *et al.*, 2000; Yang *et al.*, 2002; Takagi *et al.*, 2005b).

La stabilità dell'mRNA è un altro fattore importante. L'emivita di questa famiglia di molecole oscilla tra alcuni minuti fino a più giorni a seconda della presenza di specifici motivi presenti nella sequenza. La sequenza prossima alla struttura 5' m⁷G e la coda poli-A possono presentarsi in una varietà di forme che stabilizzano in modo diverso il trascritto. È noto poi che l'intera sequenza non tradotta in 5' è fondamentale per il reclutamento dei fattori di traduzione (Liu *et al.*, 2010), oltre che in alcuni

casi di trascrizione (De Amicis *et al.*, 2007). Un altro fattore che potrebbe incrementare significativamente il tasso di traduzione di un trascritto è l'ottimizzazione del dialetto codonico in base all'organismo ospite. In effetti, la frequenza con cui appaiono i vari codoni per ciascun amminoacido varia da organismo a organismo e può essere correlata all'abbondanza relativa dei tRNA per l'amminoacido in questione (Streatfield *et al.*, 2001).

1.2 Promotori endosperma-specifici

In questa sezione verrà introdotto il ruolo del promotore nel regolare un'espressione genica endosperma-specifica. Verrà inoltre fornita una panoramica dei meccanismi molecolari che determinano forza e specificità di questa classe di promotori, che costituiscono l'argomento di studio di questo dottorato.

1.2.1 Promotori di proteine di riserva del seme di riso

Le piante immagazzinano una rilevante quantità delle proprie riserve di azoto, zolfo e carbonio sotto forma di proteine di riserva del seme, che vengono successivamente utilizzate nella fase post-germinativa di sviluppo. Sulla base della loro solubilità, queste proteine di riserva vengono distinte in tre classi: globuline, prolamine e gluteline. Le globuline, caratterizzate dalla loro solubilità in soluzioni saline, servono come principale riserva di nutrienti nell'embrione dei semi di piante mono- e dicotiledoni, mentre le prolamine, solubili in alcool, e le gluteline, relativamente insolubili, svolgono questa funzione solo nell'endosperma delle monocotiledoni. Due delle più diffuse classi di globuline, designate 7S e 11S sulla base delle loro proprietà di sedimentazione, sono presenti in proporzioni differenti fra le dicotiledoni (Okita *et al.*, 1989). A differenza della maggior parte dei cereali che utilizza le prolamine come riserva azotata, le proteine più abbondanti nel seme di riso sono le gluteline. Queste proteine, che possono costituire l'80% delle proteine dell'endosperma, sono sintetizzate e accumulate nella fase intermedia di sviluppo del tessuto (Yamagata *et al.*, 1982). Nel seme di riso sono presenti anche le prolamine, ma questa frazione ammonta solo al 5-10% del totale delle proteine. La sintesi delle gluteline e delle prolamine

non è coordinata. Le gluteline iniziano a essere sintetizzate 4-6 giorni dopo la fioritura, mentre le prolamine cominciano a essere rilevabili solo svariati giorni dopo. La localizzazione di queste due famiglie di proteine avviene in corpi proteici morfologicamente distinti che sono frutto di due differenti processi cellulari (Tanaka *et al.*, 1980; Krishnan *et al.*, 1986). Le gluteline sono codificate da una piccola famiglia multigenica di circa 11 geni per corredo aploide (Okita *et al.*, 1989; Takaiwa *et al.*, 1987; Takaiwa *et al.*, 1990; Takaiwa *et al.*, 1991). Questi geni sono divisi in tre sottofamiglie designate GluA, GluB e GluC, composte rispettivamente da 5, 5 e 1 gene per genoma aploide (Qu *et al.*, 2008). Già alla fine degli anni ottanta erano stati sequenziati e sperimentati *in vitro* promotori di due sottofamiglie di gluteline di riso (Okita *et al.*, 1989). Nei primi anni duemila si è intensificata la ricerca sui promotori che inducono l'espressione di queste proteine, in vista di un loro possibile sfruttamento in ambito biotecnologico. Confrontando promotori di varie proteine di riserva di riso sulla base del livello di espressione del gene *reporter uidA*, codificante l'enzima beta-glucuronidasi di *E. coli* (GUS), è stato dedotto che il promotore più potente è quello della globulina 26 kDa (Wu *et al.*, 1998), risultato inatteso visto che questa proteina di riserva rappresenta solo il 2-8% della proteina del seme. Ciò è stato confermato da un esperimento di espressione transitoria in endosperma di riso in coltura; il promotore della globulina 26 kDa si è dimostrato più attivo rispetto ai promotori di tre diverse gluteline e una prolamina (Hwang *et al.*, 2001a). Inoltre, in un esperimento di trasformazione stabile di piante di riso è stato introdotto il gene che codifica per il lisozima sotto il controllo di una serie di promotori di gluteline e del promotore della globulina. Anche in questo caso il promotore della globulina si è dimostrato più efficace (Hwang *et al.*, 2002). Va rilevato che esperimenti di elettroforesi bidimensionale hanno confermato l'esistenza di un'unica forma di globulina (Pan e Reeck, 1988), mentre analisi *Southern Blotting* condotte sul genoma di riso hanno indicato la presenza di una sola copia del gene codificante questa proteina (Shorrosh *et al.*, 1992); al contrario, esistono una decina di geni codificanti per altrettante forme di glutelina. Dal complesso delle prove svolte si può concludere che non c'è

una significativa correlazione fra abbondanza di una classe di proteine di riserva e forza dei relativi promotori.

1.2.2 Classificazioni dei promotori di proteine di riserva del seme di riso

I promotori seme-specifici possono essere suddivisi in gruppi secondo vari criteri. Non solo il livello di espressione varia drasticamente da un promotore all'altro, ma anche il profilo temporale e spaziale d'espressione differiscono da caso a caso.

Usando come criterio di associazione la localizzazione della proteina espressa è possibile distinguere promotori:

- specifici per l'endosperma interno (globulina 26 kDa, allergeni da 14 kDa e 16 kDa);
- specifici per lo strato subaleuronico (prevalentemente gluteline);
- specifici per lo strato aleuronico periferico (prevalentemente prolamine);
- attivi sia nell'aleurone, sia nell'embrione (oleosine, globulina dell'embrione);
- attivi sia nello scutello, sia nell'endosperma (PPDK, AlaAT) (Qu e Takaiwa, 2004).

Prendendo in considerazione l'attività trascrizionale relativa, i promotori classificati come forti sono:

- il promotore della glutelina B4 (Qu e Takaiwa, 2004);
- il promotore della globulina 26 kDa (Qu e Takaiwa 2004);
- i promotori delle prolamine 10 kDa e 16 kDa (Qu e Takaiwa, 2004);
- il promotore della glutelina C (Qu *et al.*, 2008).

L'espressione specifica osservata nell'endosperma con l'uso di questi promotori di proteine di riserva del seme è essenzialmente regolata

dall'interazione combinata di specifici fattori di trascrizione espressi esclusivamente nella fase precoce di maturazione del seme. Questi fattori di trascrizione riconoscono specifiche sequenze del promotore dette elementi *cis*-regolatori.

1.2.3 Elementi *cis*-regolatori dei promotori di proteine di riserva del seme

Il meccanismo che regola l'espressione delle proteine di riserva del seme è stato indagato individuando elementi comuni a più promotori e studiandone la funzione mediante mutazioni puntiformi, delezioni (*loss-of-function*) o utilizzo di questi elementi in promotori sintetici (*gain-of-function*) (Wu *et al.*, 1998; Washida *et al.*, 1999; Wu *et al.*, 2000; Hwang *et al.*, 2001).

Particolare attenzione è stata dedicata ai promotori delle gluteline A2 e B1. E' stato identificato un promotore minimo per l'espressione in seme, costituito dalle prime 197 bp a monte del sito di inizio trascrizione del promotore della glutelina B1 (Wu *et al.*, 1998). All'interno di questa regione, sono stati individuati 4 motivi responsabili dell'espressione specifica in endosperma: GCN4, AACAA, AAAG e ACGT. Tra questi, il motivo GCN4 è apparso essere l'elemento chiave di regolazione. Infatti, un frammento lungo 21 bp ricavato dal promotore minimo $\Delta[-197;-1]$ contenente GCN4 e ripetuto 3 o più volte è risultato conferire specificità per l'endosperma al promotore minimo CaMV 35S $\Delta[-46;-1]$ (Wu *et al.*, 1998).

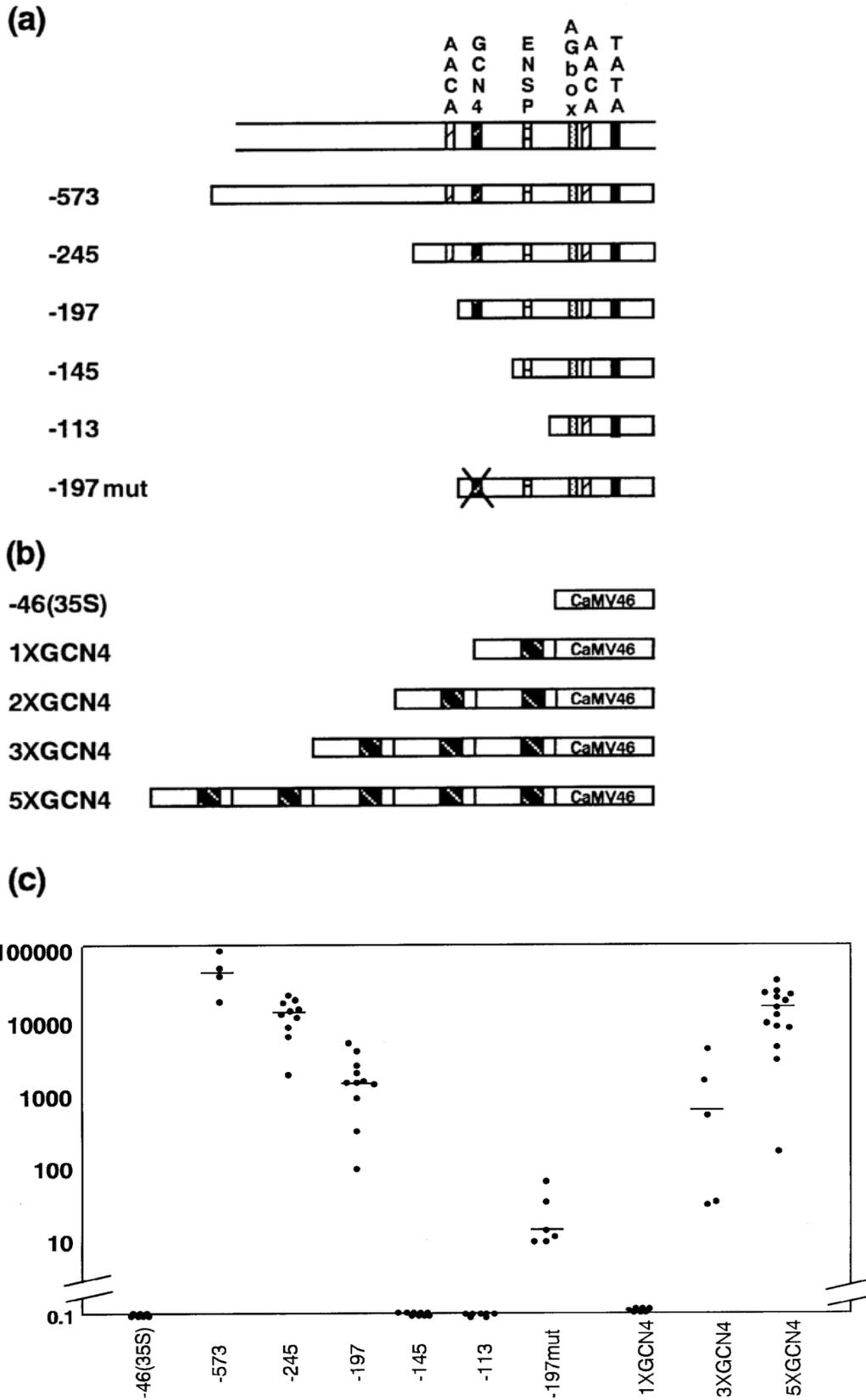


Figura 1. Esperimenti condotti da Wu *et al.* nel 1998. (a) *Loss of function*; (b) *gain of function*; (c) livello di espressione.

Per quanto concerne il motivo ACGT, la sua importanza era già stata precedentemente rilevata in mais; in questa specie, è stato dimostrato che il fattore di trascrizione O2 regola l'espressione della zeina 22 kDa mediante legame di un motivo contenente un *core* ACGT (Schmidt *et al.*, 1992). Questo motivo è peraltro diffuso in numerosi promotori di proteine di riserva del seme. Va comunque osservato che il motivo ACGT è presente anche in molti geni che mediano la risposta della pianta a vari agenti fisici e chimici, quali stimoli luminosi, fito-ormoni, stress biotici e abiotici (Izawa *et al.*, 1993). E' dunque probabile che l'attivazione specifica per l'endosperma sia dovuta al contesto in cui ACGT è inserito.

Il motivo AACA è un elemento regolatore negativo al di fuori del seme, ma nel seme agisce in combinazione con GCN4 per dirigere l'espressione specifica per l'endosperma (Takaiwa *et al.*, 1996, Yoshihara e Takaiwa, 1996). Nel promotore della glutelina A3 è presente un tratto di 45 bp contenente sia AACA, sia GCN4. Questo tratto è stato fuso al modulo base del promotore CaMV 35S $\Delta[-90;-1]$; l'elemento risultante è stato successivamente sottoposto a prove di *loss-of-function* in piante di tabacco. Una mutazione del motivo GCN4 ha ridotto l'attività del promotore, mentre una mutazione di AACA ne ha aumentato l'attività sia nell'embrione, sia nell'endosperma (Yoshihara *et al.*, 1996a). Al contrario, esperimenti simili svolti in riso con il promotore minimo $\Delta[-197;-1]$ della glutelina B1 hanno indicato che GCN4 agisce come elemento essenziale per la specificità, mentre AACA, ACGT e *prolamin-box* hanno un ruolo nella regolazione quantitativa dell'espressione; in particolare, la perdita di AACA porta a una drastica depressione dell'attività trascrizionale anche in endosperma (Wu *et al.*, 2000).

D'altro canto, ulteriori esperimenti svolti in tabacco con il promotore minimo $\Delta[-90;-1]$ hanno evidenziato che la presenza di GCN4 è condizione necessaria ma non sufficiente ad assicurare un'espressione specifica per l'endosperma, per la quale è richiesta una combinazione dei tre motivi GCN4, AACA e ACGT, tratti ad es. da GluB1 (Washida *et al.*, 1999). Sebbene gli esperimenti di trasformazione siano più facilmente realizzabili in tabacco, la scelta di questa specie nello studio di promotori endosperma-

specifici appare di dubbia validità in quanto i risultati ottenuti in piante monocotiledoni che non riassorbono l'endosperma nel corso dello sviluppo e maturazione del seme non sono evidentemente sovrapponibili.

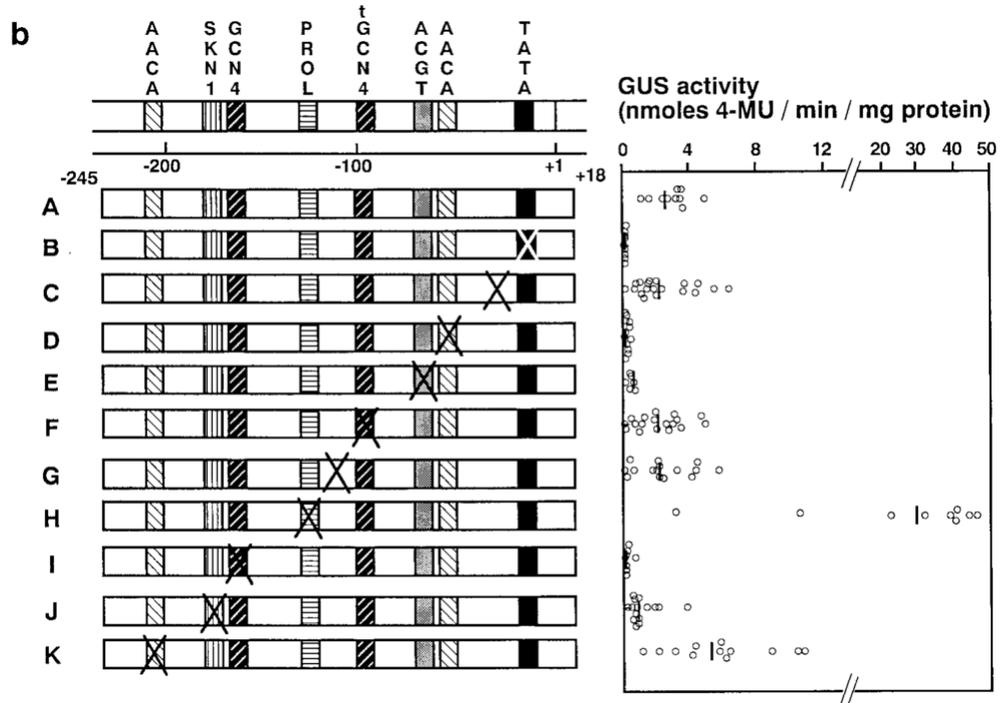


Figura 2. Esperimenti di *loss of function* condotti da Washida *et al.* nel 1999. Promotore minimo GluB1 in tabacco.

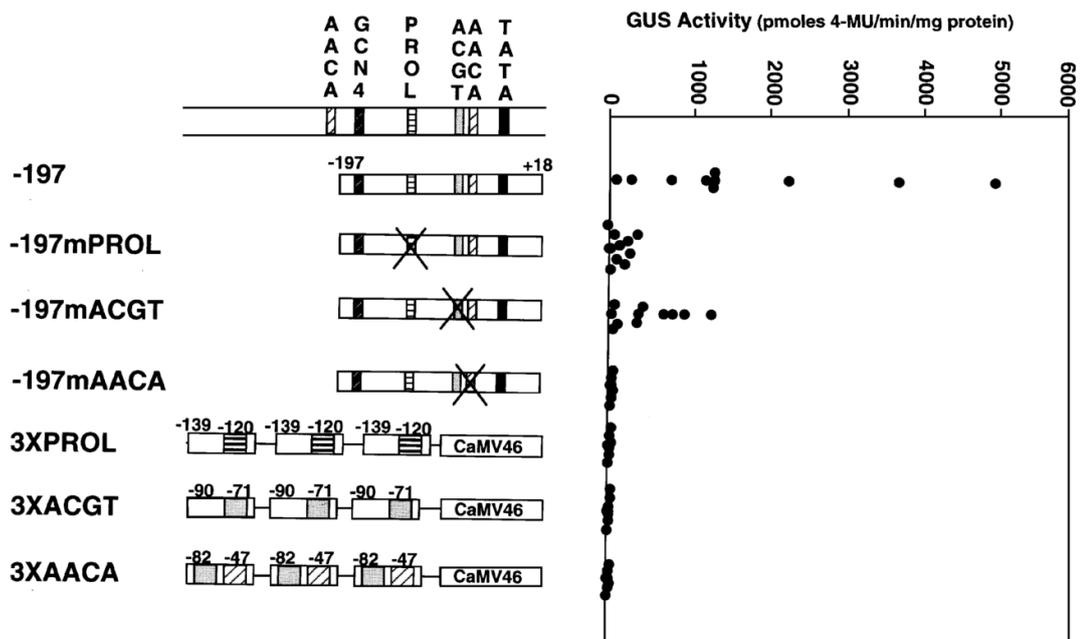


Figura 3. Esperimenti condotti da Wu *et al.* nel 2000 su GluB1 in riso.

Come precedentemente ricordato, lo studio riguardante gli elementi *cis*-regolatori di promotori endosperma-specifici è stato effettuato principalmente sui promotori relativi a glutelina A2 e B1. Va detto che altri promotori (es. promotore della glutelina C o quello della globulina 26 kDa), pur essendo in grado di indurre un'elevata e specifica espressione nell'endosperma di riso, non appaiono contenere elementi *cis*-regolatori in forma canonica (Hwang *et al.*, 2001; Qu *et al.*, 2008). A questo proposito, solo l'identificazione e la caratterizzazione dei fattori *trans*-agenti ha chiarito in modo univoco la natura e la funzione di altri elementi *cis*-regolatori dei promotori per le proteine di riserva del seme di riso.

1.2.4 Approcci sperimentali nello studio dei fattori di trascrizione

Il meccanismo di attivazione dei promotori seme-specifici è stato studiato anche caratterizzando i fattori di trascrizione presenti durante le fasi di sviluppo del seme, in quanto elementi verosimilmente implicati nell'espressione delle proteine di riserva. In questi lavori, si è proceduto al vaglio di librerie a cDNA ottenute da seme di riso e alla successiva clonazione in vettore di espressione delle sequenze candidate.

Per associare le proteine nucleari espresse durante le fasi di sviluppo del seme agli elementi *cis*-regolatori presenti lungo la sequenza dei promotori delle proteine di riserva, sono state utilizzate specifiche tecniche di biologia molecolare e biochimica.

Qui di seguito vengono brevemente elencate le metodiche applicate nello studio dei siti di riconoscimento dei fattori di trascrizione:

- *gel-retardation assay*;
- EMSA (*Electrophoretic Mobility Shift Assay*);
- *methylation interference*;
- *DNase footprinting* (Suzuki *et al.*, 1998; Onodera *et al.*, 2001; Yamamoto *et al.*, 2006).

Per associare determinati elementi *cis*-regolatori o regioni di un promotore a specifici fattori di trascrizione, sono state inoltre utilizzate tecniche di co-trasformazione mirate alla sovraespressione di un fattore di

trascrizione in presenza di un gene *reporter* posto sotto il controllo di un promotore putativamente associato al fattore in esame (Suzuki *et al.*, 1998; Yang *et al.*, 2001; Onodera *et al.*, 2001; Huang *et al.*, 2002; Yamamoto *et al.*, 2006). In taluni casi, è stato possibile individuare il promotore del fattore di trascrizione mediante sequenziamento della regione di DNA fiancheggiante la CDS. Ponendo un gene *reporter* sotto il controllo di tale promotore, è stato studiato nel dettaglio il profilo di espressione del fattore di trascrizione in esame (Onodera *et al.*, 2001; Yamamoto *et al.*, 2006).

1.2.5 Caratterizzazione di alcuni fattori di trascrizione

Le tecniche descritte nel paragrafo precedente hanno permesso di indagare l'azione di 4 fattori di trascrizione specifici per l'endosperma di riso:

1. REB (*Rice Endosperm B-zipper*) o RISBZ2; fa parte di una classe di proteine (*b-zipper*) che legano il DNA mediante un dominio basico che presenta una serie di leucine ripetute. REB lega una serie di motivi ACGT nella porzione distale del promotore della globulina 26 kDa, ma non dimostra affinità verso nessun altro motivo ACGT presente lungo i promotori della glutelina A2 e del gene PG5 della prolamina 13 kDa (Nakase *et al.*, 1997). In uno studio che ha preso in considerazione tutte le proteine *b-zipper* espresse durante lo sviluppo del seme di riso, è stato dimostrato che REB, come altri membri della classe *b-zipper*, lega il motivo GCN4, ma non è in grado di attivare l'espressione nell'endosperma solamente attraverso questo legame (Onodera *et al.*, 2000). Ad oggi l'unico promotore di cui è stata dimostrata l'attivazione da parte di REB è il promotore della globulina 26 kDa, Glb. In piante di riso transgeniche in cui REB risulta sovraespresso, un gene eterologo posto sotto il controllo di Glb risulta parimenti sovraespresso (Yang *et al.*, 2001).
2. OSMYB5; fa parte di un gruppo di proteine dette MYB, i cui geni codificanti sono stati isolati da una libreria a cDNA di seme immaturo di riso. Poiché questo fattore presenta una significativa analogia con proteine di altre specie che riconoscono il motivo AACAA, in uno studio

condotto in riso, OSMYB5 è stato scelto come candidato al riconoscimento specifico di tale elemento *cis*-regolatorio. Nell'ambito di questo studio, è stato effettivamente dimostrato che OSMYB5 lega in modo specifico 2 motivi AACAA presenti nella parte prossimale del promotore GluB1. Il sito di legame è apparso consistere in un *consensus* lungo 12 bp circa, incorporante il motivo AACAAA. I risultati di un saggio di *gel-retardation* hanno tuttavia suggerito che solo una parte della sequenza lunga 12 bp sia necessaria al riconoscimento, poiché il resto del motivo può variare senza che l'affinità di legame decada. Questo fattore di trascrizione è espresso in numerosi tessuti e sembra avere un'azione inibitoria della trascrizione in tutti i tessuti diversi dal seme. Un ulteriore elemento degno di nota è che OSMYB5 non è di per sé in grado di attivare un promotore sintetico contenente AACAA. In effetti, un promotore CaMV 35S $\Delta[-46;-1]$ con una porzione contenente AACAA e GCN4 non ha risposto all'azione effettrice di OSMYB5 (Suzuki *et al.*, 1998).

3. RISBZ-1 (*Rice Seed B-zipper*); questo fattore è stato caratterizzato insieme ad altri 4 RISBZ in uno studio approfondito effettuato sulle proteine di questa classe espresse durante lo sviluppo del seme. In tale ambito, 5 geni *b-zipper* presenti in una libreria di cDNA di seme immaturo sono stati clonati e sovraespressi in riso al fine di determinare la specificità di riconoscimento e l'attivazione esercitata sul motivo GCN4. Sebbene tutti i fattori in prova abbiano mostrato di legare GCN4, solamente RISBZ-1 è risultato in grado di attivare la trascrizione del gene bersaglio. La singolarità della sua azione è stata attribuita alla differente sequenza amminoacidica al terminale amminico. Va rilevato che RISBZ-1 può formare eterodimeri con altri membri della sua classe (Onodera *et al.*, 2000). In un successivo lavoro è stato dimostrato che RISBZ-1 lega non solo GCN4 ma anche talune varianti con sequenza fortemente divergente. Questa flessibilità nel riconoscimento ha portato a riconsiderare quanto noto su questo elemento regolatore e spiega inoltre perché promotori privi di un motivo considerato indispensabile per l'espressione in

endosperma, come GCN4, siano in realtà estremamente attivi in questo tessuto (Yamamoto *et al.*, 2006).

4. RPBF (*Rice Prolamin-box Binding Factor*); appartiene alla classe di proteine *Dof* (*DNA Binding with one finger*), un gruppo di fattori di trascrizione che condividono un dominio digitiforme unico, che complessa un singolo ione Zn^{2+} . Tale dominio è altamente conservato e proprio grazie ad esso è stato possibile individuare una proteina *Dof* espressa in seme di riso. In coltura di protoplasti, la presenza di RPBF ricombinante ha indotto l'espressione di un gene *reporter* posto sotto il controllo di 9 diversi promotori di proteine di riserva del seme di riso. Questo fattore di espressione è risultato legare i motivi AAAG delle *prolamin-box* putative di tre promotori, nonché riconoscere ulteriori siti di legame precedentemente non individuati dall'analisi comparativa delle sequenze di tali promotori. RPBF sembra intervenire nella regolazione quantitativa dell'espressione ed è in grado di riconoscere in modo funzionale la propria sequenza bersaglio anche sul filamento opposto. Anche per RPBF, i nucleotidi che costituiscono la sequenza minima di riconoscimento sono in numero inferiore a quello ritenuto necessario sulla base dell'analisi comparativa tra promotori. RPBF e RISBZ-1 hanno un forte effetto sinergico, e l'effetto sembra essere inversamente proporzionale alla distanza fra i rispettivi motivi riconosciuti (Yamamoto *et al.*, 2006).

Da quanto descritto emerge che lo studio dei fattori di trascrizione è assolutamente necessario per individuare natura e ruolo dei motivi *cis*-regolatori dei promotori endosperma-specifici e che il meccanismo che regola l'espressione delle proteine di riserva del seme, lungi dall'essere chiarito, necessita di ulteriori, numerosi approfondimenti.

1.3 Promotori sintetici

In questa sezione verranno presentate le potenzialità di una tecnologia basata sul *design ad hoc* di promotori sintetici. Verranno presentati alcuni lavori che introducono in linea teorica il concetto di promotore sintetico come combinazione di operatori logici costituiti da

motivi *cis*-regolatori. Verrà inoltre fornita una panoramica degli studi eseguiti fino ad ora su questo argomento per ottimizzare l'espressione di proteine eterologhe in pianta.

1.3.1 Strumenti per l'analisi di sequenza e manipolazione dei promotori

I miglioramenti tecnologici, conseguiti principalmente nell'area della genomica funzionale e della bioinformatica, hanno portato a una miriade di nuovi dati biologici e a una migliore conoscenza del metabolismo vegetale. Questo progresso senza precedenti pone le basi per lo sviluppo di una piattaforma volta alla manipolazione mirata dell'attività trascrizionale mediante ingegnerizzazione di promotori sintetici. I mezzi da acquisire e la strategia da seguire per controllare l'espressione di una proteina eterologa consistono nella capacità di progettare accuratamente l'uso degli interruttori molecolari e dei moduli che regolano la trascrizione confinata nello spazio e nel tempo di un transgene. L'uso di promotori sintetici può portare beneficio non soltanto alle future applicazioni biotecnologiche, ma anche alla comprensione del meccanismo di azione degli elementi *cis*- e *trans*-regolatori in pianta (Venter, 2007).

Alcuni studi, condotti in vari organismi, hanno tentato di fornire nuovi strumenti per predire la funzione dei geni sulla base della presenza e della combinazione di particolari motivi presenti nelle regioni fiancheggianti in 5' la CDS.

L'emergere di banche dati che catalogano una moltitudine di siti di legame per i fattori di trascrizione ha facilitato l'analisi su larga scala di promotori vegetali. Le tre principali banche dati di questo tipo sono PLACE, PlantCARE e TRANSFACT (Matys *et al.*, 2003). Quest'ultima viene costantemente aggiornata e fornisce una serie di possibili motivi *cis*-regolatori.

Per prevedere la funzione di un gene, o anche per individuare i motivi da combinare al fine ottenere un determinato profilo di espressione, è essenziale distinguere le sequenze riconosciute dai fattori di trascrizione, in genere sovra-rappresentate, dal rumore di fondo. Diversi metodi

computazionali sono stati sviluppati per questo scopo (Van Helden, 2003; Steffens *et al.*, 2004; Holt *et al.*, 2006). Una pubblicazione di Tompa *et al.* (2005) ha passato in rassegna i più importanti metodi e strumenti statistici di rilevamento e di classificazione dei motivi *cis*-regolatori. L'identificazione del codice di regolazione rimane però complessa, dato che all'interno di un promotore costituito da 500 a 5000 bp, le sequenze consenso riconosciute dai fattori di trascrizione hanno una lunghezza intorno a 8 bp. Una delle prime strategie seguite per risolvere questo problema prevedeva di individuare con opportuni algoritmi, i motivi sovra-rappresentati. Ne sono un esempio il campionamento di Gibbs (Lawrence *et al.*, 1993) e l'algoritmo *Expectation-Maximization* (EM), un algoritmo iterativo ampiamente utilizzato per calcolare le stime di massima verosimiglianza nel caso di dati incompleti (Bailey e Elkan, 1995). Questi metodi si basano sul presupposto che una sequenza consenso per un fattore di trascrizione abbia maggiore probabilità di essere sovra-rappresentata rispetto a qualunque altra sequenza casuale.

Ora sono disponibili numerosi altri metodi che si basano su differenti principi; alcuni sono versioni migliorate che, rispetto ai precedenti metodi, combinano algoritmi basati su modelli statistici, ad es. il modello di Markov nascosto. A questo proposito, è interessante citare un lavoro che affronta il problema del riconoscimento dei motivi *cis*-regolatori da un punto di vista steganografico (Wang *et al.*, 2006). La steganografia è una tecnica utilizzata per nascondere l'esistenza di un'informazione incorporando il messaggio che deve essere protetto in un testo di copertura in modo da ottenere il cosiddetto "*stegoscript*". La steganalisi decifra lo *stegoscript* rivelando il messaggio nascosto. Sono stati sviluppati numerosi algoritmi in grado di condurre autonomamente e in maniera altamente processiva un processo di steganalisi. Le regioni regolatrici del genoma contengono informazioni sul controllo trascrizionale, servendosi di elementi regolatori incorporati in una sequenza che fa da sfondo. Partendo da questo presupposto, gli autori hanno sviluppato un algoritmo che tratta la sequenza del genoma come uno *stegoscript* che nasconde un messaggio segreto (gli elementi *cis*-regolatori) in un testo di copertura (il resto della sequenza). Il messaggio segreto è tipicamente più conservato e statisticamente sovra-rappresentato. Ciò è

particolarmente vero per le sequenze regolatrici, dove un piccolo numero di fattori di trascrizione controlla un ampio numero di geni. Questo algoritmo cerca i motivi come se si trattasse di parole, basandosi su un vero e proprio dizionario che viene perfezionato nel corso dell'analisi. Con questo metodo sono state analizzate le regioni che regolano il ciclo cellulare di *S. cerevisiae* e di *Arabidopsis thaliana* ottenendo risultati superiori a quelli forniti da tutti gli altri metodi di individuazione.

Il limite in cui incorre questo algoritmo è tuttavia l'incapacità di una adeguata valutazione del contesto. La probabilità che una "parola" sia effettivamente un motivo regolatore viene determinata a prescindere dall'associazione con altri motivi. In altre parole, questo algoritmo non è in grado di prendere in considerazione una grammatica complessa. Questo difetto non va sottovalutato poiché l'individuazione dei motivi *cis*-regolatori è solo un primo passo verso la comprensione dell'intero sistema di attivazione trascrizionale. Per predire il profilo di espressione indotto da un determinato promotore, è in effetti necessario comprendere i principi della regolazione combinatoria. E' in altre parole necessario identificare la rete di regolazione mediante un'analisi combinatoria degli elementi del promotore (Pilpel *et al.*, 2001). Nel corso di questa analisi, devono essere identificati i motivi generici che concorrono all'attivazione di un gene in sinergia con i motivi *cis*-regolatori specificatamente indotti in particolari condizioni. La possibilità di analizzare contemporaneamente un gran numero di promotori permette, attraverso algoritmi statistici, di affinare l'individuazione dei motivi caratteristici di ciascuna famiglia di promotori. L'ottimizzazione dell'architettura del promotore su base probabilistica è seguita dalla costruzione di una mappa che descrive la sinergia dei motivi *cis*-regolatori mediante gli operatori logici "AND-NOT-OR" ecc.

1.3.2 Approccio sintetico combinatorio

Un approccio sintetico combinatorio è già stato usato in biologia dei sistemi per studiare la mutua regolazione di complessi multigenici e prevederne la risposta a stimoli provenienti dall'esterno (Guet *et al.*, 2002). In questo caso, i promotori costituiscono i nodi di una rete costituita da un sistema multigenico. Purtroppo, i modelli ottenuti con questo approccio non

sempre riescono a prevedere il fenotipo del sistema preso in esame. Le cose cambiano se si procede in modo analogo a livello di singolo promotore. Un singolo promotore è a sua volta un sistema, i cui nodi sono costituiti dai diversi motivi regolatori che competono a determinarne la risposta globale. Questo sistema diviene nodo di un sistema multigenico, gerarchicamente superiore, alla stessa maniera in cui un circuito integrato si inserisce nel circuito stampato di un computer. Seguendo questo approccio è stata studiata la regolazione trascrizionale combinatoria in singoli promotori batterici (Kinkhabwala e Guet, 2008). Il promotore *s70* di *E. coli* è stato usato come sistema modello per esplorare la capacità di un singolo promotore di integrare *input* regolatori multipli. Attraverso la ricombinazione di una serie di motivi *cis*-regolatori, è stata costruita una libreria di promotori sintetici. Il risultato è una notevole correlazione tra la geometria dei promotori e la loro funzione, che ha permesso di spiegare il fenotipo ottenuto con promotori sintetici complessi mediante regole basilari di attivazione o repressione. Queste regole possono essere rappresentate mediante l'algebra booleana.

Anche se questo risultato è stato ottenuto in un sistema batterico estremamente semplificato, comprendente solo 4 fattori di regolazione trascrizionale, non di meno questo esperimento apre la strada a una migliore comprensione dei meccanismi di regolazione dell'espressione genica negli organismi superiori.

Lo studio del ruolo e dell'interazione dei motivi regolatori è sempre maggiormente caratterizzato dall'ampio uso di strumenti bioinformatici. La creazione di sofisticati modelli *in silico* ha permesso una più accurata previsione dell'espressione genica (Fessele *et al.*, 2002; Werner *et al.*, 2003).

Uno studio di fondamentale importanza sull'organismo modello *Saccharomyces cerevisiae* ha rivelato come una logica basata sull'osservazione di motivi regolatori conservati garantisca una predizione relativamente accurata (73%) di un particolare profilo di espressione in risposta a uno specifico stimolo (Beer e Tavazoie, 2004). I risultati di questo studio suggeriscono dunque che, in una applicazione biotecnologica,

sarebbe possibile progettare un modello di promotore sintetico dotato di un particolare profilo di espressione, qualora fossero disponibili sufficienti informazioni sui singoli motivi regolatori. In linea teorica la strategia è semplice. Tuttavia la complessità e l'interazione sinergica di una moltitudine di meccanismi regolatori (per quanto singolarmente ben caratterizzati) costituisce un ostacolo ancora insormontabile. Data questa complessità è evidente che l'integrazione di una migliore comprensione dei meccanismi regolatori a livello dell'intero genoma di organismi multicellulari complessi, insieme allo sviluppo della biologia dei sistemi, costituiranno le basi del futuro studio dei promotori naturali e della progettazione di analoghi sintetici (Andrianantoandro *et al.*, 2006; Barrett *et al.*, 2006)

1.3.3 Verifica *in vivo* dell'attività dei promotori sintetici

La previsione *in silico* degli effetti della regolazione genetica alla fine deve comunque essere validata sperimentalmente. La realizzazione di promotori sintetici riveste un ruolo fondamentale nel confermare *in vivo* la logica che scaturisce dall'interazione dei motivi regolatori (Guido *et al.*, 2006; Vilar, 2006).

La maggior parte dei promotori possiede un nucleo fondamentale necessario a iniziare la trascrizione. A questo nucleo sono associati elementi regolatori che ne modulano l'attività attraverso il legame di fattori di trascrizione fase-, tessuto- o risposta-specifici. Lo sviluppo di tecnologie altamente processive per l'identificazione a tutto campo di siti di legame per fattori di trascrizione e l'archiviazione di questi ultimi in banche dati facilmente consultabili, si sono rivelati strumenti insostituibili per la caratterizzazione dei promotori naturali di pianta (Rombauts *et al.*, 2003).

Per quanto riguarda specificatamente l'ambito vegetale, l'accesso ad un elevato numero di dati genetici risultanti dal sequenziamento dell'intero genoma di due organismi modello, *Arabidopsis thaliana* e *Oryza sativa*, ha permesso la rapida identificazione di un ampio gruppo di promotori. Negli ultimi anni, si è registrato un significativo progresso nella comprensione dell'architettura dei promotori in pianta e dei meccanismi messi in atto per il reclutamento dei fattori di trascrizione generici, cioè di quelli necessari alla

trascrizione di quasi tutti i geni, a prescindere dalla loro funzione (Guilfoyle, 1997; Schwechheimer e Bevan, 1998; Singh, 1998; Liu *et al.*, 1999). Come risultato, molti promotori e fattori di trascrizione sono stati scelti per studi di espressione mirata di proteine eterologhe. I promotori utilizzati in questi studi comprendono:

- sequenze naturali;
- sequenze tronche, cioè ridotte ai soli elementi necessari per ottenere il desiderato profilo d'espressione;
- sequenze sintetiche, cioè frutto di ricombinazione di elementi regolatori di varia provenienza.

I promotori ottenuti si possono classificare in quattro categorie: costitutivi, inducibili, tessuto-specifici e stadio-dipendenti.

In un approfondito studio, Rushton *et al.* (2002), hanno combinato in promotori sintetici vari motivi regolatori, inducibili in risposta a patogeni, per verificare *in vivo* l'effetto dell'associazione. Questo studio ha chiaramente indicato che specificità e forza del promotore dipendono dalla distanza dei motivi regolatori dalla TATA-box. Inoltre, la forza del promotore non è apparsa aumentare linearmente con l'aumentare del numero di ripetizioni di un motivo *cis*-attivatore, anzi, in diverse occasioni una singola copia è risultata sufficiente per l'induzione della massima risposta a un patogeno.

1.3.4 Studi basati sul promotore della subunità 35S del virus del mosaico del cavolfiore

La regione del nucleo fondamentale del promotore, conosciuta anche come regione minima, contiene normalmente la TATA-box per l'aggancio della RNA polimerasi II e una combinazione di elementi atti a organizzare la formazione ordinata del complesso di preavvio della trascrizione a fronte di un riconoscimento da parte di fattori di trascrizione generici (Venter, 2007). Il nucleo fondamentale del promotore 35S di CaMV (promotore largamente utilizzato in passato per la produzione di piante GM) è stato ingegnerizzato in combinazione con elementi *cis*-regolatori mutuati da altri promotori, per ottenere una espressione inducibile da vari stimoli (Pietrzak *et al.*, 1989;

Comai *et al.*, 1990; Last *et al.*, 1991; Ni *et al.*, 1995; Mitsuhashi *et al.*, 1996). Non deve pertanto sorprendere se tale modulo sia stato e sia tuttora una sequenza di riferimento nella costruzione di promotori sintetici con profilo di attività trascrizionale specifico.

1.3.5 Utilizzo di elementi *cis*-regolatori per l'espressione specifica nell'endosperma

La letteratura non riporta esempi di *design ex novo* di promotori endosperma-specifici. Sono invece ben descritte una serie di delezioni, modifiche e combinazioni eseguite su sequenze naturali allo scopo di indagare la funzione di vari elementi putativamente coinvolti nella regolazione dell'espressione genica. Come già accennato al paragrafo 1.2.3, i motivi regolatori per l'espressione specifica nell'endosperma sono stati studiati manipolando proprio la sequenza di promotori naturali di seme. In questo modo, è stato possibile identificare la sequenza minima necessaria a garantire una espressione specifica per l'endosperma del gene codificante la glutelina B1 (Wu *et al.*, 1998). Manipolando questa forma minima sono stati costruiti vari promotori sintetici allo scopo di valutare il ruolo degli elementi in essa presenti. I moduli così ottenuti sono stati usati per indurre l'espressione di GUS sia in tabacco, sia in riso.

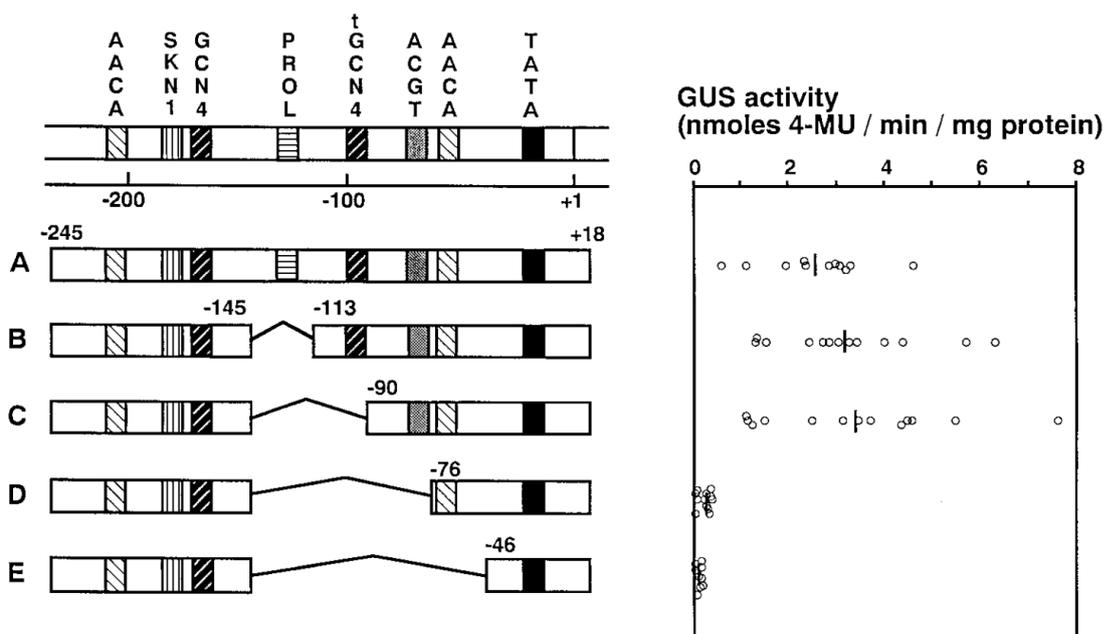


Figura 4. Promotori sintetici costruiti su base GluB1 e relativi livelli di espressione.

In base ai risultati si può dedurre che sussiste un'interazione complessa e combinatoria fra i vari elementi. L'elemento GCN4 è sufficiente a produrre un'espressione endosperma-specifica, ma la regolazione quantitativa è strettamente correlata alla presenza di altri tre elementi, ovvero i motivi ACGT, AACA e *Prol-box* (Washida *et al.*, 1999; Wu *et al.*, 2000).

Sempre considerando il promotore per la glutelina B1 sono stati prodotti 5 omologhi sintetici tronchi e 4 omologhi deleti. In particolare, utilizzando il nucleo fondamentale del promotore 35S di CaMV sono stati costruiti ibridi utilizzando elementi *cis*-regolatori tratti dal promotore minimo di GluB1 ovvero $\Delta[-197;-1]$.

I promotori chimerici così ottenuti sono stati utilizzati per indurre l'espressione transitoria di GUS in endosperma immaturo mediante tecniche biolistiche. Questo metodo consente di verificare in tempi relativamente rapidi la funzionalità di un costrutto ma non fornisce informazioni sulla specificità di tessuto e sulla fase-dipendenza (Hwang *et al.*, 2001b).

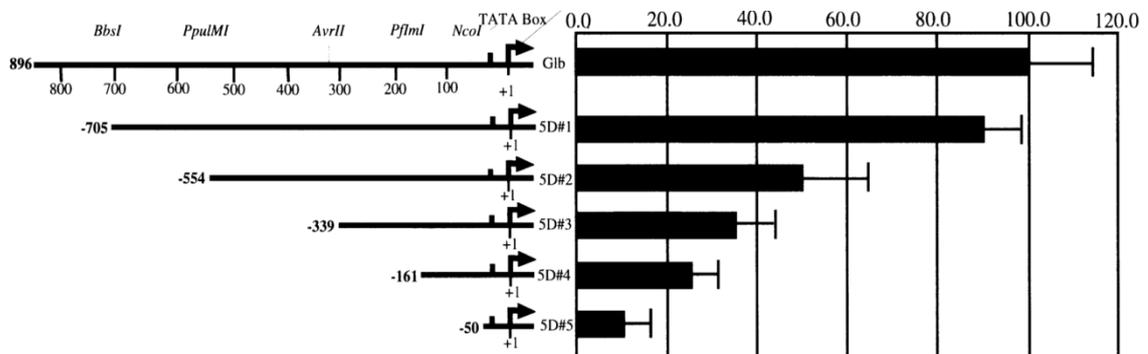


Figura 5. Promotori sintetici tronchi costruiti su base del promotore della globulina.

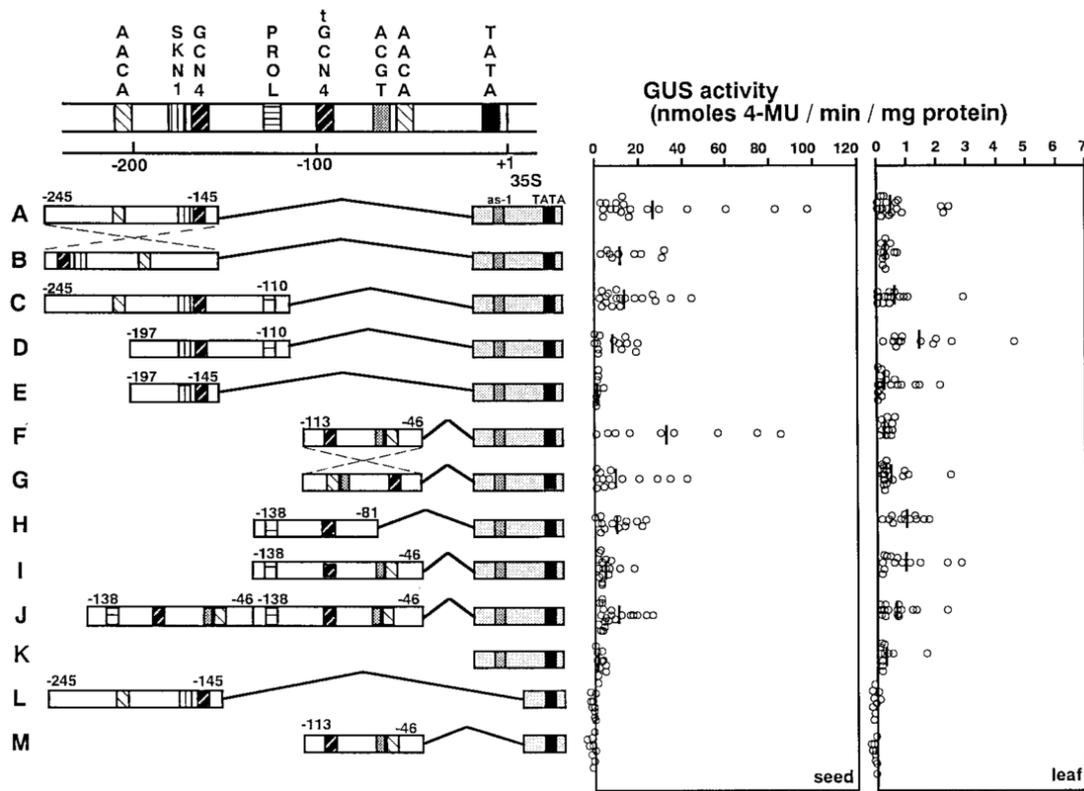


Figura 6. Promotori sintetici ottenuti dalla ricombinazione di porzioni appartenenti al promotore minimo GluB1. A destra sono riportati i livelli di espressione specifici (seed) e aspecifici (leaf).

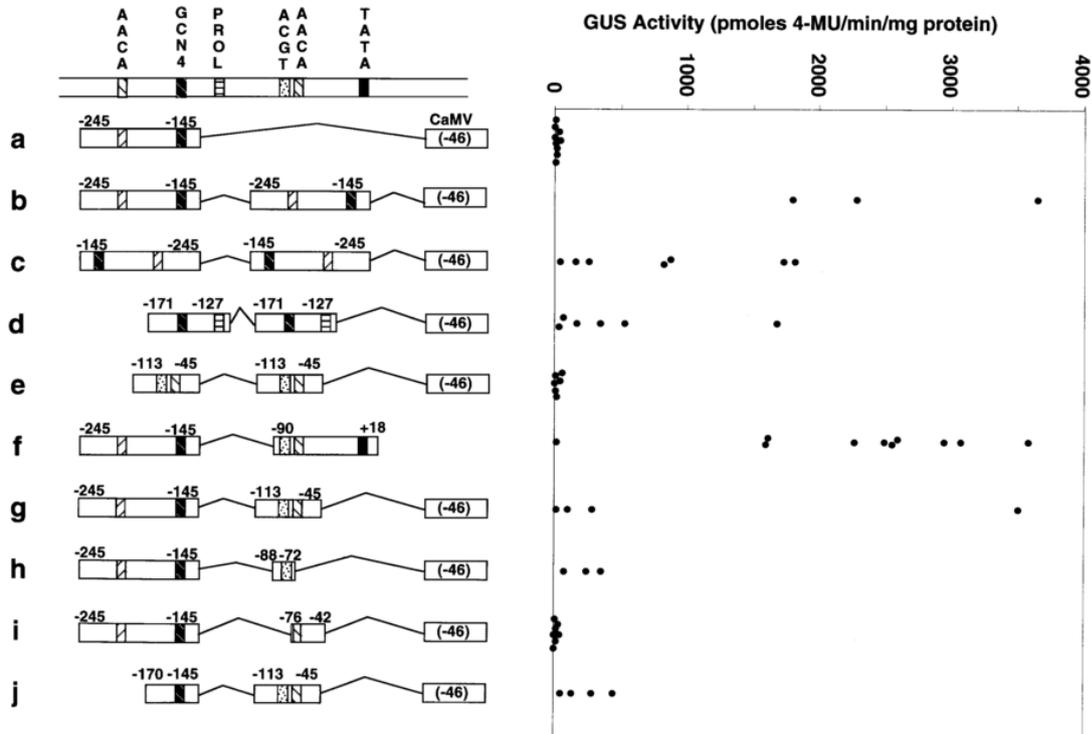


Figura 7. Promotori sintetici costruiti su base 35S CaMVΔ[-46;-1] per ricombinazione di motivi regolatori per l'espressione in endosperma.

A seconda della scelta e della disposizione dei motivi di GluB1, sono stati ottenuti livelli variabili di espressione e tessuto-specificità. Decine di promotori sintetici derivanti dall'unione di CaMV 35S $\Delta[-46;-1]$ con segmenti di GluB1 sono stati utilizzati per produrre piante transgeniche di riso o tabacco allo scopo di individuare il meccanismo di regolazione e di interazione dei motivi *cis*-regolatori (Wu *et al.*, 1998; Washida *et al.*, 1999; Wu *et al.*, 2000). Nel corso di tali studi, sono state spesso impiegate sequenze ripetute di uno specifico elemento *cis*-regolatore in modo da verificare l'effettiva attivazione da parte di particolari fattori di trascrizione. Per lo studio di OSMYB5, è stato utilizzato ad es. un dimero della porzione $\Delta[-245;-145]$ del promotore GluB1 (Suzuki *et al.*, 1998). Per valutare il grado di sinergia fra due fattori è stato ad es. usato un promotore costituito dal modulo minimo $\Delta[-46;-1]$ di CaMV 35S fuso a una porzione di 27 bp del promotore della prolamina 13 kDa NRP33, contenente la prolamin-box (*Prol-box*) e il motivo GCN4 a una distanza di 7 bp (Yamamoto *et al.*, 2006). In genere, alla regione minima del promotore sono stati addizionati a monte segmenti lunghi da 20 a 100 bp contenenti elementi *cis*-regolatori noti, insieme al loro naturale contesto di sequenza, cioè assieme alle sequenze fiancheggianti rinvenute nei promotori naturali interi; come già accennato, per GluA2 e GluB1 l'insieme degli studi effettuati ha consentito di produrre un'ampia e affidabile base di dati.

2 SCOPO DEL DOTTORATO DI RICERCA

La comprensione dei meccanismi che regolano l'espressione specifica in seme è molto cresciuta negli ultimi anni. Nonostante ciò, la mole di informazioni disponibili non è ancora in grado di produrre un salto di qualità nella produzione di proteine ricombinanti in seme. Pertanto, la piattaforma tecnologica di produzione basata su organismi vegetali soffre ancora di limitazioni dovute a un'insufficiente resa della biomassa. Ciò che l'industria farmaceutica richiede per abbandonare i sistemi attualmente in uso in favore della produzione in pianta è un aumento della resa di proteina ricombinante per unità di biomassa da processare. Solo in questo modo si possono abbattere i costi di estrazione e purificazione che, come è noto, costituiscono la maggior parte dell'onere economico nel caso dei bioreattori verdi (la produzione della biomassa ha infatti un costo assai contenuto). Molte proteine di interesse industriale e farmaceutico sono state prodotte in seme di riso (Evangelista *et al.*, 1998) e queste potenzialità di sintesi accrescono ulteriormente l'esigenza di affrontare e risolvere il problema della resa specifica.

Alla luce di queste considerazioni, obiettivo di questo lavoro è stato quello di implementare le prestazioni del sistema di espressione seme-specifico di riso attraverso interventi di ingegneria genetica sul promotore. A tale proposito, si è cercato di razionalizzare, in un quadro di insieme, il gran numero di informazioni ricavate dallo studio di promotori endosperma-specifici condotto da diversi autori. Sulla base di questo quadro d'insieme, si è scelto di affrontare un ambito di indagine ancora relativamente inesplorato, ma molto promettente: l'ingegnerizzazione di promotori seme-specifici sintetici in grado di coniugare livelli di attività trascrizionale maggiori di quelli rinvenibili nei promotori naturali con il mantenimento di un elevato grado di specificità. È stato pertanto sviluppato un metodo di lavoro iterativo suddiviso in cicli sperimentali; il metodo dovrebbe permettere di ottenere a ogni ciclo nuove informazioni, necessarie a proseguire con il minor dispendio di tempo ed energia al ciclo successivo. In altre parole, a ogni ciclo dovrebbe essere prodotta una serie di promotori sintetici, ma soprattutto una serie di dati necessari alla progettazioni di

nuovi promotori. In questa tesi viene descritta l'impostazione del metodo iterativo e la produzione di promotori efficaci come prova della validità degli assunti. Il progetto complessivo ovviamente non si è concluso con il lavoro argomento di questa tesi, ma prosegue con altri cicli di selezione, con il duplice obiettivo di affinare il metodo e ottenere sequenze sempre più efficienti e specifiche.

Per quanto riguarda le ricadute applicative di questo dottorato è bene tenere presente che esso si inserisce in un contesto di indagine più ampio. Il gruppo di ricerca presso cui è stato svolto il presente lavoro è composto da ricercatori accademici e industriali che indagano ormai da un decennio sulle possibilità offerte dagli organismi vegetali nella produzione di biofarmaci ricombinanti. Tale gruppo sta compiendo un continuo sforzo nella direzione di una migliore comprensione dei meccanismi che regolano l'espressione in un organismo vegetale complesso. Lo scopo finale è quello di integrare nei nuovi bioreattori i progressi ottenuti relativamente a ogni singolo componente che influenza la resa, mantenendo il sistema di espressione a un livello di avanguardia. Ciò significa che anche un progresso di limitata entità può efficacemente inserirsi in un quadro complessivo di miglioramento della piattaforma tecnologica. L'aumento del livello di espressione in rapporto alla biomassa può avere un impatto decisivo sulla competitività di una piattaforma produttiva, poiché oltre alla riduzione delle superfici utilizzate per la coltivazione di piante GM e dei volumi di biomassa da avviare al processamento, si otterrebbero sicuramente incrementi nella resa di estrazione e purificazione.

3 APPROCCIO METODOLOGICO

In questa sezione, che funge da raccordo fra introduzione e scopo del lavoro da un lato e la descrizione dei materiali e metodi dall'altro, vengono descritte le strategie e i processi adottati per la produzione dei dati. La sezione relativa ai materiali e metodi si limita infatti a elencare e descrivere reagenti, protocolli, procedure e metodi di elaborazione matematica dei dati. Più precisamente, la presente sezione:

- dà ragione del modo in cui è stata progettata la ricerca a partire dalla sistemazione in un quadro coerente del complesso di informazioni raccolte in letteratura;
- descrive il percorso di ricerca attraverso il quale sono stati raggiunti gli scopi prefissati;
- motiva le scelte che hanno portato alla definizione di questo percorso.

In particolare, a partire dalla definizione dell'approccio verso il problema di *design* delle sequenze, sono elencati i criteri e i principi generali alla base del metodo combinatorio di progettazione di nuovi promotori sintetici secondo un processo iterativo.

3.1 Pianificazione del flusso delle attività

Da quanto esposto in introduzione, è chiaro che l'attuale livello di conoscenza dei meccanismi che regolano l'espressione endosperma-specifica non è sufficiente a progettare con certezza una sequenza con prestazioni superiori a quelle offerte dalle sequenze naturali. Per questo motivo il progetto di ricerca è stato concepito come un processo iterativo: il raggiungimento dello scopo è subordinato al possesso di alcune informazioni che non sono disponibili, ma che possono essere ottenute attraverso i tentativi (anche mancati) di raggiungere lo scopo stesso. Lavorando per approssimazioni successive, un primo ciclo sperimentale è stato effettuato con l'intento principale di raccogliere informazioni. Con queste informazioni si è avviato il successivo ciclo sperimentale che è stato concepito in modo che fornisse nuove informazioni in grado di aumentare la probabilità di

successo al terzo ciclo sperimentale, anche in caso di mancato raggiungimento degli obiettivi specifici.

Un ciclo sperimentale consiste nelle attività che vanno dalla progettazione all'ottenimento dei risultati utili per la pianificazione di nuove sequenze. Nell'ottica di iterazione del processo, ogni ciclo deve ovviamente concludersi prima dell'inizio del successivo.

L'attività di progettazione del promotore, fulcro dell'intero lavoro, si basa su una strategia che prevede la combinazione di varie sequenze naturali.

La costruzione del promotore può avvenire in vari modi, dalla sintesi dell'intera sequenza mediante ricorso a una ditta specializzata alla combinazione di più sequenze naturali utilizzando siti di restrizione già disponibili;

I promotori sono le uniche variabili dell'esperimento ovvero il confronto tra promotori viene effettuato a parità di tutte le altre componenti della cassetta di espressione, specificatamente: la sequenza codificante (CDS) la proteina *reporter* ed eventuali segnali all'N-terminale e al C-terminale, il terminatore, la sequenza non tradotta in 3'; ad esse possono aggiungersi altri elementi invariabili quali 5'-UTR, sito di inizio trascrizione e TATA-box.

3.1.1 Scelta del tipo di trasformazione e del vettore, numero di cicli sperimentali e numero di trasformazione per ciclo

Il numero di cicli sperimentali che si può prevedere nell'ambito di un progetto di ricerca dipende dalla lunghezza del progetto e dalla durata del singolo ciclo. Il fattore che più influenza la durata di un ciclo sperimentale è il tipo di trasformazione che si utilizza. L'espressione transitoria in endosperma immaturo (Hwang *et al.*, 2001) garantisce l'ottenimento dei risultati dopo appena 48 ore dall'intervento biolistico, fatto che almeno apparentemente renderebbe la costruzione del vettore il passaggio maggiormente limitante. Per contro, questo metodo non fornisce informazioni riguardo alla tessuto-specificità, produce livelli di espressione molto bassi e variabili e perciò necessita di uno standard interno inserito in

ogni sessione biolistica; inoltre, l'individuazione del giusto stadio di sviluppo del tessuto bersaglio e la sua preparazione richiedono una notevole esperienza oltre che tempo (va comunque programmata una semina delle piante da cui verranno tratti gli espianti; in riso, la maggior parte delle cv. fiorisce dopo 90-100 giorni dalla semina). Nell'unico protocollo disponibile in letteratura, l'espressione del gene *reporter uidA* è rilevabile solo con metodi luminometrici. Ad oggi questo metodo appare solamente nei lavori pubblicati dal gruppo che lo ha sviluppato. Il concreto rischio di incontrare ostacoli nell'applicare questo tipo di trasformazione, di ottenere dati poco solidi e di rinunciare a priori di trarre alcuna informazione riguardo all'effettiva tessuto-specificità hanno indotto a optare per un protocollo consolidato di trasformazione stabile.

Il vettore di espressione utilizzato è stato scelto in quanto già ampiamente studiato presso il gruppo di ricerca dove il presente lavoro è stato svolto. Il protocollo di trasformazione prevede una selezione positiva dei trasformati secondo una procedura ormai routinaria, priva di incognite.

Il processo che va dalla trasformazione all'analisi della progenie dura circa 8-9 mesi. Per esigenze tecniche, le date di trasformazione sono state fissate tenendo conto delle esigenze legate alla stagionalità della coltura. Infatti, per favorire una sufficiente produttività di seme, le piante rigenerate *in vitro* devono essere trapiantate in serra nel periodo che va dall'inizio di febbraio alla fine di luglio. Considerato questo vincolo, nel corso del dottorato sono stati programmati due cicli di trasformazione con obiettivi distinti: nel primo sono stati posti a confronto due promotori naturali, nel secondo una serie di promotori sintetici e naturali, utilizzando come controllo negativo piante rigenerate ma non trasformate della stessa varietà utilizzata negli esperimenti di trasformazione (CR W3).

Il primo ciclo è stato programmato con l'intento di valutare l'efficienza del promotore GluC nella sua versione 1,3 kb. Va rilevato che nella pubblicazione in cui questo promotore è stato caratterizzato in termini di attività trascrizionale, sono stati utilizzati come termine di confronto promotori di gluteline mai precedentemente considerati a tal scopo (Qu *et al.*, 2008). Conseguentemente, non è possibile stabilire se la superiorità

dimostrata da GluC verso tali promotori sia mantenuta in ambiti competitivi diversi, cioè in concorrenza con promotori di cui è nota la forza. E' stato dunque necessario risolvere anticipatamente questo dubbio per procedere con cognizione di causa alla progettazione di un gruppo di promotori candidati. Il primo ciclo è inoltre servito a mettere a punto il processo di sviluppo e valutazione di un promotore, specialmente in relazione alla numerosità dei trasformati primari e alla ripetibilità dei dati analitici.

Per quanto concerne il secondo ciclo, si è presentata la possibilità di inserire in pianta almeno 10 costrutti con altrettanti promotori, fino a un massimo di 15. Poiché il laboratorio ha una capacità massima di 10 trasformazioni in contemporanea, l'attività di trasformazione è stata disgiunta in due sessioni di lavoro distanziate di 3 mesi. La seconda sessione di trasformazione è stata utilizzata per ripetere alcuni esperimenti già condotti nella prima sessione con l'intento di valutare la ripetibilità del metodo.

3.1.2 Scelta delle componenti della cassetta di espressione

Visto il numero notevole di sequenze da inserire nella cassetta di espressione, si è deciso di procedere con un unico modulo accettore uguale per tutti i promotori sintetici. L'inserimento dei promotori nella cassetta di espressione così come l'introduzione delle varie cassette così originate nel vettore di espressione è stato eseguito in parallelo. Di seguito sono elencate le componenti fisse della cassetta:

- il gene *reporter* utilizzato codifica per l'enzima lisosomiale umano beta-glucosidasi acida (GCasi). Tale gene è stato scelto per due motivi: a. era disponibile per questo enzima un protocollo di estrazione ottimizzato e un saggio immunologico validato; b. l'enzima possiede una nota azione fitotossica, è quindi un ottimo candidato per valutare la tessuto-specificità;
- la sequenza codificante il peptide segnale già associata al gene *reporter* è quella relativa al peptide navetta di glutelina B4 (GluB4). Questa scelta è stata dettata da ragioni inerenti

l'indirizzamento all'opportuno compartimento subcellulare di accumulo (vacuoli proteici di riserva);

- il terminatore scelto è stato tratto dal gene per la nopalina sintetasi di *A. tumefaciens*. E' un terminatore "neutro", cioè non si riferisce ad alcun promotore o porzione di promotore utilizzato in questo lavoro. Ciò nell'intento di non favorire alcuna sequenza;
- è sempre stato utilizzato un 5'-UTR sintetico, brevettato dal gruppo di ricerca;
- l'abbinamento 5'-UTR e 3'-UTR è dunque rimasto invariato in tutti i casi. E' infatti noto che il particolare accoppiamento fra queste due sequenze può avere un effetto diverso di stabilizzazione del messaggero;
- invariata è stata altresì l'intera porzione del modulo minimo del promotore che va da 6 bp a monte della TATA-box al 5'-UTR.

In tutti i lavori citati nell'introduzione, con il termine promotore si intende una regione in genere lunga 1000 o 2000 bp, terminante in 3' con il sito di inizio traduzione. In questo modo però, si prendono in esame contemporaneamente diversi elementi, quali la regione non tradotta al 5' dell'RNA messaggero, il sito di inizio trascrizione, la TATA-box e il suo contesto, gli elementi *cis*-regolatori propri del promotore, eventuali *enhancer* situati parecchie centinaia di basi a monte della TATA-box, ma in grado di moltiplicare l'azione degli elementi prossimi al sito di inizio trascrizione. Mancano però in letteratura lavori che prendano in esame solo la regione a monte della TATA-box, mantenendo quest'ultima e la regione non tradotta al 5' dell'RNA messaggero immutate. Dal momento che sia la TATA-box che, soprattutto, il 5'-UTR non sono influenti sul livello di espressione, in questo lavoro si è deciso di fissare queste variabili e focalizzare l'attenzione sugli elementi che li precedono. Tutti i promotori saggiati hanno la medesima sequenza da 6 bp a monte della TATA-box fino al codone ATG (AUG) di inizio traduzione e precisamente hanno la sequenza propria del promotore della glutelina B4 (GluB4).

3.2 Progettazione dei promotori

In questo capitolo verrà descritto il metodo con cui sono state analizzate le sequenze di 14 promotori seme-specifici, verranno chiariti i criteri di scelta dei promotori naturali e il metodo di combinazione di questi ultimi in sequenze ibride sintetiche. Verranno poi chiarite le motivazioni che hanno portato alla scelta di dare la precedenza, nella sperimentazione *in vivo*, ad alcuni candidati rispetto ad altri.

3.2.1 Metodo di analisi delle sequenze dei promotori delle proteine di riserva del seme di riso (PRSR)

Per la costruzione di promotori sintetici, la ricerca è stata focalizzata sugli elementi *cis*-regolatori di promotori endosperma-specifici. A tal fine sono stati esaminati i principali promotori delle PRSR. Questa classe di promotori è nota per indurre una massiva espressione nell'endosperma e nello strato aleuronico, ma non negli altri tessuti compreso l'embrione. Sono stati scelti 14 promotori su cui eseguire un'analisi di sequenza per determinare presenza e disposizione di specifici elementi regolatori e per mettere in relazione questi ultimi con il livello e il profilo spazio-temporale di espressione.

Le principali PRSR sono le prolamine, le gluteline e le globuline. Le prolamine di riso sono una famiglia di proteine composta da 4 unità, dove ogni polipeptide è codificato da un gene dotato di un proprio promotore. Questi promotori sono identificati come 10 kDa, PG5 13 kDa, NRP 33 13 kDa e 16 kDa dal peso dei polipeptidi di cui inducono l'espressione. Sono note due prolamine da 13kDa codificate da due diversi geni, appunto PG5 e NRP 33. I due promotori sono stati studiati in due diversi lavori. Quindi è stato ritenuto opportuno prenderli in esame tutti e due.

Sono stati inclusi nell'analisi 9 promotori appartenenti alle 3 sottofamiglie di glutelina (A, B e C). Il promotore della subunità da 26kDa della alfa globulina è stato esaminato con particolare attenzione per l'alta espressione che induce rispetto alle ridotte dimensioni.

Tabella 2. Promotori analizzati, abbreviazioni utilizzate in questa tesi di dottorato, numeri di accessione e riferimento bibliografico da cui sono state tratte informazioni

Promotore	Abb.	Accessione	Riferimento bibliografico
Prolamina 10kDa	Prol10	AY427572	Qu e Takaiwa, 2004
Prolamina 13kDa, PG5	PG5	AY427573	Qu e Takaiwa, 2004
Prolamina 13kDa, NRP33	NRP33	AD63901	Sha <i>et al.</i> , 1996
Prolamina 16kDa	Prol16	AY427574	Qu e Takaiwa, 2004
Glutelina A1	GluA1	EU264102	Qu <i>et al.</i> , 2008
Glutelina A2	GluA2	EU264103	Qu <i>et al.</i> , 2008
Glutelina A3	GluA3	EU264104	Qu <i>et al.</i> , 2008
Glutelina B1	GluB1	AY427569	Qu e Takaiwa, 2004
Glutelina B2	GluB2	AY427570	Qu e Takaiwa, 2004
Glutelina B3	GluB3	EU264105	Qu <i>et al.</i> , 2008
Glutelina B4	GluB4	AY427571	Qu e Takaiwa, 2004
Glutelina B5	GluB5	EU264106	Qu <i>et al.</i> , 2008
Glutelina C	GluC	EU264107	Qu <i>et al.</i> , 2008
Globulina 26kDa	Glb	AY427575	Qu e Takaiwa, 2004

Su questi 14 promotori è stata condotta un'analisi di sequenza con l'obiettivo di individuare gli elementi *cis*-regolatori che conferiscono specificità e che regolano la velocità di trascrizione.

Gli elementi regolatori sono stati annotati sulle sequenze utilizzando il *software* di gestione di sequenza *BioEdit* (§4.1). Le sequenze con le annotazioni sono state salvate come file in formato *.gb. Nel capitolo 5.1, alcuni grafici a barre riassumono i risultati dell'analisi sotto forma di mappa fisica lineare.

In un primo momento, era stato previsto di ottenere un risultato idoneo solo con l'ausilio delle specifiche banche dati disponibili in rete. Poiché questo approccio non è apparso sufficiente, è stato applicato un metodo di analisi più approfondito che verrà descritto di seguito in questo capitolo.

3.2.2 Analisi basata sull'interrogazione delle banche dati

Nella parte iniziale di questo dottorato sono state utilizzate tre banche dati relative ai promotori vegetali: PLACE, TRANSFACT e PlantCARE (§4.1).

PLACE restituisce un gran numero di possibili elementi *cis*-regolatori, per la maggior parte poco o per nulla correlati con la reale funzione del promotore. L'analisi svolta dal programma sulla banca dati lascia all'utente il compito di distinguere le informazioni utili dal rumore di fondo. Oltre a ciò la banca dati non è più aggiornata dal 2007.

TRANSFACT e PlantCARE forniscono un risultato molto più semplice da consultare. Entrambi però non sono particolarmente adatti per studiare i promotori endosperma-specifici. Le informazioni che forniscono sono ottenibili anche da una superficiale ricerca bibliografica. In conclusione non è stato possibile procedere nel lavoro limitandosi al solo uso di questi strumenti.

3.2.3 Sviluppo di un metodo basato sui dati reperiti in letteratura

Vista l'impossibilità di analizzare in maniera soddisfacente le sequenze dei promotori basandosi solo su strumenti informatici, si è proceduto a un lavoro d'integrazione delle informazioni pubblicate negli articoli che trattano l'argomento specifico. Diversi algoritmi sono disponibili per l'individuazione di motivi *cis*-regolatori (§1.3.1), ma in genere essi sono concepiti per lavorare su un intero genoma e non su un ridotto numero di promotori. Poiché 14 promotori rappresentano un campione troppo esiguo per l'applicazione fruttuosa degli algoritmi statistici citati in introduzione, si è preferito effettuare la ricerca manualmente, tenendo in considerazione i principi base che regolano il funzionamento degli strumenti di analisi di sequenza automatici.

In questo paragrafo viene descritto il metodo seguito per l'analisi delle 14 sequenze. Questo metodo, messo a punto nella fase iniziale del dottorato, ha portato all'individuazione delle possibili sequenze consenso dei quattro principali elementi *cis*-regolatori implicati nell'espressione in endosperma di riso. Le sequenze consenso individuate sono state classificate secondo il grado di probabilità che esse svolgano un'effettiva funzione *in vivo*. Per stilare una lista dei possibili consensi sono stati principalmente osservati i risultati di alcuni esperimenti pubblicati nel corso della caratterizzazione della sequenza costituita dai 197 nucleotidi a monte

del sito di inizio trascrizione del promotore della glutelina B1, $\Delta[-197;-1]$ (Wu *et al.*, 1998; Washida *et al.*, 1999; Wu *et al.*, 2000). Inoltre, sono state determinanti le osservazioni dei risultati degli esperimenti di *methylation interference* condotti sul promotore della prolamina 13kDa NRP33 $\Delta[-300;-1]$ e sul promotore della globulina 26kDa $\Delta[-190;-1]$. Questi esperimenti sono stati condotti al fine di caratterizzare la specificità di legame di due fattori di trascrizione espressi nel seme di riso (Yamamoto *et al.*, 2006). Il quadro è stato completato ricercando in letteratura qualsiasi informazione che permettesse di desumere le sequenze consenso legate da fattori *trans*-agenti caratteristici delle prime fasi della maturazione del seme di riso. In questo modo è stato creato una sorta di "dizionario" in modo analogo a quello che avviene negli algoritmi per l'individuazione automatica dei consensi. L'identificazione dei motivi costituenti elementi regolatori si è basata su 3 criteri. Nel dettaglio, sono state accettate come sequenze consenso:

1. Motivi che causano una perdita di attività del promotore qualora mutati o deleti.
2. Motivi che vengono legati da particolari fattori di trascrizione in esperimenti di protezione dalla metilazione.
3. Motivi che inseriti in un promotore lo rendono suscettibile all'attivazione di particolari fattori di trascrizione.

Tabella 3. Sequenze consenso identificate in base ai dati reperiti in letteratura. Il codice di ambiguità è: R=A/G, Y=C/G, K=G/T, S=G/C, M=A/C, W=A/T, B=G/C/T, V=A/G/C, D=A/G/T, H=A/C/T, N=A/G/C/T

	GCN4	Prol-Box	AACA	ACGT
GluB1	TGAGTCA	TGCAAAG AGGAAAG	YAACAAAC	GTACGTGC
NRP33	TGACTCA	CATAAAG ATCAAAG TCTAAAG		
Glb	TGMCACA TGTCACC	YAAAAAG AKAAAAG CAAAAAG		GCCACGTMAG
GluA2		ACGAAAG		
GluA3		TTTAAAG		

Queste sequenze consenso hanno permesso di localizzare alcuni elementi *cis*-regolatori all'interno dei 14 promotori delle PRSR. Questo metodo, però, non ha permesso di condurre a compimento l'analisi. Infatti, alla fine del lavoro sono rimaste evidenti lacune. In particolare:

- non è stato possibile identificare il motivo GCN4 nei promotori delle gluteline B5 e C;

- non è stato possibile individuare il motivo AACAA nel promotore delle prolamine da 13kDa e 16kDa;

- le sequenze identificate come consenso per il motivo ACGT sono apparse insufficienti a individuare i relativi elementi *cis*-regolatori nei promotori delle PRSR. In effetti, il fatto che ACGT sia un motivo diffuso nei più svariati promotori vegetali rende difficile l'individuazione certa degli elementi *cis*-regolatori specifici dei promotori delle PRSR.

La mancanza di elementi importanti come GCN4, AACAA e ACGT nei promotori delle PRSR è del resto altamente improbabile. Per completare l'analisi di sequenza è stato ipotizzato che le sequenze consenso elencate in Tabella 3 non siano sufficienti a individuare tutti gli elementi regolatori. Il procedimento di ricerca è stato reiterato e per rendere più elastici i criteri di identificazione sono stati analizzati i risultati di alcuni esperimenti pubblicati nel corso della caratterizzazione dei fattori di trascrizione OSMYB5, RISBZ1, RISBZ2 e RPBF (Nakase *et al.*, 1997; Suzuki *et al.*, 1998; Onodera *et al.*, 2001; Yamamoto *et al.*, 2006). In questi esperimenti, un frammento di DNA contenente la sequenza consenso per un determinato fattore viene marcato. La mobilità di questo frammento in gel di poliacrilammide viene ritardata dalla presenza del fattore di trascrizione. L'aggiunta di un eccesso del medesimo frammento di DNA non marcato abolisce per competizione il ritardo nella mobilità elettroforetica del DNA marcato. Qualora invece il frammento competitore contenga mutazioni, la capacità di quest'ultimo di competere con la sequenza originale marcata dipenderà dall'effetto che le mutazioni introdotte hanno sull'affinità per il fattore di trascrizione. Se le mutazioni introdotte non coinvolgono nucleotidi fondamentali per il riconoscimento della sequenza consenso, il competitore si sostituirà al frammento marcato nel legame con il fattore. In questo modo non si

osserverà alcuna riduzione di mobilità del frammento marcato. Se le mutazioni introdotte nel frammento competitore coinvolgono nucleotidi fondamentali per l'affinità di legame con il fattore di trascrizione, anche un largo eccesso di competitore non riuscirà a distrarre il fattore di trascrizione dal frammento marcato. Si osserverà in questo caso una riduzione della mobilità del DNA marcato sovrapponibile a quella ottenuta in assenza di competitore. I risultati di questi esperimenti suggeriscono l'esistenza di sequenze minime necessarie al riconoscimento dei fattori di trascrizione. L'interpretazione data dagli autori non è sempre stata rigorosa dal punto di vista logico e spesso le conclusioni hanno fornito criteri troppo stringenti per l'individuazione del consenso. Di seguito viene descritto un esempio di come una non corretta interpretazione dei dati ha portato a conclusioni che sono state smentite dai dati di un successivo lavoro. L'identificazione della sequenza consenso per GCN4 condotta nel lavoro di Onodera *et al.* (2001) mediante un esperimento di EMSA (*Electrophoretic Mobility Shift Assay*) ha fatto concludere agli autori che mutazioni introdotte in qualunque posizione della sequenza TGAGTCA causassero la perdita di affinità da parte del fattore RISBZ1 nei confronti della sequenza mutata e che dunque l'elemento specificatamente riconosciuto da RISBZ1 fosse TGAYTCA.

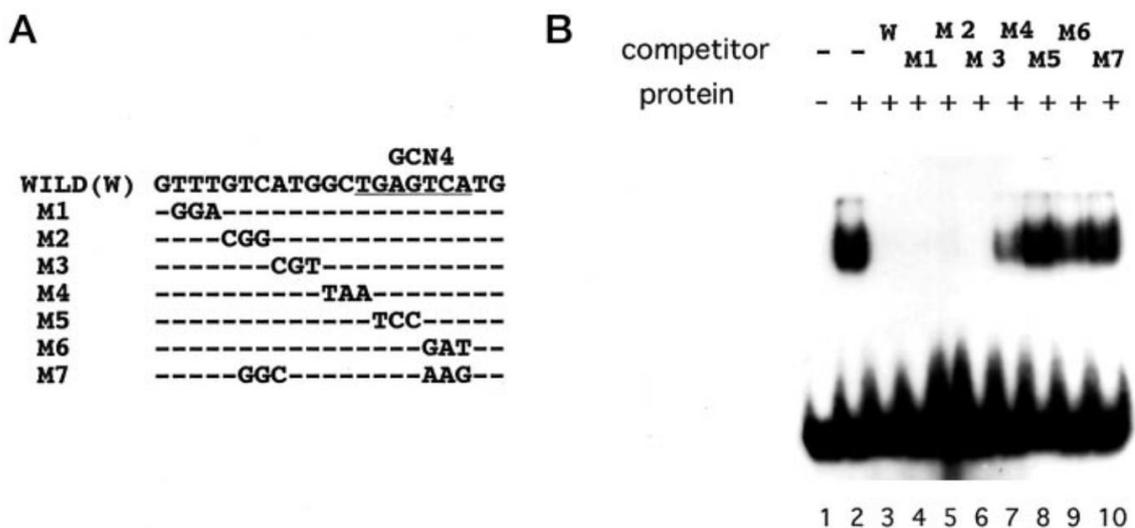


Figura 8. EMSA condotto da Onodera *et al.* (2001); qualunque mutazione perturbante il motivo TGAGTCA nel competitore appare causare l'assenza di legame con RISBZ1.

Questa conclusione è poi stata smentita nei fatti da esperimenti di *methylation interference* condotti nel 2006 da Yamamoto *et al.* sullo stesso fattore di trascrizione. Gli esperimenti pubblicati nel 2006 hanno dimostrato chiaramente che pur mancando un motivo TGAGTCA nel promotore della globulina 26kDa $\Delta[-190;-1]$, RISBZ1 è in grado di legare sequenze compatibili con i risultati di EMSA, ma non con le conclusioni, troppo stringenti, tratte dagli autori. Analogamente a quanto avvenuto per l'elemento GCN4, anche per altri elementi *cis*-regolatori può essere stata individuata una sequenza consenso troppo stringente. Va detto tuttavia che, a differenza di quanto avvenuto nel caso di GCN4, l'errata determinazione non è stata ancora segnalata in lavori successivi. I dati di tutti gli EMSA pubblicati sono stati dunque rielaborati determinando tutte le sequenze che palesemente non sono compatibili con il legame da parte del rispettivo fattore di trascrizione. L'insieme che contiene tutte le sequenze tranne queste è stato chiamato "sequenza necessaria", perché rappresenta la sequenza al di fuori della quale non è possibile il legame. A tal proposito è bene ricordare che un insieme di sequenze è rappresentato da tutte le combinazioni che soddisfano una sequenza degenerata.

E' stata definita "sequenza sufficiente" l'insieme delle sequenze che sicuramente conducono al riconoscimento da parte del fattore di trascrizione. In genere "sufficienti" sono proprio le sequenze proprie dei frammenti di promotori su cui sono stati eseguiti gli studi pubblicati. In tal caso, esistendo un'evidenza bibliografica diretta, il legame con il fattore di trascrizione è fuori questione.

Tabella 4. Riassunto delle conclusioni tratte dalla elaborazione dei risultati di esperimenti di EMSA. R=A/G, Y=C/G, K=G/T, S=G/C, M=A/C, W=A/T, B=G/C/T, V=A/G/C, D=A/G/T, H=A/C/T, N=A/G/C/T

Fattore di trascrizione	Elemento <i>cis</i>-regolatore	Sequenza necessaria	Sequenza sufficiente
RISBZ1	GCN4	TGASYBM TV DSTCA	TGASTCA
RISBZ2	ACGT	YACGTNND	GCCACGTMAG
RPBF	Prolamin-box	AAAG	YRYAAAG
OSMYB5	AACA	AACNNA	YAACRRAYY

Utilizzando queste sequenze è stato possibile completare la mappatura dei più noti elementi *cis*-regolatori dei promotori delle PRSR. I motivi che rispondono ai criteri delle sequenze sufficienti sono stati annotati come elementi certi, mentre i motivi che rientrano nell'insieme delle sequenze necessarie sono stati annotati come elementi regolatori putativi.

3.2.4 Possibili approcci alla progettazione di promotori sintetici

Una volta individuati i motivi che vengono riconosciuti dai fattori di trascrizione endosperma-specifici e le regioni che li contengono, è stato studiato un modo per combinarli in una sequenza sintetica. Il fattore che ha maggiormente condizionato l'approccio al *design* di un promotore sintetico è l'orizzonte temporale in cui si prevede di ottenere dati che diano conto dell'efficacia della sequenza progettata. In altre parole, il numero di cicli entro cui si intende arrivare a un risultato definitivo determina l'approccio con cui verranno progettati i promotori candidati. Le considerazioni che hanno portato alla definizione del metodo sono riportate di seguito.

Impiegando 4 o 5 cicli, ciascuno dei quali fornisce informazioni su una decina di sequenze è possibile operare con un approccio esplorativo: validare cioè i dati provenienti dalle analisi e integrarli con ulteriori informazioni prima di procedere al disegno del promotore definitivo

Dal momento che i tempi tecnici non hanno permesso di completare più di 2 cicli sperimentali è stato deciso di strutturare la ricerca di dottorato in modo da raggiungere un primo risultato al termine del secondo ciclo sperimentale, con l'obiettivo di produrre una prova di concetto della validità del progetto di ricerca. Progetto che, però, non è stato concepito per concludersi con questo primo risultato, ma per proseguire con almeno un terzo ciclo. Di conseguenza i promotori chimerici presentati in questa tesi sono stati progettati nell'intento di ottenere informazioni utili per il prossimo ciclo di sperimentazione. La scelta del metodo di progettazione ha tenuto conto delle lacune informative che era necessario colmare con i risultati del secondo ciclo e allo stesso tempo della necessità di produrre una prova di concetto entro la fine del corso di dottorato. Per soddisfare entrambe le

istanze è stato scartato l'approccio esplorativo che avrebbe comportato la necessità di saggiare a tappeto una serie di modificazioni su sequenze naturali. E' stato invece preferito un approccio combinatorio che consente di ottenere il massimo delle informazioni da un set di promotori ibridi. I progetti dei promotori chimerici frutto di questo approccio non coprono tutte le combinazioni possibili, ovviamente la rosa teorica dei candidati è stata sottoposta al rigoroso vaglio delle nozioni acquisite in letteratura. In caso contrario il numero delle sequenze da saggiare sarebbe cresciuto a dismisura. In tal senso si può parlare di approccio combinatorio informato.

La progettazione è stata sviluppata in maniera integrata alla costruzione, seguendo un metodo "taglia-e-cuci" che ha permesso di produrre velocemente un gruppo assortito di promotori ibridi.

3.2.5 Scelta dei promotori da combinare

Il primo ciclo non ha presentato problemi di progettazione, poiché è consistito nella costruzione di un solo promotore e nel saggio di confronto con il promotore naturale della glutelina B4 (controllo positivo). Il promotore Δ GluC saggiato si può definire un promotore sintetico a tutti gli effetti, dal momento che è costituito dalla TATA-box di GluB4 e dalla regione a monte mutuata da GluC, ma troncata 1040 bp dall'estremità 5' della sequenza che si trova depositata in banca dati. I due promotori comparati in questa prova hanno la stessa lunghezza e il confronto diretto ha avuto lo scopo di valutare se la porzione Δ GluC possa fungere da base di partenza per i promotori successivi.

Per il secondo ciclo è stato necessario progettare una serie di promotori chimerici sintetici; per fare ciò si è proceduto a scegliere candidati naturali da cui trarre le regioni da combinare.

Il primo criterio di scelta è stato il livello di espressione desumibile dai dati pubblicati. Il profilo di espressione all'interno del seme è stato utilizzato come criterio di esclusione. Promotori caratterizzati da un accentuato accumulo di proteina ricombinante nello strato aleuronico a scapito dell'endosperma non sono stati considerati come basi di partenza per la costruzione delle sequenze chimeriche.

Come si vedrà nelle prossime sezioni, la validità del metodo utilizzato è stata verificata sulla base dei risultati ottenuti. In questo modo i dati raccolti sui promotori introdotti stabilmente in riso hanno avuto una duplice applicazione. Fornire nuove informazioni sulla funzione delle varie regioni all'interno dei promotori delle PRSR e confermare o smentire le ipotesi su cui è stato basato il metodo di progettazione presentato in questa sezione. La possibilità di validare a ogni ciclo sperimentale il metodo adottato permette di correggere gli eventuali errori di impostazione compiuti.

In questo modo si realizza l'idea iniziale di una piattaforma di sviluppo che a ogni ciclo sperimentale si affina nei metodi e si arricchisce di informazioni.

4 MATERIALI E METODI

4.1 Supporti informatici

Tutte le operazioni di analisi e allineamento di sequenze nucleotidiche sono state condotte con l'ausilio del *software: BioEdit Sequence Alignment Editor. Copyright © 1997-2005 (Hall, 1999).*

Il lavoro di costruzione dei promotori sintetici è stato pianificato servendosi del *software: Clone Manager 5, versione 5.05. Copyright © 1995-1998, Scientific & Educational Software.*

Parte dell'analisi di sequenza dei promotori naturali è stata svolta interrogando le seguenti banche dati:

- PLACE (<http://www.dna.affrc.go.jp/PLACE/>),
- TRANSFACT (<http://www.gene-regulation.com/pub/databases.html>)
- PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>)

Il disegno dei *primers* utilizzati per l'amplificazione dei promotori naturali e per il sequenziamento è stato realizzato utilizzando l'applicazione *Primer3* (<http://frodo.wi.mit.edu/>) disponibile in rete.

4.2 Materiali e metodi di biologia molecolare

4.2.1 Estrazione di DNA genomico da foglia di riso

Per estrarre il DNA da foglie di riso trasformato è stato utilizzato il protocollo di Doyle e Doyle (1987).

4.2.2 Metodiche di biologia molecolare

Tutte le reazioni di amplificazione sono state effettuate utilizzando il termociclatore MyCycler™ Thermal cycler (Bio-Rad) o il termociclatore PCR Sprint® (Hybrid). Gli oligonucleotidi sintetici utilizzati, di lunghezza variabile da 15 a 53 basi, sono stati sintetizzati dalla ditta Sigma Genosys.

Le separazioni elettroforetiche di DNA su gel di agarosio (GellyPhor® LE, EuroClone) sono state effettuate con l'apparato BioRad alle condizioni di corsa 4 V/cm per circa 1 h. Il recupero da gel di amplificati PCR e di prodotti

ottenuti da digestione enzimatica è stato realizzato mediante il kit Wizard® SV Gel & PCR Clean-up System (Promega).

Nota: i marcatori utilizzati nella visualizzazione dei frammenti di DNA in gel di agarosio sono 1 kb DNA ladder (NEB) (Fig. 2.1 A), 25 bp DNA step ladder Promega G4511 (Fig. 2.1.B) e 100 bp DNA ladder (NEB) (Fig. 2.1C).

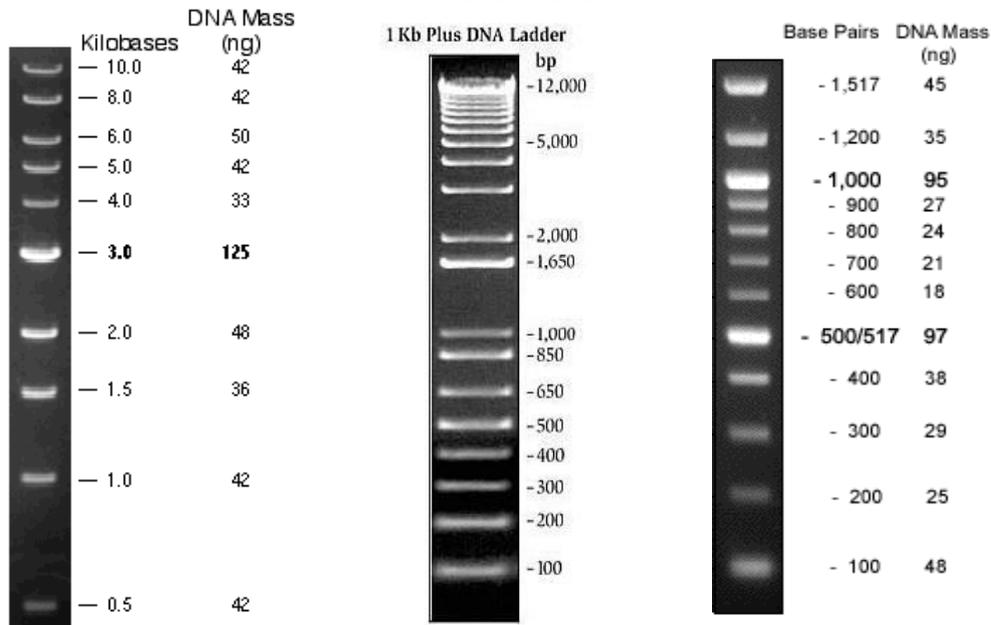


Figura 2.1 A, B, C. Marcatori molecolari utilizzati.

Le digestioni enzimatiche sono state effettuate utilizzando enzimi di restrizione e tamponi di digestione forniti da New England Biolabs (NEB).

Ciascuna reazione di digestione è stata allestita utilizzando i seguenti componenti:

Componenti	Quantità
Tampone opportuno 10X	1.6 µL
BSA 100X	0.2 µL
Enzimi di restrizione (uno o più)	5-10 U
DNA da digerire	<=500 ng
H ₂ O	a volume
Volume finale	16.0 µL

Le reazioni di digestione sono state singole, doppie o triple a seconda dell'uso rispettivamente di 1, 2 o 3 enzimi di restrizione. Le digestioni condotte con più di un enzima hanno richiesto necessariamente una compatibilità di tampone tra i diversi enzimi di restrizione impiegati.

Le reazioni sono state lasciate in incubazione a 37°C per tempi variabili, da 2 a 14 ore (O/N, over-night) in funzione di diversi fattori, quali la quantità di DNA da digerire, l'efficienza dell'enzima e il possibile rischio di *star activity*. In digestioni prolungate e in presenza di enzimi caratterizzati da *star activity*, alla miscela di reazione è stata aggiunta come stabilizzante albumina di siero bovino (BSA)(NEB). La digestione con gli enzima *ApaI* e *ApoI* sono state condotte ad una temperatura di 25°C e 50°C rispettivamente.

Le reazioni di saldatura tra inserti e vettori sono state realizzate mediante l'enzima T4 ligasi (Promega) in presenza del tampone specifico, con incubazioni a 4°C overnight (*ligation* lenta) oppure a temperatura ambiente per almeno un'ora (*ligation* rapida). I quantitativi di inserto e vettore necessari per ciascuna reazione sono stati calcolati secondo la seguente formula:

$$\frac{\text{ng vettore} \times \text{bp inserto}}{\text{bp vettore}} \times \frac{3}{1} = \text{ng inserto}$$

considerando per ogni reazione una quantità di 25 ng di vettore.

Ligation "rapida", incubazione a RT, minimo 1h:

Vettore	25.0 ng
Inserto	ng calcolati
Tampone 2x (Promega)	5.0 µL
T4 ligasi	1.0 µL
Volume finale	10.0 µL

Ligation "lenta", incubazione a 4°C O/N:

Vettore	25.0 ng
Inserito	ng calcolati
Tampone 2x (Promega)	1.0 µL
T4 ligasi	1.0 µL
Volume finale	10.0 µL

Per la trasformazione di *E. coli* è stato utilizzato il ceppo JM101, reso competente con CaCl₂ (Sambrook *et al.*, 1989); la trasformazione di cellule competenti di *A. tumefaciens* (ceppo EHA105), preparate secondo il protocollo di Lin *et al.* (1995), è stata effettuata mediante elettroporazione con l'elettroporatore Micro Pulser™ (Bio-Rad).

Il DNA plasmidico dei ceppi trasformati è stato estratto mediante Wizard® plus minipreps (Promega) o QIAprep® spin miniprep Kit (Qiagen). Il sequenziamento dei tratti di DNA sintetizzati e/o subclonati è stato attuato da IGA Services srl o Primm srl.

4.3 Vettori utilizzati

Nell'ambito di questa tesi, per le operazioni di realizzazione e assemblaggio delle cassette geniche da trasferire nei vettori finali di espressione sono stati utilizzati i vettore commerciali pUC18, pGEM-T e pCAMBIA1300_PMI.

4.3.1 pUC18

pUC18 è un plasmide di *E. coli* ad alto numero di copie. Porta il gene per la resistenza all'ampicillina e presenta un MCS all'interno della regione codificante per il frammento α della beta-galattosidasi, rendendo possibile lo *screening* secondo colore delle colonie trasformate.

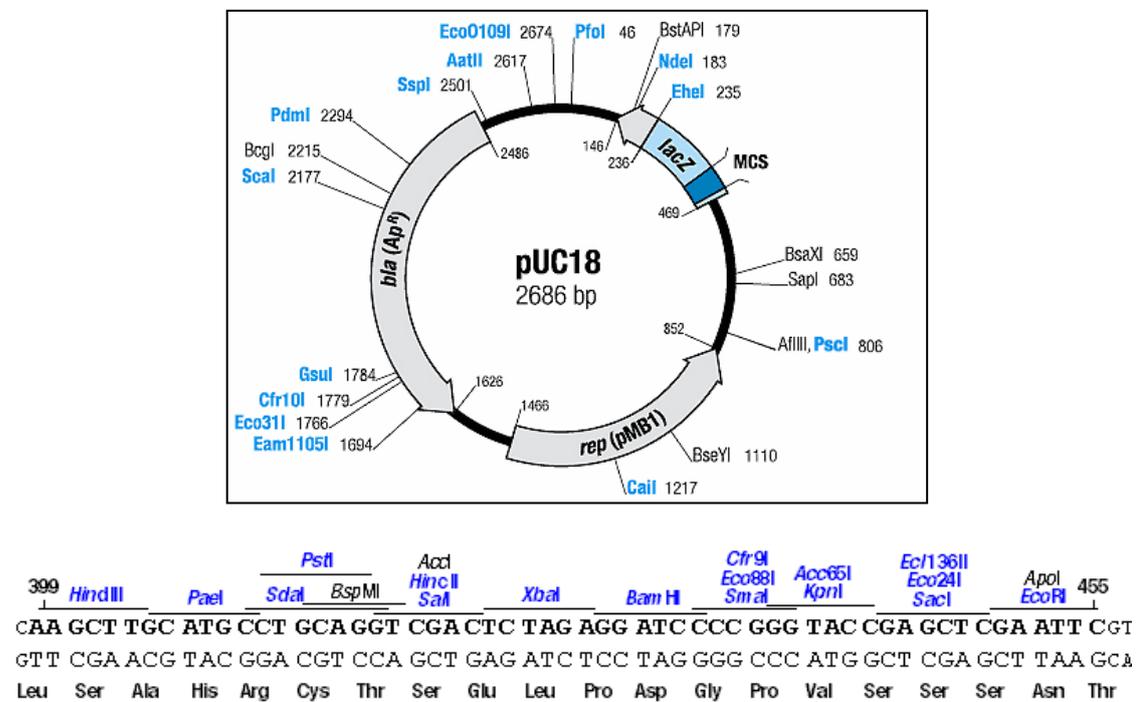


Figura 9. Rappresentazione schematica del vettore pUC18 con in dettaglio il MCS.

Infatti, l'inattivazione inserzionale permette ai cloni ricombinanti di essere identificati direttamente tramite α -complementazione: grazie alla presenza di un substrato cromogeno per la β -galattosidasi, le colonie che non presentano l'inserito appaiono blu, mentre quelle correttamente trasformate presentano il caratteristico colore bianco crema di *E. coli*.

4.3.2 pGEM-T

Il vettore pGEM[®]-T presenta ai lati del MCS i siti di legame per diversi *primer*, fra cui M13 *forward* e *reverse*, utilizzabili per l'amplificazione o il sequenziamento di inserti (www.promega.com).

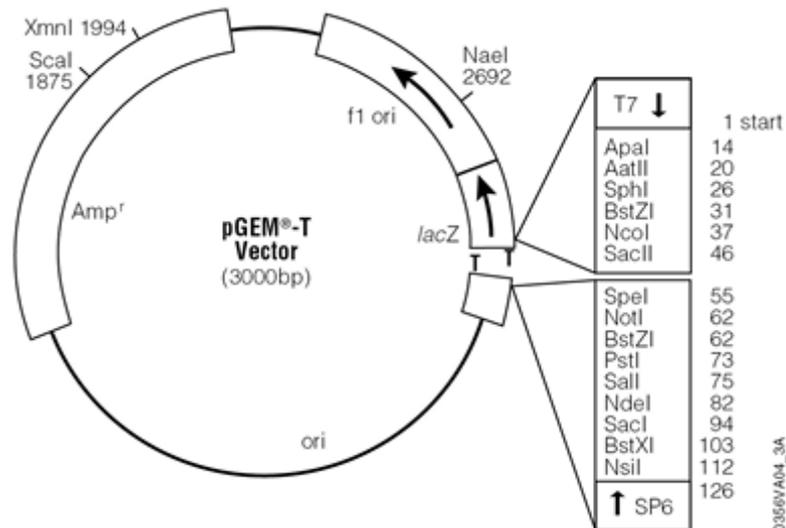


Figura 10. Rappresentazione schematica di pGEM-T con in dettaglio il MCS.

Il vettore pGEM[®]-T è un plasmide high copy *number* derivato da pUC18 contenente i promotori T7 e SP6 nella regione fiancheggiante il sito di clonazione multiplo (MCS, *multiple cloning site*); quest'ultimo mappa all'interno della regione codificante per il frammento α dell'enzima β -galattosidasi (Fig. 2.3). L'inattivazione inserzionale permette ai cloni ricombinanti di essere identificati direttamente tramite α -complementazione.

4.3.3 pCAMBIA

Come base per il vettore finale di espressione in riso è stato utilizzato un vettore pCAMBIA opportunamente modificato, in cui sono state inserite le cassette geniche di interesse.

I vettori appartenenti alla famiglia pCAMBIA sono stati sviluppati appositamente per la trasformazione di piante mono- e dicotiledoni e,

sebbene vadano ancora perfezionati, presentano alcune caratteristiche che li rendono adatti a questo tipo di lavoro:

- vengono replicati in alto numero di copie in *E. coli*;
- presentano un'elevata stabilità in agrobatterio grazie al replicone di pVS1;
- hanno dimensione relativamente ridotta (7-12kb); presentano un gene per la resistenza antibiotica utilizzabile per la selezione di cellule trasformate batteriche (kanamicina).

La versione utilizzata in questo lavoro è stata modificata rispetto a quella fornita dall'istituto CAMBIA (Camberra, Australia). Infatti, in un precedente lavoro condotto dal gruppo di ricerca presso cui è stato svolto il dottorato, è stato sostituito il marcatore selezionabile in pianta. In particolare, il gene che conferisce la resistenza ad un antibiotico (igromicina) è stato sostituito con il gene batterico PMI che codifica per l'enzima fosfomannosio isomerasi (GenBank acc. n° M15380). L'uso di questo marcatore permette una selezione positiva delle piante trasformate (Miles e Guest, 1984).

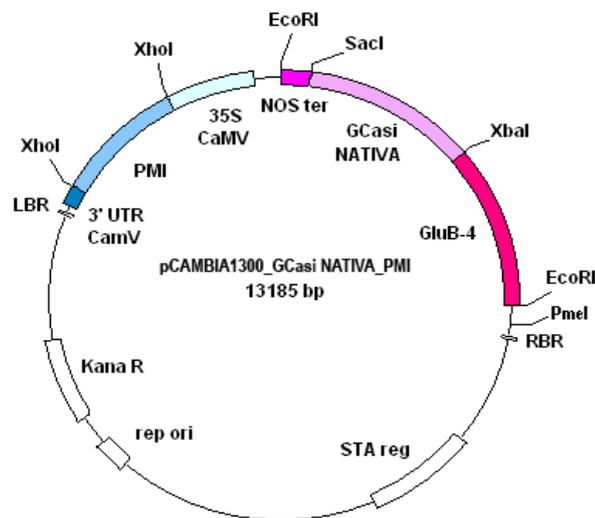


Figura 11. Struttura del vettore pCambia con il marcatore selezionabile PMI. Nell'immagine è riprodotto il vettore che contiene il promotore di GluB4.

Il vettore pCambia 1300 PMI (Fig. 2.4) è un vettore binario contenente un *polylinker* derivato da pUC18 e i geni *Kan^r*(nptIII) per la

resistenza all'antibiotico kanamicina e PMI per la selezione positiva delle piante trasformate, rispettivamente. Il gene *Kan^r* codifica per una neomicina fosfotransferasi in grado di inattivare gli antibiotici appartenenti alla famiglia della neomicina, in particolare kanamicina e amikomicina, conferendo un elevato livello di resistenza alle cellule batteriche che possiedono il plasmide portante tale gene e il promotore batterico. Il gene per la PMI codifica per la fosfomannosio isomerasi e permette di selezionare le piante trasformate utilizzando un substrato povero di saccarosio, ma ricco di mannosio come agente di selezione. Questo gene è posto sotto il controllo di un promotore attivo in pianta.

Il vettore utilizzato in questo lavoro è un pCAMBIA1300 modificato, indicato pCAMBIA13xx. Esso diverge dal vettore commerciale nella zona del MCS, che risulta quasi totalmente rimossa (delezione di 50 bp) al fine di impedire interferenze con i siti di restrizione unici presenti nella cassetta di interesse.

4.4 Promotori naturali da genoma di riso CR W3

4.4.1 Promotore della Glutelina B4 con *leader* sintetico LLTCK

Il promotore della glutelina B4 fuso al *leader* sintetico LLTCK è stato costruito in un lavoro precedente e l'inserito è stato clonato con orientamento inverso in pGEM-T. Di seguito è riportata la sequenza inserita nel MCS:

```
5'-NsiI-BstXI-SacI-NdeI-SalI-PstI-EagI-NotI-SpeI-TGCATGCGAATTCTACAGGGTT  
CCTTGCGTGAAGAAGGGTGGCCTGCGGTTACCATTAACGGTCACGACTACTTCCAGCTAGTAC  
TGGTGACCAACGTCGCGGCGGCAGGGTCAATCAAGTCCATGGAGGTTATGGGTTCCAACACAGC  
GGATTGGATGCCGATGGCAGTAACTGGGGCGCCCAATGGCACTCACTGGCCTACCTCACCGGT  
CAAGGTCTATCCTTTAGGGTCACCAACACAGATGACCAAACGCTCGTCTTCACCAACGTCGTGC  
CACCAGGATGGAAGTTTGGCCAGACATTTGCAAGCAAGCTGCAGTTCAAGTGAGAGGAGAAGCC  
TGAATTGATAACCGAGCGTTTCTTTGGGAGTAACATCTCTGGTTGCCTAGCAAACATATGATT  
GTATATAAGTTTCGTTGTGCGTTTATTCTTTTCGGTGTGTAATAACATACATGCTTTCTCTGAT  
ATTTTCTTGTATATATGTACACACACACGACAAATCCTTCCATTTCTATTATTATTGAACAATT  
TAATTGCGAGGGCGAGTACTTGTCTGTTTACCTTTTTTTTTTCAGATGGCATTTTATAGTTTAAC  
CTTTTCATGGACCGGCAGTAGTTCTAACCATGAAATGAAAAGAAATCATAGTCCACACCACGCAGG  
GACATTGTGGTCATTTTAGACAAGACGATTTGATTAATGTCTTGTATGATATGGTCGACAGTGA
```

GGACTAACAAACATATGGCATATTTTATTACCGGCGAGTTAAATAAATTTATGTCACAGTAATA
 AACTGCCTAATAAATGCACGCCAGAAAATATAATGATAAAAAAAGAAAAGATACATAAGTCCA
 TTGCTTCTACTTTTTTAAAAATTAAATCCAACATTTTCTATTTTTTGGTATAAACTTGGAAAGTA
 CTAGTTGGATATGCAAAATCATCTAACCTCCATATATTTTCATCAATTTGTTTACTTTACATATG
 GGAGAGGATAGTATGTCAAAGAAAATGACAACAAGCTTACAAGTTTCTTATTTTTAAAAGTTCCG
 CTAAC TTATCAAGCATAGTGTGCCACGCAAACTGACAACAAACCAACAAATTTAAGGAGCGCC
 TAAC TTATCATCTATGACATACCGCACAAAATGATAACATACTAGAGAACTTTATTGCACAAA
 AGGAAATTTATCCATAAGGCAAAGGAACATCTTAAGGCTTTGGATATACATTTACCAACAAGCA
 TTGTTTGTATTACCCCTAAAGCGCAAGACATGTCATCCATGAGTCATAGTGTGTATATCTCAAC
 ATTGCAAAGCTACCTTTTTTCTATTATACTTTTTCGCATTATAGGCTAGATATTATCTATACATG
 TCAACAAACTCTATCCCTACGTCATATCTGAAGATTCTTTTCTTCACTATATAAGTTGGCTTCC
 CTGTCATTGAACTCACATCAACCAGCCCAACACGTATTTTTACAACAATACCAACAACAACAAC
 AACAAACAACATTACAATTACGTATTTCTCTCTCTAGAA-SacII-NcoI-EagI-SphI-3'

Nella sequenza sono sottolineati i siti di restrizione *SphI* e *EcoRI* aggiunti in 5' e *XbaI* aggiunto in 3'. Il *leader* sintetico è indicato in corsivo.

4.4.2 Sequenze di riferimento per gli altri promotori naturali

Le sequenze dei promotori utilizzati in questo lavoro sono depositate in banca dati Genbank (NCBI) e sono accessibili in rete.

Tabella 5. Elenco dei promotori utilizzati con i relativi numeri di accessione.

Promotore	Accessione	Riferimento bibliografico
Globulina 26kDa	AY427575	Qu e Takaiwa, 2004
Glutelina B1	AY427569	Qu e Takaiwa, 2004
Glutelina B4	AY427571	Qu e Takaiwa, 2004
Glutelina C	EU264107	Qu <i>et al.</i> , 2008
Prolamina 10kDa	AY427572	Qu e Takaiwa, 2004
Prolamina 16kDa	AY427574	Qu e Takaiwa, 2004

4.4.3 Primers utilizzati per l'amplificazione e il sequenziamento dei promotori naturali

Le sequenze dei *primers* sono state disegnate in modo da ottenere una T_m prossima ai 60°C. I *primers reverse* appaiano esattamente 6nt a monte della TATA-box e aggiungono un sito *XhoI* all'amplificato. Questo sito di restrizione è stato utilizzato per legare i promotori al modulo accettore. Il modulo accettore contiene la TATA-box e il sito iniziatore del promotore

della glutelina B4. Questo modulo viene descritto nel capitolo successivo. Le coppie di *primers forward* e *reverse* per ogni singolo promotore sono state disegnate, laddove possibile, in modo da amplificare l'intera regione del promotore, esclusa la TATA-box. In alcuni casi è stato necessario fissare un *primer forward* leggermente più interno rispetto all'estremità 5'. Glb 26kDa *for* e Prol 16kDa *for* appaiano 4bp e 26bp più a valle rispetto all'estremità 5' della sequenza depositata. Tutti gli altri *forward* coincidono con l'inizio del promotore. Per ogni coppia di *primers* sono state condotte almeno due reazioni di amplificazione. I prodotti amplificati sono stati direttamente controllati mediante sequenziamento prima di essere clonati.

Tabella 6. Elenco dei *primers* e relative caratteristiche.

Promotore	Nome	Sequenza 5'>3'	Tm °C
Glutelina B1	GluB1 <i>for</i>	TTCTTGCTACCAACAACCTTCACA	60
	GluB1 <i>rev</i>	CTCGAGAAAGCATCTTGGGATACA	61
Globulina 26kDa	Glb <i>for</i>	CCTGGAGGGAGGAGAGG	58
	Glb <i>rev</i>	CTCGAGGTGGCCATGGTT	62
Prolamina 10kDa	Prol10 <i>for</i>	TCAGTTAAAATTGAAAATAATGCAA	57
	Prol10 <i>rev</i>	CTCGAGTGGTAGCTGTGTAAG	55
Prolamina 16kDa	Prol16 <i>for</i>	GTGCTACGCTGGGTTAATTAG	57
	Prol16 <i>rev</i>	CTCGAGATAGCATGATCTCTATG	56
Glutelina C	GluC <i>for</i>	GCGTTCAAGATTTATTTTGGTA	58
	GluC <i>rev</i>	CTCGAGGTAGACGAGAATATATGC	59

4.4.4 Sequenziamento degli amplificati

Gli amplificati ottenuti sono stati recuperati da gel: 5 µL del purificato da gel (10/20 ng/µL) sono stati uniti a 1.6 µL di opportuno *primer* ad una concentrazione di 2 µM e conferiti a IGA Service per il servizio di sequenziamento.

Per il sequenziamento dei promotori sono stati utilizzati *primers* che appaiano all'interno delle sequenze, a intervalli tali da coprire l'intera sequenza. Le sequenze ottenute sono state confrontate con quelle depositate in banca dati utilizzando *BioEdit*. Le sequenze ottenute che differivano da quelle attese sono state riamplicate e risequenziate per confermare le differenze rilevate. Le varie sequenze dei promotori sono

state poi clonate in pGEM-T come descritto nel capitolo 4.2. Le colonie ottenute sono state analizzate per stabilire l'orientamento d'inserzione. I cloni con l'orientamento desiderato sono stati risequenziati per conferma.

Tabella 7. *Primers* utilizzati per il sequenziamento.

Nome	Sequenza 5'>3'	Tm °C
GluB-1 2.3kb Rev2	GCTTTTCCTTGGTCGTTTTG	60
GluB-1 2.3kb Rev3	TTACACATCTGACCACGGAGA	59
GluB-1 2.3kb For2	ATTCACATAGAGCGGGCAGA	61
GluB-1 2.3kb Rev4	ATGGTTTCTCGGTCTTGCTC	59
GluB-1 2.3kb Rev5	TCAGACAGTAAGGCATATTGGTG	59
GluB-1 2.3kb For3	CAATATGCTTTATTGTCCGGTA	57
GluB-1 2.3kb For4	GATGTGGCCTAGGACGAGTG	61
Glb-1 26kDa For2	CATATATCCGTGGGCTTCCA	60
Glb-1 26kDa rev2	TCTTTGTGACGTGGTGGAAA	60
Glb-1 26kDa rev3	CGGGTCCCTCAACTATCAAA	60
GluC Rev2	GACTGGTGGCATGCATCTAT	59
GluC Rev3	TGCAATCAGCAGCAAGGTA	59
GluC For2	TTTGAGCAAATGTGGCTTTG	60
GluC Rev4	GGTCACAGATAAATCCGCAAG	60
GluC For3	GACACTCTCTCACGGCACTCT	60
Prol 10kDa For2	TTTCCCTTTACAACAATGTTCTTTA	60
Prol 16kDa For2	GGGCAATGCTTTTGTGAAAT	60
Prol 16kDa For_bis	TGGATCAACTAGAGTAGTTAAACGAG	57

4.4.5 Inversione dell'orientamento di inserzione dei promotori di GluB4 e Glb

Il plasmide pGEM-T contenente l'inserto da invertire è stato digerito con *EagI*. Vettore e inserto sono stati separati su gel di agarosio. Le bande corrispondenti alla lunghezza attesa sono state recuperate e purificate come descritto nel capitolo 4.2. I prodotti ottenuti sono stati legati e la miscela di reazione è stata utilizzata per la trasformazione di *E. coli* come descritto nel capitolo 4.2.2. Le colonie ottenute sono state saggiate per verificare l'orientamento dell'inserto mediante PCR.

4.5 Costruzione del modulo accettore

4.5.1 TATA-box e leader

Per inserire un sito *Xho*I a monte della TATA-box sono stati usati i seguenti *primers*:

Tabella 8. *Primers* per l'amplificazione del tratto contenente TATA-box e leader

Nome	Sequenza 5'>3'	Tm°C
TATA <i>for</i>	CTCGAGCTTCACTATATAAGTTGGCTTC	63
TATA <i>rev</i>	GAATTCATGCAGCAGGATGTG	62

Come stampo è stato usato un plasmide contenente il promotore della Glutelina B4, il *leader* sintetico LLTCK, e la sequenza codificante per l'alfa-glucosidasi acida umana (hGAA) ottimizzata per ottenere un'elevata espressione in endosperma di riso (sequenza nativa GenBank NM_000152). Tale plasmide è già stato impiegato in precedenti lavori di clonazione presso il gruppo di ricerca.

Il *primer_for* aggiunge un sito *Xho*I al 5', mentre il *primer_rev* inserisce un sito GAATTC all'interno della sequenza di GAA. Di seguito si riporta la sequenza dell'amplificato atteso:

```
CTCGAGCTTCACTATATAAGTTGGCTTCCCTGTCATTGAACTCACATCAACCAGCCCA  
ACACGTATTTTTACAACAATACCAACAACAACAACAACAACAACATTACAATTACGT  
ATTTCTCTCTCTAGAAATG GCCACCATTGCGTTCTCCCGGCTGTCCATCTACTTCTGCG  
TGCTGCTGCTGTGCCACGGCTCCATGGCCGGCCACATCCTGCTGCATGAATTC
```

Il prodotto di amplificazione è stato clonato in pGEM-T ed è stato verificato l'orientamento di inserzione. Il prodotto pGEM-T-TATA-LLTCK è stato controllato mediante sequenziamento.

4.5.2 Inserimento del terminatore

Per la realizzazione del modulo accettore è stato utilizzato il terminatore del gene per la nopalina sintasi (Depicker *et al.*, 1982). Il tratto corrispondente al terminatore era già disponibile come inserto clonato in un vettore pUC18-NOS tra i siti *Sac*I ed *Eco*RI. Di seguito è riportata la sequenza inserita nel MCS:

XbaI-BamHI-SmaI-KpnI-SacI-GAATTTCCCCGATCGTTCAAACATTT
GGCAATAAAGTTTCTTAAGATTGAATCCTGTTGCCGGTCTTGCGATGATTATCATATA
ATTTCTGTTGAATTACGTTAAGCATGTAATAATTAACATGTAATGCATGACGTTATTT
ATGAGATGGGTTTTTATGATTAGAGTCCCGCAATTATCATTTAATACGCGATAGAAAA
CAAAATATAGCGCGCAAACACTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTA
GATCGG-*EcoRI*

L'inserto relativo al terminatore è stato recuperato da pUC18-NOS mediante digestione con gli enzimi *XbaI* ed *EcoRI* e successivamente clonato all'interno di pGEM-T-TATA-LLTCK, aperto con i medesimi enzimi, generando pGEM-T-TATA-LLTCK-NOS.

4.5.3 Inserimento di un secondo *EcoRI* a monte della cassetta di espressione

Per rendere possibile la movimentazione della cassetta di espressione nel vettore finale (pCAMBIA) è stato necessario inserire un secondo sito *EcoRI* a monte della cassetta di espressione. Questa operazione è stata realizzata sostituendo parte del MCS con un inserto sintetico composto da due oligonucleotidi fosforilati complementari:

Tabella 9. Coppia di oligonucleotidi utilizzati per aggiungere un sito *EcoRI* nel MCS

Nome	Sequenza 5'>3'
<i>EcoRI-for</i>	CGAATTCGACGTGCGATGCTCCCGGCCGCCATGGCCGCGGGATTC
<i>EcoRI-rev</i>	TCGAGAATCCCGCGGCCATGGCGGCCGGGAGCATGCGACGTCAATTCGGGCC

I due oligonucleotidi sono stati mescolati in proporzione equimolare e sono stati appaiati mediante incubazione a 95°C per 45 sec seguita da un progressivo raffreddamento della miscela fino a raggiungere la temperatura ambiente. L'inserto così ottenuto è stato purificato in gel di agarosio 1,5%. Il purificato è stato clonato in pGEM-T-TATA-LLTCK-NOS, tra i siti di restrizione *ApaI* e *XhoI*. Il prodotto pGEM-T-2x*EcoRI*-TATA-LLTCK-NOS è stato controllato mediante PCR per escludere possibili casi di inserzione multipla.

4.5.4 Inserimento del gene *reporter*

Nel presente lavoro sono stati utilizzati due geni *reporter* codificanti due enzimi lisosomiali umani attualmente utilizzati come biofarmaci nella

terapia enzimatica sostitutiva di malattie rare. In un esperimento iniziale come gene *reporter* è stato usato il gene per l'alfa glucosidasi acida (GAA), la cui mutazione difettiva è causa della glicogenosi di tipo II. In tutto il resto del lavoro è stato, invece, impiegato il gene per la beta glucosidasi acida umana, in breve GCasi (GenBank D16328); la riduzione parziale o totale nell'attività enzimatica della GCasi è responsabile dell'insorgenza della malattia di Gaucher. In entrambi i geni, per favorire l'espressione in riso, la sequenza codificante il peptide segnale (PS) naturale è stata sostituita con quella relativa alla glutelina 4 (PS Glub4) ottimizzata per un'elevata espressione endosperma-specifica.

ATG GCC ACC ATT GCG TTC TCC CGG CTG TCC ATC TAC TTC TGC
GTG CTG CTG CTG TGC CAC GGC TCC ATG GCC GCC CGC CCC...

Per inserire il gene GCasi non è stato possibile utilizzare gli enzimi *XbaI* e *SacI*, poiché quest'ultimo è già presente nel MCS di pGEM-T. Si è deciso, quindi, di partire da un plasmide che già conteneva la CDS di GCasi legata al terminatore NOS. Mediante digestione con *XbaI* e *AflII* è stato ottenuto l'insero contenente la GCasi seguita dalla porzione non tradotta in 3', e un frammento di NOSTer. L'insero è stato clonato all'interno di pGEM-T-2x*EcoRI*-TATA-LLTCK-NOS, tra *XbaI* e *AflII*. Il modulo accettore pGEM-T-2x*EcoRI*-TATA-LLTCK-PSGluB-4-GCasi-NOS è stato sequenziato prima di essere assemblato con i vari promotori sintetici.

4.6 Costruzione di pCAMBIA-ΔGluC-GAA

Il plasmide pGEM-T GluC è stato digerito con *SphI* e *XhoI* per ottenere il promotore della Glutelina C privo delle prime 1043 bp (ΔGluC). L'insero è stato clonato in pGEM TATA LLTCK PSGluB-4 GAA NOSTer, tra *SphI* e *XhoI*. La cassetta di espressione è stata subclonata in pUC18 tra *SphI* ed *EcoRI* ottenendo il plasmide pUC18-ΔGluC-GAA. Tale vettore è stato sottoposto a digestione enzimatica con gli enzimi di restrizione *HindIII* ed *EcoRI* al fine di recuperare la cassetta di espressione e consentirne l'inserimento nel vettore pCAMBIA opportunamente aperto con i medesimi enzimi.

4.7 Costruzione dei promotori sintetici

In questo capitolo verrà introdotta una nomenclatura specifica per designare i promotori ibridi in funzione dei promotori da cui sono state tratte opportune regioni. Qui di seguito sono elencati i nomi abbreviati dei promotori naturali utilizzati.

Tabella 10. Lista delle abbreviazioni utilizzate per costruire i nomi dei promotori sintetici composti da più sezioni di promotori naturali.

Abbreviazione	Proteina/subunità	N° accessione
P10	Prolamina 10kDa	AY427572
P16	Prolamina 16kDa	AY427574
B1	Glutelina B-1	AY427569
B4	Glutelina B-4	AY427571
C	Glutelina C	EU264107
Glb	Globulina 26kDa	AY427575

4.7.1 Derivati di Glb

L'inserto Glb è stato prodotto digerendo il plasmide pGEM-T/Glb-r con *SpeI* e *XhoI*.

NOTA: per indicare il fatto che il promotore è inserito in senso inverso nel MCS verrà da qui in poi utilizzato un suffisso **-r** (Es. P10-r, B4-r).

Tabella 11. Enzimi utilizzati per la costruzione dei derivati del promotore della globulina

Vettore	Tra i siti	Nome prodotto	Lunghezza
P16	<i>XhoI XbaI</i>	P16-Glb	1427
B1	<i>XhoI BlnI</i>	B1-Glb	2124
C	<i>XbaI XhoI</i>	C-Glb- <u>1278</u>	1278
C	<i>NheI XhoI</i>	C-Glb- <u>2695</u>	2695

NOTA: i suffisso **-1278** e **-2695** indicano la lunghezza in bp del prodotto ottenuto. Questa notazione è stata usata in tutti quei casi in cui la nomenclatura basata sulle sezioni dei promotori naturali (§Tabella 10) non è sufficiente a risolvere ambiguità fra due promotori sintetici.

4.7.2 Derivati di Prol10

I 2 derivati di P10 sono stati prodotti aprendo il vettore pGEM-T/Prol10 con *EcoRI* e *XhoI*.

Tabella 12. Enzimi utilizzati per la costruzione dei derivati del promotore della prolamina 10kDa

Inserto	Digerito con	Lunghezza	Nome
Glb	<i>ApoI XhoI</i>	993	P10-Glb
C	<i>ApoI XhoI</i>	882	P10-C

4.7.3 Derivati di GluB4

Nella seguente tabella viene illustrato come sono stati costruiti i 5 derivati sintetici a partire dal vettore pGEM-T/GluB4::LLTCK:

Tabella 13. Enzimi utilizzati per la i costruzione dei derivati del promotore della glutelina B4

Inserto	Digerito con	Clonato fra	Lunghezza	Nome
B1	<i>SphI BsaAI</i>	<i>SphI EheI</i>	2181	B1-B4
Glb-r	<i>ApoI HindIII</i>	<i>EcoRI HindIII</i>	921	Glb-B4
C	<i>AatII ApoI</i>	<i>AatII EcoRI</i>	2592	C-B4
C-Glb	<i>SphI ApoI</i>	<i>SphI EcoRI</i>	2180	C-Glb-B4
C-Glb-2695	<i>AatII HindIII</i>	<i>AatII HindIII</i>	3087	C-Glb-B4-3087

4.8 Assemblaggio della cassetta di espressione e del vettore di trasformazione

4.8.1 Cassetta di espressione

Tutti i promotori che hanno il sito *XhoI* all'estremità 3' sono stati inseriti nel modulo accettore utilizzando questo stesso sito congiuntamente a *SphI* o *AatII* a seconda della disponibilità di siti.

I derivati di GluB4, che sono privi di *XhoI*, ma già fusi al LLTCK, sono stati movimentati nel modulo accettore utilizzando *XbaI* come sito in 3'. Quando *XbaI* era presente all'interno del promotore si è ricorsi a *SnaBI* che si trova 14 bp a monte di *XbaI*. Segue una tabella che riassume la strategia di assemblaggio utilizzata:

Tabella 14. Siti di restrizione utilizzati per inserire i vari promotori nella cassetta di espressione

Promotore	Siti utilizzati
Glb	<i>SphI XhoI</i>
C	<i>AatII XhoI</i>
P16-Glb	<i>SphI XhoI</i>
B1-Glb-	<i>SphI XhoI</i>
C-Glb	<i>SphI XhoI</i>
C-Glb	<i>AatII XhoI</i>
B1-B4	<i>SphI XbaI</i>
Glb-B4	<i>SphI XbaI</i>
C-B4	<i>AatII SnaBI</i>
C-Glb-B4	<i>SphI XbaI</i>
C-Glb-B4-3087	<i>AatII SnaBI</i>
P10-Glb	<i>SphI XhoI</i>
P10-C	<i>SphI XhoI</i>

4.8.2 Vettore di trasformazione

Tutte le cassette di espressione assemblate in pGEM-T sono state inserite nel vettore finale pCAMBIA-PMI per la trasformazione stabile di riso servendosi della coppia di siti *EcoRI* presenti ai terminali 5' e 3' della cassetta di espressione. Le colonie ottenute dalla trasformazione di *E. coli* con il vettore finale sono state saggiate in PCR per individuare quelle con orientamento corretto.

Tabella 15. *Primers* utilizzati per determinare l'orientamento di inserzione

Nome	Sequenza	Tm °C
NOS <i>for</i>	TCGTTCAAACATTTGGCAAT	58
Space <i>rev</i>	CAGGAAACAGCTATGACCATG	60
RBR <i>rev</i>	GCCTGTCGCGTAACTTAGGA	58

4.9 Trasformazione di *Oryza sativa* mediante *Agrobacterium tumefaciens*

Per la trasformazione di riso, varietà CR W3, è stato utilizzato il protocollo di Hiei *et al.* (1994), modificato da Hoge (Rice Research Group, Institute of Plant Science, Leiden University) e Guiderdoni (programma

Biotrop, Cirad, Montpellier, France) fino all'ottenimento dei calli trasformati. Per la successiva fase di selezione è stato applicato, invece, il protocollo di Datta e Datta (2006). Qui di seguito sono brevemente descritte le principali fasi dell'intera procedura.

4.9.1 Preparazione e sviluppo di calli embriogenici da scutello di riso

La trasformazione di riso è avvenuta su calli embriogenici derivati da scutello.

Per indurre la proliferazione di calli da tessuto scutellare è stato utilizzato il seguente protocollo operativo:

- è stata eseguita la sbramatura (eliminazione delle glume) di 200 semi di riso;
- per eliminare potenziali patogeni e saprofiti contaminanti che potrebbero interferire con la coltura di calli, si è proceduto quindi alla disinfezione delle cariossidi private delle glume:
 - il primo trattamento di disinfezione ha previsto la permanenza dei semi sbramati per 2 min in una soluzione di etanolo al 70%;
 - dopo il passaggio in etanolo, i semi sono stati trasferiti in una soluzione di ipoclorito di sodio al 5% con 2 gocce di detergente Tween-20 e ivi mantenuti in agitazione lenta per 30 min;
 - per eliminare ogni traccia di ipoclorito di sodio che avrebbe potuto inibire l'induzione a callo degli scutelli, sono stati eseguiti una serie di lavaggi in H₂O sterile della durata di 15 min ciascuno;
- dopo aver effettuato l'ultimo lavaggio, i semi sono stati asciugati su carta bibula sterile;
- sulla superficie del substrato per l'induzione a callo (CIM, *callus-induction medium*), dispensato nel volume di 25 mL all'interno di piastre Petri (Ø 90 mm), sono stati posizionati 12 semi per piastra;
- le piastre così ottenute sono state incubate al buio, a una temperatura di 28°C per 21 giorni; dopo 1 settimana di incubazione

si è proceduto all'eliminazione dell'endosperma e della radichetta per favorire lo sviluppo del callo proveniente dallo scutello (lo scutello si riconosce per la sua massa compatta, parzialmente inclusa nell'endosperma di colorazione gialla);

- terminate le 3 settimane di induzione, si è operato il trasferimento del callo su substrato CIM rinnovato, a cui ha fatto seguito la frammentazione delle masse callose senza l'utilizzo di bisturi, seguendo le linee di fratturazione naturalmente presenti sul callo;
- la sub-coltura è stata fatta proseguire per altri 10 giorni in modo da sviluppare il callo embriogenico e renderlo idoneo alla trasformazione.

4.9.2 Co-coltura dei calli con *A. tumefaciens*

Per ottenere quantità sufficienti di *A. tumefaciens* per la trasformazione, i ceppi portanti il plasmide contenente la cassetta di interesse e il gene per la fosfomannosio isomerasi sono stati incubati per 3 giorni a 30°C in LB-agar.

Ottenute le colture di agrobatterio, le relative patine di crescita batterica sono state prelevate e sospese nel mezzo liquido di co-coltivazione (CCML, *co-cultivation medium liquid*), fino a ottenere una O.D.₆₀₀ di circa 1.0, corrispondente a $3-5 \cdot 10^9$ cellule/mL.

I calli migliori, cioè quelli con un diametro di circa 2 mm, compatti e dal colore tendente al bianco, sono stati trasferiti in una piastra Petri contenente 35 mL di sospensione batterica e lasciati in immersione per 15 min in agitazione; si è proseguito quindi all'asciugatura del callo utilizzando carta bibula sterile.

E' stato disposto un numero massimo di 20 calli per piastra Petri high-edge (Sarstedt) contenente il substrato solido per la co-coltura (CCMS, *co-cultivation medium solidified*).

I calli sono stati quindi incubati in ambiente buio, a una temperatura di 25°C per 3 giorni.

4.9.3 Selezione dei calli basata su PMI

Dopo aver effettuato la co-coltura dei calli embriogenici di riso con agrobatterio, si è proceduto alla selezione dei tessuti trasformati sfruttando il sistema di selezione positiva basato su PMI (fosfomannosio isomerasi) come *marker* selezionabile e mannosio come agente selettivo. Tale metodo prevede l'impiego di substrati di coltura contenenti concentrazioni crescenti di mannosio e concentrazioni decrescenti di saccarosio.

La procedura impiegata è stata la seguente:

- trasferimento dei calli provenienti dalla co-coltura con *A. tumefaciens* su substrato PSM (*pre-selection medium*) privo di mannosio e contenente il 3% di saccarosio; incubazione per 1 settimana al buio a una temperatura di 28°C;
- trasferimento dei calli su substrato di selezione SMI (*selection medium I*) contenente 2% saccarosio e 1.5% mannosio e incubazione al buio per 2 settimane a una temperatura di 28°C;
- trasferimento dei calli su substrato di selezione SMII (*selection medium II*) contenente 1% saccarosio e 2% mannosio e incubazione al buio per 2 settimane a una temperatura di 28°C;
- segue rigenerazione.

4.9.4 Rigenerazione di piantine di riso da calli trasformati

Dopo che il T-DNA portato dall'agrobatterio ha avuto modo di inserirsi in maniera stabile nel genoma di riso, si è proceduto alla rigenerazione delle piantine putativamente trasformate. Questo processo è avvenuto grazie a un'opportuna stimolazione ormonale del callo trasformato seguendo la procedura qui riportata:

- i calli embriogenici di riso selezionati sono stati trasferiti in piastre Petri *high-edge*, su substrato per la pre-rigenerazione PRM (*pre-regeneration medium*) contenente 0.5% saccarosio e 2.5% mannosio e incubazione al buio per 2 settimane a una temperatura di 28°C;

- dopo il passaggio sul substrato PRM i calli sono stati trasferiti sul substrato per la rigenerazione RM (*regeneration medium*), privo di mannosio, nel numero massimo di 8-10 unità per piastra Petri *high-edge*. La crescita delle piantine è avvenuta in presenza di luce, a 28°C per 3-4 settimane;
- quando le piantine sono risultate sufficientemente grandi da poter essere separate dal callo (≥ 3 cm di altezza), si è proceduto al loro trasferimento in tubi di coltura contenenti 25 mL del substrato per la radicazione rm (*rooting medium*);
- la sub-coltura all'interno dei tubi è proseguita per circa 3 settimane sempre a 28°C alla luce;
- a conclusione del processo rigenerativo, le piante sono state trasferite in torba e allevate in serra.

Tabella 16. Composizione di ciascun substrato impiegato nel processo di trasformazione genetica mediato da *Agrobacterium tumefaciens*.

Composizione	CIM	CCML	CCMS	PSM	SMI	SMII	PRM	RM	rm
N6 Macroelementi I (mL)	50				50	50	50	50	
N6 Macroelementi II (mL)	50				50	50	50	50	
MS FeNaEDTA (mL)	10				10	10	10	10	
B5 Vitamine (mL)	10				10	10	10	10	
B5 Microelementi (mL)	1				1	1	1	1	
Prolina (mg)	500				500	500	500	500	
Glutammina (mg)	500				500	500	500	500	
CEH (mg)	300				300	300	300	300	
MES (mg)	500				500	500	500	500	500
RA (mL)		25	25	25					
RB (mL)		25	25	25					
RC (mL)		25	25	25					
Tiamina (mg)		1	1	1					0,1
2,4-D (mg)		2,5	2,5	2,5	2,5	2,5			
Glucosio (g)		10	10						
MES (mg)		500	500	500					
Acetosiringone (mM)		0,1	0,1						
Agarosio SPI (g)			7	7	7	7			
Phytigel (g)							4,5	4,5	2,5
Saccarosio (g)	30			30	20	10	5	30	50
Mannosio (g)					15	20	25		
Cefotaxime (mg)				400	400	400	250	50	
ABA (mg)							5		
BAP (mg)							2	3	
NAA (mg)							1	0,5	
MS Sali (mL)									100
Glicina (mg)									2
Acido nicotinico (mg)									0,5
Piridossina (mg)									0,5
Inositolo (mg)									100
H ₂ O	q.b. a 1 L								

Tabella 17. Dettaglio della composizione degli ingredienti dei substrati elencati in Tabella 16.

N6 macroelementi II	Dose per 1 L
CaCl ₂ ·2H ₂ O	3,32 g

N6 macroelementi I	Dose per 1 L
KNO ₃	56,60 g
(NH ₄) ₂ SO ₄	9,26 g
KH ₂ PO ₄	8,00 g
MgSO ₄ ·7H ₂ O	3,70 g

B5 microelementi	Dose per 1 L
MnSO ₄ ·H ₂ O	10 g
KI	0,75 g
H ₃ BO ₃	3 g
Zn ₄ ·7H ₂ O	2 g
CuSO ₄	0,025 g
Na ₂ MoO ₄ ·2H ₂ O	0,25 g
CoCl ₂ ·6H ₂ O	0,025 g

B5 Vitamine	Dose per 1 L
Inositolo	10 g
Tiamina-HCl	1 g
Acido nicotinico	0,1 g
Piridossina-HCl	0,1 g

Composizione R _A	Dose per 1 L
KNO ₃	162 g

Composizione R _C	Dose per 1 L
CaCl ₂ ·2H ₂ O	6 g
H ₃ BO ₃	114,4 mg
Na ₂ MoO ₄ ·2H ₂ O	5,2 mg
FeSO ₄ ·7H ₂ O	496 mg
Na ₂ EDTA·2H ₂ O	668 mg

MS FeNaEDTA	Dose per 1 L
FeSO ₄ ·7H ₂ O	2,784 g
Na ₂ EDTA·2H ₂ O	3,724 g

Composizione MS sali	Dose per 1 L
MnSO ₄ ·H ₂ O	16,9 mg
CuSO ₄ ·5H ₂ O	0,025 mg
ZnSO ₄ ·7H ₂ O	8,63 mg
CaCl ₂ ·2H ₂ O	440 mg
KH ₂ PO ₄	170 mg
KI	0,83 mg
NH ₄ NO ₃	1,650 g
KNO ₃	1,9 g
NaMoO ₄ ·2H ₂ O	0,25 mg
EDTAFeNa	40 mg
H ₃ BO ₃	6,2 mg
MgSO ₄ ·7H ₂ O	370 mg
CaCl ₂ ·6H ₂ O	0,025 mg

Composizione R _B	Dose per 1 L
MgSO ₄ ·7H ₂ O	10 g
(NH ₄) ₂ SO ₄	13,2 g
NaH ₂ PO ₄ ·H ₂ O	11 g
ZnSO ₄ ·7H ₂ O	88 mg
MnSO ₄ ·H ₂ O	80 mg
CuSO ₄ ·5H ₂ O	8 mg

4.9.5 Condizioni di coltivazione in cella climatica e in serra

Ciascuna pianta è stata trapiantata in vaso avente capacità di 7 L. Le piante sono state allevate in serra con una temperatura minima impostata a

20°C. Il substrato di coltura è costituito da torba bionda svedese media (ALCE, Agrochimica) e bentonite (50kg/m³) miscelate in ragione di 1 m³ di torba ogni 100kg di bentonite, e concimate opportunamente. Il ph del substrato è stato portato a 6 con CaCO₃

4.10 Verifica della presenza del gene GAA o GCasi nelle piante rigenerate

Per verificare la presenza del gene GAA o GCasi nelle piante di riso putativamente trasformate, è stato estratto DNA genomico da foglia come descritto nel paragrafo 4.2.1; tale DNA è stato successivamente sottoposto ad analisi PCR.

4.10.1 PCR di controllo per saggiare l'amplificabilità del DNA estratto

Per verificare se il DNA estratto fosse effettivamente amplificabile e dunque prevenire errate o dubbie interpretazioni dei risultati ottenuti con PCR, è stata programmata l'amplificazione di un tratto di DNA presente a livello di cpDNA di riso.

Tabella 18. *Primers* utilizzati per saggiare l'amplificabilità del DNA genomico estratto.

Nome	Sequenza	Tm °C
CPC	CGAAATCGGTAGACGCTACG	60
CPD	GGGGATAGAGGGACTTGAAC	58

4.10.2 PCR per l'amplificazione del gene GAA e del gene GCasi

Per verificare la presenza dei gene *reporter* GAA e GCasi le piante sono state saggiate tramite PCR.

Tabella 19. *Primers* utilizzati per l'amplificazione dei geni *reporter*.

Nome	Sequenza	Tm °C
GAA <i>for</i>	GTCCATCTACTTCTGCGTGC	60
GAA <i>rev</i>	AGGTAGAAGGGATGGGAGC	62
hGCasi <i>for</i>	TCTGAGGCAAAGAGCATGG	63
hGCasi <i>rev</i>	GCTGCTGCTCTCAACATCC	62

4.11 Analisi proteica dei trasformati primari

4.11.1 Estrazione di proteine totali da semi GAA e GCasi

I semi di riso trasformati sono stati sottoposti a sbramatura e, quando desiderato, a sbiancatura meccanica con apposita sbiancatrice Hercules (Baragioli, Vercelli).

Per la quantificazione del contenuto di enzima ricombinante nell'endosperma tramite analisi DAS-ELISA, le proteine totali sono state estratte da semi non sbiancati.

Per ottenere estratti di proteine totali da saggiare mediante DAS-ELISA è stato messo a punto un protocollo di estrazione che ha previsto le seguenti fasi:

- Raccolta di 1-2 pannocchie mature da ciascun individuo.
- Essiccazione delle pannocchie in locale asciutto e areato per circa 3 giorni fino al raggiungimento di una umidità relativa del seme pari al 14%.
- Campionatura casuale di 40 semi per ciascuna linea.
- Sbramatura dei semi con sbramino manuale da tavolo per riso.
- Macinazione del campione con mulino a sfere Retsch MM2 alla velocità 30 per 2 minuti nel caso di seme non sbiancato; 20 per 2 minuti nel caso di riso sbiancato.
- Prelievo di 70 mg della farina ottenuta.
- Omogeneizzazione in mortaio della farina con 1 mL di opportuno tampone di estrazione scelto in funzione della proteina di interesse da saggiare.
- Successiva diluizione con ulteriori 7 mL del medesimo tampone.
- Incubazione in agitazione continua secondo le seguenti condizioni:
 - 4°C per 1 h nel caso di estratti GCasi
 - 15 minuti a RT per i campioni GAA
- Prelievo di 1 mL e centrifugazione a 20000xg per 40 min a 4°C

- Recupero della fase liquida contenente le proteine e conservazione a -20°C.

I tamponi di estrazione sono stati i seguenti:

- Tris-HCl 50 mM, NaCl 0.5 M pH=7,0 per l'estrazione di proteine totali da semi GCasi
- Idrogenofosfato disodico 250 mM + NaCl 250 mM pH=7,4 per l'estrazione di proteine totali da campioni GAA.

4.11.2 Saggio immunoenzimatico (DAS-ELISA)

Il saggio DAS-ELISA si basa su un doppio riconoscimento immunologico. Un anticorpo specifico per la proteina che si vuole rilevare, chiamato *capturing*, viene ancorato al fondo dei pozzetti di una piastra di polistirene (reazione di *coating*). Terminata la reazione di *coating*, ciascun pozzetto viene messo a contatto con una soluzione di BSA per saturare la residua capacità legante del polistirene (*blocking*). Dopo una serie di lavaggi, i campioni che si suppongono contenere la proteina di interesse vengono seminati nei pozzetti. Dopo successivi lavaggi, un secondo anticorpo, specifico per la proteina di interesse, coniugato a un enzima in grado di catalizzare una reazione che permette di rilevare la presenza della proteina, viene dispensato in tutti pozzetti. Dopo i lavaggi, il complesso *capturing*-proteina-coniugato viene rilevato tramite contatto con una soluzione contenente un adeguato substrato cromogenico, come ad es. il 3,3',5,5'-tetramethylbenzidine (TMB).

L'anticorpo di *capturing* è stato prodotto dalla purificazione su proteina A dell'antisiero di coniglio anti-Myozyme[®] o dalla purificazione per affinità dell'antisiero di coniglio anti-Cerezyme[®] a seconda che la proteina di interesse da rilevare fosse rispettivamente GAA o GCasi. Entrambi gli anticorpi sono stati prodotti dalla ditta Davids Biotechnologie.

L'anticorpo policlonale anti-GCasi o anti-GAA coniugato alla perossidasi di rafano (HRP) è stato pure fornito dalla ditta Davids Biotechnologie.

Il saggio è stato svolto nei seguenti passaggi:

- *Coating*: vengono dispensati in ciascun pozzetto 100 μL di PBS diluito 1:5, sodio azide 0,01% + anti-Myozyme[®] o anti-Cerezyme[®] purificato a una concentrazione finale rispettivamente di circa 4 ng/ μL e 2 ng/ μL . La piastra è incubata O/N a 4°C.
- *Blocking*: vengono dispensati in ciascun pozzetto 300 μL PBS + BSA 2,5% + sodio azide 0.01% e ivi mantenuti per almeno 20 minuti a RT prima di procedere al passaggio successivo.
- I campioni (estratti proteici grezzi contenenti GAA o GCasi) sono opportunamente diluiti in tampone di diluizione (PBS + Tween20 0,1% + BSA 1%). Una volta diluiti, i campioni sono seminati nel volume di 50 μL nei pozzetti e incubati per 30 min a 37°C in agitazione.
- Dopo successivi lavaggi, l'anticorpo anti-Myozyme[®] o anti-Cerezyme[®] coniugato a HRP viene opportunamente diluito in tampone di diluizione (stesso tampone utilizzato per i campioni) e seminato (50 μL) nei pozzetti. Dopo 30 minuti di incubazione a 37°C in agitazione, si procede con i lavaggi e con la rilevazione del segnale.
- Sviluppo del segnale: si impiega il substrato TMB, tetrametilbenzidina (Ultra-TMB, Pierce oppure SureBlue, Kpl) (100 μL per pozzetto) bloccando la reazione con 100 μL HCl 1 M, dopo circa 5 min di incubazione. L'assorbanza viene misurata a 450 nm utilizzando il lettore di piastra Modulus II (Promega) oppure Tecan.

Per calcolare la concentrazione della proteina di interesse nei campioni, in ciascun saggio è stata costruita una curva di taratura con concentrazioni note di Myozyme[®] o Cerezyme[®].

Per l'elaborazione dei dati si è proceduto all'immissione dei valori di assorbanza nel software *Curve Fitting Data Analysis* (Promega) e all'assegnazione dei valori di concentrazione nota degli standard. I valori di concentrazione dei campioni ottenuti come *output* del software utilizzando una curva lineare a quattro parametri sono stati successivamente elaborati considerando il fattore di diluizione adottato al fine di ottenere le concentrazioni reali degli estratti.

4.11.3 Validazione del saggio DAS-ELISA

Allo scopo di verificare la validità del metodo DAS-ELISA in termini di specificità e selettività, sia per GAA sia per GCasi, sono stati allestiti tre distinti test ELISA, ciascuno con il medesimo estratto proteico saggiato a tre diverse diluizioni. A ciascuna diluizione è stato successivamente aggiunto lo standard alla concentrazione nota di 0.8 ng/ μ L (Myozyme o Cerezyme a seconda che la molecola da quantificare fosse α - o β -glucosidasi). Il metodo DAS-ELISA è stato validato quando i risultati relativi ai campioni con aggiunta dello standard sono ricaduti all'interno dell'intervallo di validità stabilito, corrispondente a +/- 15% del valore atteso.

4.11.4 Determinazione del livello di espressione del gene *reporter*

Il livello di espressione del gene *reporter* è stato determinato mediante analisi di un campione costituito dal prodotto di macinazione di 40 semi T_2 prelevati da ciascun trasformato primario (T_1). Nel primo ciclo sperimentale il numero di piante saggiate è stato volutamente limitato a venti individui PCR-positivi, scelti a caso. Nelle altre due sessioni è stato misurato il livello di espressione di tutti gli individui PCR-positivi, che avessero dato almeno 40 semi. Ne deriva che le popolazioni saggiate sono numericamente diverse da promotore a promotore. Gli estratti relativi a ciascun trasformato primario sono stati diluiti 1:10 nel caso del *reporter* GAA e 1:5 nel caso di GCasi. Con questo rapporto di diluizione il segnale ELISA è sempre caduto entro l'intervallo di affidabilità del saggio. Quasi tutte le piante trasformate hanno dato un livello espressione rilevabile di proteina *reporter*.

4.12 Analisi di tessuto-specificità e profilo d'espressione

Per evidenziare eventuali perdite di tessuto-specificità o variazioni nel profilo di espressione, sono stati preparati estratti proteici da organi e tessuti diversi dall'endosperma: foglia, seme abortito, seme immaturo, glume. E' stato inoltre analizzata la frazione di farinaccio corrispondente al materiale di scarto risultante dal processo di sbiancatura (pericarpo, germe, strato aleuronico).

4.12.1 Estrazione di proteine totali da foglia

50 mg di foglia sono stati omogeneizzati in mortaio con 1 mL di tampone di estrazione (50 mM Tris-HCl 0,5 M NaCl pH7,0) a RT. L'omogenato così ottenuto è stato centrifugato alla massima velocità per 40 min a 4°C. Il surnatante è stato raccolto e aliquotato; una aliquota è stata diluita 1:10 in tampone di diluizione per ELISA mentre il restante estratto è stato conservato a -20°C previa quantificazione della proteina totale con metodo spettrofotometrico a 280nm (NanoDrop, ThermoScientific), ponendo per convenzione un coefficiente di estinzione molare pari a quello di un'immunoglobulina.

4.12.2 Estrazione di proteine totali da semi abortiti e semi immaturi.

Il materiale è stato raccolto in base ai seguenti criteri:

- Seme abortito: prelievo da una pannocchia con semi in stato di maturazione cerosa di spighe abortite.
- Seme immaturo: seme raccolto nel periodo che va dal termine della fase di botticella a 72 ore dall'emissione delle antere.

Per valutare il profilo di espressione del gene *reporter*, è stata eseguita un'estrazione di proteine totali da singoli semi. A tal scopo si è ricorsi a una diluizione tessuto/tampone di 1:20 utilizzando lo stesso tampone descritto per l'estrazione di proteine totali da foglia. Vista la difficoltà a reperire sufficiente materiale al giusto stadio di maturazione, il volume di estrazione è stato regolato di volta in volta in base alla disponibilità di seme. Le estrazioni sono state eseguite in eppendorf da 1,5 mL con micropestello. L'omogenato è stato centrifugato alla massima velocità per 20 min a 4°C e il surnatante recuperato. Gli estratti proteici ottenuti dai singoli semi sono stati diluiti in opportuno tampone (PBS, BSA 2%) e analizzati in DAS-ELISA. La quantificazione della proteina totale è stata eseguita con microspettrofotometro a 280nm (NanoDrop, ThermoScientific), come già descritto in precedenza (4.11.1).

4.12.3 Prove di sbiancatura

Per eseguire tali prove, sono stati utilizzati per ciascun promotore 10 g di seme proveniente da un campione di piante che presentavano un elevato livello di espressione. Il seme è stato essiccato come descritto nel paragrafo 4.11.1. Successivamente è stato privato delle glume, che sono state pesate e conservate; 40 semi sbramati sono stati prelevati e conservati. Circa 8 g di seme privato delle glume sono stati sottoposti a sbiancatura meccanica con apposita sbiancatrice.

Ogni campione è stato trattato per un tempo pari a 30 sec. Quindi il farinaccio e il seme sono stati recuperati smontando in ogni sua parte la camera di sbiancatura. Il materiale recuperato è stato pesato. I campioni di seme sono stati estratti come descritto nel paragrafo 4.11.1. Per determinare il livello di espressione della proteina *reporter* nel farinaccio, è stato adottato il seguente protocollo. Circa 200 mg di farinaccio recuperato a seguito del processo di sbiancatura sono stati prelevati e polverizzati finemente in mortaio; il materiale ottenuto è stato sottoposto a estrazione di proteine totali seguendo il medesimo protocollo utilizzato per la farina di riso (§ 4.11.1).

4.12.4 Estrazione di proteine totali dalle glume

Un campione omogeneo di 100 mg di glume è stato estratto in mortaio utilizzando una diluizione tessuto/campione di 1:20. L'omogenato così ottenuto è stato centrifugato alla massima velocità per 40 min a 4°C. Il surnatante è stato separato e una parte di esso è stata immediatamente diluita 1:4 in tampone ELISA 1% BSA.

4.12.5 Elaborazione dei dati ottenuti

Il livello di espressione in endosperma è stato espresso in termini di quantità di proteina *reporter* per unità di biomassa. Per biomassa si intende la farina da cui è stata effettuata l'estrazione (§4.11.1). Questo valore è stato calcolato a partire dalla concentrazione di proteina *reporter* nell'estratto proteico. Una semplice formula mette in relazione i due valori;

$$E=[reporter] \times C$$

dove E rappresenta il livello di espressione sulla biomassa e C il coefficiente di proporzionalità. Questo coefficiente è caratteristico della preparazione degli estratti. Nel caso del procedimento descritto in 4.11.1

$$C=114,3 \text{ mL/g}$$

Il livello di espressione negli altri tessuti è stato espresso come quantità di proteina *reporter* su proteina totale estratta. Questo valore è stato espresso in ppm, ovvero $\mu\text{g/g}$. Per quantificare il grado di tessuto-specificità, il livello di espressione nei diversi tessuti è stato rapportato al livello di espressione in endosperma. La quantità di proteina *reporter* presente per unità di farina di seme sbiancato e non sbiancato è stata calcolata come descritto precedentemente. La quantità di proteina *reporter* persa per rimozione dello strato aleuronico è stata ottenuta per determinazione della proteina *reporter* presente nel farinaccio recuperato.

5 RISULTATI

5.1 Mappatura degli elementi *cis*-regolatori

5.1.1 Elementi *cis*-regolatori nei promotori delle prolamine

Tutti i promotori delle prolamine sono abbastanza corti. L'analisi di questa famiglia di promotori è stata facilitata dal fatto che un membro di questa famiglia è stato utilizzato nello studio di due fattori di trascrizione, RISBZ1 e RPBPF (Yamamoto *et al.*, 2006). In forza di questo studio, sono stati identificati e localizzati con precisione gli elementi *cis*-regolatori riconosciuti da questi due fattori.

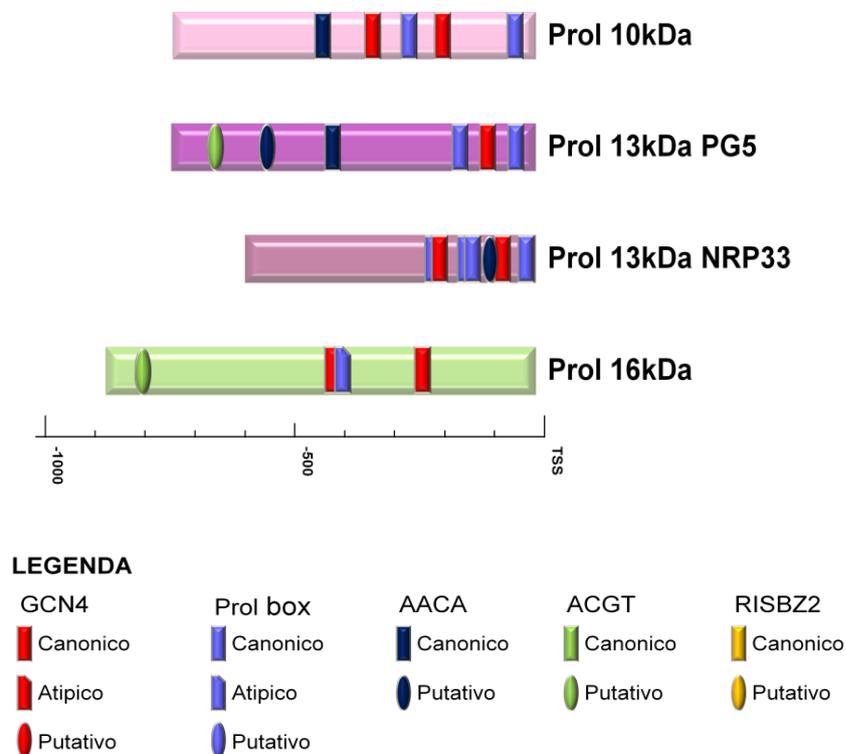


Figura 12. Mappa degli elementi *cis*-regolatori nei promotori delle prolamine.

In tutti i membri della famiglia si trova il motivo GCN4 nella sua forma canonica TGACTCA. Questo motivo si trova 333 nt a monte della TATA-box in Prol10, 92 nt in PG5, 69 nt in NRP33 e 410 nt e 230 nt in Prol16 (§Figura 12). E' poi presente un motivo equivalente TGAGTCA in NRP33 187 nt a monte della TATA-box. In prossimità dei motivi GCN4 in 5' di Prol16 e NRP33 si trovano motivi AAAG/CTTT che insieme a GCN4

formano la tipica *endosperm-box*. Altri elementi Prol-box sono disseminati lungo le sequenze di questi promotori. Non sono stati individuati motivi AACAA in Prol16. Sebbene vi siano vari motivi ACGT lungo PG5 e Prol16, solo uno di essi ha una sequenza che può essere ricondotta al consenso studiato in GluB1.

5.1.2 Elementi *cis*-regolatori nei promotori delle gluteline

La famiglia dei promotori delle gluteline comprende 9 sequenze, tutte molto lunghe. L'unica a non superare 2 kb è quella di GluB4. Il termine di riferimento che è stato utilizzato per l'analisi di queste sequenze è il promotore della glutelina B1, forse il più studiato fra i promotori delle PRSR.

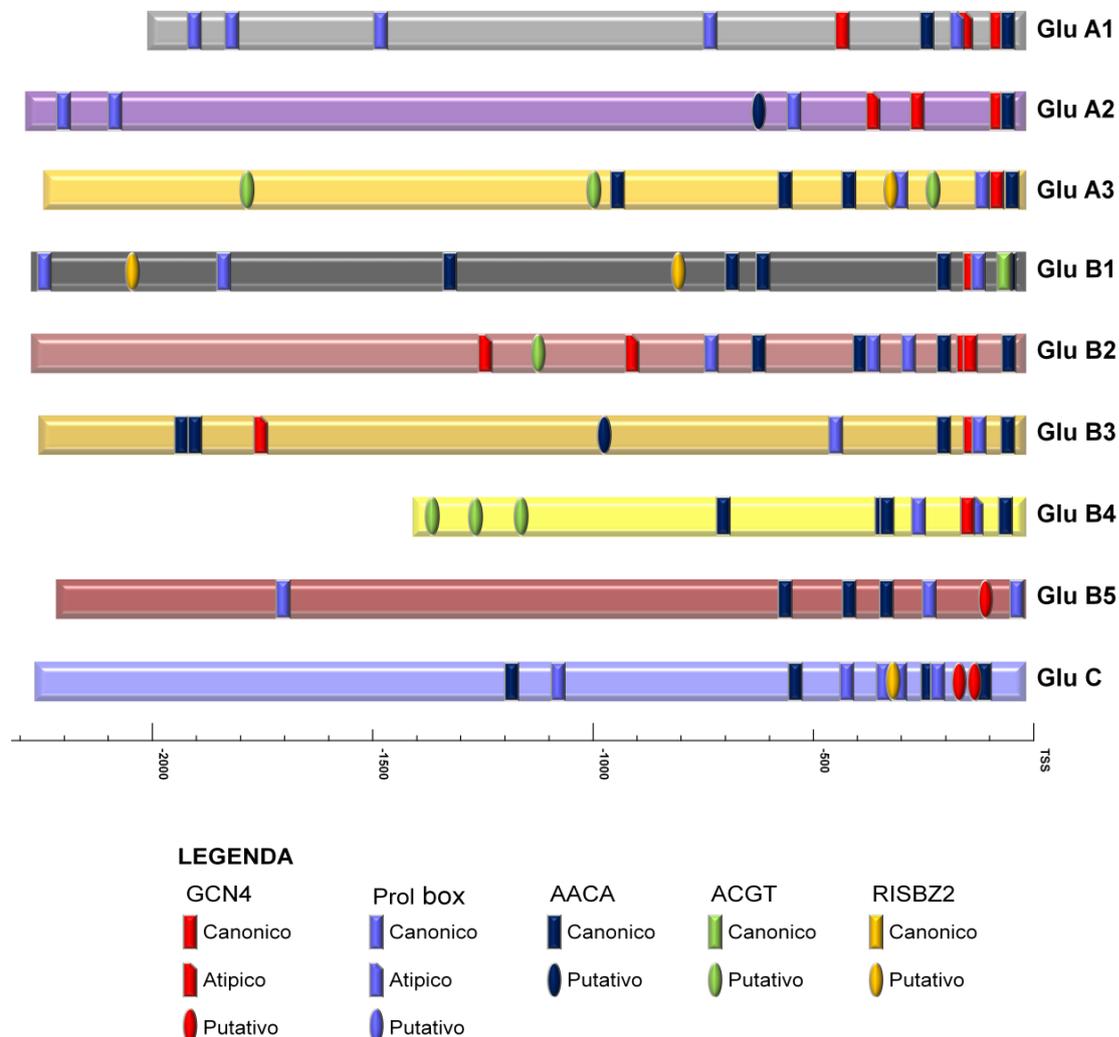


Figura 13. Mappa degli elementi *cis*-regolatori nei promotori delle gluteline.

I promotori della glutelina A1, A2 e A3 sono piuttosto simili fra loro e mostrano una significativa omologia nella loro porzione in 3'. Essi sono accomunati dalla presenza di un motivo AACCA 45 nt a monte della TATA-box e di un elemento GCN4 nella sua forma canonica TGAGTCA in posizione -70 nt. GluA3 si distingue per l'abbondanza di motivi AACCA, mentre gli altri 2 membri di questa sottofamiglia sono più ricchi di Prol-box ed elementi GCN4. Nella sottofamiglia GluB si nota una caratteristica disposizione dei motivi AACCA. Uno di questi elementi si trova circa 40 nt a monte della TATA-box, mentre un secondo circa 190 nt in 5'. In mezzo ad essi, in posizione -130 ca. si trova l'elemento GCN4. Differisce leggermente da questo schema il promotore della glutelina B4, in cui bisogna risalire fino a 322 nt a monte della TATA-box per trovare il secondo motivo AACCA. Il promotore GluB5 è un caso del tutto anomalo. Non presenta alcuna somiglianza con gli altri membri della sottofamiglia ed è persino difficoltoso individuare con certezza un singolo motivo GCN4. Prendendo in esame le regioni prossime all'estremità 5', si nota una serie di motivi ACGT putativi in GluB4 e un potenziale sito di legame per il fattore RISBZ2 in GluB1. Il promotore della glutelina C presenta, come gli altri membri della famiglia, un elemento GCN4 racchiuso fra due AACCA. Tutto è però traslato più a monte e la sequenza identificata come potenziale GCN4, è del tipo TGASYBM. Risalendo la sequenza si incontrano svariati motivi AAAG/CTTT e un potenziale sito di riconoscimento per RISBZ2.

5.1.3 Elementi *cis*-regolatori nel promotore della globulina 26kDa

L'analisi del promotore della globulina si è basata fondamentalmente su due pubblicazioni: Yamamoto *et al.* (2006) hanno indagato la capacità che i due fattori RISBZ1 e RPBF hanno di legare la regione prossimale del promotore; Nakase *et al.* (1997) hanno dimostrato che RISBZ2 riconosce in modo specifico un motivo GCCACGT MAG presente nella parte distale del promotore, circa 700 nt a monte della TATA-box. Questo è l'unico bersaglio di RISBZ2 finora citato in letteratura.

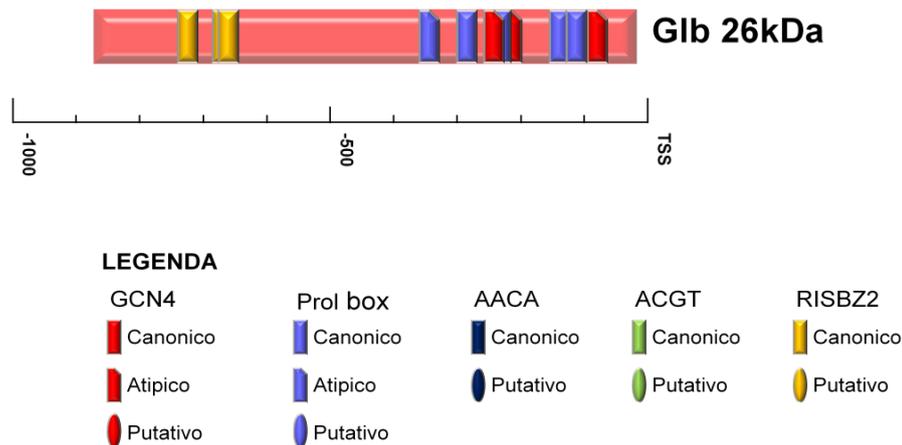


Figura 14. Mappa degli elementi *cis*-regolatori nel promotore della globulina.

Il promotore della globulina 26kDa possiede 3 elementi GCN4. Benché questi tre elementi abbiano sequenze atipiche, è stato dimostrato il loro riconoscimento da parte di RISBZ1 (Yamamoto *et al.*, 2006). I tre motivi GCN4 sono inframmezzati da varie Prol-box. Totalmente assente è il motivo AACA.

5.2 Scelta e clonazione dei promotori necessari al lavoro

5.2.1 Scelta dei candidati

Per la costruzione dei promotori sintetici è stato necessario scegliere alcuni promotori naturali da combinare in nuove sequenze. Questa scelta è stata innanzitutto condotta sulla base del livello e del profilo di espressione dei promotori candidati. E' stato composto un gruppo eterogeneo per profilo temporale di espressione e per elementi *cis*-regolatori presenti. In tutto sono stati scelti 6 promotori: 2 da utilizzare come base (GluB4 e Glb), 4 come fonte di elementi utili a potenziare i precedenti due, possibilmente senza sconvolgerne il profilo di espressione.

Tabella 20. Promotori utilizzati per costruire gli ibridi e relative caratteristiche.

Nome	Livello di espressione	Profilo di espressione	Lunghezza
GluB4	Altissimo	Endosperma	1401
GluB1	Medio	Endosperma	2263
GluC	Altissimo	Endosperma	2263
Glb	Alto	Endosperma interno	875
Prol 10kDa	Alto	Endosperma/aleurone	726
Prol 16kDa	Medio	Endosperma/aleurone	831

5.2.2 Promotori amplificati da DNA genomico

Il promotore della glutelina B4 era già stato clonato in un precedente lavoro. Per gli altri cinque promotori si è proceduto all'amplificazione da DNA genomico estratto dalla varietà CR W3 utilizzando opportuni *primers*. Tutte le coppie di *primers* hanno fornito un amplificato della lunghezza attesa. La coppia di *primers* per il promotore della globulina ha prodotto una seconda banda di 1300 bp oltre a quella attesa a 900 bp. Per ciascun promotore sono state eseguite due reazioni distinte di amplificazione. I prodotti di entrambe sono stati sequenziati.

5.2.3 Sequenziamento degli amplificati

In generale, le sequenze dei promotori isolati dalla varietà di riso utilizzata in questi studi (CR W3) sono risultate identiche o quasi identiche a quelle depositate in banca dati. L'unica eccezione fa riferimento al promotore della glutelina B1 che ha manifestato differenze di maggiore entità, coinvolgenti anche motivi *cis*-regolatori. Per il resto, il promotore della glutelina C amplificato da CR W3 si è dimostrato totalmente uguale a quello depositato per la cv. Taizhong 65, utilizzata da Qu *et al.*, (2008). Le sequenze di tutti i rimanenti promotori sono risultate divergere solo per alcuni nucleotidi da quelle depositate per la varietà Nipponbare. Tali differenze possono ricondursi alla normale variabilità intra-specifica. Si rimanda all'appendice per una più dettagliata descrizione delle differenze fra sequenze.

5.2.4 Clonazione

Per quasi tutti i promotori è stato possibile individuare con facilità una colonia che presentasse l'orientamento desiderato e la sequenza corretta. Solo il promotore della globulina ha dimostrato una spiccata tendenza a inserirsi con orientamento inverso, tant'è vero che è stato necessario digerire con *EagI* e risaldare vettore e inserto per ottenere l'orientamento desiderato. Si rimanda all'appendice per i dettagli del procedimento.

5.3 Costruzione dei promotori sintetici e delle relative cassette di espressione

Per non appesantire la lettura, gli argomenti relativi a questo capitolo verranno qui solo accennati. Si rimanda all'appendice per una trattazione esaustiva dei risultati ottenuti dal lavoro di costruzione delle cassette di espressione.

5.3.1 Costruzione di pCAMBIA-ΔGluC-GAA

Questo vettore di trasformazione è stato usato nel primo ciclo sperimentale. Il promotore GluC privato delle prime 1000 bp è stato legato al gene *reporter* tramite la TATA-box di GluB4. La cassetta di espressione è stata successivamente inserita nel vettore finale.

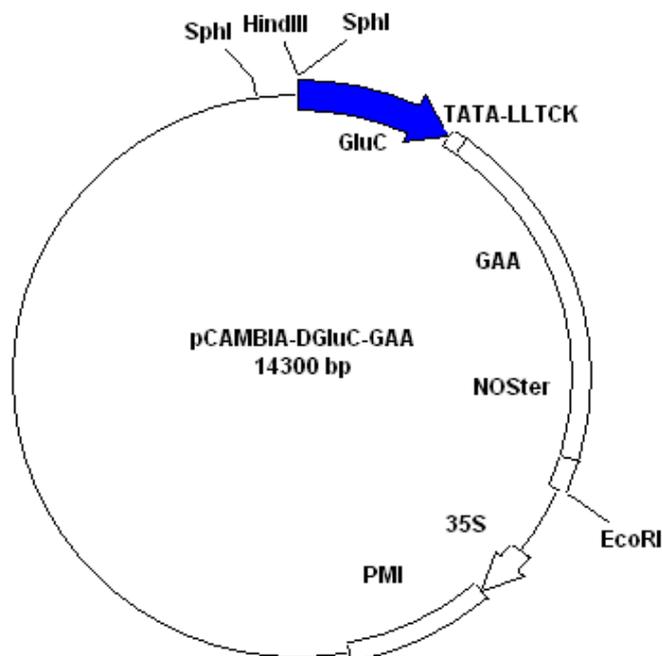


Figura 15. Mappa del vettore di trasformazione utilizzato nel primo ciclo sperimentale.

5.3.2 Costruzione del modulo accettore

Questo modulo è stato realizzato in occasione del secondo ciclo sperimentale. Esso costituisce la parte invariante della cassetta di espressione ed è stato costruito in modo da poter accogliere i vari promotori sintetici e successivamente essere trasferito nel vettore finale. La costruzione di questo modulo sono state assemblate, una dopo l'altra tutte le componenti necessarie all'espressione del gene *reporter* in pianta.

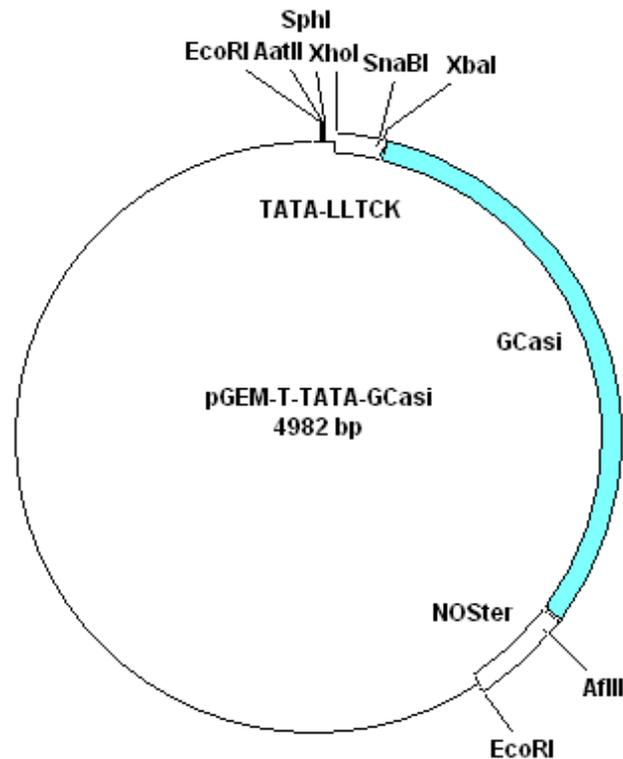


Figura 16. Schema del modulo accettore; a monte del gene *reporter* sono evidenziati i siti necessari all'inserimento del promotore, e i 2 siti *EcoRI* laterali per la sua movimentazione nel vettore finale di espressione.

5.3.3 Costruzione dei promotori sintetici

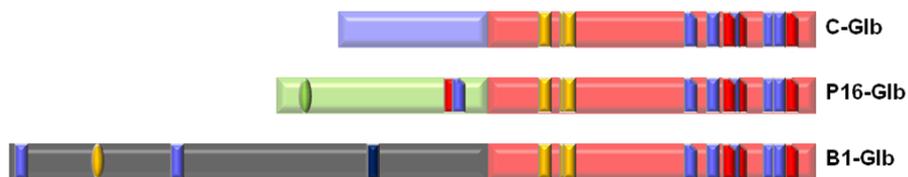
I dettagli inerenti la costruzione dei promotori sintetici possono essere reperiti in appendice. In questo paragrafo verrà richiamata l'attenzione sulla struttura dei sette promotori ibridi che sono stati valutati nel secondo ciclo sperimentale. Verrà esaminata con particolare riguardo la distribuzione dei motivi *cis*-regolatori lungo le sequenze ottenute introducendo sigle, architetture e mappe che verranno ampiamente richiamate nel seguito della sezione.

I promotori chimerici sono stati costruiti sulla base di due promotori naturali: Glb e GluB4.

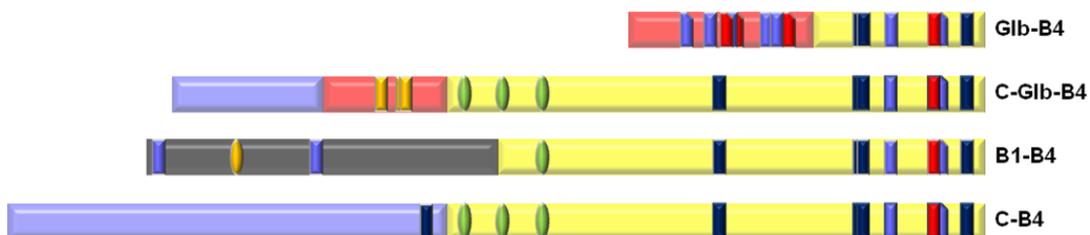


In tutti i casi, l'uno o l'altro dei due promotori costituisce la regione in 3'. Dunque è possibile suddividere i costrutti in due gruppi a seconda della provenienza della sequenza che costituisce la regione immediatamente a monte della TATA-box.

Un gruppo comprende i 3 derivati del promotore della globulina.

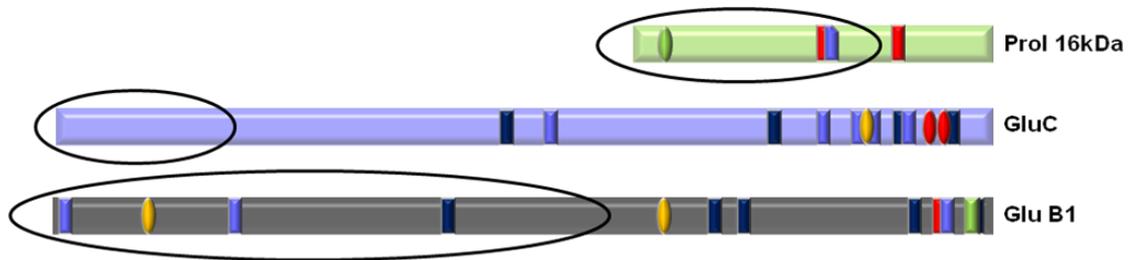


L'altro gruppo include invece i 4 derivati del promotore della glutelina B4.



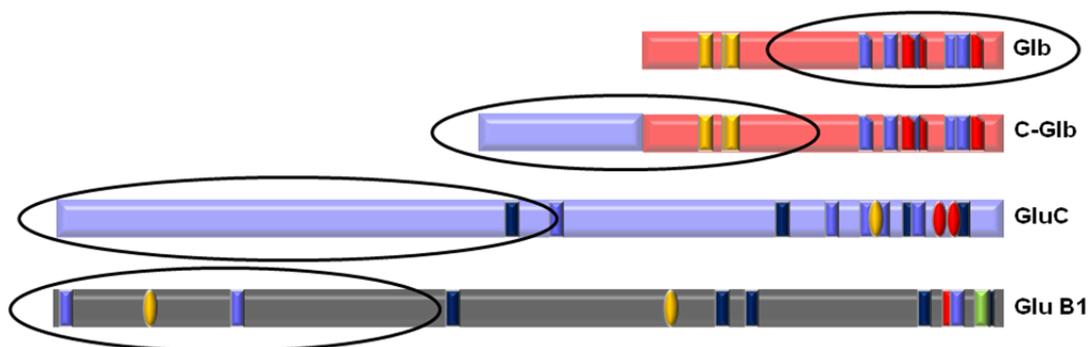
Nel primo gruppo, Glb è stato sempre utilizzato per intero, al contrario la lunghezza della porzione di GluB4 utilizzata nei promotori del secondo gruppo varia da caso a caso.

I promotori che sono stati combinati per costruire i derivati ibridi di Glb sono: Prol16, GluB1 e GluC. Di questi promotori sono state utilizzate le regioni che si trovano all'estremità 5'.



In questo modo alcune *prol-box* di GluB1 e una *endosperm-box* (*prol-box*+*GCN4*) derivante da Prol16 sono state associate al promotore della globulina. Va sottolineato che l'analisi di sequenza non ha rilevato alcun motivo nella porzione di GluC utilizzata. La valutazione *in vivo* di questo tratto di DNA è stata pianificata anche in vista di una validazione del metodo di individuazione di motivi *cis*-regolatori internamente a promotori di geni codificanti PRSR.

Per realizzare i 4 costrutti con base GluB4 sono state utilizzate a seconda dei casi la porzione 3' del promotore della globulina, una porzione di uno degli ibridi appartenenti al primo gruppo (C-Glb) e infine le regioni in 5' di GluB1 e GluC.



Il tratto mutuato da Glb contiene una fitta rete di importanti siti di legame per fattori di trascrizione. Nel promotore ibrido GlbB4, una porzione di GluB4 apparentemente povera di motivi regolatori risulta deleta a favore di quest'ultimo; scopo di tale sostituzione è la verifica della funzionalità del tratto mutuato da Glb in posizione traslata di 450 bp.

Il frammento proveniente da C-Glb contiene siti multipli di riconoscimento per il fattore RIZBZ-2. Questo fattore di trascrizione riconosce un motivo GCCACGT MAG e fa parte di una classe di proteine (*b-zipper*) che legano il DNA mediante un dominio basico che presenta una serie di leucine ripetute. Il legame avviene in tre punti, nella porzione distale del promotore della globulina 26 kDa. Al contrario, questo fattore di trascrizione non mostra affinità verso alcun altro motivo ACGT presente lungo altri promotori delle PRSR, quali quello della glutelina A2 e della prolamina PG5 13 kDa (Nakase *et al.*, 1997). La combinazione della porzione di Glb contenente i siti di legame per il fattore RISBZ-2 con il promotore della glutelina A2 non è risultata aumentare il livello di espressione del promotore ibrido ottenuto rispetto al suo analogo naturale. Un sensibile aumento del livello di espressione è stato ottenuto solamente mediante la sovra-espressione del fattore di trascrizione stesso (Yang *et al.*, 2001). Anche se i risultati pubblicati non depongono a favore dell'utilizzo di questa porzione di Glb nell'ottica di un aumento dell'espressione, è stato deciso ugualmente di ricorrere a questa combinazione.

La porzione in 5' di GluB1 è stata fusa al promotore GluB4 poiché è stato rilevato che il promotore della glutelina B1 privato di questa porzione perde i 9/10 della propria efficacia. E' stato dunque ipotizzato che in questa porzione risiedano alcuni importanti motivi *cis*-regolatori che possono svolgere funzione di enhancer in un promotore sintetico.

La porzione 5' di GluC è stata scelta con il mero intento di capire se in essa vi fossero elementi utili non evidenziati dall'analisi di sequenza.

5.3.4 Assemblaggio delle cassette di espressione e del vettore di trasformazione

I dettagli sulla realizzazione delle cassette sono riportati in appendice. In questo paragrafo viene solo schematizzato il metodo a modulo accettore universale, concepito per lavorare in parallelo, senza vincoli di siti di restrizione.

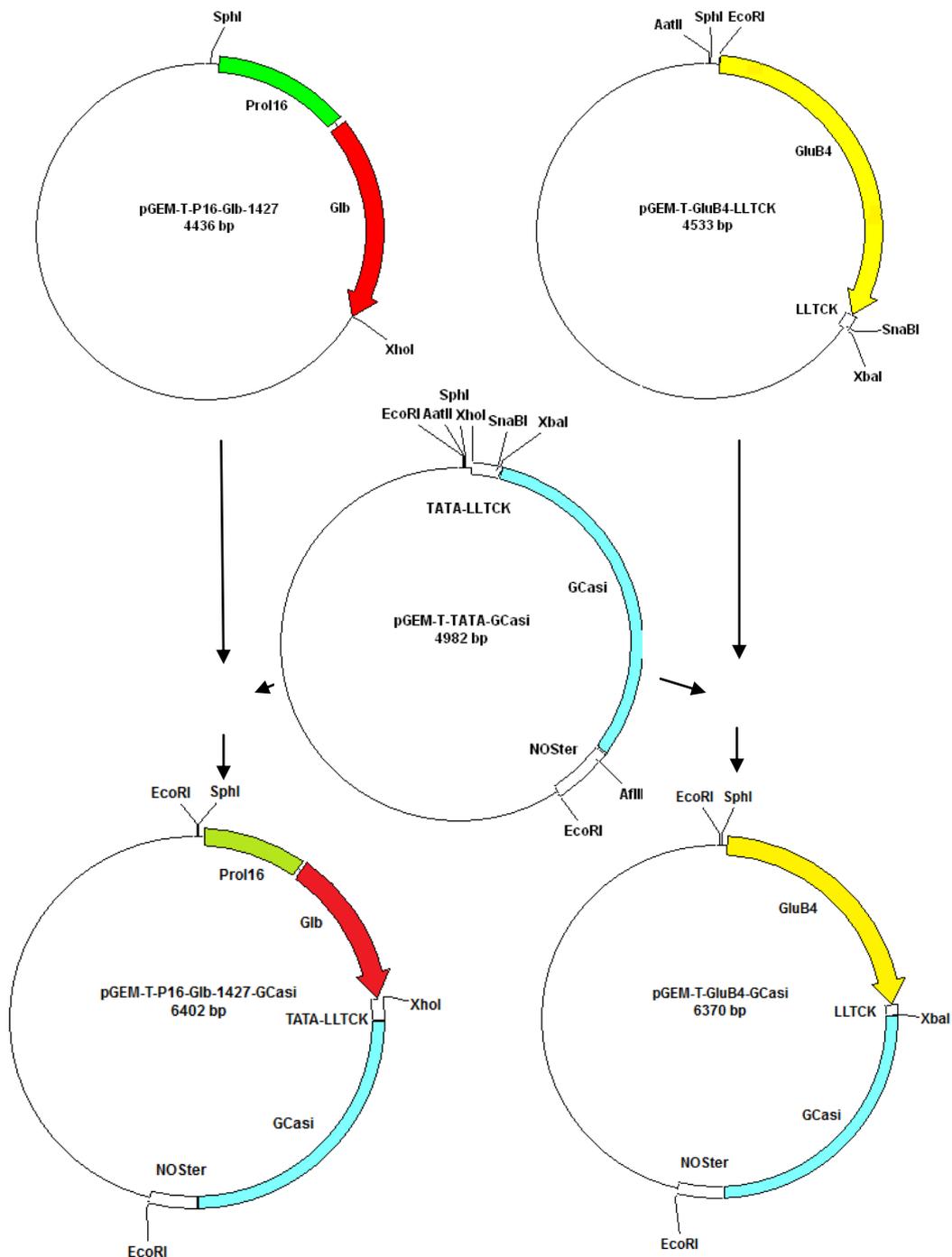


Figura 17. Due esempi di come i promotori sintetici sono stati inseriti nel modulo accettore.

5.4 Produzione delle piante transgeniche

Come già accennato, la produzione di una popolazione di piante transgeniche è il passaggio limitante nello studio di un promotore sintetico. Sono infatti necessari circa 9 mesi per ottenere dalla sintesi del costrutto piante analizzabili. Nel corso del dottorato è stato possibile concludere due cicli sperimentali.

Nel primo ciclo sperimentale sono state prodotte due popolazioni di piante transgeniche. Una popolazione è stata trasformata con Δ GluC legata al gene *reporter* tramite la TATA-box di GluB4. L'altra popolazione è stata trasformata con GluB4 ed ha assolto alla funzione di controllo. Nel corso di questo ciclo sperimentale è stato possibile ottimizzare le condizioni di trasformazione in modo da ottenere una quantità sufficiente di piante analizzabili. Oltre a ciò, questo ciclo ha avuto la funzione di validare il metodo di analisi e quantificazione del livello di espressione sulla progenie.

Nel secondo ciclo sperimentale, sono stati valutati 10 promotori dei quali 1 ibrido ternario, 6 ibridi binari, 2 naturali con la TATA-box sostituita e uno (GluB4) completamente naturale.

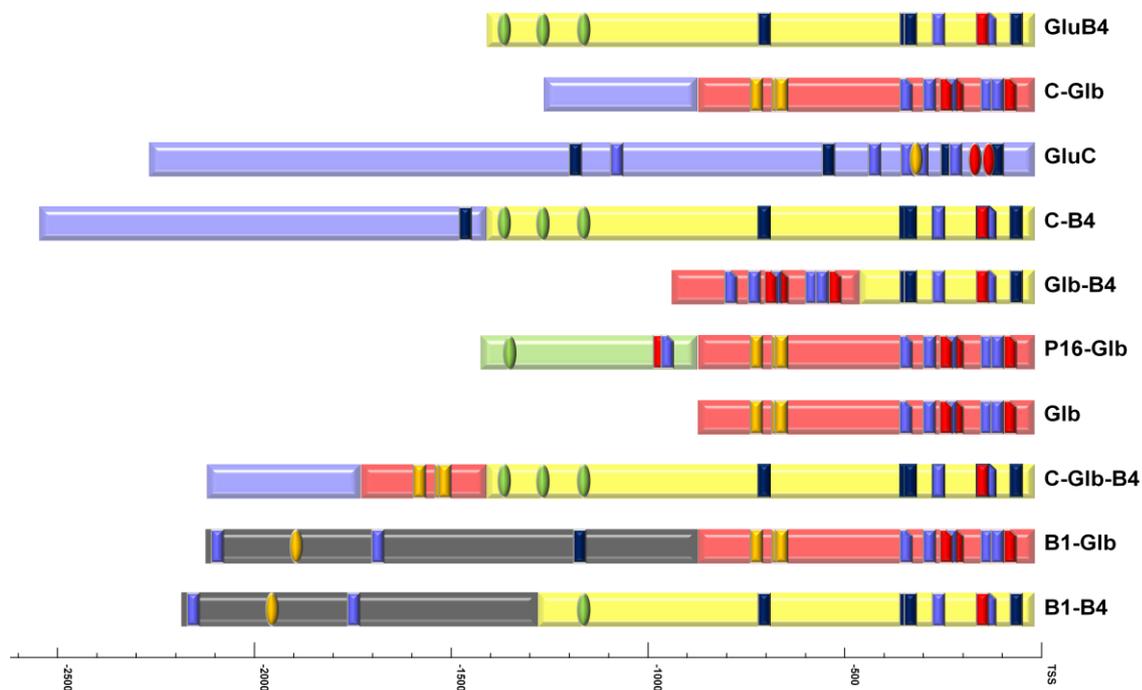


Figura 18. Schema riassuntivo rappresentante i promotori studiati in riso. Sono riportati graficamente gli elementi *cis*-regolatori individuati nel lavoro di analisi delle sequenze.

Confrontare efficacemente le prestazioni di questi candidati ha posto alcuni problemi organizzativi. Ai fini di una migliore gestione del lavoro di trasformazione, si è pertanto deciso di operare in 2 sessioni distinte (§ Appendice).

5.5 Produzione di seme e livello di espressione del gene *reporter* nei trasformati primari

In questo capitolo vengono presentati i risultati ottenuti nei due cicli sperimentali, con riferimento all'architettura dei promotori. Per un'esposizione dettagliata di tutti i dati raccolti su produttività di seme e livello di espressione del transgene, si rimanda all'appendice.

5.5.1 Produzione di seme

Al momento della raccolta è stato effettuato un censimento del seme prodotto dai trasformati primari. La produttività delle piante ottenute non è stata omogenea, né per costrutto, né per periodo di crescita in serra. In linea generale, le piante trasformate, più lungamente manipolate hanno prodotto meno seme del gruppo di controllo (rigenerato *in vitro* ma non trasformato). Come atteso, è stata notata una significativa differenza di produttività tra piante poste a dimora in serra precocemente (mese di febbraio) e piante allevate in serra nel periodo tardo-primaverile/estivo.

In particolare, in due popolazioni trasformate con due derivati di GluB4, B1-B4 e con C-Glb-B4, la produzione media di seme per pianta è risultata 4 volte inferiore alla media di tutte le piante trasformate e 8 volte inferiore al controllo.

5.5.2 Livelli di espressione indotti dai promotori saggiati nel primo ciclo sperimentale

In tale ambito, sono state prodotte due popolazioni di trasformati primari: una contenente la forma ridotta del promotore del gene codificante la glutelina C, l'altra contenente il promotore della glutelina B4, utilizzato come controllo e termine di paragone.

Da ciascuna popolazione è stato estratto a caso un campione di 20 individui PCR positivi, sui quali sono stati eseguiti saggi immunologici in

doppio per la quantificazione del livello di espressione del gene *reporter* GAA in seme. Sulla base delle classifiche ottenute, è stato ricavato per ciascun promotore:

- il valore medio dei livelli di espressione dei singoli trasformati primari analizzati;
- il valore medio dei livelli di espressione dei trasformati primari occupanti il primo quarto di classifica;
- il valor massimo di classifica.

Un primo risultato degno di nota è consistito nella dimostrazione che esiste un'espressione della proteina *reporter* in piante trasformate con pCAMBIA/ Δ GluC-GAA. Questa evidenza indica che la fusione tra la porzione in 3' del promotore GluC e la TATA-box di GluB4 (§4.5.1) origina un ibrido del tutto funzionale.

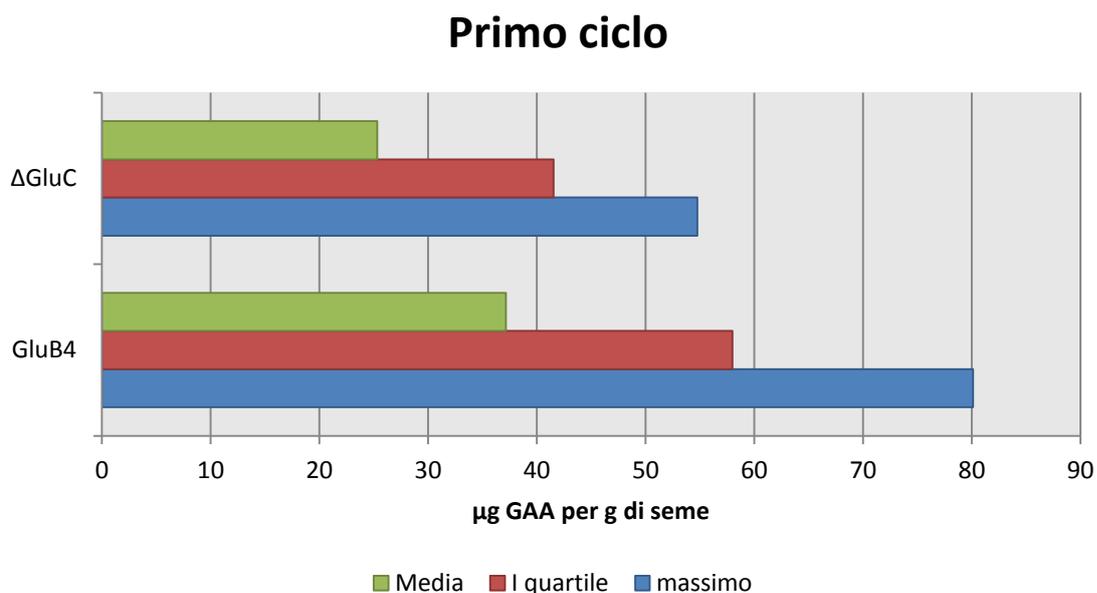


Grafico 1. I tre parametri per definire l'efficienza di un promotore sono la media di tutto il campione, la media del primo quarto di classifica e il livello del miglior espressore.

Utilizzando i tre parametri anzidetti nella valutazione dell'efficacia del promotore Δ GluC, si può inferire che esso presenta un'efficacia pari a circa 2/3 rispetto al promotore della glutelina B4.

E' interessante notare che dopo l'asportazione della porzione distale di GluC due promotori hanno all'incirca la stessa lunghezza ($\Delta\text{GluC} = 1.3 \text{ kb}$, $\text{GluB4} = 1.4 \text{ kb}$). In forza di questi risultati si è deciso che ΔGluC è certamente un promotore interessante, ma non tale da essere utilizzato come base di partenza per costruire serie di promotori sintetici. Infatti, per svolgere questo ruolo appare certamente più consono GluB4. Una terza conclusione che si è potuta trarre da questa prova riguarda l'efficacia del metodo di campionamento e analisi del seme, tale da suggerire la non necessità di una determinazione in doppio del contenuto di proteina *reporter* (§ dati sulla deviazione riportati nell'appendice). Infine, i dati raccolti sulle due popolazioni di trasformati primari hanno suggerito che la valutazione di un promotore non sembra essere influenzata dal tipo di parametro preso in considerazione.

5.5.3 Livelli di espressione indotti dai promotori saggiati nel secondo ciclo sperimentale

In modo analogo a quanto eseguito nel primo ciclo, il contenuto in proteina *reporter* è stato quantificato nel seme raccolto da singole piante T_1 . Va sottolineato che nel secondo ciclo è stato usato un gene reporter diverso che codifica per la beta-glucosidasi umana. I dati non sono dunque confrontabili in termini assoluti con quelli raccolti nel primo ciclo.

Per ogni promotore è stata stilata una classifica che ordina gli individui che compongono la popolazione di trasformati primari sulla base del livello di espressione del transgene. Per valutare l'attività trascrizionale di ciascun promotore, sono stati ricavati i tre parametri che fanno riferimento al massimo, alla media e alla media del primo quarto di classifica. In questo paragrafo viene presentato il risultato complessivo dell'esperimento. Il dettaglio di tutti i dati ottenuti è consultabile in appendice. Per tre costrutti la trasformazione è stata ripetuta una seconda volta nella successiva sessione. I dati ottenuti hanno dimostrato che la variabilità intersessione è ridotta, specie se confrontata con quella riscontrata tra le popolazioni a confronto entro ciascuna sessione. In considerazione di ciò, i risultati ottenuti nel secondo ciclo vengono

presentati come blocco unico e riportando per i tre costrutti saggiati in doppio il valore medio dei tre parametri.

Come già evidenziato nel primo ciclo sperimentale, all'interno di ciascuna popolazione sono notevoli le differenze fra gli individui. Questo risultato era in qualche modo atteso sulla base dell'effetto che ha la posizione in cui si integra il transgene all'interno del genoma ospite. Importanti differenze sono state riscontrate anche fra le varie popolazioni prese in esame e ciò conferma che la manipolazione della sequenza del promotore porta profondi effetti sull'attività del promotore stesso. Dalla classifica si evince immediatamente che alcuni promotori ibridi inducono un livello di espressione più alto rispetto al migliore promotore naturale (GluB4). Oltre a C-Glb-B4, anche P16-Glb e Glb-B4 si posizionano ad un livello significativamente superiore a GluB4, qualunque sia il parametro che si prende in esame.

Al capo opposto della classifica si situano i promotori derivati da GluB1. La porzione 5' di questo promotore sembra avere una azione negativa o nulla sul livello di espressione dei promotori a cui viene fuso.

Osservando i risultati fatti registrare dai promotori composti da GluC si nota che anche questo promotore, almeno per la porzione che è stata utilizzata in questo ciclo, non è in grado di influenzare significativamente il livello di espressione del promotore a cui viene fuso. C-Glb ha caratteristiche sovrapponibili a Glb da solo e C-B4 non differisce molto da GluB4 da solo.

Questo risultato è da considerarsi tutt'altro che fallimentare, perché aggiunge particolari elementi a quanto noto a riguardo di GluC e conferma in pieno le previsioni formulate servendosi del metodo per l'analisi di sequenza.

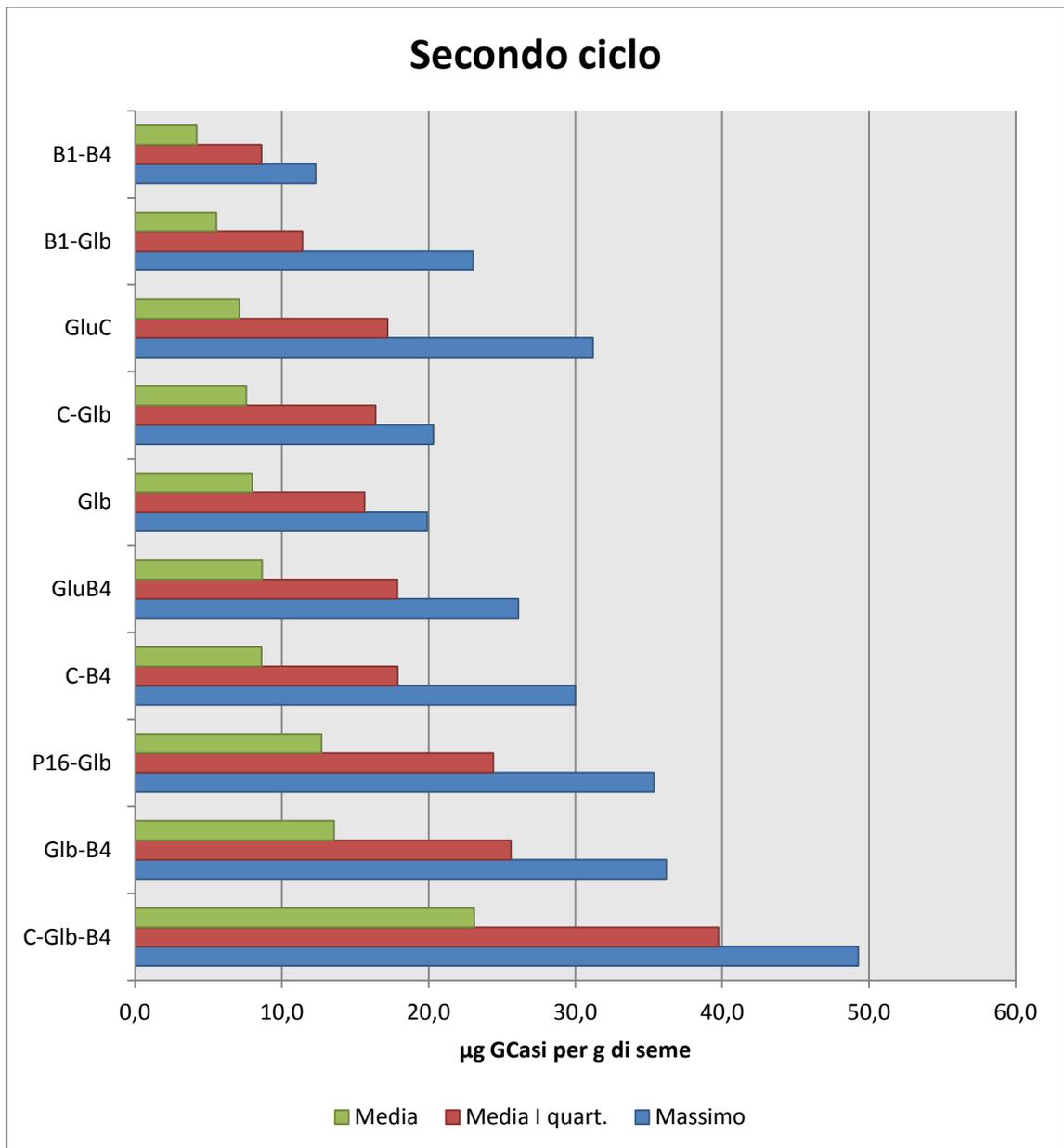


Grafico 2. Grafico riportante i valori assunti dai parametri di valutazione dei promotori posti a confronto nel secondo ciclo sperimentale. Le popolazioni prodotte nelle due sessioni di trasformazione sono state accorpate in un unico grafico. Per i promotori che sono stati trasformati in replicato è riportata la media dei parametri ottenuti. Nella pagina successiva è riportata una figura che illustra la struttura dei promotori ordinati secondo un livello crescente di espressione indotta.

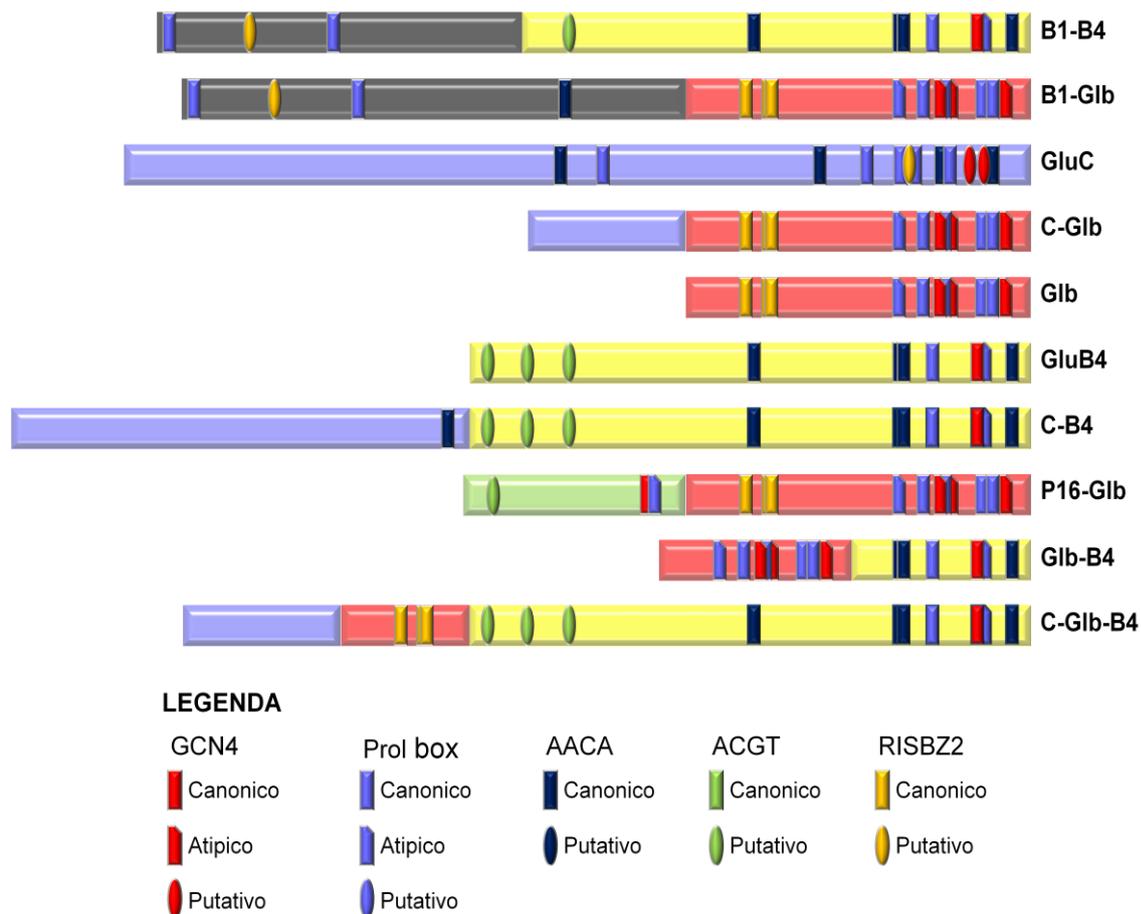


Figura 19. Struttura dei promotori saggiati nel secondo ciclo sperimentale. L'ordine rispecchia quello della classifica in Figura 18.

5.6 Tessuto-specificità

Il mantenimento di un'adeguata tessuto-specificità è requisito fondamentale per l'utilizzo di un promotore sintetico a fini biotecnologici. Per questo motivo, nel secondo ciclo sperimentale è stata utilizzata una proteina *reporter* con spiccata azione fitotossica.

Il fatto che il promotore caratterizzato dal livello medio di espressione più elevato sia anche quello la cui popolazione ha prodotto in media meno seme, ha portato a formulare l'ipotesi che la ridotta produttività potesse essere causata da uno stato di sofferenza delle piante dovuto all'espressione di GCasi in tessuti diversi dall'endosperma.

Sono stati impostati diversi esperimenti nell'intento di rilevare un'espressione di GCasi in tessuti diversi dall'endosperma. In questo capitolo verranno presentati i risultati di questi esperimenti e verranno anticipate alcune conclusioni. Anche in questo caso la gran parte dei dati

grezzi e delle note rilevate durante l'esecuzione degli esperimenti possono essere consultati in appendice.

5.6.1 Livello di espressione di GCasi in foglia

Per quantificare il livello di espressione di GCasi nel tessuto fogliare sono stati eseguiti due esperimenti. In prima istanza sono state prese in esame tutte le piante appartenenti alle due popolazioni che avevano dato maggiori problemi di produttività. In seguito, si è proceduto ad analizzare per ciascun costrutto le foglie dei migliori espressoristi.

Nella prova in cui erano oggetto di esame le popolazioni a bassa produttività di seme, è stato rilevato un contenuto di GCasi peraltro modesto solo nelle foglie di alcune piante C-Glb-B4, mentre il livello di espressione aspecifica per B1-B4, sebbene rilevabile in alcuni casi, è sempre stato dell'ordine delle unità di ppm sulla proteina totale estratta. I valori di proteina *reporter* rilevati in foglia in alcune piante trasformate con il promotore C-Glb-B4 hanno superato 100 ppm sulla proteina totale estratta. Questo dato non è apparso trascurabile e ha indotto l'avvio di un'indagine sui rapporti tra livello di espressione in foglia e in endosperma nonché tra espressione aspecifica e produttività di seme. I risultati ottenuti non sono apparsi supportare l'ipotesi di una perdita di specificità di espressione da parte di C-Glb-B4 o di qualsiasi altro promotore sintetico. rendono meno nitido il quadro della situazione. L'effetto posizione, infatti dovrebbe influenzare allo stesso modo il livello di espressione in tutti i tessuti. Ci si attende poi che i segni dovuti alla fitotossicità siano proporzionali all'accumulo della proteina fitotossica nei tessuti vegetativi.

La necessità di valutare l'accumulo di proteina *reporter* non solo in termini assoluti, ma anche in relazione all'attività del promotore in seme ha portato ad una specifica elaborazione dei dati: per correlare meglio il livello di espressione in foglia e in endosperma è stato utilizzato un indice che dà ragione del rapporto fra i due livelli.

$$Leakage = \frac{ppm \text{ GCasi su proteina totale}}{\mu g \text{ GCasi/g endosperma}}$$

L'esperimento in cui sono state prese in esame i migliori trasformati primari ottenuti per ciascun costrutto è di più facile interpretazione, per la presenza di controlli costituiti dai promotori naturali.

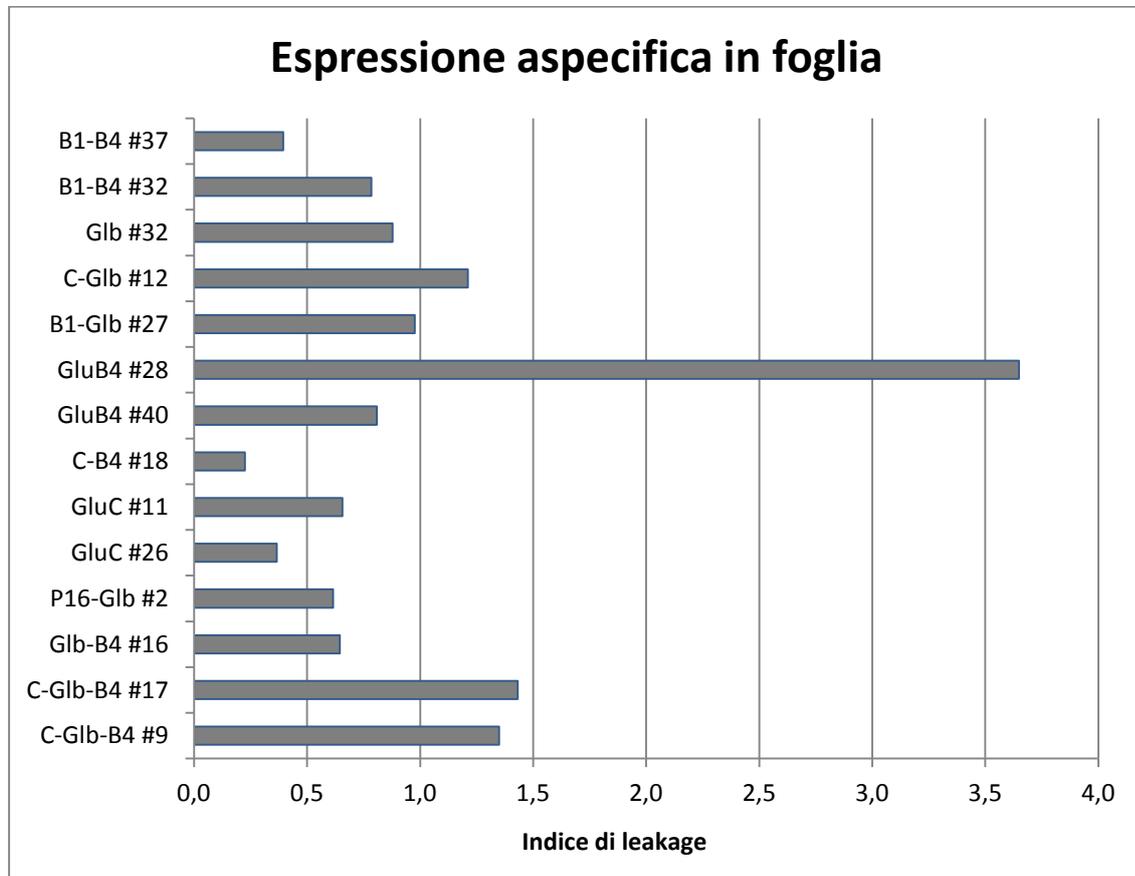


Grafico 3. Livelli di espressione del gene *reporter* nelle foglie dei migliori espressorì per ciascun costrutto.

L'indice di *leakage* dei 2 migliori espressorì per C-Glb-B4, cioè del promotore che ha dato la più elevata espressione in seme, risulta più elevato rispetto alla media dei valori registrati per i migliori espressorì ma molto inferiore al valore registrato per GluB4#28, una pianta trasformata con un costrutto portante il promotore naturale GluB4. Difficilmente spiegabili nei termini dell'ipotesi anzidetta sono anche le differenze riscontrate fra individui dello stesso gruppo, come GluC #11 e #26 o GluB4#28 e Glub4#40.

5.6.2 Livello di espressione di GCasi in vari tessuti

L'esperimento è stato finalizzato a rilevare eventuali differenze significative fra il profilo di espressione di C-Glb-B4 e quello di Glb che giustifichino la ridotta produttività del promotore sintetico rispetto alla controparte naturale. Non si rilevano differenze significative fra i tessuti prelevati dalle due popolazioni. I risultati dettagliati di questo esperimento vengono riportati in appendice.

5.6.3 Distribuzione della proteina *reporter* all'interno del seme

Per separare lo strato aleuronico è stata impiegata un'apparecchiatura per la sbiancatura del riso. La ridotta entità dei campioni disponibili per questo processo non ha permesso un recupero quantitativo delle varie frazioni prodotte. Ciò nonostante, in questo modo è stato possibile valutare in maniera rapida e conveniente la distribuzione della proteina *reporter* all'interno del seme.

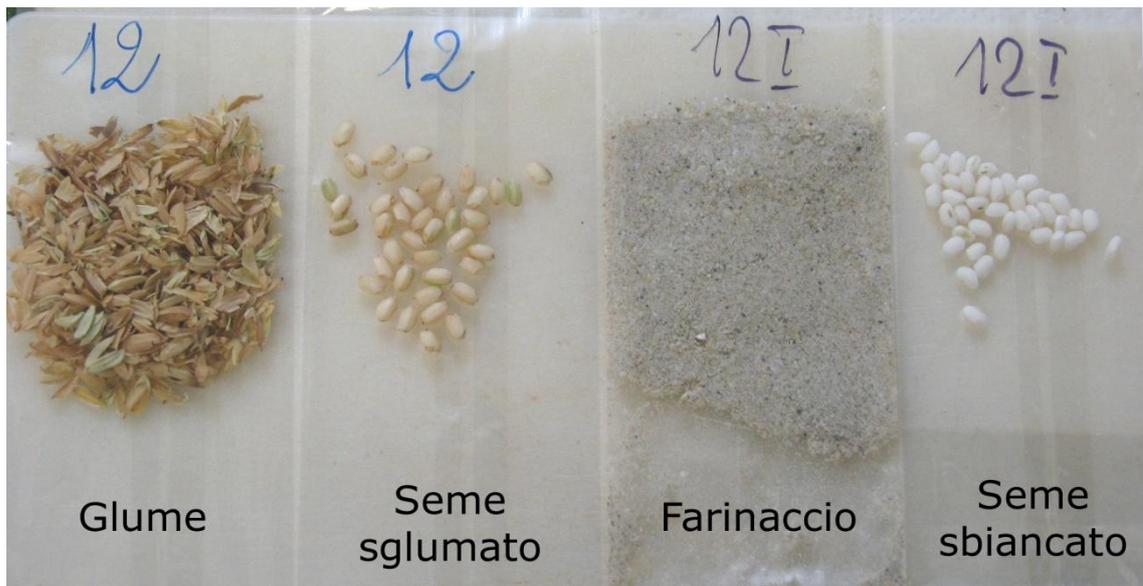


Figura 20. Frazioni prodotte con il processo di sbiancatura del seme di riso.

Per valutare la specificità per l'endosperma di ciascun promotore si è tenuto conto della frazione di GCasi che va persa dopo il processo di sbiancatura. In appendice sono riportati i dati grezzi e il processo di elaborazione che ha reso possibile questa valutazione.

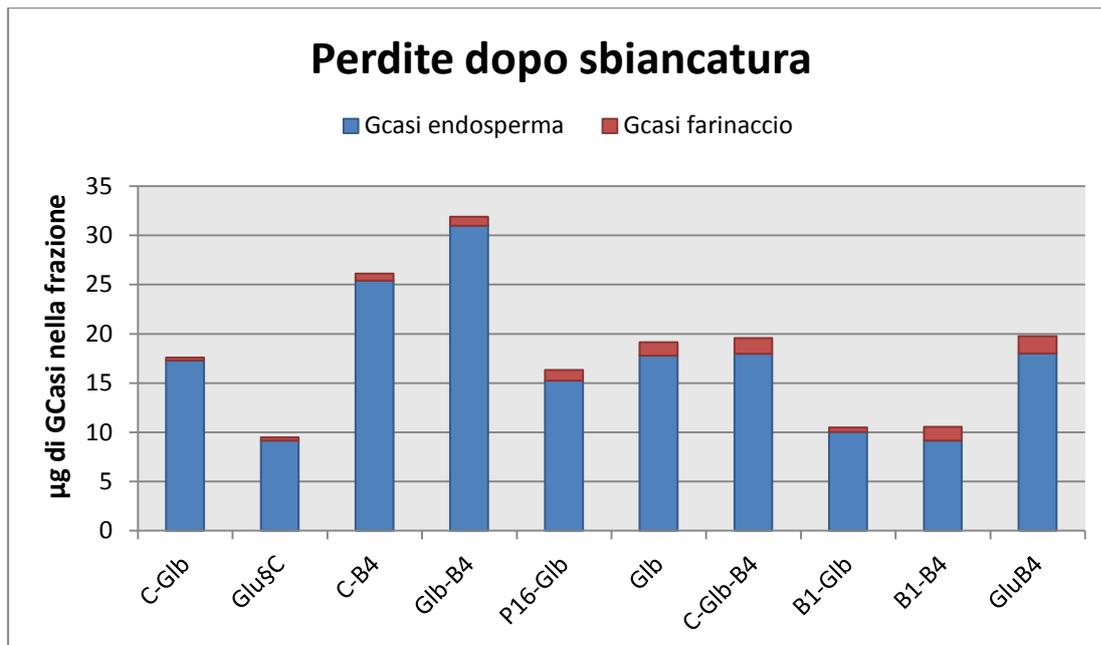


Grafico 4. Distribuzione della proteina reporter tra endosperma e tessuto aleuronico (farinaccio).

In tutti i casi la frazione di GCasi persa nel farinaccio è risultata modesta al punto che pericarpo e strato aleuronico non sembrano essere tessuti interessati dalla sintesi di beta-glucosidasi.

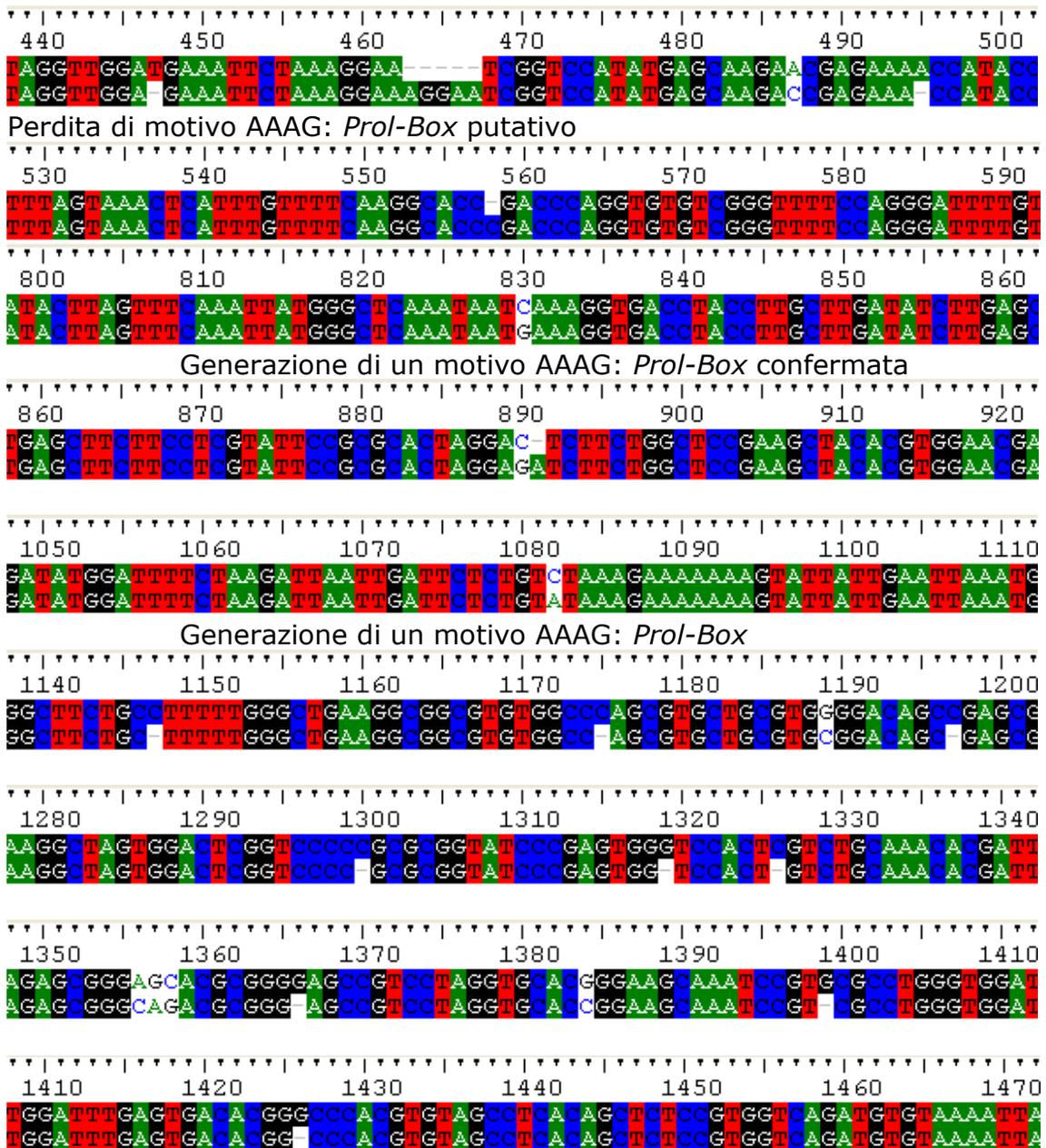
5.7 Appendice alla sezione dei risultati

5.7.1 Risultati dei sequenziamenti dei promotori naturali

Le sequenze ottenute sono state appaiate a quelle depositate in banca dati per rilevare eventuali differenze. Nonostante l'utilizzo di una DNA polimerasi ad alta fedeltà sono state spesso riscontrate differenze tra le sequenze derivanti dalle 2 PCR così come tra quest'ultime e il riferimento in banca dati. Al fine di distinguere fra errori della polimerasi ed effettivi polimorfismi varietali sono state eseguite altre due reazioni di amplificazione per ciascun promotore e i prodotti sono stati sequenziati. I risultati dei sequenziamenti hanno permesso di individuare le effettive discrepanze tra le sequenze depositate e quelle effettivamente presenti in CR W3.

Il promotore della glutelina C è risultato avere sequenza identica a quella depositata. Il promotore della glutelina GluB1di CR W3 presenta invece 11 inserzioni di singoli nucleotidi lungo la sequenza nonché varie

sostituzioni e le seguenti delezioni: 5 nt in posizione 463, 4 in posizione 2060 e 2 in posizione 2094. Va sottolineato che rispetto alla sequenza depositata in banca dati, quella di CR W3 manca di una prol-box putativa in posizione -460 ma è dotata di 2 prol-box aggiuntive in posizione -830 (ATCAAAG) e -1082 (TCTAAAG); entrambi i motivi rilevati hanno mostrato la capacità di legare il fattore RPBF nel promotore della prolamina 13kDa NPR33 (Yamamoto *et al.*, 2006).



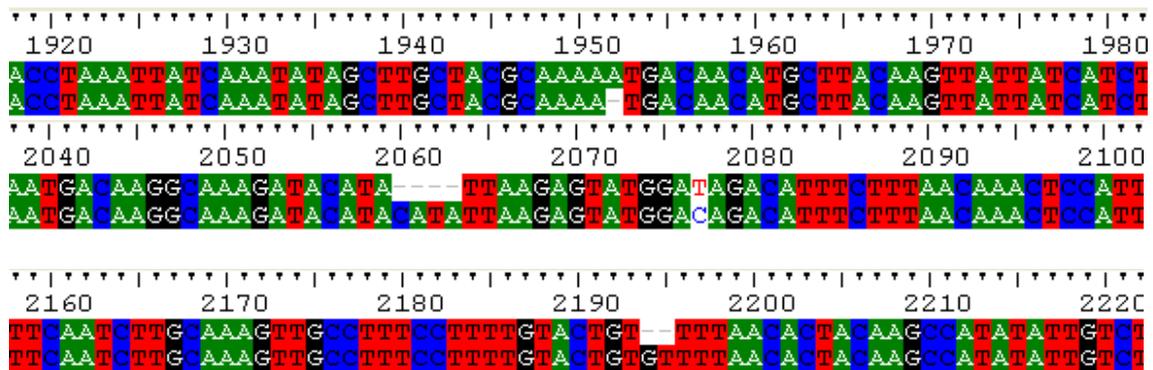


Figura 21. Promotore della glutelina B1: differenze fra la sequenza in CR W3 (sopra) e la sequenza depositata in banca dati, gi|42742291|gb|AY427569.1| *Oryza sativa*, var. Nipponbare (japonica cultivar-group).

Nella varietà di riso utilizzata in questo lavoro, la sequenza del promotore della globulina 26kDa presenta solo qualche discrepanza nella parte iniziale. Tutti i motivi *cis*-regolatori individuati nella sequenza depositata sono perfettamente conservati in CR W3.

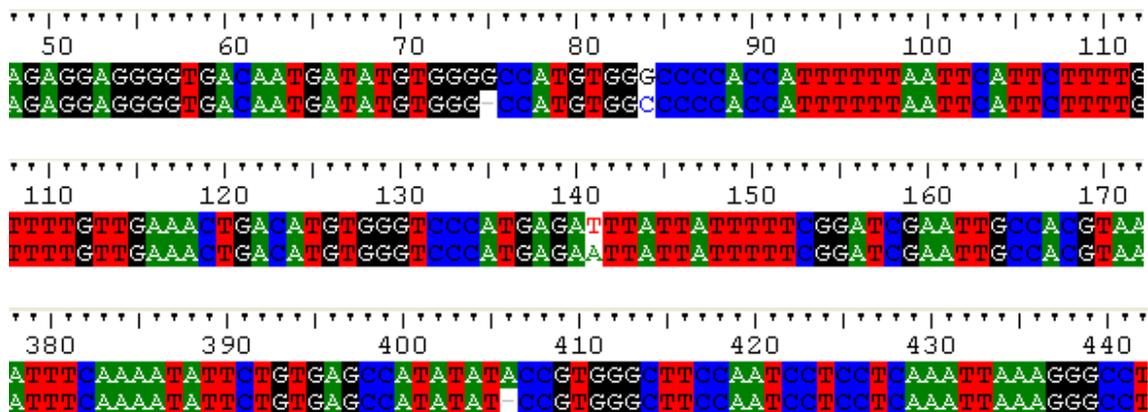


Figura 22. Promotore della globulina: differenze fra la sequenza in CR W3 (sopra) e la sequenza depositata, gi|42742297|gb|AY427575.1| *Oryza sativa*, var. Nipponbare (japonica cultivar-group).

Rispetto alla sequenza di riferimento, il promotore della prolamina 10 kDa isolato da CR W3 presenta come unica differenza un'inserzione in posizione -420, mentre il promotore della prolamina 16 kDa presenta due inserzioni. E' inoltre importante ricordare che l'amplificato utilizzato in questo lavoro manca delle prime 12 bp.

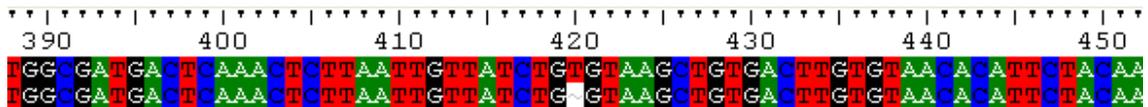


Figura 23. Promotore della prolamina 10kDa, differenze fra la sequenza in CR W3 (sopra) e la sequenza depositata, gi|42742294|gb|AY427572.1| *Oryza sativa*, var. Nipponbare (japonica cultivar-group).

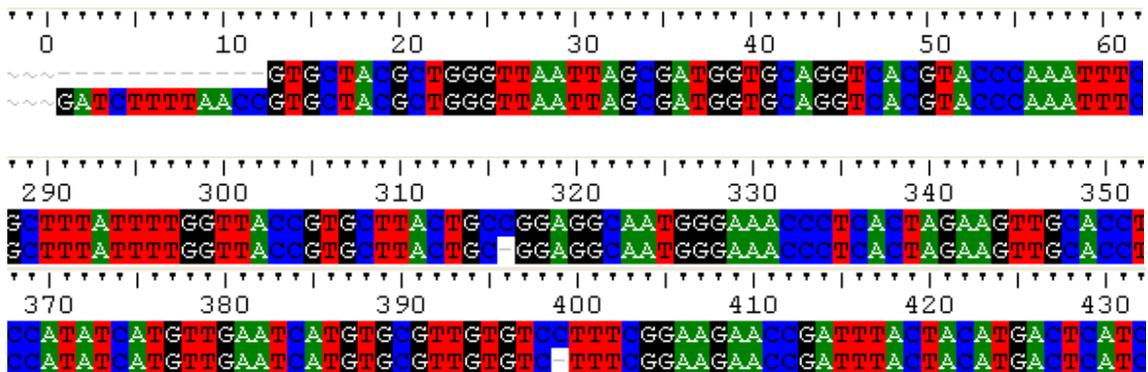


Figura 24. Promotore della prolamina 16kDa: differenze fra la sequenza in CR W3 (sopra) e la sequenza depositata, gi|42742296|gb|AY427574.1| *Oryza sativa*, var. Nipponbare (japonica cultivar-group)

5.7.2 Costruzione del modulo accettore

Per ottenere la porzione che va dalla TATA-box alla fine del *leader* sintetico LLTCK, si è proceduto ad una amplificazione con specifici *primers* su un plasmide che conteneva gli elementi desiderati. L'amplificato clonato in pGEM-T è stato usato per trasformare cellule di *E. coli*. Le colonie ottenute dalla trasformazione sono state controllate mediante PCR e ciò per determinare presenza, dimensione e orientamento dell'inserto. Le colonie positive sono state sequenziate usando i *primers* universali M13 *for* e *rev*. Il sequenziamento ha confermato l'identità dell'amplificato con la sequenza desiderata.

Servendosi del sito *EcoRI* inserito all'estremità 3' dell'amplificato con il *primer* TATA *rev*, e del sito *XbaI* presente fra LLTCK e PS, è stato possibile inserire il terminatore NOS con il 3'-UTR, ottenuto mediante digestione di pUC18-NOS con *XbaI* e *EcoRI*. Con gli stessi enzimi è stato digerito pGEM-T/TATA::LLTCK::PS::GAA. L'inserto corrispondente a NOS è stato legato nel vettore e da 3 colonie ottenute dalla trasformazione di *E. coli* è stato estratto il DNA plasmidico da sottoporre a digestione di controllo con *EcoRI* e *XbaI*. Individuato un clone desiderato positivo, si è proceduto all'aggiunta

di un secondo sito *EcoRI* all'inizio del MCS. La saldatura dell'adattatore sintetico composto dagli oligonucleotidi *EcoRI for* ed *EcoRI rev* nel vettore pGEM-T/TATA::LLTCK::NOS è stata condotta come descritto in 4.4.3.

L'inserto relativo a GCasi è stato ottenuto mediante digestione con *XbaI* e *AflII* di un plasmide contenente la porzione PS::GCasi::NOS. L'inserto è stato saldato all'interno del vettore pGEM-T-2x*EcoRI*-TATA::LLTCK::NOS, aperto con gli stessi enzimi. Il vettore ingegnerizzato è stato inserito in *E. coli* e da 3 colonie è stato preparato il DNA plasmidico da sottoporre a digestione di controllo con gli enzimi *EcoRI* e *XhoI*. Due colonie hanno dato il pattern di restrizione atteso e sono state sequenziate per l'intera lunghezza del modulo accettore prima di procedere con il lavoro di assemblaggio delle cassette di espressione. Una volta confermata l'identità della sequenza con quella desiderata, il plasmide è stato utilizzato come modulo accettore dei promotori sintetici prodotti nel corso di questo studio.

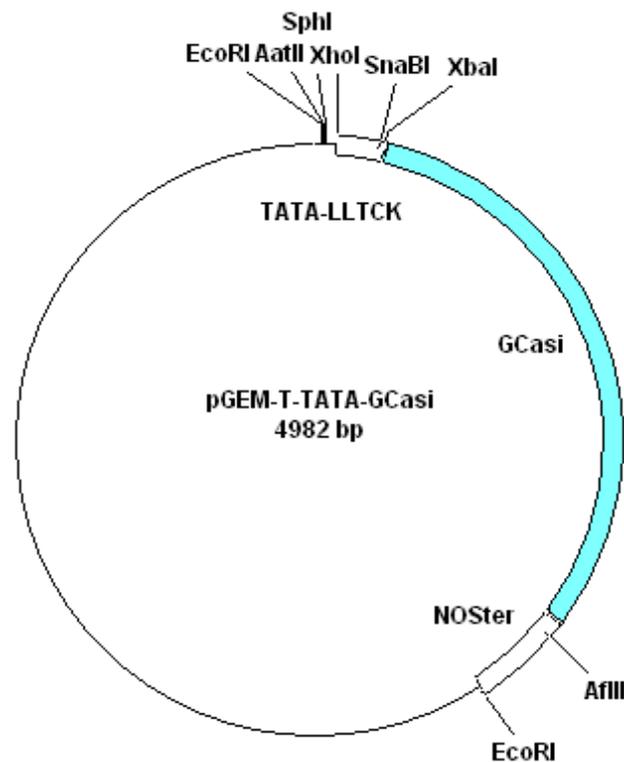


Figura 25. Modulo accettore: a monte del gene *reporter* sono evidenziati i siti necessari all'inserimento del promotore, e i 2 siti *EcoRI* laterali per la sua movimentazione nel vettore di espressione.

5.7.3 Costruzione di pCAMBIA-ΔGluC-GAA

Questo vettore di trasformazione è stato usato nel primo ciclo sperimentale, prima che fosse costruito il modulo accettore definitivo con 2 siti *EcoRI*. In una fase iniziale, è stato prodotto pGEM-T-TATA::LLTCK::PS::GAA::NOS, successivamente impiegato come modulo accettore per la costruzione di una serie di plasmidi da utilizzare in esperimenti di biolistica. In questo modulo, fra *SphI* e *XhoI* è stato inserito ΔGluC ottenendo la cassetta di espressione ΔGluC-TATA::LLTCK::GAA::NOS.

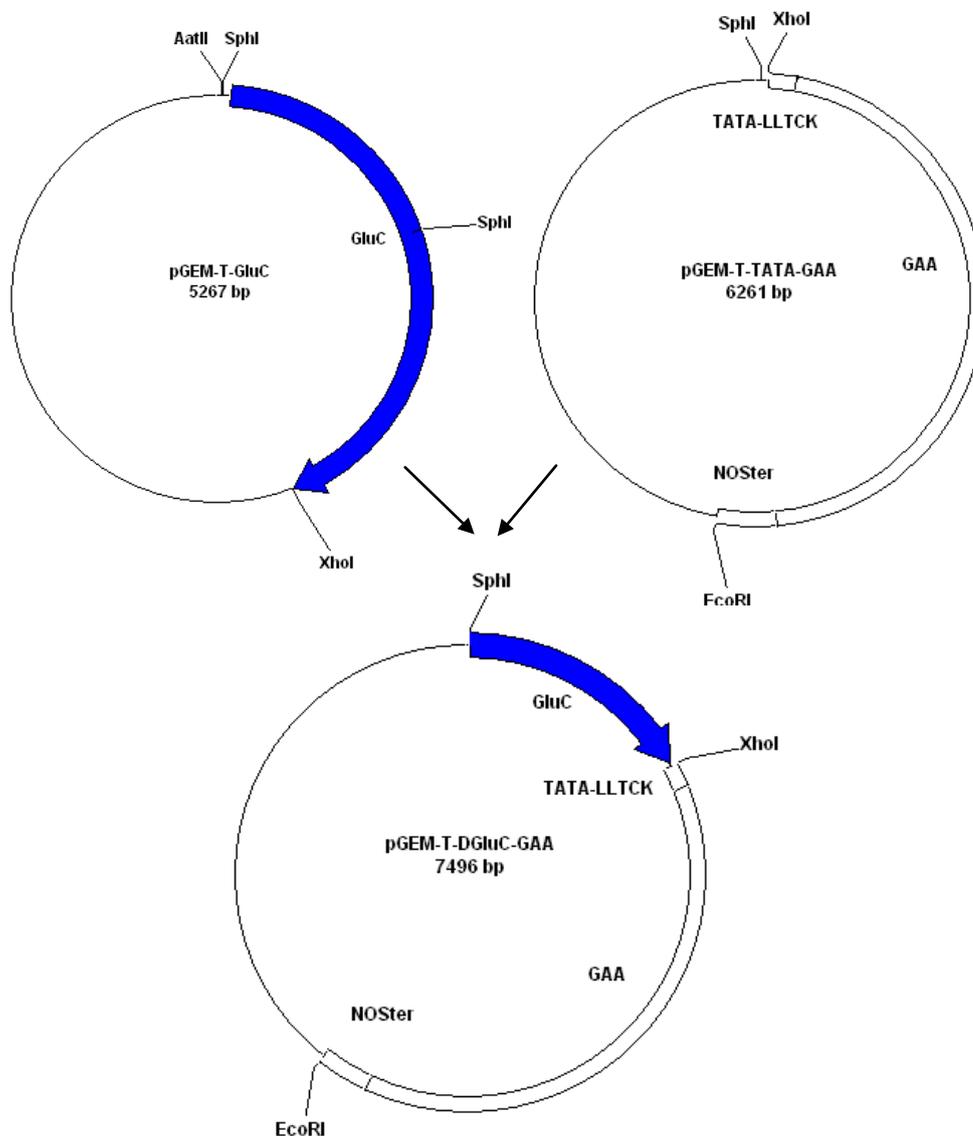


Figura 26. Inserimento della porzione di GluC a monte della TATA-box nel modulo accettore pGEM-T/TATA::LLTCK::GAA::NOS.

La cassetta di espressione è stata sub-clonata in pUC18 fra *Sph*I ed *Eco*RI e poi inserita in pCAMBIA, fra *Hind*III ed *Eco*RI.

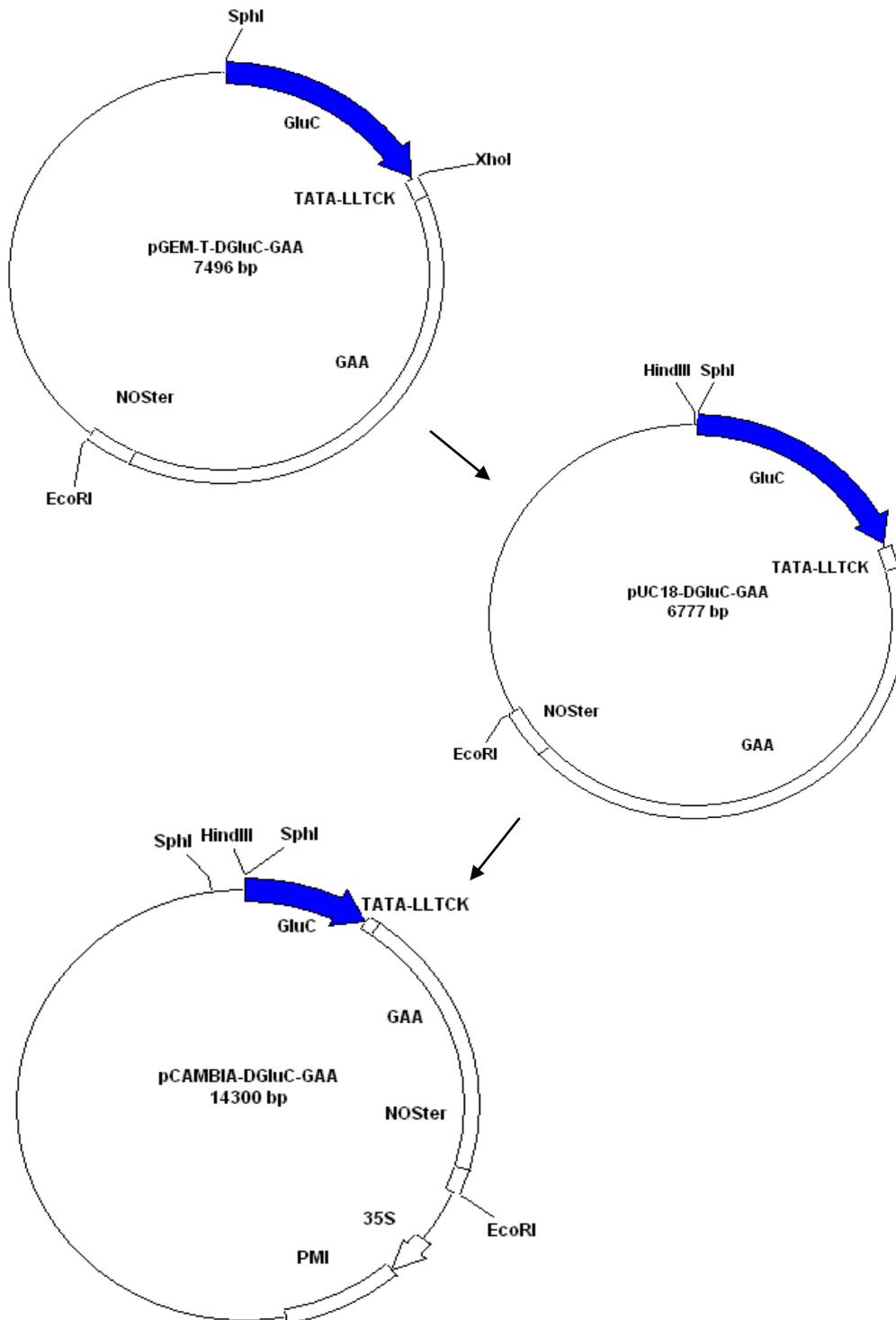


Figura 27. Inserimento della cassetta Δ -GluC-GAA nel vettore di espressione tramite sub-clonazione in pUC18. Il sito *Hind*III è stato tratto da pUC18 e sfruttato per l'inserimento in pCAMBIA.

5.7.4 Costruzione dei promotori sintetici

Per realizzare una serie di promotori sulla base di Glb si è proceduto alla digestione di pGEM-T/Glb-r, con *SpeI* e *XhoI*. In questo modo si è ottenuto un inserto coincidente con l'intero promotore della globulina. Nel contempo i vettori contenenti Prol16, GluB1 e GluC sono stati aperti con *XhoI* e un enzima, *XbaI* o *NheI* che producesse un'estremità 5'-CTAG, compatibile con *SpeI*.

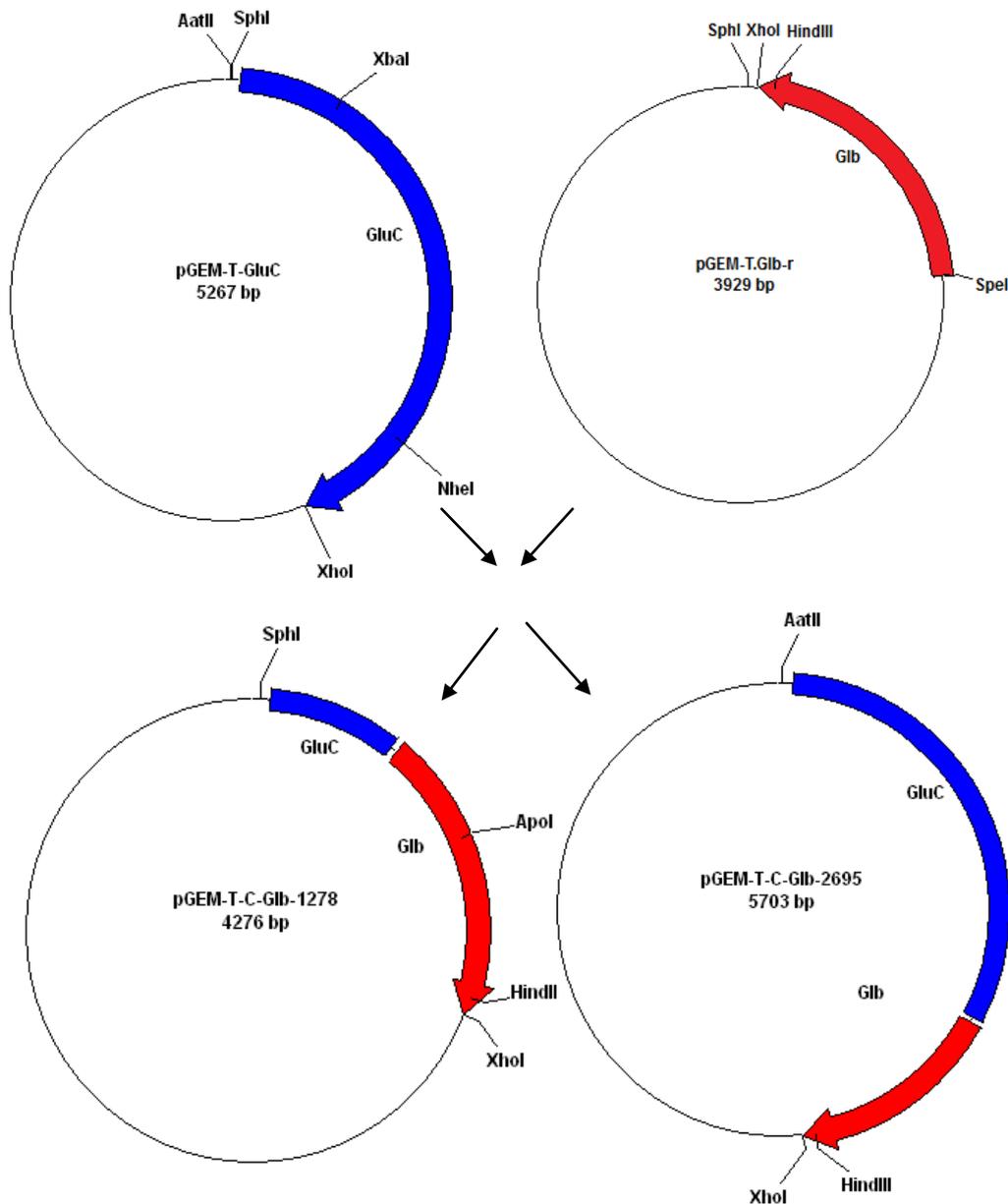


Figura 28. Costruzione di C-Glb mediante digestione del vettore pGEM-T-GluC con *XbaI*; costruzione di C-Glb_2695 mediante digestione del vettore pGEM-T-GluC con *NheI*.

Successivamente alla saldatura dell'inserto Glb nei vettori recuperati dalle digestioni sopra descritte e successivamente alla trasformazione di cellule JM101 di *E. coli*, 6 colonie putativamente trasformate sono state verificate mediante PCR su coltura batterica per confermare la presenza dell'inserto Glb. A tal fine, sono state usati i *primers* GluC *for* e Glb *rev2* che appaiano ai lati della regione di giunzione fra GluC e Glb. La presenza dell'amplificato di 958 bp ha confermato l'inserzione di Glb in GluC nelle colonie 1, 2, 3 e 5. Come controprova è stato utilizzato il *primer* M13 *rev* che in combinazione con GluC *for* produce un amplificato di circa 464 bp in caso di *self-ligation*.

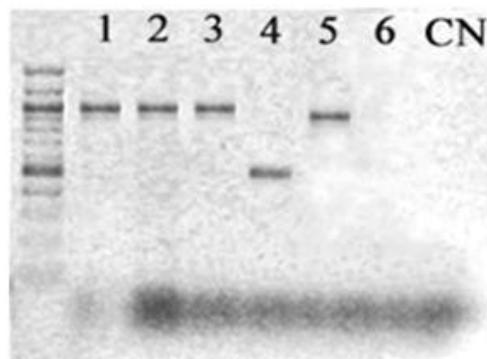


Figura 29. Gel di agarosio in cui sono stati caricati i prodotti PCR duplex ottenuti su coltura batterica con i *primers* GluC *for*, Glb *rev2* e M13 *rev* su 6 colonie di C-Glb.

Tutte e sei le colonie del costrutto C-Glb-2695 hanno prodotto un amplificato di circa 720 bp, ovvero di una lunghezza simile a quella attesa per il frammento compreso fra GluC *for3* e Glb *rev2* in C-Glb-2695.

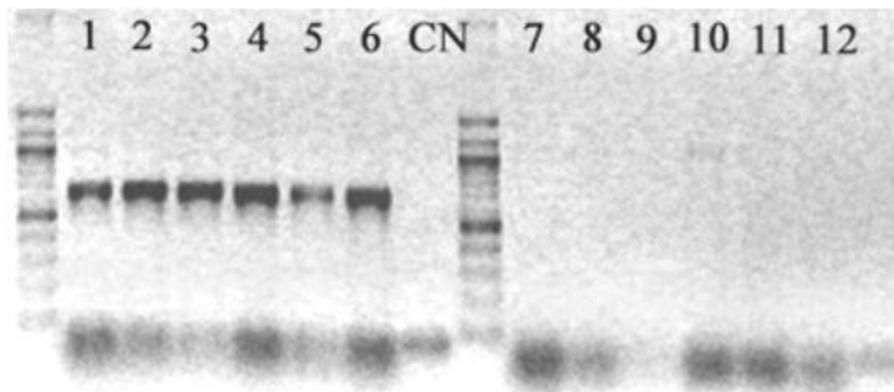


Figura 30. Gel di agarosio in cui sono stati caricati i prodotti PCR su coltura batterica. Corsie 1-6: colonie 1-6 con *primers* GluC *for3* e Glb *rev2*. Corsie 7-12: colonie 1-6 con *primers* GluC *for3* e M13 *rev*.

Per inserire l'inserto Glb in pGEM-T/Prol16 si è proceduto alla digestione di questo plasmide con *Xba*I e *Xho*I. *Xba*I ha prodotto un'estremità coesiva compatibile con quella lasciata da *Spe*I in Glb.

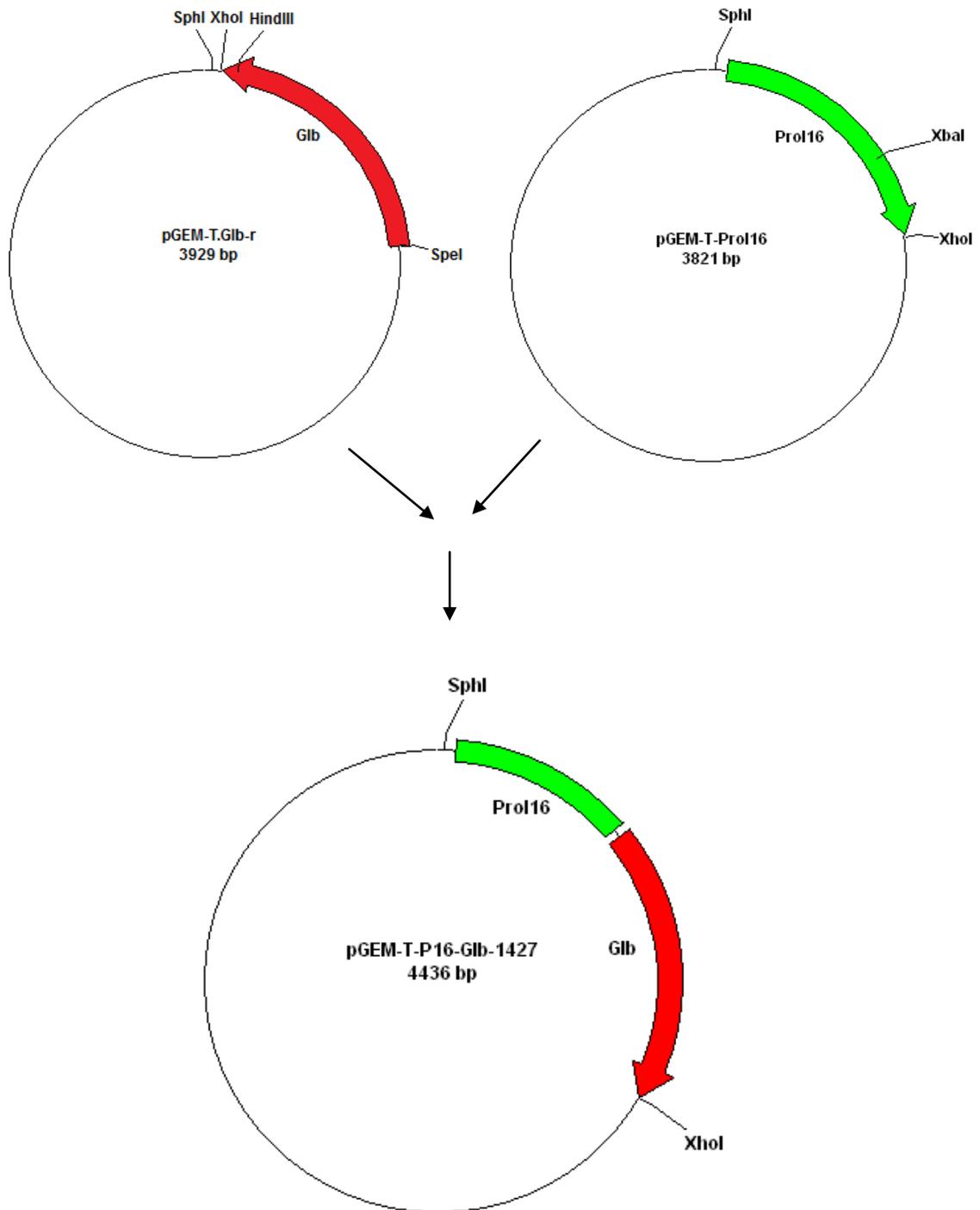


Figura 31. Sostituzione della porzione prossimale di Prol 16 con Glb. *Xba*I e *Spe*I producono estremità compatibili.

Per la costruzione di B1-Glb il vettore pGEM-T/GluB1 è stato digerito con *BlnI* e *XhoI*. *BlnI* genera una estremità compatibile con quella generata da *SpeI* in Glb-r.

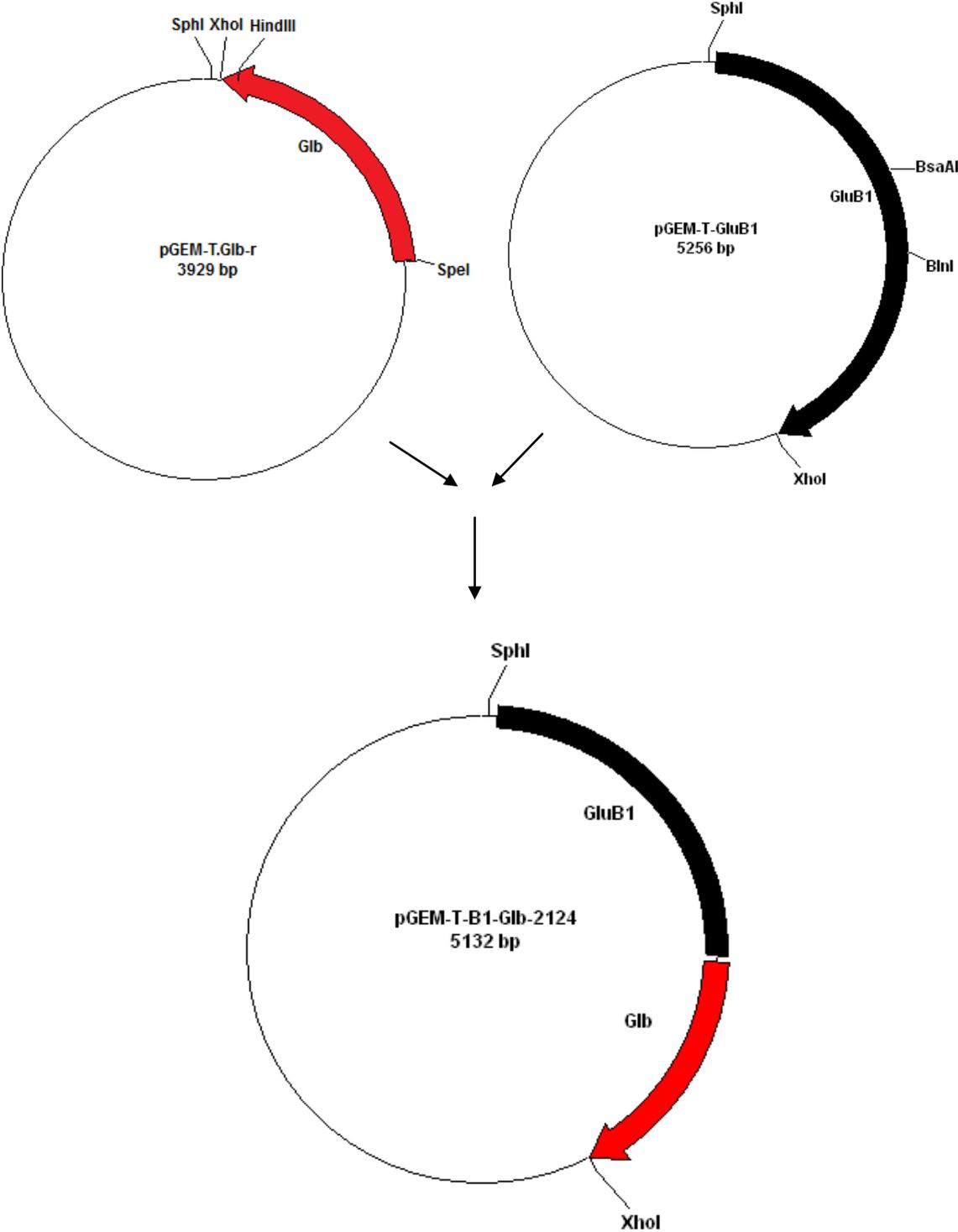


Figura 32. Costruzione del promotore ibrido B1-Glb. L'apertura del vettore avviene tramite taglio con *BlnI* e *XhoI*.

Per procedere alla costruzione dei promotori sintetici è stato necessario invertire il senso di inserzione dei promotori GluB4 e Glb. Il promotore GluB4 era già stato clonato in un precedente lavoro. L'amplificato relativo al promotore della globulina 26kDa ha dimostrato una spiccata preferenza ad inserirsi in pGEM-T nell'orientamento inverso. Non è stato possibile ottenere Glb clonato nel senso desiderato, dunque si è proceduto alla sua inversione.

Per invertire gli inserti è stato utilizzato il sito *EagI*, presente nel MCS ad entrambe le estremità del sito di inserzione. Dopo digestione con questo enzima, l'inserto ricavato è stato nuovamente legato al vettore. La miscela di saldatura è stata usata per trasformare cellule JM101 di *E. coli*. Le colonie ottenute sono state saggiate per il corretto orientamento mediante PCR su coltura.

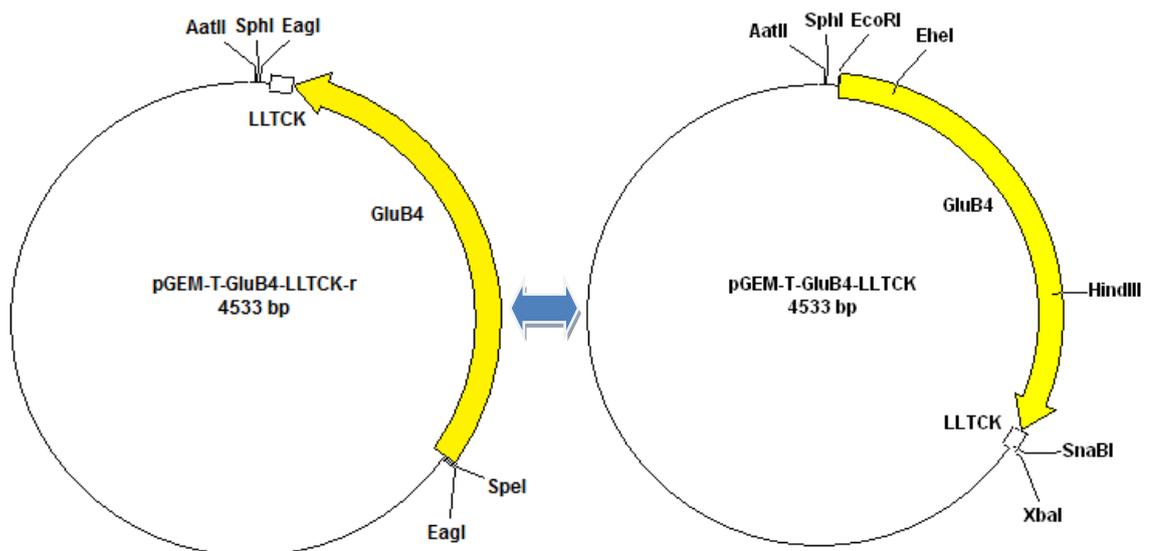


Figura 33. Modalità di inversione dell'inserto clonato nel MCS di pGEM-T. Questa procedura è applicabile a qualunque inserto clonato fra le 2 T, purchè privo di siti di riconoscimento per *EagI*.

A partire dal promotore GluB4 sono stati derivati una serie di promotori ibridi che ne mantengono una porzione 5' più o meno lunga. Nella costruzione di questi promotori, pGEM-T/GluB4::LLTCK è stato utilizzato sempre come vettore, mentre gli inserti sono stati prodotti dalla digestione di Glb, C-Glb, C-Glb, GluB1 e GluC.

Da pGEM-T/Glb-r è stato ricavato un inserto corrispondente alla porzione prossimale di Glb mediante digestione sequenziale con *Hind*III e con *Apo*I. L'estremità coesiva lasciata da *Apo*I, compatibile con quella di *Eco*RI, è stata sfruttata per legare l'inserto in pGEM-T/GluB4::LLTCK.

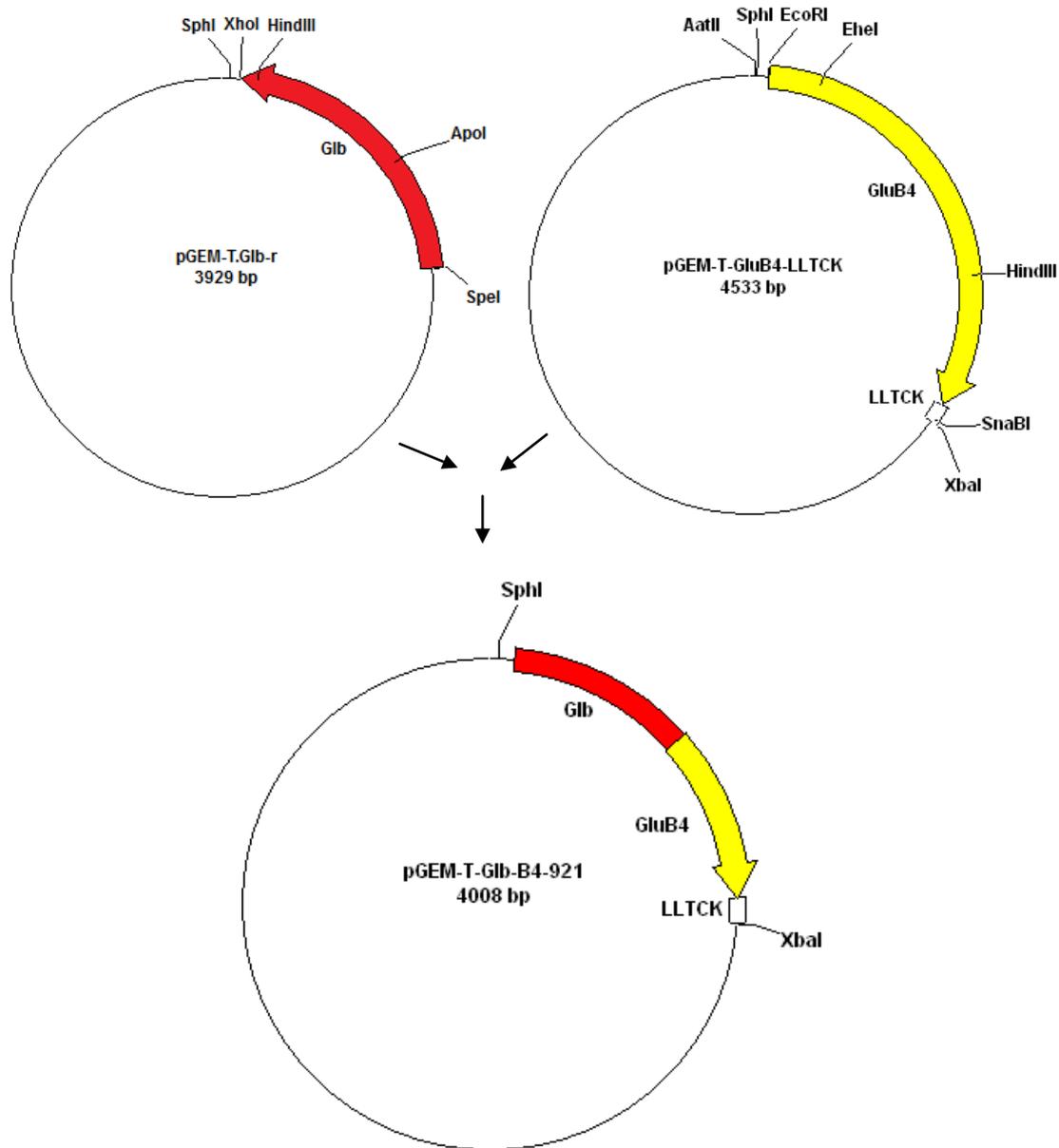


Figura 34. Costruzione di un promotore ibrido con base GluB4.

Per la costruzione di un promotore ibrido composto da GluB4 e GluC, si è proceduto a creare un inserto GluC linearizzando il plasmide pGEM-T/GluC con *AatII*. Il prodotto linearizzato è stato poi digerito con *ApoI*. L'inserto purificato da gel è stato inserito in pGEM-T/GluB4::LLTCK fra *AatII* e *EcoRI*.

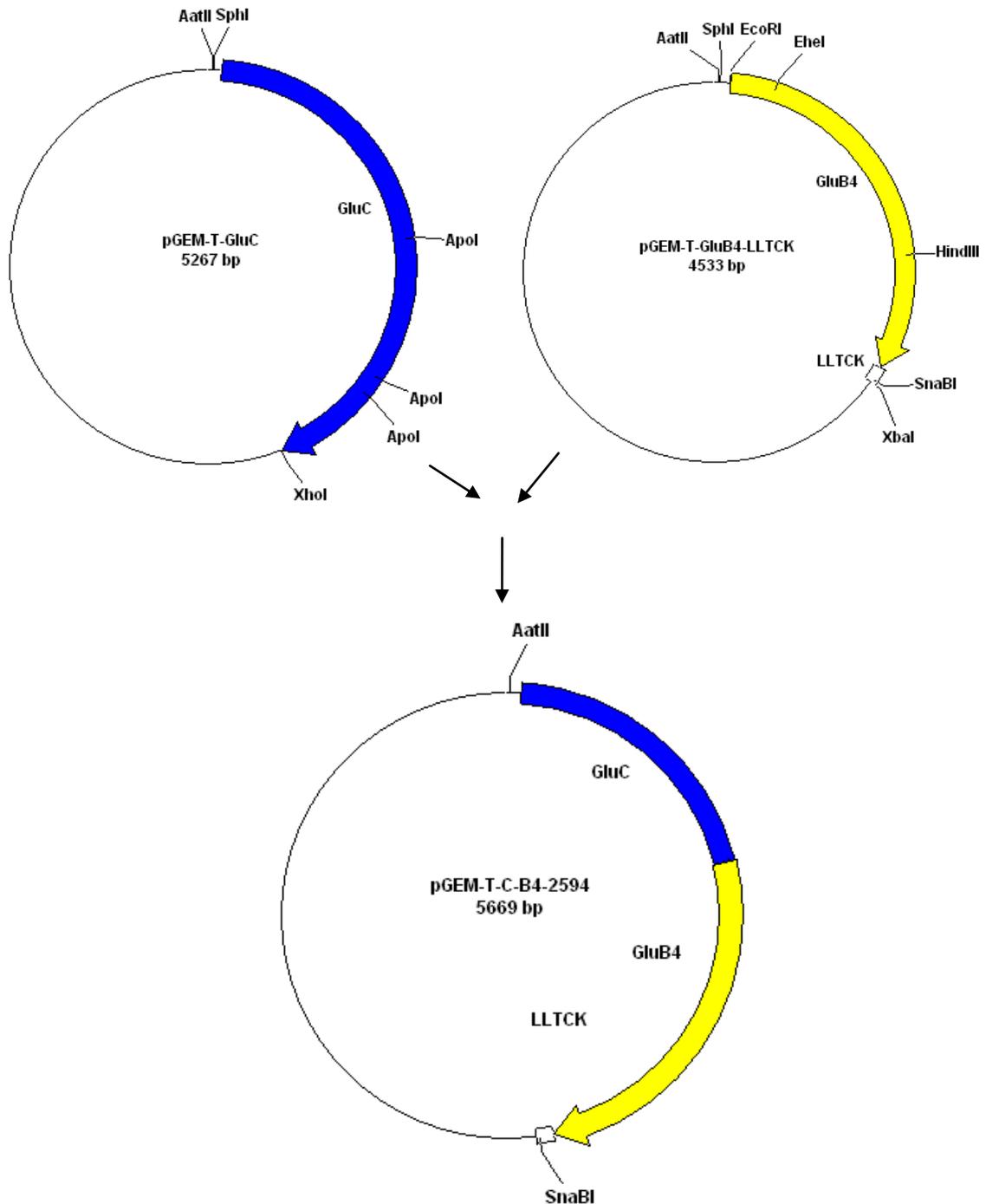


Figura 35. Inserimento dell'elemento GluC per digestione sequenziale con *AatII* e *ApoI*. L'inserto viene saldato immediatamente a monte di GluB4, in corrispondenza del sito *EcoRI*.

Analogamente al caso precedente, per costruire il promotore ternario C-Glb-B4, la porzione di C-Glb compresa fra *SphI* e *ApoI* è stata inserita in pGEM-T/GluB4::LLTCK, tra *EcoRI* e il sito *SphI* presente nel MCS.

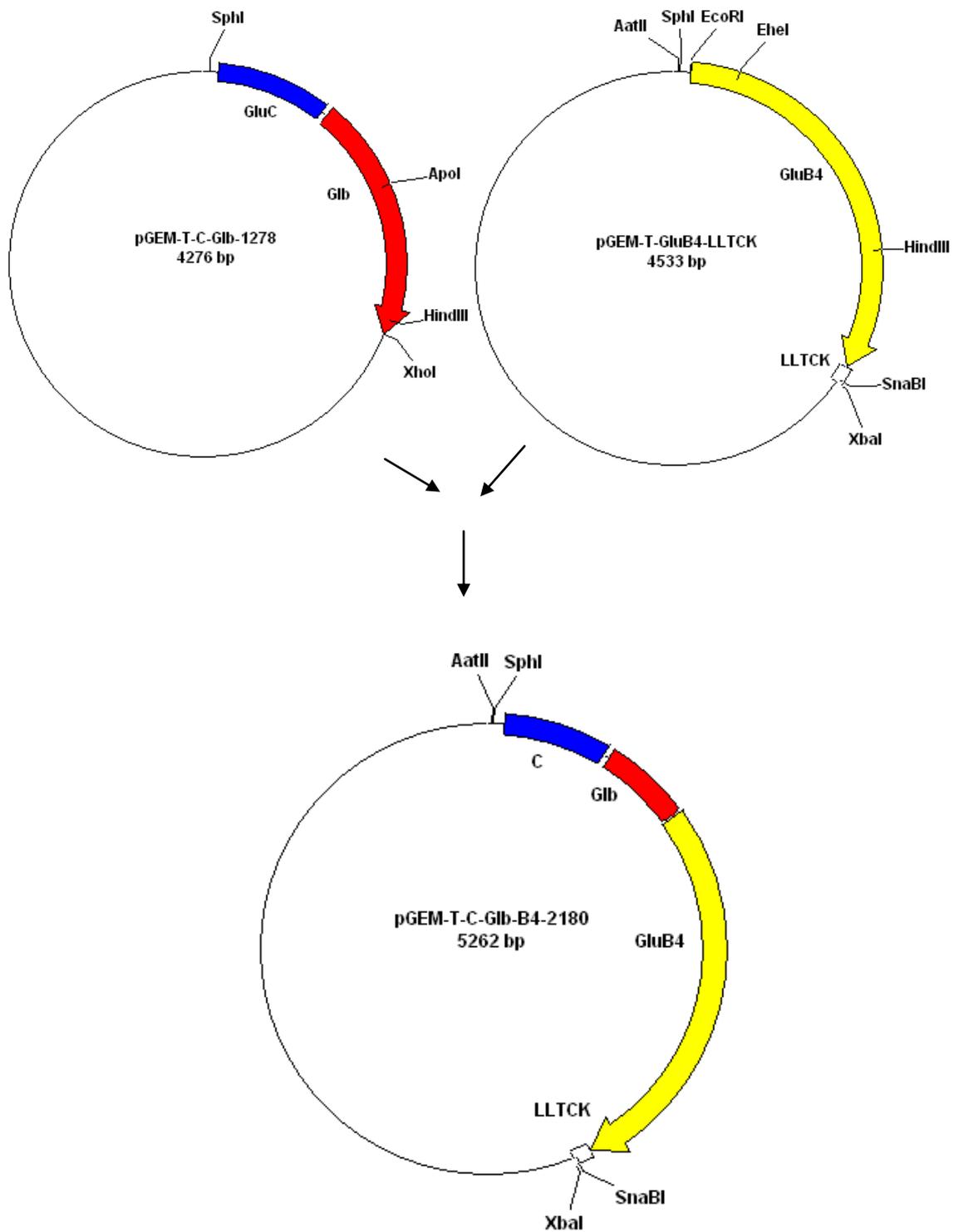


Figura 36. Costruzione del promotore ternario C-Glb-B4.

Per costruire il promotore ternario C-Glb-B4-3087, la porzione distale di GluB4 è stata sostituita con l'inserto ricavato dalla digestione con *AatII* e *HindIII* del promotore ibrido C-Glb. L'inserto è stato saldato nel vettore aperto con la medesima coppia di enzimi.

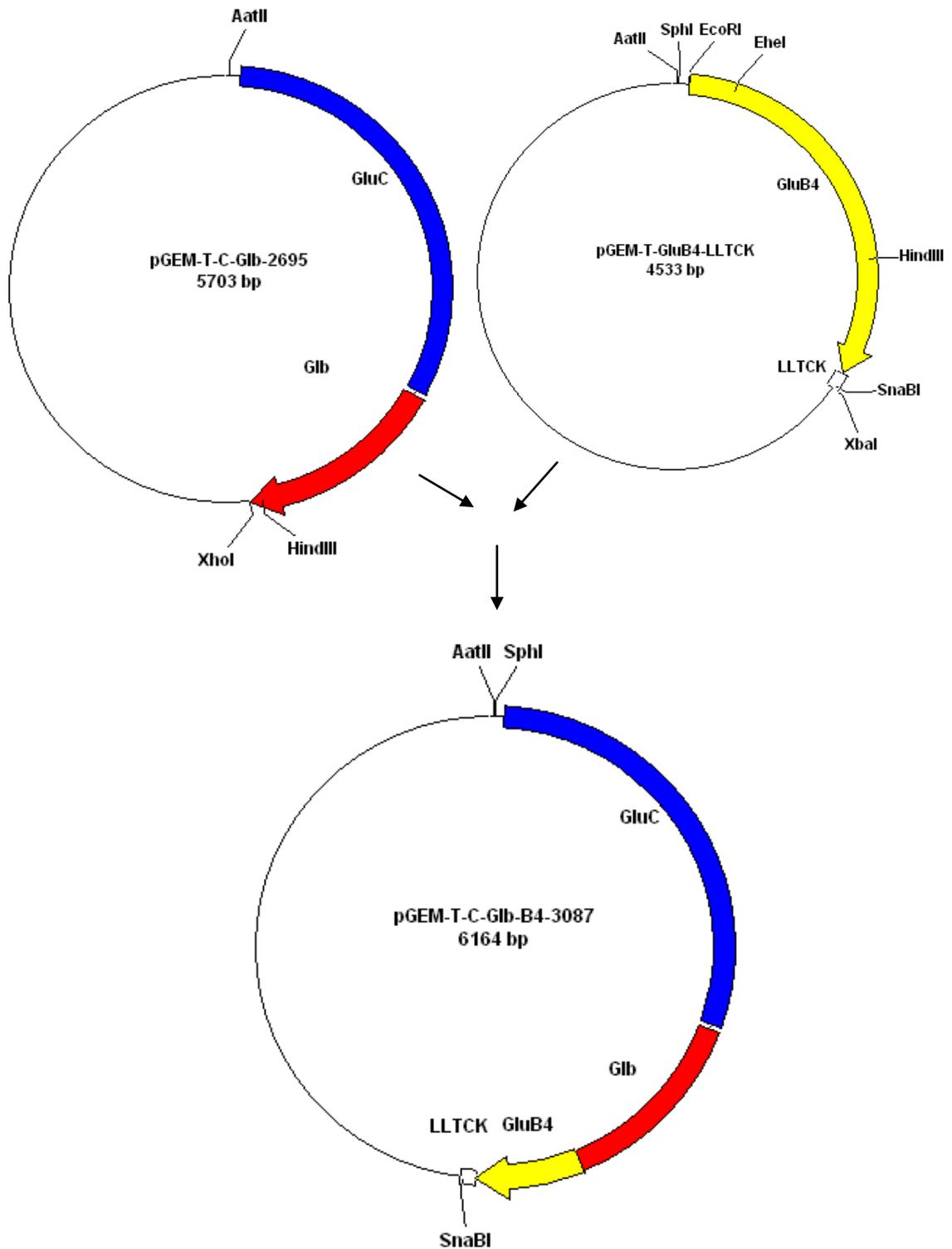


Figura 37. Sostituzione della porzione distale di GluB4, sita a monte di *HindIII*.

Per saldare la porzione distale di GluB1 al promotore della glutelina B4, sono stati sfruttati due siti che producono estremità piatte. *Bsa*AI ed *Ehe*I tagliano a 1 kb e a 200 bp circa dall'inizio dei due promotori rispettivamente.

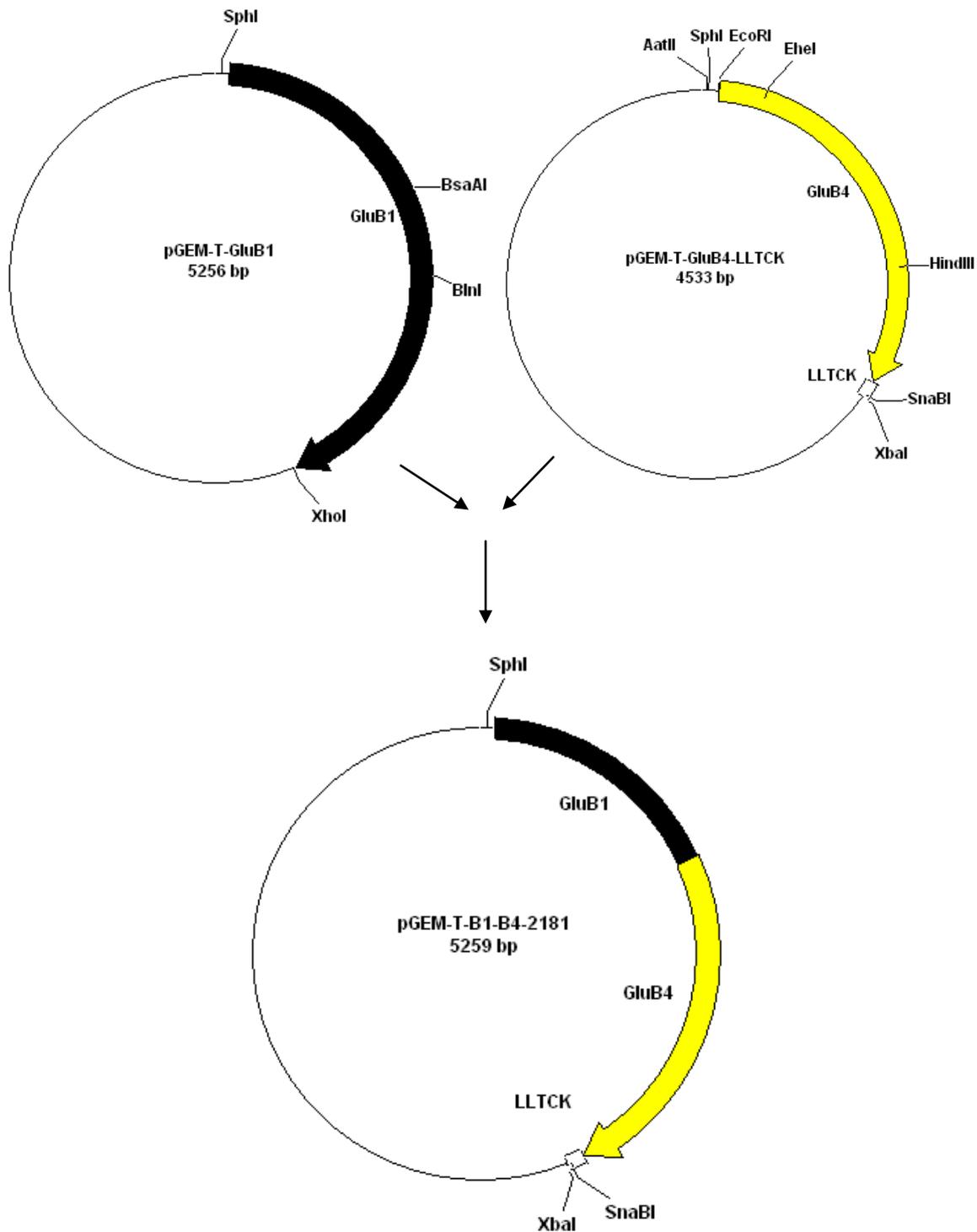


Figura 38. Costruzione di B1-B4. L'inserto derivato da GluB1 ha un'estremità piatta in 3', prodotta dall'enzima *Bsa*AI. Il vettore derivato da pGEM-T-GluB4-LLTCK viene aperto con *Sph*I ed *Ehe*I.

Successivamente alla saldatura dei vari inserti all'interno del vettore pGEM-T/GluB4::LLTCK e alla trasformazione di cellule JM101 di *E. coli*, 3 colonie putativamente trasformate per costrutto sono state verificate mediante PCR su coltura batterica per confermare la presenza dell'inserto Glb.

Per le colonie relative al promotore B1-B4 è stata utilizzata la coppia di *primers* GluB1*for* e GluB1*rev4*, che appaia nella regione iniziale di GluB1. La presenza di una banda di lunghezza appena inferiore a 500 bp ha confermato la presenza dell'inserto nelle colonie 1 e 3, mentre nella colonia 2 la mancanza di amplificato ha suggerito la presenza di un vettore richiuso.

Per le colonie relative al promotore C-B4 è stata utilizzata la coppia di *primers* GluC*for* e GluC*rev4*, che appaia nella regione iniziale del GluC. La presenza di una banda di poco superiore a 500 bp ha confermato la presenza dell'inserto nelle 3 colonie analizzate.

Per le colonie relative al promotore C-Glb-B4 è stata utilizzata la coppia di *primers* Glb*for2* e Glb*rev2*, che appaia nella regione centrale del Glb. La presenza di una banda intorno alle 150 bp ha confermato la presenza di Glb in tutte 3 le colonie. Per ulteriore conferma è stata utilizzata la coppia di *primers* GluB1*for* e GluC*rev4*. La presenza di una banda poco sopra le 500 bp ha confermato la presenza dell'inserto in tutte 3 le colonie.

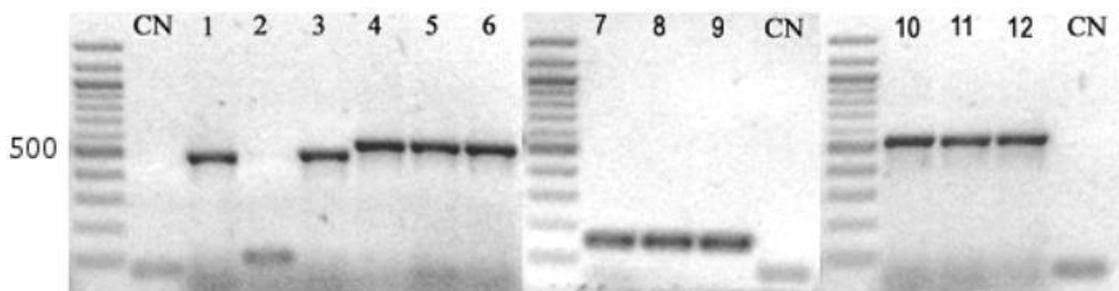


Figura 39. Gel di agarosio in cui sono stati caricati i prodotti ottenuti mediante PCR effettuata direttamente su coltura batterica. Corsia 1-3: B1-B4, colonie 1-3 *primers* GluB1 *for* e GluB1 *rev4*. Corsia 4-6: C-B4, colonie 1-3 *primers* GluC *for* GluC *rev4*. Corsia 7-9: C-Glb-B4, colonie 1-3 *primers* Glb *for2* e Glb *rev2*. Corsia 10-12 C-Glb-B4, colonie 1-3 *primers* GluC *for* e GluC *rev4*.

Da una colonia PCR-positiva per ciascun costrutto è stato preparato il DNA plasmidico per procedere a un'analisi di restrizione.

Tutte le digestioni hanno prodotto il *pattern* di restrizione atteso.

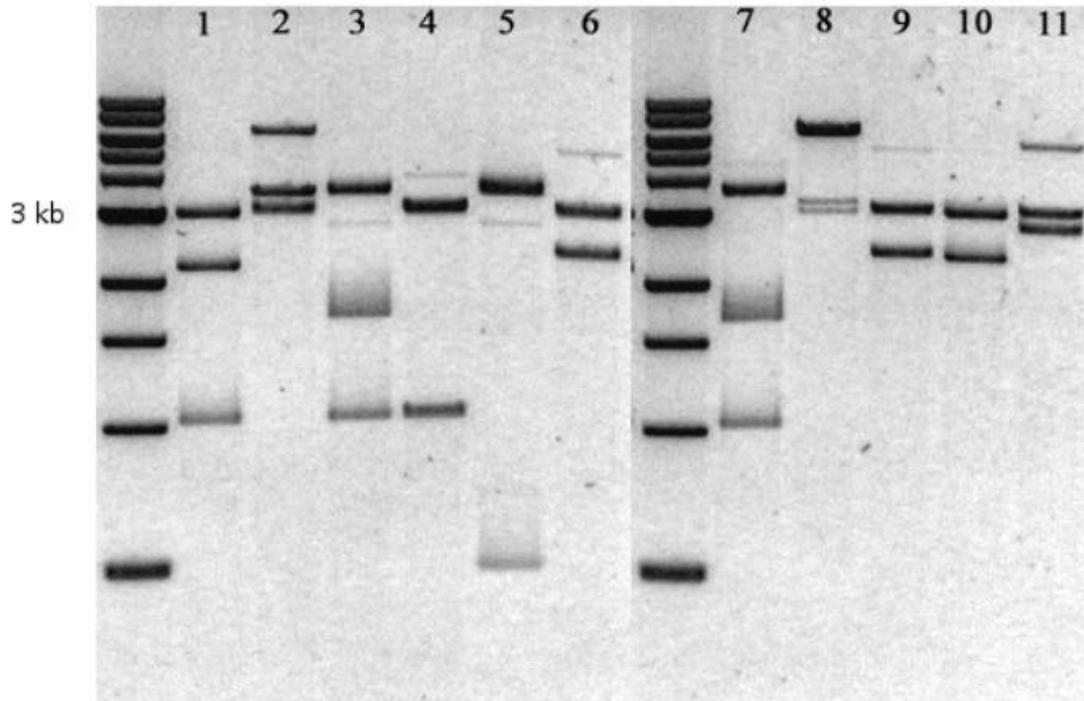
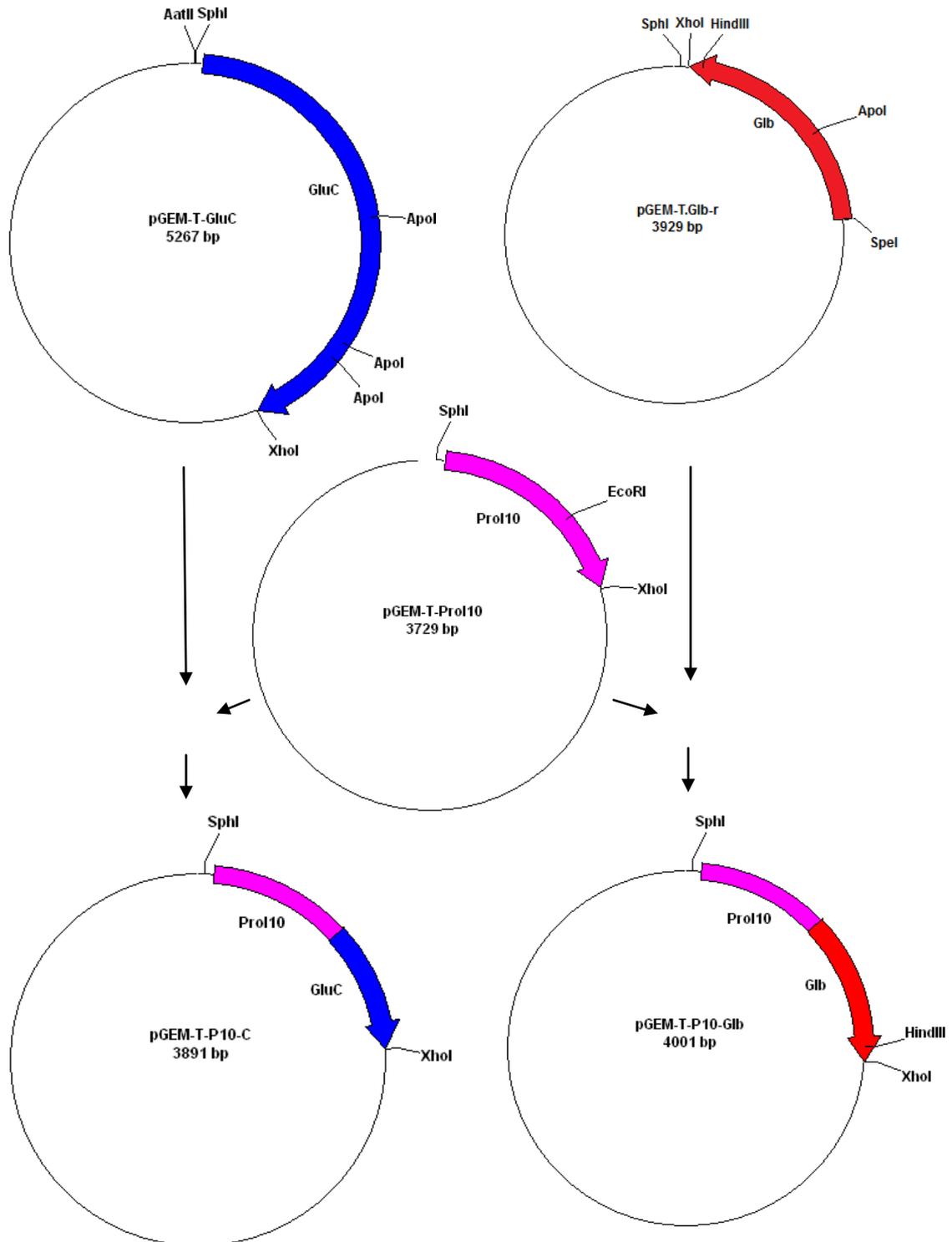


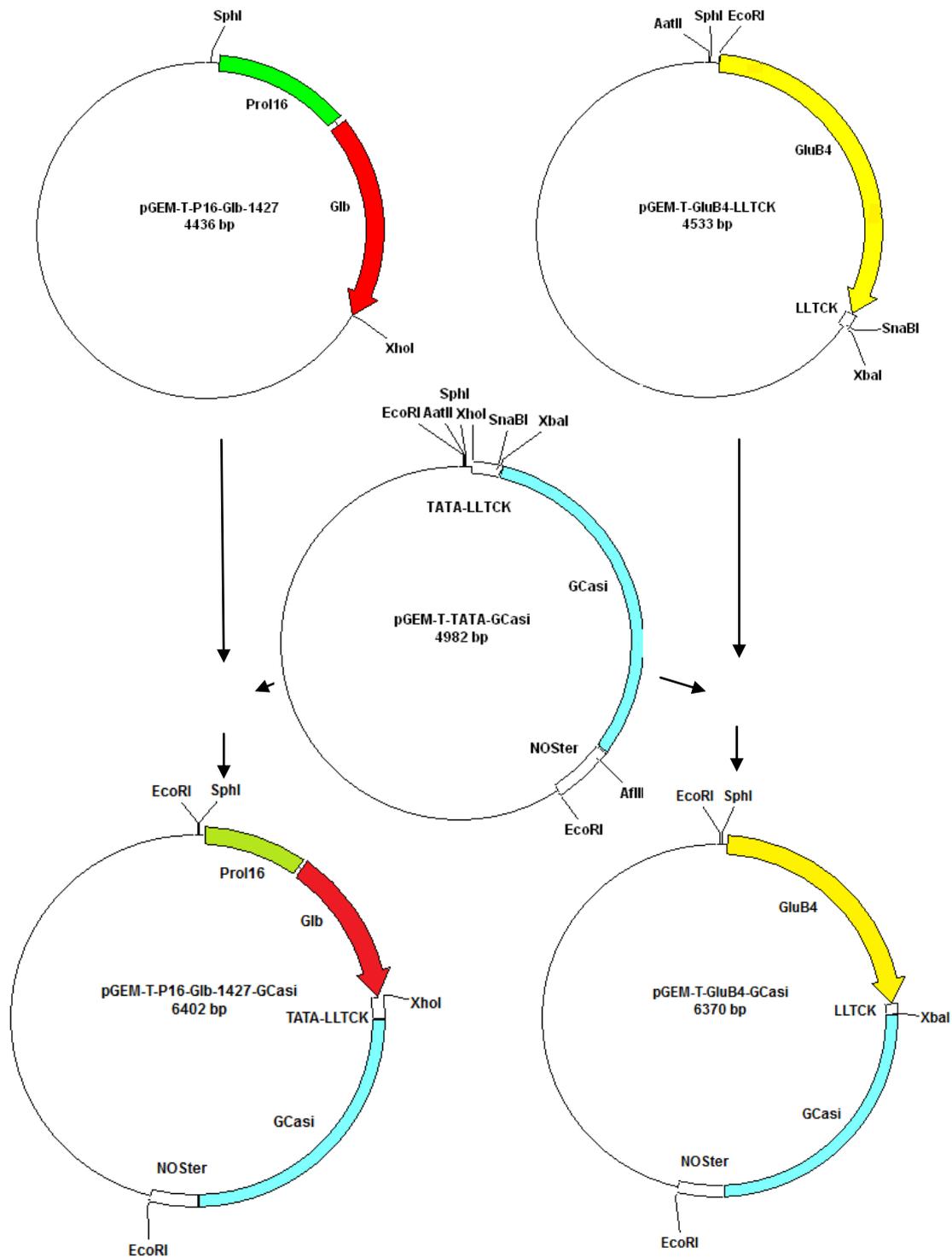
Figura 40. Corsia 1: C-Glb-B4-3087 digerito con *SphI* e *SnaBI*. Corsia 2: C-Glb-B4-3087 digerito con *AatII* e *MluI*. Corsia 3: C-Glb-B4-3087 digerito con *SphI* e *HindIII*. Corsia 4: Glb-B4 digerito con *AatII* e *SnaBI*. Corsia 5: Glb-B4 digerito con *SphI* e *HindIII*. Corsia 6: C-Glb-B4 digerito con *AatII* e *SnaBI*. Corsia 7: C-Glb-B4-3087 digerito con *SphI* e *HindIII*. Corsia 8: C-Glb-B4-3087 digerito con *AatII* e *SnaBI*. Corsia 9: B1-B4 digerito con *AatII* e *SnaBI*. Corsia 10: B1-B4 digerito con *SphI* e *SnaBI*. Corsia 11: C-B4 digerito con *AatII* e *SnaBI*.

Per ottenere i promotori sintetici derivati da Prol10, il plasmide pGEM-T/Prol10 è stato digerito con *EcoRI* e *XhoI*. GluC e Glb sono stati linearizzati con *XhoI* e successivamente digeriti con *ApoI*. Gli inserti ottenuti dal taglio tra *ApoI* e *XhoI* sono stati legati nel vettore tra *EcoRI* e *XhoI*.

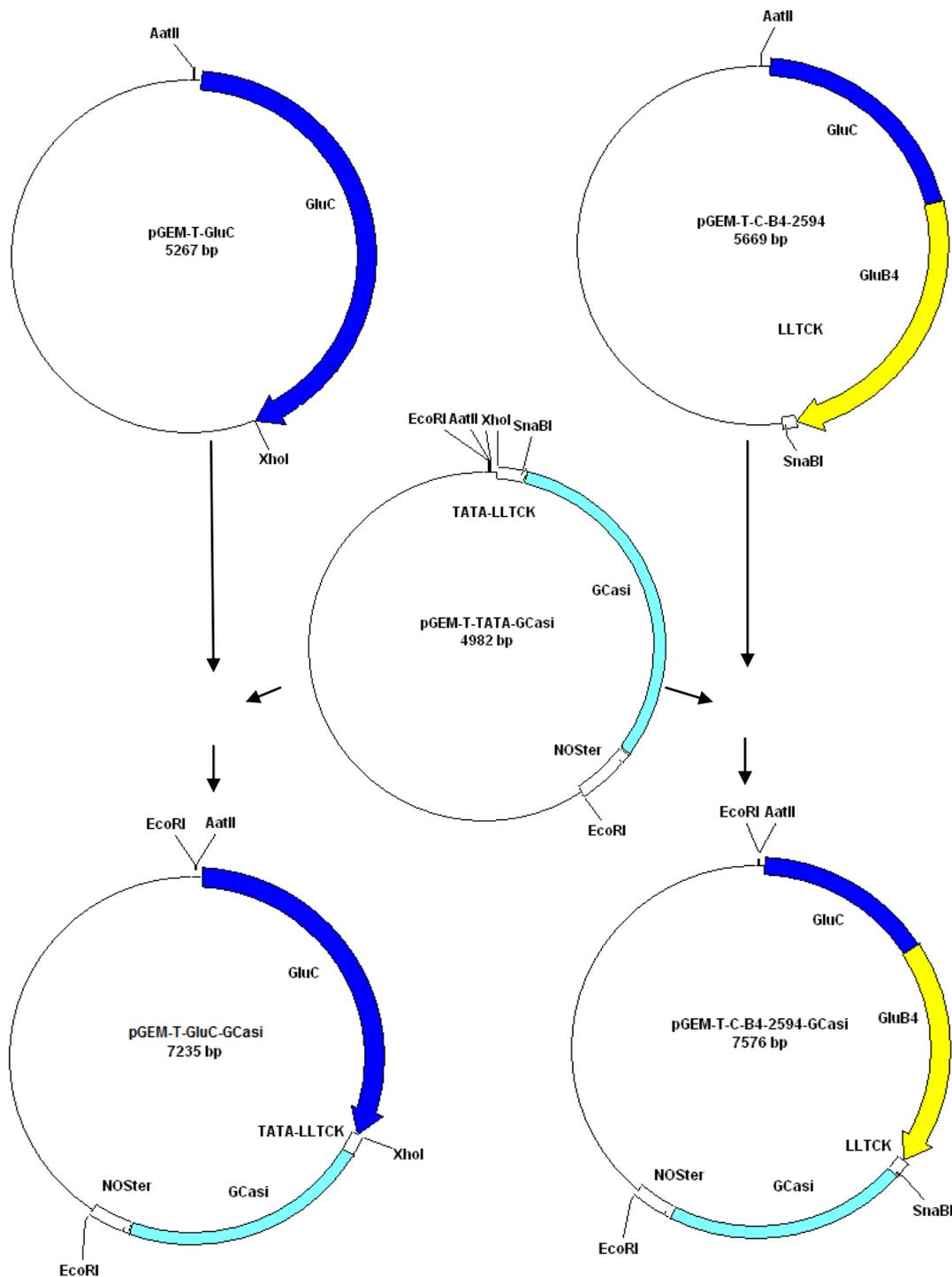


5.7.5 Assemblaggio delle cassette di espressione e del vettore di trasformazione

Per inserire i promotori sintetici nel modulo accettore sono state sfruttate diverse coppie di enzimi a seconda delle necessità.



Le sequenze prive di TATA-box sono state inserite fra *Sph*I e *Xho*I. I derivati di GluB4, che ne conservano TATA-box e 5'-UTR, sono stati inseriti direttamente a monte della CDS, fra *Sph*I e *Xba*I. Poiché il promotore della glutelina C contiene nella sua sequenza sia *Sph*I, sia *Xba*I, sono stati utilizzati enzimi di restrizione alternativi quali *Aat*II e *Sna*BI, rispettivamente.



Il modulo accettore è stato aperto in digestioni distinte effettuate con 4 diverse coppie di enzimi di restrizione per tenere conto delle diverse modalità di clonazione degli inserti nei rispettivi vettori: *SphI-XhoI*, *SphI-XbaI*, *AatII-XhoI* e *AatII-SnaBI*. I prodotti di digestione sono stati recuperati da gel e privati dei terminali fosforilati per evitare il fenomeno dell'autosaldatura. Nel contempo, i plasmidi contenenti i promotori naturali e sintetici sono stati digeriti, ciascuno con l'opportuna coppia di enzimi. Gli inserti sono stati recuperati da gel e purificati.

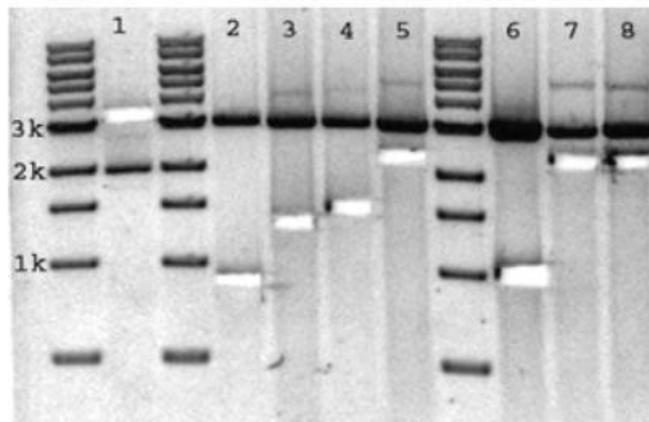


Figura 41. Gel da cui sono state recuperate le bande relative ai vari promotori. Corsia 1: pGEM-T/TATA::LLTCK::NOS digerito con *XbaI* e *SacI*. Corsia 2: pGEM-T/Glb digerito con *SphI* e *XhoI*. Corsia 3: pGEM-T/C-Glb digerito con *SphI* e *XhoI*. Corsia 4: pGEM-T/P16-Glb digerito con *SphI* e *XhoI*. Corsia 5: pGEM-T/B1-Glb, digerito con *SphI* e *XhoI*. Corsia 6: pGEM-T/Glb-B4 digerito con *SphI* e *XbaI*. Corsia 7: pGEM-T/C-Glb-B4 digerito con *SphI* e *XbaI*. Corsia 8: pGEM-T/B1-B4 digerito con *SphI* e *XbaI*.

Ciascun inserto è stato saldato nella opportuna forma di vettore. Successivamente alla trasformazione di cellule JM101, 6 colonie putativamente trasformate per costrutto sono state verificate mediante PCR su coltura batterica per confermare la presenza dell'inserto. Le colonie risultate positive, sono state poste in coltura *overnight* e i batteri ottenuti dalla coltura sono stati utilizzati per purificare il DNA plasmidico necessario all'analisi di restrizione. Il *pattern* ottenuto con gli enzimi *EcoRI* e *XhoI*, *XbaI*, o *SnaBI* è risultato quello atteso. *XhoI* ha sempre dato una banda di 2971 bp (vettore) e una di 1968 bp (gene *reporter*). Come atteso, la terza banda è risultata caratteristica del promotore: 1471 bp per P16-Glb, 2167 bp per B1-Glb, 1311 bp per C-Glb, 2302 bp per GluC, 906 bp per Glb. La

digestione con *EcoRI* e *XbaI* ha prodotto 2 bande di 2971 bp e 1843 bp più una banda caratteristica avente lunghezza: 2294 bp per B1-B4, 2297 bp per C-Glb-B4, 1012 bp per Glb-B4. Effettuando la digestione con *SnaBI* su C-B4, sono state ottenute le seguenti bande: 2971 bp, 1855 bp e la banda caratteristica di 2692 bp.

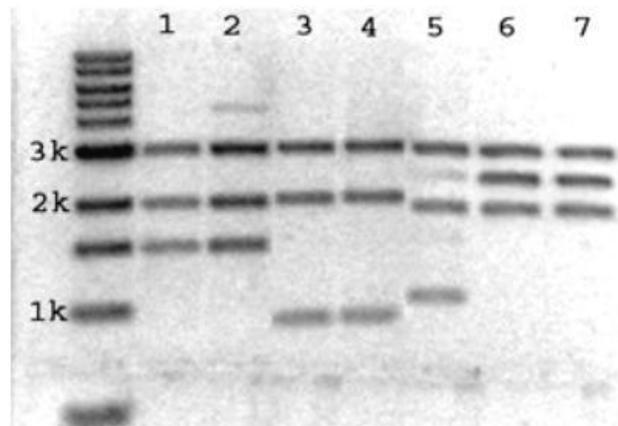


Figura 42. Analisi di restrizione delle cassette di espressione (prima parte). Corsie 1 e 2: pGEM-T/P16-Glb::*GCasi*, colonie 1 e 2, mp digerite con *XhoI* e *EcoRI*. Corsie 3 e 4: pGEM-T/Glb::*GCasi*, colonie 1 e 2, mp digerite con *XhoI* e *EcoRI*. Corsia 5: pGEM-T/Glb-B4::*GCasi*, colonia 1, mp digerita con *XhoI* e *EcoRI*. Corsie 6 e 7: pGEM-T/B1-B4::*GCasi*, colonia 5, mp digerita con *XbaI* ed *EcoRI*.

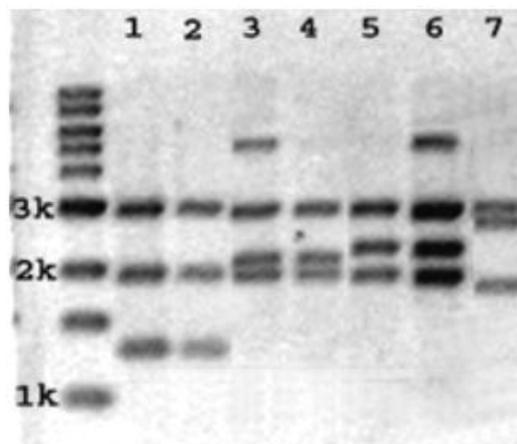


Figura 43. Analisi di restrizione delle cassette di espressione (seconda parte). Corsie 1 e 2: pGEM-T/C-Glb::*GCasi*, colonie 1 e 2, mp digerite con *XhoI* e *EcoRI*. Corsie 3 e 4: pGEM-T/B1-Glb::*GCasi*, colonie 2 e 4, mp digerite con *XhoI* e *EcoRI*. Corsie 5 e 6: pGEM-T/GluC::*GCasi*, colonie 1 e 2, mp digerite con *XbaI* e *EcoRI*. Corsia 7: pGEM-T/C-B4::*GCasi*, colonia 1, mp digerita con *SnaBI* ed *EcoRI*.

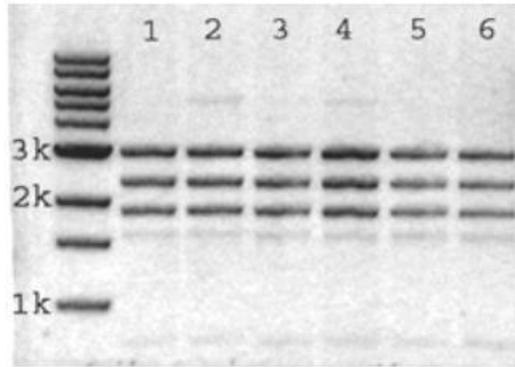


Figura 44. Analisi di restrizione della cassetta C-Glb-B4. Corsie 1-6: pGEM-T/C-Glb-B4::GCasi, colonie 1-6, mp digerite con *EcoRI* e *XbaI*.

5.7.6 Trasformazione di riso con *Agrobacterium tumefaciens*

La procedura di infezione del riso con cellule di *A. tumefaciens* portanti i costrutti di interesse si è svolta senza inconvenienti e secondo i tempi e modi previsti dal protocollo di riferimento. Di seguito vengono riportate alcune immagini fotografiche rappresentative dei punti fondamentali del processo di trasformazione genetica.

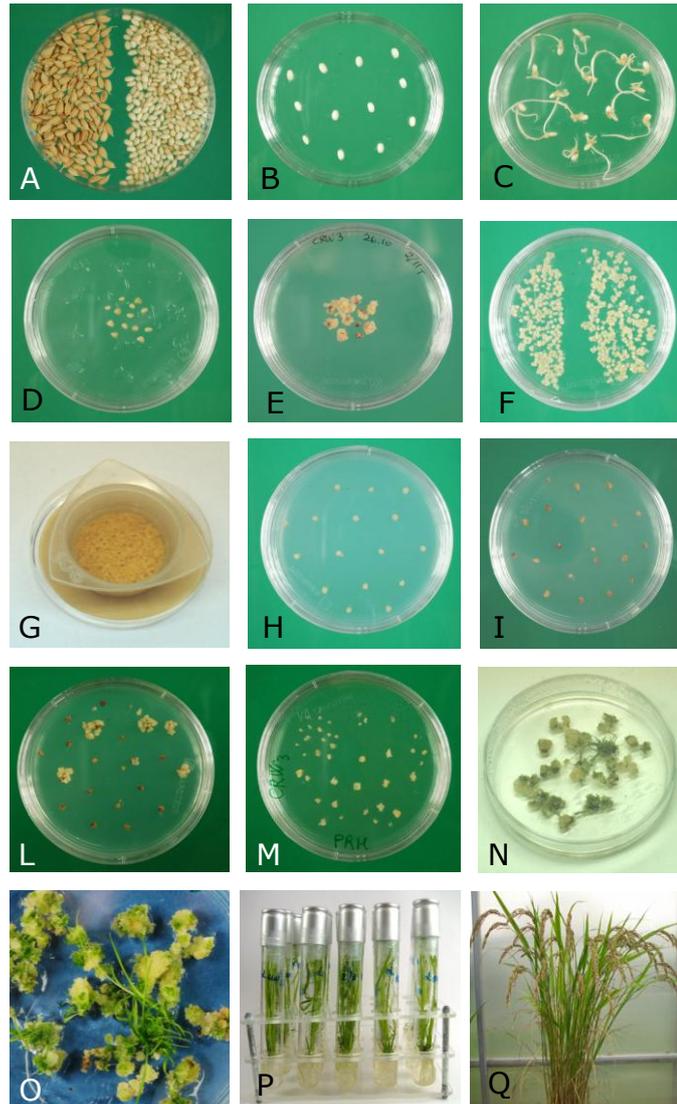


Figura 45. Tavola raffigurante le principali fasi di trasformazione del riso. A: sbramatura; B: disinfezione; C: germinazione; D: isolamento scutelli; E: callogenesi; F: selezione embrioidi; G: infezione con coltura liquida di agrobatterio; H: embrioidi su substrato CCM dopo infezione; I: embrioidi trasformati e non trasformati su substrato SMI; L: calli da embrioidi trasformati su substrato SMII; M: calli trasformati su substrato PRM; N: rigenerazione; O: differenziazione; P: radicazione; Q: pianta trasformata matura.

Nel corso di questo lavoro sono state condotte tre sessioni di trasformazione. Una sessione in occasione del primo ciclo, in cui il promotore tronco della glutelina C è stato confrontato con il promotore naturale della glutelina B4. La trasformazione dei promotori costruiti durante il secondo ciclo è stata divisa in due sessioni, una eseguita a novembre del 2011 in cui diversi promotori sintetici sono stati messi a confronto con due promotori naturali forti, GluC e Glb; un'altra, a febbraio 2012, in cui sono state ripetute tre trasformazioni della sessione principale, per verificare la ripetibilità del metodo in due periodi di coltivazione diversi. A febbraio del 2012 è stato mandato in trasformazione anche un terzo promotore naturale forte, ovvero il promotore della glutelina B4 al fine di fornire un ulteriore elemento verso cui confrontare i dati ottenuti con le sequenze artificiali.

Le trasformazioni sono state programmate con l'obiettivo di ottenere almeno 20 trasformati primari analizzabili per promotore. Le 3 sessioni di trasformazione sono state programmate tenendo conto che non tutte le piante rigenerate potranno essere utilizzate per l'analisi di progenie. È necessario infatti tener conto delle perdite in fase di coltivazione, della presenza all'interno della popolazione di rigeneranti di individui *escapes*, cioè in grado di superare la selezione pur non essendo transgenici, ed infine del fatto che alcuni individui possono avere una produttività talmente limitata da impedire una corretta analisi di progenie. Prudenzialmente è stato dunque previsto di rigenerare una trentina di piante per ogni costrutto. La quantità attesa di piante rigenerate è strettamente legata al numero di embrioidi che si infettano con ciascun costrutto. L'efficienza della fase relativa alla preparazione degli embrioidi è molto variabile. Per questo motivo la quantità di embrioidi da produrre è stata sempre calcolata in eccesso.

La trasformazione del primo ciclo è servita a individuare entro certi limiti la scala su cui operare. Questa trasformazione ha prodotto un numero di piante transgeniche largamente superiore a quanto necessario per l'analisi. Sulla base di questa esperienza si è deciso di porre a germinare 100 semi per costrutto. Sia per la sessione di novembre 2011, sia per quella

di febbraio 2012 il numero di embrioidi prodotti è stato più che soddisfacente.

Anche la quantità di piante generate per ciascun promotore è stata adeguata alle esigenze.

Sessione	Promotore	N° rigenerati
Novembre '11	C-Glb	32
	GluC	30
	C-B4	32
	Glb-B4	33
	P16-Glb	35
	Glb	36
	CN	18
	C-Glb-B4	28
	B1-Glb	33
	B1-B4	36
Febbraio '12	C-B4	40
	B1-B4	40
	GluC	41
	GluB4	43

Tabella 21. Numero di individui rigenerati per promotore.

5.7.7 Analisi molecolari sulle piante trasformate

Da tutte le piante dei tre gruppi sono stati prelevati 100 mg di lamina fogliare. Il DNA genomico è stato estratto e sottoposto ad analisi PCR come descritto nel capitolo 4.10. La percentuale di *escapes* si è attestata in media attorno al 10% ma è stata osservata a riguardo una notevole variabilità: la frequenza massima di piante *escapes* è stata registrata per GluC nella sessione di novembre 2011 (37%), mentre il valore minimo (0%) per i costrutti contenenti C-Glb e C-Glb-B4.

5.7.8 Produzione di seme

Anche tra individui appartenenti alla stessa popolazione sono state comunque rilevate variazioni di produttività tanto che in diversi casi non è stato possibile ottenere nemmeno 40 semi (1 g ca.) necessari alla valutazione del trasformato primario.

Tabella 22. Produttività media di seme registrata nelle diverse popolazioni di trasformati primari e nel controllo negativo (CN, var. CR W3 non trasformata).

Promotore	gr seme
C-Glb	8,5
GluC	8,1
C-B4	8,5
Glb-B4	6,9
P16-Glb	7,7
Glb	11,8
CN	16,7
C-Glb-B4	1,9
B1-Glb	8,6
B1-B4	2,9
Media	8,1

In tutti i casi la causa della ridotta produttività è sembrata riconducibile ad aborto del fiore o del seme in formazione e non a problemi legati allo sviluppo vegetativo. Notevole poi il fatto che anche all'interno del gruppo di controllo non trasformato molti individui non abbiano prodotto seme sufficiente, sebbene proprio in questo stesso gruppo siano stati riscontrati i picchi di produttività per pianta.

Tabella 23. Riassunto dei risultati generali ottenuti nel lavoro di trasformazione. >40 semi: numero di piante su cui è stato possibile condurre un'analisi dei livelli di espressione del gene *reporter*.

Sessione	Promotore	N° rigenerati	>40 semi	Escape	% Escape
Novembre '11	C-Glb	32	24	0	0%
	GluC	30	25	11	37%
	C-B4	32	25	1	3%
	Glb-B4	33	20	1	3%
	P16-Glb	35	27	4	11%
	Glb	36	34	4	11%
	CN	18	10	3	17%
	C-Glb-B4	28	18	0	0%
	B1-Glb	33	31	3	9%
	B1-B4	36	15	1	3%
Febbraio '12	C-B4	40	39	2	5%
	B1-B4	40	37	2	5%
	GluC	41	36	4	10%
	GluB4	43	36	10	23%
				Media	10%

5.7.9 Risultati del primo ciclo sperimentale

Per entrambe le popolazioni, è stato ottenuto un numero di piante largamente superiore a 30. Da ciascuna popolazione sono stati tratti 20 individui a caso e su ciascuno di essi sono stati effettuati due distinti campionamenti di 40 semi. E' stata determinata la concentrazione di GAA in ciascun campione secondo la procedura descritta nel capitolo 4.11. Di ciascuna replica di campionamento è stata calcolata la media e la deviazione standard di popolazione. Sulla base delle medie è stata stilata una classifica per ciascun promotore.

Tabella 24. Contenuto medio di GAA e deviazione standard rilevati su 2 repliche di campionamento della progenie. ID è il codice identificativo dell'individuo

GluB4			GluΔC		
ID	Media	Dev. standard	ID	Media	Dev. standard
12c	80,14	11,04	27	54,77	7,02
16c	63,60	1,19	46	40,75	5,92
31c	54,49	1,83	25	39,84	1,14
6c	49,72	4,95	44	38,82	3,88
54c	41,95	1,88	39	33,53	3,67
63c	40,60	2,03	26	31,37	3,39
49c	40,07	1,16	42	28,79	1,93
61c	39,36	1,78	23	27,20	3,11
28c	37,62	10,05	6	26,89	1,50
58c	35,42	4,66	34	26,51	2,75
48c	33,75	10,41	30	24,77	3,84
60c	32,62	12,13	14	21,80	1,76
36c	31,60	12,24	35	19,66	2,30
24c	29,28	0,33	31	17,94	2,96
32c	27,65	14,93	24	14,61	0,61
10c	26,49	11,27	20	13,70	1,14
41c	23,05	1,78	10	12,80	0,20
8c	21,58	1,16	47	12,32	1,23
22c	20,36	6,79	54	11,34	0,44
33c	13,53	1,61	49	8,96	2,00

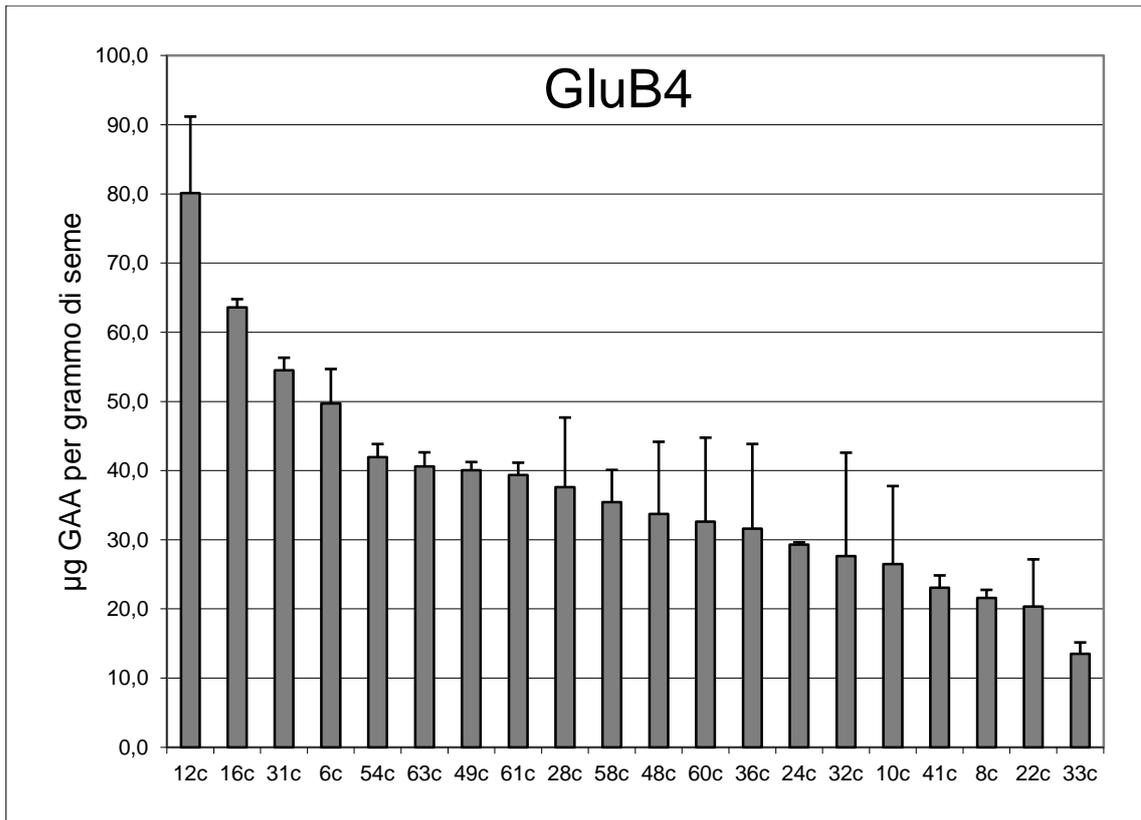


Grafico 5. Promotore GluB4: distribuzione dei livelli di espressione all'interno del campione della popolazione di trasformati primari. Le barre indicano la deviazione standard.

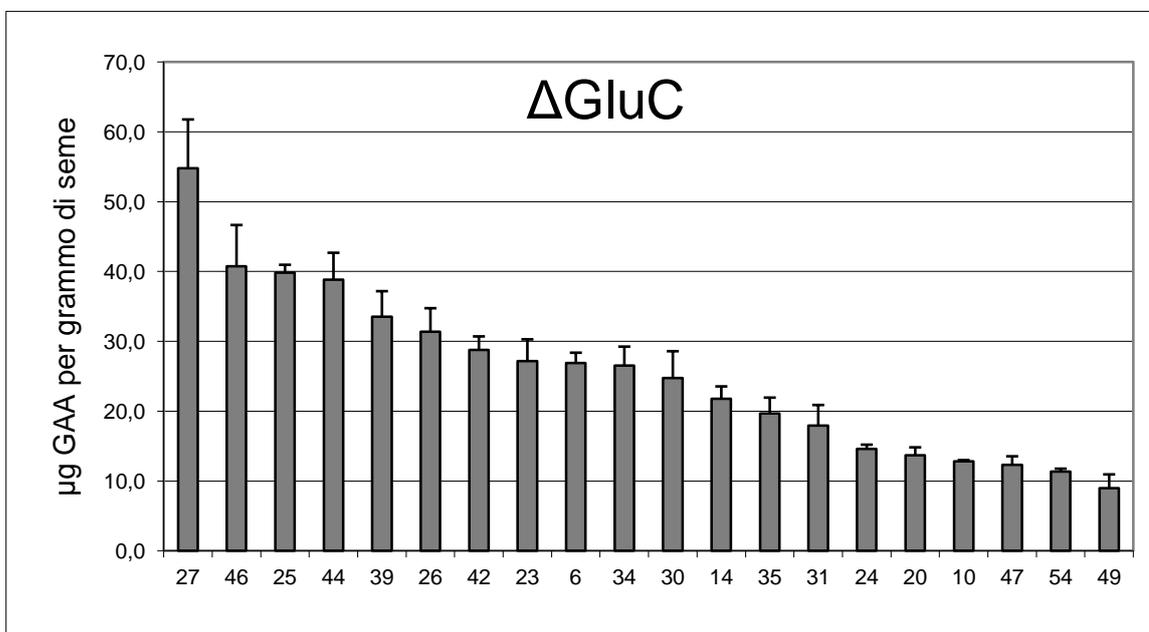


Grafico 6. Promotore Δ GluC: distribuzione dei livelli di espressione all'interno del campione della popolazione di trasformati primari . Le barre indicano la deviazione standard.

Sulla base delle due classifiche sono stati ottenuti per ciascun promotore i seguenti parametri:

- il valor medio dei livelli di espressione dei singoli trasformati primari analizzati;
- il valor medio dei livelli di espressione dei trasformati primari che occupano il primo quarto di classifica;
- il valor massimo per ciascuna delle due classifiche;

Tabella 25. Classifiche del livello di espressione del transgene in trasformati primari appartenenti ai due gruppi. Per ciascun promotore è stata calcolata la media dell'intero campione e la media del primo quartile, ovvero dell'insieme dei trasformati primari che occupano il primo quarto di classifica.

GluB4		GluΔC	
ID	μg/gr	ID	μg/gr
12c	80,14	27	54,77
16c	63,60	46	40,75
31c	54,49	25	39,84
6c	49,72	44	38,82
54c	41,95	39	33,53
63c	40,60	26	31,37
49c	40,07	42	28,79
61c	39,36	23	27,20
28c	37,62	6	26,89
58c	35,42	34	26,51
48c	33,75	30	24,77
60c	32,62	14	21,80
36c	31,60	35	19,66
24c	29,28	31	17,94
32c	27,65	24	14,61
10c	26,49	20	13,70
41c	23,05	10	12,80
8c	21,58	47	12,32
22c	20,36	54	11,34
33c	13,53	49	8,96
I° quartile	57,98	I° quartile	41,54
Media	37,14	Media	25,32

5.7.10 Risultati della sessione di trasformazione effettuata a novembre 2011

Come già accennato nel paragrafo precedente, le piante prodotte con la trasformazione di novembre 2011 hanno evidenziato problemi di produttività e ciò ha comportato una riduzione del numero di individui analizzabili.

Le piante sono state allevate in serra e sono andate in fioritura a maggio. Nel mese di giugno sono state raccolte le prime pannocchie mature. Le operazioni di sfarinatura, estrazione e quantificazione dei campioni sono state svolte il più possibile in un unico blocco temporale, per evitare di introdurre variabilità nell'esperimento.

In questa trasformazione è stato inserito anche un controllo non trasformato. La popolazione ottenuta per questo controllo è stata analizzata con le medesime modalità utilizzate per gli altri costrutti. Per tutti i membri della popolazione di controllo negativo, il segnale ELISA rilevato è risultato trascurabile. Nel caso più elevato è stato registrato un rumore di fondo pari a 0.078 $\mu\text{g GCasi}$ per g di farina estratta (pianta 7-13). Si è deciso pertanto di considerare questo dato come soglia minima di espressione e di eguagliare a zero tutti i valori uguali o inferiori ad esso. Va comunque considerato che questa evenienza si è verificata una sola volta, nel caso della pianta 9-34. Anche il livello di espressione, come la produttività di seme, è risultato variare molto all'interno delle singole popolazioni. Per ogni promotore studiato è stata stilata una classifica dei trasformati primari e calcolato il valore medio di espressione nonché la media del primo quartile.

Tabella 26. Livelli di espressione del transgene rilevati nelle piante trasformate nel mese di novembre 2011. La tabella continua nella pagina successiva.

C-Glg		GluC		C-B4		Glb-B4	
ID	µg/gr	ID	µg/gr	ID	µg/gr	ID	µg/gr
1-12	20,32	2-11	28,04	3-18	27,10	4-16	36,18
1-2	17,00	2-10	22,93	3-32	21,77	4-3	35,85
1-24	16,46	2-28	7,97	3-22	19,86	4-4	27,19
1-15	16,11	2-1	5,52	3-2	19,02	4-29	15,13
1-4	14,26	2-18	5,52	3-8	16,46	4-33	13,68
1-22	14,20	2-3	4,88	3-17	16,19	4-13	12,90
1-21	12,06	2-21	4,34	3-5	15,68	4-6	11,60
1-27	11,51	2-4	3,95	3-7	13,49	4-5	11,59
1-19	10,28	2-15	3,45	3-24	12,43	4-22	10,93
1-8	8,92	2-5	3,17	3-33	8,48	4-31	9,45
1-30	6,13	2-6	2,62	3-26	7,25	4-1	9,33
1-25	6,07	2-2	1,70	3-21	5,67	4-17	8,65
1-20	6,03	2-27	0,75	3-28	4,20	4-8	8,52
1-5	5,97	2-30	0,18	3-31	4,20	4-15	8,50
1-18	5,95			3-11	3,20	4-19	8,43
1-28	5,26			3-14	3,06	4-23	7,76
1-6	2,78			3-15	3,05	4-28	7,51
1-26	1,32			3-29	2,78	4-2	7,47
1-32	0,66			3-25	2,75	4-21	6,88
1-31	0,33			3-9	2,75		
1-11	0,00			3-12	2,54		
1-23	0,00			3-16	2,10		
1-3	0,00			3-1	1,94		
1-29	0,00			3-10	1,40		
I° quart.	16,39	I° quart.	16,11	I° quart.	20,07	I° quart.	25,61
Media	7,57	Media	6,79	Media	9,06	Media	13,55

Tabella 27. Livelli di espressione rilevati nelle piante trasformate nel mese di novembre 2011. La tabella continua nella pagina successiva.

P16-Glb		Glb		C-Glb-B4	
ID	µg/gr	ID	µg/gr	ID	µg/gr
5-2	35,36	6-32	19,91	8-9	49,28
5-1	25,43	6-18	18,39	8-17	45,05
5-22	25,31	6-22	17,71	8-16	33,66
5-13	21,28	6-33	14,63	8-19	31,04
5-30	19,66	6-36	14,32	8-14	27,60
5-14	19,38	6-19	13,57	8-23	23,32
5-35	18,08	6-25	10,88	8-7	21,64
5-10	18,06	6-7	10,64	8-28	19,91
5-7	15,16	6-20	10,06	8-1	19,51
5-24	13,66	6-27	9,09	8-21	18,15
5-9	13,29	6-23	8,88	8-4	17,89
5-16	10,92	6-10	8,46	8-25	17,33
5-26	9,06	6-4	8,29	8-11	17,09
5-18	8,97	6-21	7,89	8-3	16,78
5-15	8,72	6-28	7,63	8-24	16,48
5-28	6,71	6-29	7,60	8-27	15,88
5-12	6,53	6-24	7,51	8-15	14,18
5-20	5,16	6-12	7,15	8-6	11,19
5-25	4,30	6-31	7,10		
5-27	2,95	6-16	6,71		
5-19	2,46	6-17	6,36		
5-6	1,10	6-34	4,95		
5-4	0,45	6-3	3,82		
		6-35	3,25		
		6-5	2,28		
		6-9	1,11		
		6-30	0,87		
		6-26	0,46		
		6-8	0,00		
		6-13	0,00		
I° quartile	24,40	I° quartile	15,63	I° quartile	39,76
Media	12,70	Media	7,98	Media	23,11

Tabella 28. Livelli di espressione rilevati nelle piante trasformate nel mese di novembre 2011. La tabella continua dalle pagine precedenti.

B1-Glb		B1-B4		CN	
ID	µg/gr	ID	µg/gr	ID	µg/gr
9-27	23,04	10-32	7,29	7-6	0,00
9-24	18,95	10-37	6,77	7-17	0,00
9-5	8,09	10-14	6,04	7-2	0,00
9-18	8,08	10-2	5,62	7-18	0,00
9-23	7,72	10-3	4,29	7-5	0,00
9-16	7,34	10-24	3,50	7-16	0,00
9-26	6,55	10-35	3,48	7-15	0,00
9-17	5,86	10-31	3,43	7-10	0,00
9-21	5,31	10-33	3,07	7-12	0,00
9-3	4,92	10-28	2,75	7-13	0,08
9-8	4,89	10-36	2,69	7-1	0,06
9-20	4,85	10-11	2,63	7-4	0,06
9-31	4,75	10-5	2,53	7-8	0,00
9-6	4,69	10-20	2,08		
9-1	4,40	10-10	0,12		
9-14	4,27				
9-11	4,26				
9-30	4,19				
9-2	3,92				
9-12	3,36				
9-7	2,79				
9-33	2,60				
9-22	2,49				
9-25	2,34				
9-29	2,33				
9-13	1,87				
9-28	1,26				
9-34	0,00				
I° quartile	11,39	I° quartile	6,43		
Media	5,54	Media	3,75		

Per stilare la classifica dei promotori, è stato preso a riferimento il valore medio di ciascuna popolazione. Quale sia in realtà il parametro migliore per valutare l'efficacia di un promotore è argomento che verrà dibattuto nella sezione successiva.

Tabella 29. Classifica riassuntiva in cui i 9 promotori sono ordinati in base al valore medio di espressione registrato nei trasformati primari analizzati.

Novembre 2011	Parametri in µg/g		
Nome promotore	Massimo	Media I quart.	Media
C-Glb-B4	49,3	39,8	23,1
Glb-B4	36,2	25,6	13,6
P16-Glb	35,4	24,4	12,7
C-B4	27,1	20,1	9,1
Glb	19,9	15,6	8,0
C-Glb	20,3	16,4	7,6
GluC	28,0	16,1	6,8
B1-Glb	23,0	11,4	5,5
B1-B4	7,3	6,4	3,8

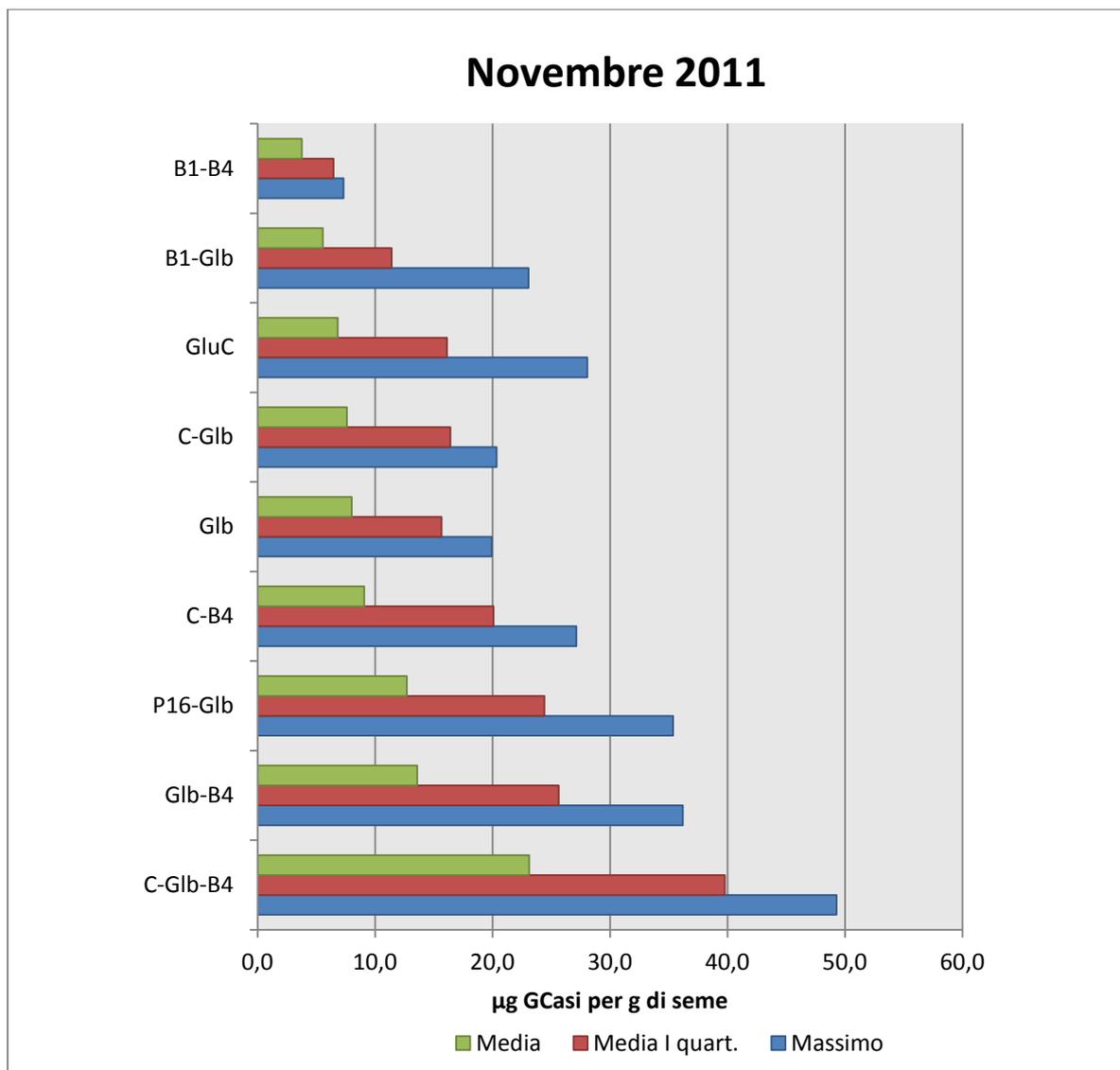


Grafico 7. Istogramma che evidenzia per ciascuno dei 9 promotori il valore assunto da media di popolazione, media del I quartile, e massimo. I promotori sono ordinati secondo un livello crescente di espressione indotta.

5.7.11 Livelli di espressione nel seme delle piante transgeniche prodotte a febbraio 2012

In questa sessione 3 promotori saggiati nella sessione precedente sono stati messi a confronto con il promotore della glutelina B4. Le piante prodotte con la trasformazione di febbraio 2012 sono state coltivate nella stagione ideale, conseguentemente la produttività dei singoli trasformati è risultata decisamente più elevata rispetto alla sessione di trasformazione precedente e ciò ha contribuito a una maggiore ampiezza delle popolazioni da saggiare.

Tabella 30. Livelli di espressione del transgene nei trasformati primari ottenuti con la seconda sessione di trasformazione.

C-B4b		B1-B4b		GluC		GluB4	
ID	µg/g	ID	µg/g	ID	µg/g	ID	µg/g
3bis-28	33,01	10bis-6	17,29	11-26	34,45	12-40	26,11
3bis-18	17,87	10bis-7	14,87	11-35	17,71	12-28	25,30
3bis-19	16,93	10bis-14	14,86	11-5	16,01	12-27	20,93
3bis-30	15,73	10bis-38	13,40	11-9	12,74	12-36	16,75
3bis-20	12,99	10bis-16	8,27	11-15	10,57	12-20	14,69
3bis-31	12,43	10bis-31	8,23	11-34	10,09	12-19	10,67
3bis-36	11,94	10bis-40	7,09	11-7	8,62	12-11	10,58
3bis-40	10,85	10bis-17	6,83	11-36	8,53	12-2	10,26
3bis-34	10,40	10bis-25	5,47	11-37	8,19	12-33	10,02
3bis-32	9,96	10bis-35	5,34	11-23	6,25	12-26	9,32
3bis-16	9,33	10bis-36	5,07	11-24	5,18	12-35	9,11
3bis-27	8,99	10bis-11	4,85	11-12	4,98	12-12	8,16
3bis-17	8,42	10bis-39	4,69	11-18	4,85	12-1	7,74
3bis-1	6,77	10bis-21	4,45	11-33	4,59	12-24	7,58
3bis-24	6,59	10bis-33	4,12	11-1	4,36	12-14	7,29
3bis-12	6,56	10bis-32	4,00	11-4	4,03	12-31	7,28
3bis-4	6,09	10bis-22	3,71	11-13	3,90	12-23	6,32
3bis-15	5,80	10bis-9	3,61	11-8	3,85	12-13	5,18
3bis-9	5,63	10bis-13	3,22	11-27	3,78	12-10	5,05
3bis-14	5,41	10bis-10	2,75	11-11	3,72	12-18	4,98
3bis-25	5,22	10bis-1	2,66	11-6	3,56	12-37	4,74
3bis-11	4,77	10bis-3	2,61	11-14	3,34	12-22	2,81
3bis-13	4,63	10bis-18	2,40	11-10	2,99	12-32	2,71
3bis-3	4,61	10bis-29	2,39	11-17	2,91	12-38	2,57
3bis-2	4,59	10bis-4	2,32	11-32	2,90	12-5	2,26
3bis-8	4,28	10bis-5	2,20	11-25	2,77	12-21	1,84
3bis-37	4,27	10bis-26	2,15	11-22	2,25	12-8	1,11
3bis-26	4,18	10bis-30	1,36	11-29	2,23	12-30	1,11
3bis-5	3,92	10bis-37	1,12	11-3	2,15		
3bis-6	3,62	10bis-26	0,39	11-38	1,65		
3bis-39	3,47	10bis-2	0,00	11-21	1,43		
3bis-22	3,33	10bis-19	0,00	11-31	0,29		
3bis-10	3,10	10bis-27	0,00				
3bis-7	2,59	10bis-12	0,00				
3bis-35	2,50	10bis-8	0,00				
3bis-29	2,03						
3bis-33	0,00						
I° qu.	15,79	I° qu.	10,70	I° qu.	14,84	I° qu.	17,86
Media	7,44	Media	4,62	Media	6,40	Media	8,66

Ciò nonostante, si sono comunque verificati casi di indisponibilità di un numero sufficiente di semi per l'analisi immunologica.

Anche in questo caso il criterio tenuto in considerazione per ordinare in classifica i promotori è il valore della media dei livelli di espressione di tutti i trasformati.

Tabella 31. Classifica riassuntiva dei quattro promotori trasformati a febbraio.

Febbraio 2012	Parametri in $\mu\text{g/g}$		
Nome promotore	Massimo	Media I quartile	Media
GluB4	26,1	17,9	8,7
C-B4	33,0	15,8	7,4
GluC	34,4	14,8	6,4
B1-B4	17,3	10,7	4,6

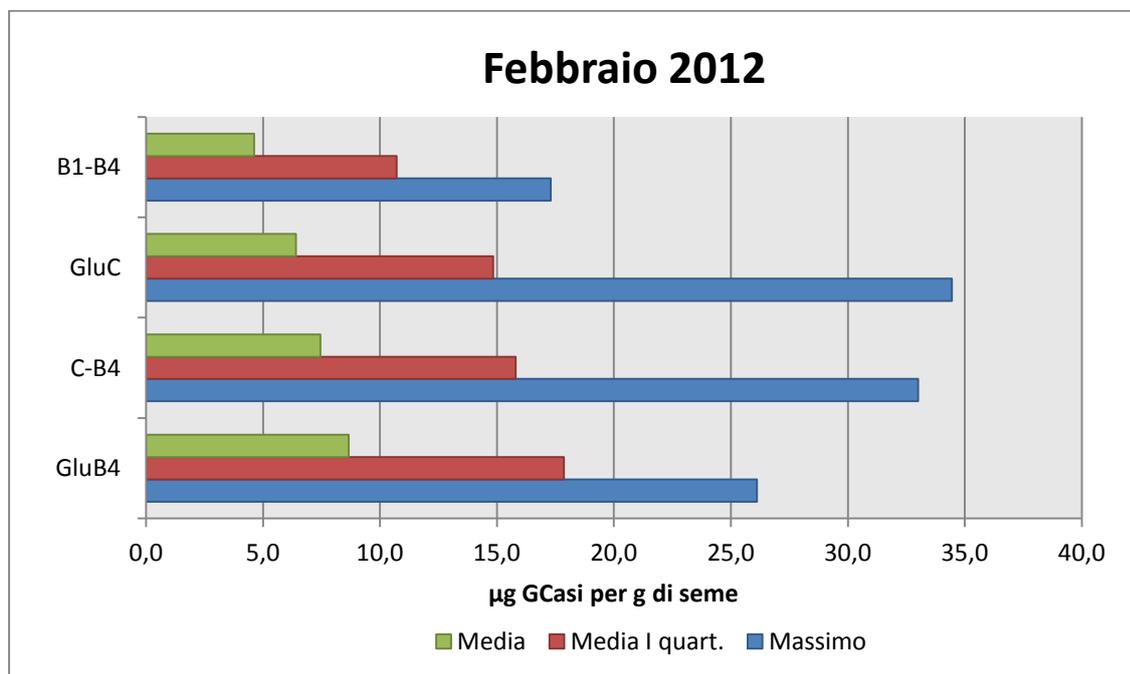


Grafico 8. Istogramma relativo alla sessione di trasformazione di febbraio 2012.

5.7.12 Estrazione di proteine da foglie dei trasformati primari

L'analisi del livello di espressione di proteina *reporter* nelle foglie di tutti i trasformati primari è stata effettuata solo per i promotori C-CIb-B4 e B1-B4.

Tabella 32. Concentrazione di proteina totale negli estratti da foglia. Valori misurati a 280nm con $e=1.7$.

Nome promotore	Pianta	Proteina mg/mL
C-GIb	1-12	1,3
GluC	2-11	2,6
	11-26	1,3
C-B4	3-18	1,8
GIb-B4	4-16	1,1
P16-GIb	5-2	1,7
GIb	6-32	1,2
CN	7-6	2,3
C-GIb-B4	8-9	3,2
	8-17	3,1
B1-GIb	9-27	2,0
B1-B4	10-32	1,4
	10-37	1,9
GluB4	12-40	1,4
	12-28	1,3

5.7.13 Estrazione di proteine da seme immaturo e da seme abortito

Sono stati preparati estratti proteici da piante trasformate con i costrutti GIb e C-GIb-B4. Il materiale su cui svolgere questo esperimento è stato ottenuto da culmi di accostamento prodotti dagli individui appartenenti ai due gruppi. I semi sono stati raccolti da piante che avevano mostrato un elevato livello di espressione del gene *reporter*. Per ottenere seme immaturo sono state individuate pannocchie all'inizio della fioritura, cioè quando le antere erano state emesse solo all'estremità apicale dell'infiorescenza.



Nelle pannocchie prescelte, scendendo dall'apice, è stato individuato l'ultimo seme ad aver emesso le antere. Sono stati raccolti singolarmente 8

152

semi per pannocchia, partendo due semi a monte dell'ultimo fiorito. Sono anche state raccolte pannocchie intere allo stadio di botticella.



Figura 46. Pannocchia allo stadio di botticella.

Tabella 33. Materiale raccolto da piante trasformate con Glb e con C-Glb-B4.

Promotore	ID pianta	Materiale prelevato
Glb	18	1 pannocchia botticella
	22	3 semi abortiti
	33	8 semi immaturi
C-Glb-B4	14	1 pannocchia botticella 7 semi immaturi
	16	4 semi abortiti
	17	2x8 semi immaturi

Ciascun seme immaturo o abortito è stato prelevato, pesato e messo in provetta eppendorf da 1.5 mL. E' stato aggiunto tampone di estrazione in ragione di 1:40, e il seme è stato omogenato con pestello. Il peso del seme immaturo è risultato compreso tra 3 e 5 mg, pertanto sono stati ottenuti 120-200 μ L di estratto proteico.

Tabella 34 Concentrazione di proteina totale negli estratti da seme immaturo e seme abortito. Valori misurati a 280nm con e=1.7.

ID	mg/ml	ID	mg/ml	ID	mg/ml
6-33-I	0,60	8-14-IV	0,41	8-17-1.II	0,40
6-33-II	0,49	8-14-V	0,63	8-17-1.III	0,54
6-33-III	0,48	8-14-VI	0,53	8-17-1.IV	0,34
6-33-IV	0,61	8-14-VII	0,58	8-17-1.V	0,55
6-33-V	0,76	8-14-VIII	0,63	8-17-1.VI	0,76
6-33-VI	0,43	8-14-IX	0,65	8-17-1.VII	0,34
6-33-VII	0,44	8-14-X	0,68	8-17-1.VIII	0,48
6-33-VIII	0,49			8-17-1.IX	0,39
ID	mg/ml	ID	mg/ml	ID	mg/ml
8-17-2.VI	1,13	6-22-VII	0,39	6-18 bott	2,81
8-17-2.VII	0,77	6-22-XI	0,42	8-14 bott	1,75
8-17-2.VIII	0,47	6-22-XII	0,73		
8-17-2.IX	0,66	8-16-III	0,92		
8-17-2.X	0,51	8-16-IV	0,59		
8-17-2.XI	0,48	8-16-V	0,51		
8-17-2.XII	0,32	8-16-VI	0,47		
8-17-2.XIII	0,54				

5.7.14 Prove di sbiancatura

Sono stati prelevati circa 10 g di seme dai migliori espressoristi per ciascun costrutto. In caso di gruppi con scarsa produttività di seme, è stato necessario scendere nella classifica fino alla decima posizione per accoppiare una sufficiente quantità di seme, lasciando comunque una scorta di sicurezza per ciascun trasformato; nei casi migliori è stato invece possibile ottenere la quantità desiderata attingendo solamente dalle 2 migliori linee. La sbiancatura è stata effettuata a mano mediante 5 passaggi attraverso i rulli gommati di uno sbiamino manuale da tavolo. L'asportazione delle glume ha portato a una perdita di circa il 20% del peso iniziale. Una parte delle glume è stata conservata per l'estrazione della proteina totale.

Il seme sbiamato è stato sottoposto a sbiancatura meccanica. Per ciascun promotore sono stati trattati circa 8 g di seme sbiamato. Questa quantità si è dimostrata sufficiente a essere processata nella sbiancatrice. E' però risultato impossibile recuperare in maniera precisa il farinaccio prodotto nei singoli passaggi. Da prove iniziali è risultato che sono sufficienti

30 secondi di trattamento per rimuovere efficacemente lo strato aleuronico. Tempi più prolungati causano un'evidente riduzione dell'endosperma.

Tabella 35. Peso in grammi dei campioni di farinaccio, riso sbramato e sbiancato.

Promotore	gr. sbramato	gr. farinaccio	gr. sbiancato
C-Glb	8,4	1,4	7,0
GluC	8,2	1,0	6,8
C-B4	8,6	1,1	7,4
Glb-B4	8,4	1,2	7,0
P16-Glb	8,9	1,2	7,5
Glb	7,7	1,1	6,5
C-Glb-B4	8,3	1,3	6,7
B1-Glb	7,7	1,1	6,6
B1-B4	7,2	1,2	5,9
GluB4	7,6	1,2	6,3

5.7.15 Estrazione della proteina totale dai prodotti di sbiancatura

Per eseguire tale estrazione, 100 mg di glume sono stati omogeneizzati in mortaio con 2 mL di tampone; 40 semi non sbiancati e 40 semi sbiancati per promotore sono stati macinati con mulino a palle. Per il materiale privato dello strato aleuronico, il protocollo è stato adattato come descritto nel 4.11.1.

Un procedimento diverso è stato seguito per il farinaccio recuperato dal processo di sbiancatura. Questa frazione presenta una granulometria molto disomogenea. In mezzo a una polvere fine si notano frammenti di qualche centinaio di micron. In una prima prova il farinaccio non è stato sottoposto a nessuna procedura di ulteriore macinazione prima di procedere all'estrazione della proteina totale. Visti i bassi livelli di espressione rilevati, successivamente si è proceduto a polverizzare il campione come descritto in 4.12.3.



Figura 47. Campioni di materiale prodotto nel processo di sbiancatura.

5.7.16 Livello di espressione di GCasi in foglia

Al saggio ELISA, tutti gli estratti da foglia hanno dato bassi segnali. Nella prova preliminare sono state testate 3 diluizioni, 1:10, 1:100, 1:1000. Solo alla diluizione 1:10 è corrisposto un segnale significativo.

Su 15 piante saggate per il promotore B1-B4, solo 9 hanno dato un segnale rilevabile. Il valore di concentrazione di GCasi ricavato dal saggio è stato normalizzato per la concentrazione della proteina totale, ottenendo le ppm di proteina *reporter* sulla proteina totale espressa in foglia. Il ppm è stato normalizzato rispetto al livello di espressione in endosperma.

Tabella 36. Espressione di GCasi in foglia nella pianta B1-B4. Nella prima colonna sono elencate le concentrazioni di GCasi nell'estratto, nella seconda la concentrazione di proteina totale nell'estratto. Il valore è stato ottenuto con saggio Bradford, utilizzando BSA per la retta di taratura. Nella terza colonna l'espressione della proteina *reporter* è espressa in parti per milione sulla proteina totale. La quarta colonna riporta i dati rilevati rispetto all'espressione di GCasi in endosperma. La quinta mette in rapporto il valore in ppm in foglia con il livello di espressione in endosperma per definire un parametro normalizzato, chiamato leakage.

Pianta	Foglia ng/μL	P Tot mg/mL	ppm	Endosp. μg/g	Leakage
10-36	0,014	1,1	13	2,69	4,7
10-28	0,023	2,5	9	2,75	3,4
10-11	0,018	2,6	7	2,63	2,6
10-24	0,017	3,3	5	3,50	1,4
10-32	0,011	2,3	5	7,29	0,7
10-5	0,010	3,2	3	2,53	1,2
10-2	0,009	2,0	5	5,62	0,8
10-20	0,006	1,9	3	2,08	1,6
10-31	0,006	2,1	3	3,43	0,8

Su 18 piante saggiate per il promotore C-Glb-B4, solo 11 hanno dato un segnale apprezzabile. La concentrazione di GCasi negli estratti da foglia per questo promotore è piuttosto alta rispetto a B1-B4.

Tabella 37. Espressione di GCasi in foglia per il promotore C-Glb-B4.

ID	Foglia ng/μL	P Tot mg/mL	ppm	Endosp. μg/g	Leakage
8-1	0,237	2,4	100	19,5	5,1
8-11	0,217	2,3	94	17,1	5,5
8-14	0,311	2,4	130	27,6	4,7
8-15	0,276	2,2	126	14,2	8,9
8-16	0,186	2,0	91	33,7	2,7
8-17	0,216	2,8	78	45,1	1,7
8-19	0,227	1,8	124	31,0	4,0
8-21	0,103	0,6	178	18,2	9,8
8-23	0,268	1,7	156	23,3	6,7
8-27	0,117	1,2	98	15,9	6,1
8-9	0,279	2,0	139	49,3	2,8

Sono stati saggiati in ELISA gli estratti da foglia dei migliori espressorii per ciascun promotore. In questo esperimento il rapporto tra

peso di foglia e volume di tampone è stato diminuito con l'intento di ottenere estratti più stabili.

Tabella 38. Espressione di GCasi in foglia nei migliori espressor di ciascun costrutto. La quarta colonna riporta i dati rilevati rispetto all'espressione di GCasi in endosperma. La quinta mette in rapporto il valore in ppm in foglia con il livello di espressione in endosperma per definire un parametro normalizzato, chiamato leakage.

Promotore	ID	Foglia µg/mL	P Tot mg/mL	ppm	Endosp. µg/g	Leakage
C-Glb	1-12	0,032	1,30	25	20,32	1,2
GluC	2-11	0,048	2,60	18	28,04	0,7
	11-26	0,016	1,30	13	34,43	0,4
C-B4	3-18	0,011	1,80	6	27,10	0,2
Glb-B4	4-16	0,049	2,10	23	36,18	0,6
P16-Glb	5-2	0,037	1,70	22	35,36	0,6
Glb	6-32	0,021	1,20	18	19,91	0,9
CN	7-6	0,002	2,30	1	0,00	n.d.
C-Glb-B4	8-9	0,213	3,20	66	49,28	1,3
	8-17	0,200	3,10	65	45,05	1,4
B1-Glb	9-27	0,045	2,00	23	23,04	1,0
B1-B4	10-32	0,008	1,40	6	7,29	0,8
	10-37	0,005	1,90	3	6,67	0,4
GluB4	12-40	0,030	1,40	21	26,11	0,8
	12-28	0,120	1,30	92	25,30	3,6

5.7.17 Livello di espressione di GCasi in vari tessuti

Tabella 39. Espressione di GCasi in singolo seme immaturo di piante trasformate con il promotore della globulina 26kDa.

Prom.	ID	Seme µg/mL	P Tot mg/mL	ppm	Endosp. µg/g	Leakage
Glb	6-33-I	0,068	0,6	113	14,6	7,7
	6-33-II	0,051	0,49	105	14,6	7,2
	6-33-III	0,028	0,48	59	14,6	4,1
	6-33-IV	0,039	0,61	63	14,6	4,3
	6-33-V	0,049	0,76	64	14,6	4,4
	6-33-VI	0,027	0,43	63	14,6	4,3
	6-33-VII	0,021	0,44	48	14,6	3,3
	6-33-VIII	0,023	0,49	48	14,6	3,3

Tabella 40. Espressione di GCasi in singolo seme immaturo di piante trasformate con il promotore C-Glb-B4.

Prom.	ID	Seme µg/mL	P Tot mg/mL	ppm	Endosp. µg/g	Leakage
C-Glb-B4	8-14-IV	0,061	0,41	148	27,6	5,4
	8-14-V	0,099	0,63	157	27,6	5,7
	8-14-VI	0,069	0,53	130	27,6	4,7
	8-14-VII	0,066	0,58	114	27,6	4,1
	8-14-VIII	0,070	0,63	111	27,6	4,0
	8-14-IX	0,047	0,65	72	27,6	2,6
	8-14-X	0,071	0,68	105	27,6	3,8
	8-17-1.II	0,035	0,40	87	47,6	1,8
	8-17-1.III	0,043	0,54	81	47,6	1,7
	8-17-1.IV	0,052	0,34	154	47,6	3,2
	8-17-1.V	0,085	0,55	155	47,6	3,3
	8-17-1.VI	0,123	0,76	162	47,6	3,4
	8-17-1.VII	0,056	0,34	163	47,6	3,4
	8-17-1.VIII	0,037	0,48	77	47,6	1,6
	8-17-1.IX	0,057	0,39	146	47,6	3,1
	8-17-2.VI	0,141	1,13	125	47,6	2,6
	8-17-2.VII	0,126	0,77	163	47,6	3,4
	8-17-2.VIII	0,081	0,47	172	47,6	3,6
	8-17-2.IX	0,097	0,66	147	47,6	3,1
	8-17-2.X	0,086	0,51	169	47,6	3,6
	8-17-2.XI	0,088	0,48	183	47,6	3,8
	8-17-2.XII	0,076	0,32	238	47,6	5,0
	8-17-2.XIII	0,140	0,54	259	47,6	5,5

L' esperimento programmato mirava a rilevare eventuali differenze significative fra il profilo di espressione di C-Glb-B4 e quello di Glb che giustificassero la ridotta produttività del promotore sintetico rispetto al naturale. Sono stati analizzati gli estratti da seme abortito. Sia nel gruppo del promotore della globulina, sia in quello del promotore ibrido C-Glb-B4, i semi abortiti sono risultati contenere una quantità residuale di GCasi, variabile tra 60 e 194 ppm.

Tabella 41. Espressione di GCasi in singolo seme abortito di piante trasformate con il promotore della globulina 26kDa e con il promotore C-Glb-B4.

Prom.	ID	Seme µg/mL	P Tot mg/mL	ppm	Endosp. µg/g	Leakage
Glb	6-22-VII	0,063	0,39	161	17,7	9,1
	6-22-XI	0,081	0,42	193	17,7	10,9
	6-22-XII	0,065	0,73	89	17,7	5,0
	6-18 bott	0,010	2,81	4	18,39	0,2
C-Glb-B4	8-16-III	0,077	0,92	83	33,7	2,5
	8-16-IV	0,049	0,59	83	33,7	2,5
	8-16-V	0,031	0,51	60	33,7	1,8
	8-16-VI	0,091	0,47	194	33,7	5,8
	8-14 bott	0,019	1,75	11	27,6	0,4

Gli estratti proteici da glume hanno mostrato un ridottissimo segnale malgrado si trattasse di campioni analizzati in ELISA alla minima diluizione possibile (1:4).

In base ai risultati riportati in tabella, si potrebbe ipotizzare che alcuni promotori inducano un'espressione del gene *reporter* nelle glume. Tuttavia, poiché in nessun caso è stata misurata un'espressione di GCasi superiore a 50 ppm, è molto verosimile che il segnale rilevato sia collegato a interferenze generate dalla particolare costituzione del campione e/o dalla presenza di tracce di altri tessuti nel campione.

Tabella 42. Espressione aspecifica di GCasi nelle glume delle piante appartenenti alle popolazioni ottenute dai promotori saggiati nel secondo ciclo.

Promotore	µg/mL	P Tot mg/mL	ppm
C-Glb	0,043	0,92	47
GluC	0,033	0,84	40
C-B4	0,044	1,04	43
Glb-B4	0,026	0,88	29
P16-Glb	0,012	0,92	13
Glb	0,007	0,33	20
C-Glb-B4	0,002	0,75	3
B1-Glb	0,006	0,79	8
B1-B4	0,003	0,56	6
GluB4	0,013	0,73	17

5.7.18 Quantificazione in ELISA di GCasi nei prodotti di sbiancatura

Tabella 43. Distribuzione di GCasi nelle varie frazioni provenienti dal processo di sbiancatura. sgl=seme sbramato; sb= seme sbiancato; far=farinaccio (residuo di sbiancatura).

Promotore	Fraz.	µg/mL	g tex	µg GCasi	%
C-Glb	sgl	0,17	8,4	17,42	
	sb	0,20	7,0	17,30	1%
	far	0,02	1,27	0,28	2%
GluC	sgl	0,10	8,18	10,38	
	sb	0,11	6,8	9,16	12%
	far	0,02	1,34	0,32	3%
C-B4	sgl	0,33	8,57	35,17	
	sb	0,27	7,4	25,39	28%
	far	0,05	1,12	0,73	2%
Glb-B4	sgl	0,33	8,36	34,03	
	sb	0,35	7,0	30,99	9%
	far	0,06	1,23	0,89	3%
P16-Glb	sgl	0,13	8,93	14,94	
	sb	0,16	7,5	15,24	0%
	far	0,07	1,18	1,07	7%
Glb	sgl	0,24	7,72	23,25	
	sb	0,22	6,5	17,79	24%
	far	0,10	1,14	1,37	6%
C-Glb-B4	sgl	0,19	8,34	19,41	
	sb	0,21	6,7	17,98	7%
	far	0,10	1,25	1,57	8%
B1-Glb	sgl	0,09	7,72	9,15	
	sb	0,12	6,6	10,02	0%
	far	0,03	1,13	0,47	5%
B1-B4	sgl	0,13	7,23	11,90	
	sb	0,12	5,9	9,16	23%
	far	0,09	1,23	1,39	12%
GluB4	sgl	0,26	7,63	24,61	
	sb	0,23	6,3	18,00	27%
	far	0,12	1,21	1,75	7%

6 DISCUSSIONE

In questa tesi sono stati presi in considerazione gli elementi che concorrono alla creazione di un sistema integrato di sviluppo di un promotore sintetico con profilo di espressione specifico per l'endosperma di riso. In particolare, sono stati approfonditi sia gli aspetti riguardanti il flusso di attività necessarie alla costruzione e alla valutazione delle prestazioni di nuovi candidati sintetici, sia quelli più specificatamente legati all'attività di progettazione.

I promotori endosperma-specifici inducono l'espressione di una classe di proteine, in genere proteine di riserva, la cui sintesi avviene in maniera circoscritta nello spazio e nel tempo. La pianta sintetizza queste proteine solo nell'endosperma, durante la maturazione del seme. L'interesse nei confronti dei promotori di questa classe di proteine è dovuto alla loro applicabilità per l'espressione efficiente e circoscritta al seme di proteine eterologhe.

Il mercato biofarmaceutico è in continua espansione e l'industria sta ricercando nuove piattaforme produttive che superino i limiti dei sistemi batterici e allo stesso tempo abbattano i costi dei sistemi di espressione animali. Una piattaforma basata sull'espressione stabile in pianta, seppur promettente in linea di principio, ha finora dimostrato alcuni limiti dovuti alla bassa resa di proteina ricombinante sulla biomassa. Con l'uso di promotori seme-specifici, negli ultimi 10 anni il problema delle basse rese è stato notevolmente attenuato. Non solo il seme è un organo naturalmente deputato all'accumulo di proteine, ma i promotori che inducono l'espressione in questo tessuto hanno permesso di raggiungere livelli promettenti di produzione di proteine ricombinanti. L'espressione nell'endosperma ha poi dimostrato di essere un ottimo modo per aggirare problemi dovuti alla fitotossicità della proteina eterologa nei confronti dell'organismo ospite (Patti *et al.*, 2012).

Il riso è fra le specie più promettenti per applicazioni biotecnologiche. La scelta di sviluppare una strategia di sintesi nell'endosperma di riso è motivata da diverse caratteristiche favorevoli riguardanti la pianta nel suo

complesso e il possibile sito di accumulo. In effetti, il riso presenta indubbi vantaggi a livello agronomico ma anche aspetti favorevoli legati a fattori di tipo genetico-molecolare; tra i primi vi sono: la brevità del ciclo biologico (140-170 gg), un sistema di unioni essenzialmente fondato sulla cleistogamia, elevate rese per ettaro (7 t/ha), una tecnica colturale semplice e scarsamente bisognosa di manodopera, considerato che molte operazioni agronomiche sono meccanizzabili. Tra i vantaggi legati alla possibilità di ingegnerizzarlo geneticamente vi sono: la disponibilità di efficienti protocolli di trasformazione, un genoma completamente sequenziato, l'esistenza di un parco varietale ampio comprendente varietà adatte alla sola utilizzazione industriale e dotate di marcatori genetici legati alla cariossida. Inoltre, a differenza di altre specie vegetali come ad es. il tabacco, il seme di riso può essere processato in modo da separare le componenti metabolicamente attive del seme, quali l'embrione, da quelle inerti, come l'endosperma. Tale aspetto risulta fondamentale nella produzione di proteine eterologhe fitotossiche; le proteine di interesse possono infatti essere accumulate in modo specifico nell'endosperma evitando possibili interferenze con il metabolismo della pianta e, in particolare, dell'embrione. Inoltre, nell'endosperma le proteine ritrovano un ambiente naturalmente predisposto al loro accumulo e a quello delle sostanze di riserva in una biomassa altamente conservabile e ricca di compartimenti subcellulari (Takaiwa *et al.*, 2007). In questo senso numerosi sforzi sono attualmente rivolti ad aumentare il livello di espressione di transgeni in endosperma di riso.

Qualunque sia la specie o il tessuto, i due principali fattori che influenzano l'accumulo di una proteina sono la velocità di sintesi e la velocità di degradazione. La velocità di sintesi è prevalentemente influenzata dalla velocità di trascrizione. Il rateo di trascrizione è fortemente determinato dall'efficienza del promotore. Questi presupposti fanno capire quanto sia importante la scelta del promotore nella creazione di una piattaforma tecnologica basata su piante transgeniche. I promotori più utilizzati in passato per l'espressione di proteine eterologhe in endosperma di riso sono quello della glutelina A2 e quello della globulina 26kDa (Yang *et al.*, 2001; Huang *et al.*, 2002); nel gruppo di ricerca presso cui ho svolto

l'attività di dottorato, si è invece puntato sul promotore del gene codificante la glutelina B4 in quanto esso induce l'espressione unicamente nell'endosperma interno, e a un livello superiore a quello raggiunto dagli altri promotori seme-specifici (Qu e Takaiwa, 2004). Allo stato dell'arte il promotore GluB4 sembra cioè essere la scelta migliore per l'espressione di una proteina eterologa in endosperma, ma l'ottenimento di rese ancora più elevate renderebbe il sistema riso definitivamente competitivo nei confronti delle più consolidate strategie di produzione attualmente in uso a livello industriale. E' in questo contesto che si inserisce il progetto a cui fa capo questo dottorato. Poiché non risultano disponibili sequenze di promotori naturali atte ad aumentare ulteriormente il livello di trascrizione di transgeni, si è deciso di ricorrere allo sviluppo di sequenze artificiali, con lo scopo di superare le prestazioni offerte dagli analoghi naturali.

L'intento di progettare un promotore che funzioni meglio di promotori analoghi naturali è piuttosto ambizioso e comporta la necessità di comprendere, per lo meno a grandi linee, i meccanismi che sottendono la regolazione dell'espressione genica. Per comprendere il meccanismo che regola l'espressione delle proteine di riserva del seme, vari autori si sono impegnati nell'individuazione di motivi altamente conservati in diversi promotori seme-specifici. Gli elementi regolatori putativi sono stati verificati mediante mutazioni puntiformi, delezioni (*loss-of-function*) o utilizzo di questi elementi in promotori sintetici (*gain-of-function*) (Wu *et al.*, 1998; Washida *et al.*, 1999; Wu *et al.*, 2000; Hwang *et al.*, 2001). Sono stati individuati 4 motivi responsabili dell'espressione specifica in endosperma: GCN4, AACA, AAAG e ACGT. Il motivo GCN4 appare un elemento chiave di regolazione nel promotore GluB1. Il motivo ACGT è diffuso in numerosi promotori di proteine di riserva del seme ma anche in molti promotori di geni che mediano alla varie risposte, come stimoli luminosi, fito-ormoni e vari stress biotici e abiotici (Izawa *et al.*, 1993). E' stato appurato che l'attivazione specifica per l'endosperma sia dovuta al contesto in cui ACGT è inserito (Nakase *et al.*, 1997). Il motivo AACA è un elemento regolatore negativo al di fuori del seme, ma nel seme agisce in combinazione con GCN4 per dirigere l'espressione specifica nell'endosperma (Takaiwa *et al.*, 1996, Yoshihara e Takaiwa, 1996). Il motivo AAAG, detto anche *Prolamin-*

box, è un elemento molto diffuso nei promotori delle proteine di riserva. Spesso una copia di questo elemento si trova a breve distanza da GCN4, con cui ha un'azione sinergica (Yamamoto *et al.*, 2006).

Il meccanismo di attivazione dei promotori seme-specifici è stato meglio compreso caratterizzando i fattori di trascrizione responsabili dell'espressione delle proteine di riserva nel seme. Il cDNA di un fattore di trascrizione, inserito in un vettore di espressione, consente di eseguire gli esperimenti utili a determinare quali sono i motivi riconosciuti all'interno di un promotore, qual è la sequenza necessaria al riconoscimento, quali elementi *cis*-regolatori e quali promotori il fattore è in grado di attivare in coltura di protoplasti e infine qual è l'azione combinata di due fattori di trascrizione (Suzuki *et al.*, 1998; Onodera *et al.*, 2001; Yamamoto *et al.*, 2006).

6.1 Studi preliminari alla progettazione

Nella fase iniziale di questo lavoro è stato reperito quanto disponibile in letteratura a proposito degli elementi *cis*-regolatori dei promotori delle PRSR. Le informazioni raccolte sono state elaborate in una lista di sequenze consenso dei motivi regolatori. Le sequenze sono state distinte in certe o putative. Le sequenze consenso degli elementi *cis*-regolatori sono state così riunite in una classificazione sistematica, che non era stato possibile né reperire in letteratura, né ottenere attraverso le banche dati disponibili in rete.

Sulla base delle sequenze consenso è stata svolta l'analisi su 14 promotori di proteine di riserva del seme. I risultati di questa analisi sono stati annotati lungo le sequenze utilizzando il *software BioEdit*. In questo modo la disposizione e la natura degli elementi regolatori per ogni promotore è stata resa consultabile in un *file* in formato .gb. Per facilitare la visione di insieme sono state disegnate mappe a barre che rappresentano i 14 promotori annotati.

Il lavoro svolto sulle sequenze dei promotori naturali ha portato a conclusioni utili alla progettazione dei promotori sintetici. Tuttavia, poiché il

metodo applicato per l'individuazione dei motivi regolatori non era stato precedentemente validato *in vivo*, si è preferito non basarsi unicamente sui risultati del lavoro di analisi delle sequenze. È importante notare che in più occasioni, le previsioni fatte sulla base dell'analisi di sequenza non sono state poi confermate dai dati scientifici raccolti.

6.2 Discussione sull'approccio metodologico seguito in base ai risultati ottenuti dall'analisi di sequenza

L'organizzazione delle fasi temporali della ricerca è stata dettata dalla necessità di saggiare *in vivo* i risultati dell'analisi di sequenza sui promotori delle PRSR. In quest'ottica il lavoro è stato diviso in due cicli sperimentali, secondo un metodo iterativo in cui i risultati del primo ciclo forniscono informazioni per il ciclo successivo.

La produzione delle piante transgeniche ha costituito il passaggio limitante. Nel corso del dottorato sono stati completati con successo due cicli sperimentali e sono state poste le basi per l'ulteriore sviluppo del progetto. Nel primo ciclo il promotore della glutelina C è stato confrontato con il promotore della glutelina B4. I dati raccolti hanno permesso di stabilire una gerarchia fra i due promotori candidati. Il metodo iterativo si è rivelato prezioso, perché ha evitato di compiere una scelta sbagliata o di partire con un numero di promotori candidati troppo elevato.

Nel secondo ciclo sperimentale il problema della progettazione dei nuovi promotori sintetici è stato affrontato con un approccio combinatorio. Dai 14 promotori di proteine di riserva, sono stati scelti i candidati che rispondevano a determinati criteri, riguardo al livello e al profilo di espressione. Regioni provenienti dai promotori selezionati sono state combinate per fondere in un'unica sequenza più elementi regolatori. È stato seguito un metodo integrato di progettazione e costruzione, che ha consentito di ridurre i tempi e i costi della realizzazione delle cassette di espressione. La pianificazione del lavoro di biologia molecolare è stata notevolmente facilitata dall'uso di *Clone Manager 5*. Fornendo le sequenze di partenza e le coordinate delle regioni da combinare, questa programma ha restituito una serie di percorsi di sintesi atti a raggiungere lo scopo.

Procedendo in questo modo, è stato possibile costruire tutti i promotori avvalendosi unicamente di siti di restrizione già presenti. Sono stati usati come basi di partenza i promotori della glutelina B4 e della globulina 26kDa. A queste sequenze sono state combinate parti dei promotori GluB1, GluC, Prol10 e Prol16.

La trasformazione con i costrutti del secondo ciclo è stata effettuata in due tempi per esigenze tecniche. I promotori Pr10-Glb, Pr10-C, C-Glb-2692 e C-Glb-B4-3087, sebbene già inseriti nella cassetta di espressione, sono stati esclusi da questo ciclo di sperimentazione poiché le informazioni ottenibili dall'impiego di tali promotori non sono apparse così utili nel contesto del progetto.

Nel complesso, l'attività di trasformazione condotta nel corso di questo dottorato ha prodotto risultati adeguati alle esigenze del progetto. Il numero di piante rigenerate è sempre stato superiore a quello pianificato. Anche la percentuale di *escapes* si è mantenuta quasi sempre su valori accettabili. L'unica eccezione è costituita dal gruppo di piante trasformate con il promotore GluC nella sessione di novembre 2011 dove circa un terzo delle piante trasformate non è risultato PCR negativo.

Potrebbe sembrare opportuno affiancare alla trasformazione stabile un metodo rapido per indurre l'espressione del gene *reporter* in endosperma in maniera transitoria. La possibilità di ottenere in pochi giorni indicazioni, anche approssimative, sulle caratteristiche di una sequenza neo-sintetizzata, moltiplicherebbe il numero di promotori candidati saggiabili e ridurrebbe il ruolo della trasformazione stabile a strumento di verifica dei candidati più promettenti. Purtroppo, non è possibile sfruttare tale metodo nella valutazione di promotori sintetici poiché è impossibile ad oggi stimare con certezza la forza, tessuto-specificità e fase-dipendenza di un promotore dalle caratteristiche ignote.

Nella sessione di trasformazione di novembre 2011, molte piante sono state afflitte da un problema di scarsa produttività e non è stato possibile analizzare alcuni individui per il mancato raggiungimento del numero minimo di semi necessario alla determinazione del livello di espressione. Questo inconveniente ha imposto di procedere nello studio con

2 popolazione che non raggiungevano il numero minimo di individui che era stato stabilito in partenza. Nonostante ciò, l'ampiezza dei campioni di piante trasformate sottoposte ad analisi è apparsa assolutamente adeguata anche in raffronto ad altri studi precedentemente condotti su promotori endosperma-specifici.

Qu e Takaiwa (2004) hanno basato le loro conclusioni su popolazioni che contavano tra 6 e 21 individui; lo stesso Qu (2008) ha comparato i promotori di varie gluteline analizzando popolazioni che contavano tra 5 e 22 individui. In questo lavoro il campione analizzato non è mai sceso al di sotto dei 14 individui e di norma sono stati analizzati almeno 20 trasformati per costrutto.

6.3 Livello di espressione del gene *reporter*

In questo capitolo verranno discussi i risultati ottenuti con i promotori chimerici progettati e costruiti per il saggio *in vivo*.

Come già affermato, in questo lavoro per ogni promotore è stata prodotta una popolazione di circa 20 individui. All'interno di ogni popolazione il livello di espressione può variare molto da pianta a pianta, a seconda del numero di copie di transgene inserite e alla posizione in cui tali copie si integrano all'interno del genoma. Questo pone un problema di interpretazione dei dati analitici, qualora si intenda caratterizzare un promotore sulla base del livello di espressione che è in grado di indurre in un determinato tessuto. Bisogna tenere presente ad esempio che, mentre il livello di espressione medio di una popolazione è un parametro che resta indipendente dal numero di individui, il valore massimo ne è invece fortemente influenzato: più ampia è la popolazione, più è probabile che essa contenga un individuo caratterizzato da un evento di integrazione raro, tale da garantire un livello di espressione particolarmente elevato. D'altra parte la media è un parametro di scarso interesse dal punto di vista delle applicazioni biotecnologiche di un promotore, dal momento che sono i migliori espressor ad essere selezionati e fissati per generare la linea produttiva. Solo nell'ipotesi che i livelli di espressione interni alle popolazioni di trasformati primari seguano la stessa distribuzione, la media può essere considerata un parametro predittivo del valore massimo. In questo lavoro, è

stato calcolato per ciascun promotore il rapporto fra il valore massimo e la media dei livelli di espressione registrati e ciò nell'intento di descrivere l'ampiezza della distribuzione. Questo parametro ha assunto in media il valore di 3,3 ed è risultato più basso nelle popolazioni numericamente meno consistenti. Tuttavia, la popolazione relativa a GluC prodotta con la trasformazione di novembre 2011, pur essendo molto ristretta, ha fornito un livello massimo di espressione quadruplo rispetto alla media. Nella trasformazione di febbraio 2012, il rapporto fra i due valori per lo stesso promotore è salito a 5,4, confermando la tendenza di questo promotore a dare picchi di espressione.

Novembre 2011			Massimo / Media	
	Massimo	Media		Popolazione
C-Glb-B4	49,3	23,1	2,1	18
Glb-B4	36,2	13,6	2,7	19
P16-Glb	35,4	12,7	2,8	23
C-B4	27,1	9,1	3,0	24
Glb	19,9	8,0	2,5	30
C-Glb	20,3	7,6	2,7	24
GluC	28,0	6,8	4,1	14
B1-Glb	23,0	5,5	4,2	28
B1-B4	7,3	3,8	1,9	15
Febbraio 2012				
GluB4	26,1	8,7	3,0	28
C-B4	33,0	7,4	4,4	38
GluC	34,4	6,4	5,4	32
B1-B4	17,3	4,6	3,7	35
		Media	3,3	

Per ottenere un quadro più completo dell'andamento dei livelli di espressione all'interno della popolazione è stata calcolata anche la media dei valori che occupano il primo quarto della classifica. Questo parametro rende conto della capacità del promotore di produrre picchi di espressione, ma mette al riparo dalla variabilità che caratterizza il valore massimo.

Nell'ordinare in una classifica i promotori sintetici è stato scelto come parametro di riferimento la media di tutti i livelli di espressione, poiché,

trattandosi di un confronto fra popolazioni di diversa ampiezza, è stato comunque ritenuto opportuno basarsi sull'unico parametro indipendente dal numero di individui. Va comunque sottolineato che se si fosse tenuto conto del migliore espressore, la classifica avrebbe avuto un ordine diverso.

Nel confrontare i dati ottenuti in questo lavoro con quelli precedentemente pubblicati sull'argomento è bene ricordare che questo studio si focalizza sulla porzione che precede la TATA-box. Ciascun promotore, senza eccezione, condivide la medesima sequenza TATA-box, mutuata dal promotore della glutelina B4. In questo, il presente lavoro si distingue dagli studi precedentemente effettuati da altri autori, in cui i promotori sono stati valutati anche insieme al proprio *leader*.

Fatte queste premesse, si può notare che il promotore della globulina presenta livelli di espressione di GCasi inaspettatamente elevati. Prendendo in esame la media, Glb risulta 1,2 volte più potente di GluC, mentre era attesa una *performance* attorno all'80% di quest'ultimo (Qu and Takaiwa, 2004; Qu *et al.*, 2008). Similmente, il rapporto tra i livelli di espressione di GluB4 e Glb è sceso da 1,6 (dati pubblicati) a 1,1 (dato da noi rilevato).

Il promotore della globulina fuso alla TATA-box del promotore della glutelina B4 e al *leader* sintetico LLTCK appare decisamente più forte della versione saggiata da Qu e Takaiwa (2004). Per spiegare la differenza tra i dati pubblicati e quanto rilevato in questa tesi, si può ipotizzare che la sostituzione della TATA-box e del 5'-UTR nativi abbia avuto l'effetto di incrementare il livello di espressione di Glb. Il motivo potrebbe consistere nella scarsa compatibilità del 5'UTR di Glb con il 3'-UTR relativo al terminatore NOS. E' noto che l'associazione fra 5'-UTR e 3'-UTR può avere importanti effetti sulla stabilità del mRNA e di conseguenza sul livello di espressione (Yang *et al.*, 2009). Negli studi precedenti il promotore della globulina potrebbe effettivamente essere stato sottovalutato a causa del particolare terminatore utilizzato.

Prendendo in considerazione i dati ottenuti sui trasformati primari di febbraio 2012, il livello di espressione medio indotto da GluC risulta pari al 74% circa di quello di GluB4. Per questo valore non esiste un riferimento in letteratura poiché questi due promotori non sono stati in alcun caso

utilizzati contemporaneamente. Tuttavia, facendo riferimento a dati prodotti con lo stesso protocollo ma pubblicati in articoli diversi, si conferma quanto osservato in questo lavoro.

L'inserimento della regione 5' di GluB1 a monte di Glb e GluB4 non ha sortito gli effetti sperati. E' stato rilevato che il promotore GluB1 nella sua versione 2,3 kb induce un livello di espressione 10 volte superiore a quello indotto dall'analogo privato delle 1000 bp in 5', pur mantenendo il medesimo profilo di espressione (Qu e Takaiwa, 2004). Ciò ha fatto sperare che nella porzione 5' di questo promotore ci fosse un motivo in grado di agire da *enhancer*. Nessuno dei due promotori sintetici che contengono questa porzione ha presentato tuttavia un elevato livello di espressione, al contrario è apparso che la presenza della regione appartenente a GluB1 abbia un effetto negativo. B1-Glb è inferiore a Glb qualora si consideri indifferentemente la media di popolazione e la media del primo quartile. Diverso è il caso di B1-B4, in cui alla sequenza originale di GluB4 mancano le prime 160 bp, ma analoghe le conclusioni. Per giustificare il calo di espressione rispetto al promotore GluB4 nella sua forma integrale, sarebbe necessario ipotizzare la presenza in questa porzioni di elementi *cis*-regolatori ignoti di fondamentale importanza. In alternativa, l'effetto negativo andrebbe attribuito interamente alla presenza di GluB1.



Figura 48. I due promotori che comprendono la porzione 5' di GluB1. Per la legenda dei motivi regolatori vedere pagina successiva.

In entrambi i casi, i dati ottenuti sono in contrasto con quanto pubblicato. Si può ipotizzare che nel promotore GluB1 amplificato da CR W3 le differenze riscontrate rispetto alla sequenza depositata abbiano portato alla soppressione di uno o più importanti elementi regolatori. Un'altra ipotesi è che il fattore di trascrizione che riconosce la porzione 5' di GluB1 non venga espresso in modo omogeneo in tutte le varietà di *Oryza sativa* e che quindi l'attivazione di GluB1 non sia uguale in CR W3 e in Kitaake.

I due promotori sintetici in cui GluC costituisce il terminale in 5' non sembrano godere di nessun effetto potenziante. C-Glb induce livelli di espressione sovrapponibili a quelli di Glb. C-B4 induce la produzione di proteina *reporter* in misura simile a GluB4. Queste evidenze suggeriscono che nella porzione 5' del promotore GluC non siano contenuti elementi in grado di influenzare l'attività trascrizionale di un promotore. Questa ipotesi era già stata avanzata al termine dell'analisi di sequenza di questo promotore (§5.1.2) ed è confermata dal confronto dei livelli di espressione indotti dall'intero promotore GluC e dalla sua forma tronca. Come indicato in precedenza, nel secondo ciclo sperimentale il livello di espressione medio indotto da GluC è stato pari al 74% circa di quello di GluB4 (§5.7.11), valore in linea con quello registrato nel primo ciclo con Δ GluC (68%) (§5.7.9). Nella pratica, l'effetto dell'asportazione di 1 kb in 5' si è rivelato pressoché trascurabile.

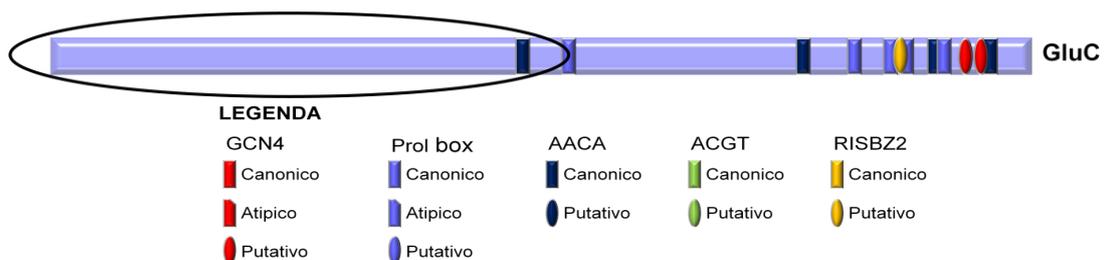


Figura 49. Porzione del promotore GluC che, in base ai risultati ottenuti appare inattiva. Notare che anche l'analisi di sequenza conferma questo dato.

Quando la porzione 5' del promotore della prolamina 16kDa viene combinata con il promotore della globulina 26 kDa, si ottiene un promotore chimerico decisamente più potente di Glb. In effetti, Prol16-Glb induce un livello di espressione superiore al migliore promotore naturale, GluB4, qualunque sia il parametro preso in considerazione cioè valore massimo, media del primo quartile e media di popolazione. Si sottolinea che con questo promotore chimerico è stata realizzata la prova di concetto del metodo combinatorio di progettazione (§3.2.4). Se si osserva la mappa degli elementi regolatori di Prol16 (§5.1), si nota che rispetto a PG5 e NRP33, gli elementi fondamentali sono spostati più a monte lungo il promotore. Come si osserva in Fig. 50, le 500bp in 5' fuse alla sequenza di

Glb comprendono anche gli elementi GCN4 e Prol-box associati in quella che viene detta *endosperm-box*. E' ipotizzabile che questi motivi conferiscano a Glb un sensibile potenziamento dell'attività trascrizionale.

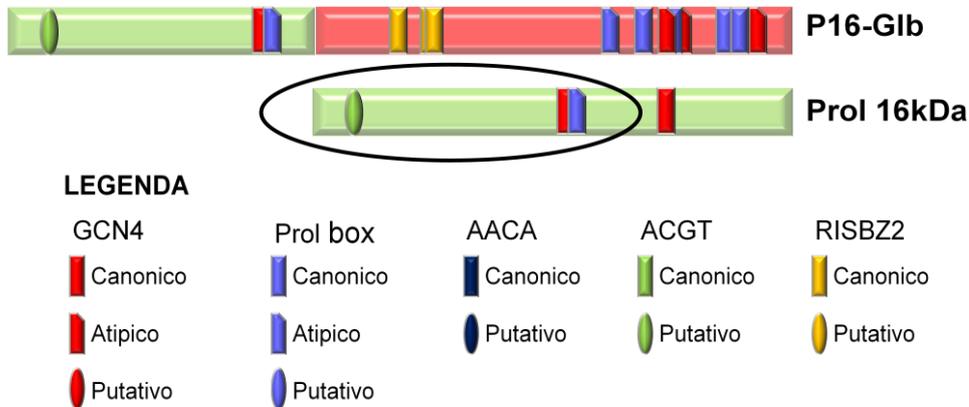


Figura 50. Porzione di Prol16 in grado di aumentare il livello di espressione di Glb. Notare come la posizione degli elementi regolatori sia stata traslata in direzione 5' nel promotore chimerico.

Glb-B4 rappresenta il prodotto di fusione delle porzioni 3' dei 2 promotori impiegati come base strutturale nel secondo ciclo sperimentale.

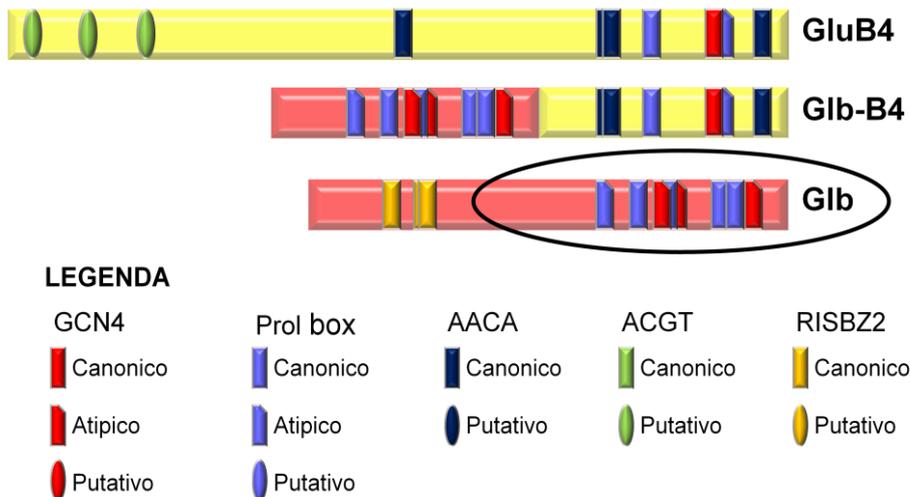


Figura 51. Il promotore Glb-B4 e i due promotori naturali che lo compongono. Si noti che in questo caso è stata utilizzata la porzione 3' di entrambi gli elementi.

Questo promotore si è dimostrato anch'esso più potente di GluB4. Il livello di espressione ottenuto è notevole soprattutto in relazione alle ridotte dimensioni del promotore (< 1 kb). Questo fatto rende Glb-B4 un ottimo candidato per ulteriori manipolazioni. Osservando la mappa è facilmente

rilevabile la combinazione di due regioni a elevata densità di elementi *cis*-regolatori. In tal senso, le mappe tracciate in seguito all'analisi di sequenza si dimostrano in assoluto accordo con i dati sperimentali. Ciò che maggiormente sorprende è piuttosto la funzionalità del fitto sistema di elementi regolatori di Glb anche in posizione traslata di circa 450bp a monte della sua naturale posizione, prossima alla TATA-box.

In Glb-B4 manca la porzione di Glb contenente i siti ACGT riconosciuti dal fattore RISBZ-2. Questa porzione, da quanto riportato in letteratura è risultata essere determinante nel garantire un elevato livello di espressione (Yang *et al.*, 2001; Nakase *et al.*, 1997). D'altra parte, la fusione di un segmento contenente i siti per RIZBZ-2 all'estremità 5' di GluA2 non ha modificato il livello di espressione in maniera significativa (Yang *et al.*, 2001). In C-Glb-B4 questa porzione si trova fusa all'estremità 5' di GluB4. Infatti, C-Glb-B4 è costituito dall'intera sequenza di GluB4 fusa alla metà distale di Glb, contenente i siti di legame per RISBZ2. A questa regione è a sua volta fusa la regione di 400 bp ca. in 5' di GluC. In base a quanto visto precedentemente, è tuttavia poco probabile che la porzione mutuata da GluC abbia una reale influenza sull'attività trascrizionale. In questo caso, in contrasto con quanto riportato da Yang *et al.* (2001), il livello di espressione ottenuto con il promotore sintetico è significativamente più elevato rispetto al promotore naturale di derivazione e probabilmente anche più elevato della somma dei livelli che caratterizzano ciascun promotore utilizzato nella sua costruzione.

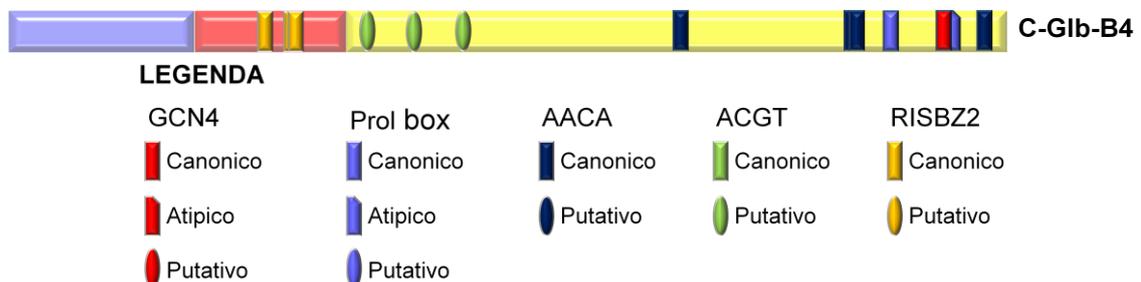


Figura 52. Struttura di C-Glb-B4, unico ibrido ternario fra i promotori saggati nel corso del secondo ciclo sperimentale.

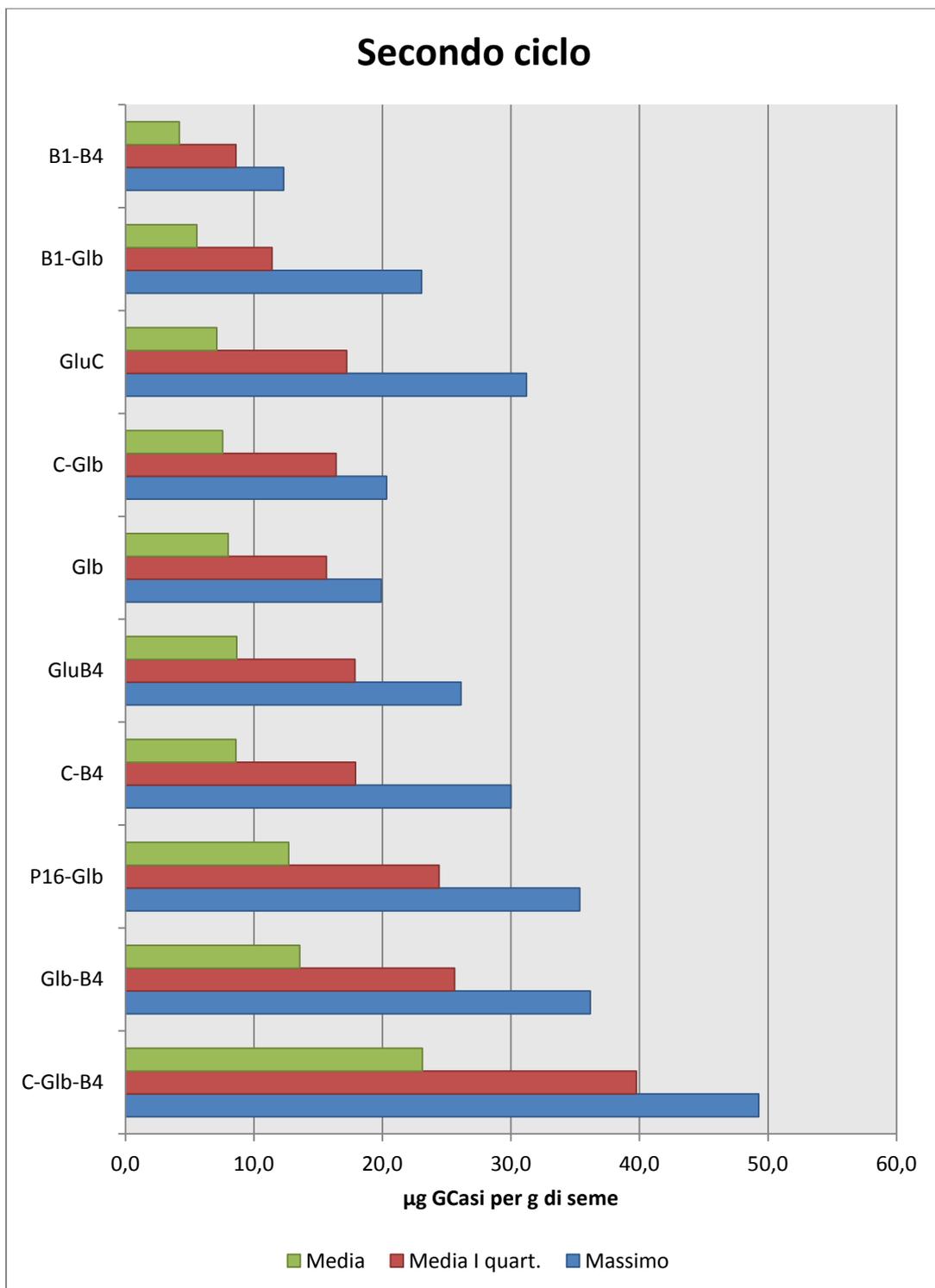


Grafico 9. Si riportano qui i risultati per una più agevole consultazione

6.4 Considerazioni sulla riproducibilità del saggio *in vivo*

Sono state prodotte due popolazioni di trasformati primari per ciascuno di questi promotori: B1-B4, GluC, C-B4. Le due sessioni di trasformazione che hanno generato le due popolazioni hanno avuto luogo a novembre 2011 e a febbraio 2012, rispettivamente. Ciascuna popolazione costituisce una replica sperimentale del saggio *in vivo* dei tre promotori sintetici. Il fatto che in questo lavoro siano stati previsti tre esperimenti in replicato si è dimostrato prezioso sotto diversi aspetti. La popolazione B1-B4 prodotta in occasione della seconda sessione di trasformazione non ha sofferto gli stessi problemi di produttività che hanno caratterizzato la popolazione ottenuta a novembre 2011. In ragione di questo fatto è stata scartata l'ipotesi riguardante una perdita di specificità del promotore con conseguente accumulo della proteina *reporter* fitotossica in tessuti diversi dall'endosperma.

La ripetizione della trasformazione con il promotore della glutelina C ha dato conferma della tendenza di questo promotore a produrre popolazioni in cui il livello di espressione del transgene si distribuisce in un intervallo più ampio rispetto agli altri promotori.

Il fatto che nelle trasformazioni di novembre 2011 e febbraio 2012 siano stati inclusi tre controlli incrociati ha permesso di confrontare indirettamente il promotore GluB4 con i vari promotori sintetici. Infatti, GluB4 è stato inserito solo nella seconda sessione di trasformazione e, in mancanza di controlli incrociati, non sarebbe stato possibile confrontarne il livello di espressione con i promotori trasformati nella prima sessione di trasformazione. In tal caso, sarebbe venuto a mancare il confronto con uno dei migliori promotori naturali delle PRSR fino a ora studiati.

Per valutare l'effetto della stagionalità sui risultati del saggio *in vivo*, sono state calcolate le medie e le deviazioni standard (in assoluto e in percentuale) per i tre parametri di valutazione dei promotori.

Infine, per ogni singolo parametro è stata calcolata la media delle tre deviazioni percentuali relative ai tre promotori. In questo modo si è inteso

verificare se le deviazioni tendono ad annullarsi fra loro e quindi sono casuali o se c'è un "effetto stagionalità" che influisce con regolarità sul livello di espressione dei trasformati primari.

Tabella 44. Dati grezzi ottenuti dalle tre trasformazioni replicate.

	Massimo					
	Novembre	Febbraio	Media	Dev. St.	dev %	media dev%
B1-B4	7,3	17,3	12,3	5,0	41%	
GluC	28,0	34,4	31,2	3,2	10%	
C-B4	27,1	33,0	30,1	3,0	10%	20%
	Media I quartile					
	Novembre	Febbraio				
B1-B4	6,4	10,7	8,6	2,1	25%	
GluC	16,1	14,8	15,5	0,6	-4%	
C-B4	20,1	15,8	17,9	2,1	-12%	3%
	Media					
	Novembre	Febbraio				
B1-B4	3,8	4,6	4,2	0,4	10%	
GluC	6,7	6,4	6,6	0,2	-3%	
C-B4	9,1	7,4	8,3	0,8	-10%	-1%

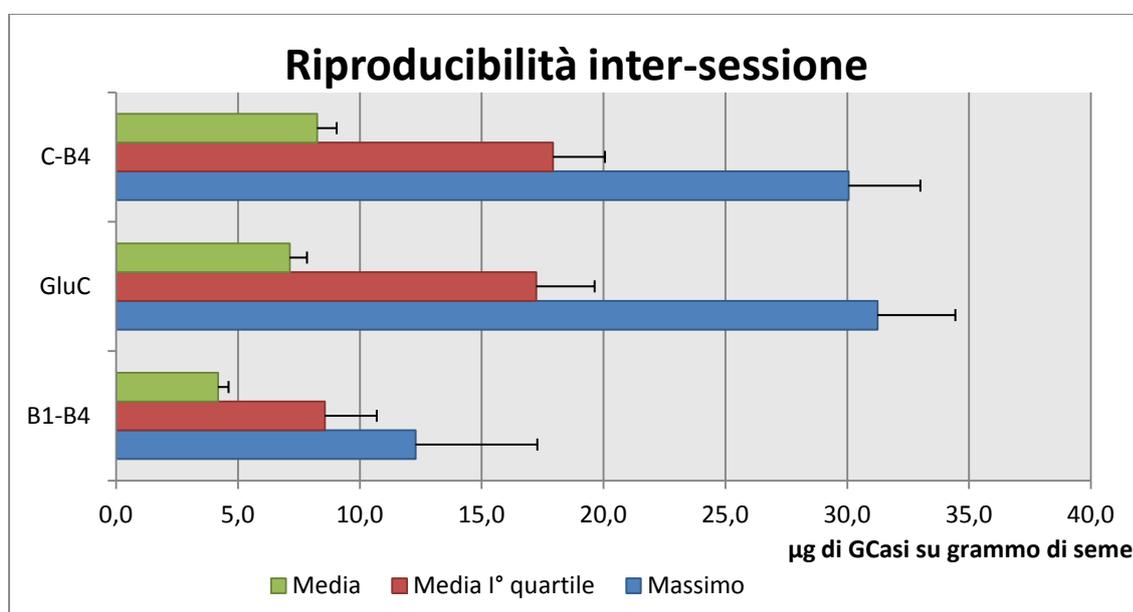


Grafico 10. Dati riguardanti le trasformazioni replicate a febbraio 2012.

Quasi tutti i parametri hanno mostrato una deviazione percentuale inferiore al 15%. Gli unici valori che eccedono questo limite sono quelli che si riferiscono al massimo e alla media del primo quartile nelle due popolazioni ottenute con B1-B4, fatto in parte prevedibile considerata l'ampiezza molto diversa delle due popolazioni. E' comunque interessante notare che il livello medio di espressione si è mantenuto su valori paragonabili sebbene le due trasformazioni abbiano presentato tra loro differenze radicali in termini di numero di piante, produttività di seme e stato generale degli individui. Inoltre, nonostante il valore medio di espressione ottenuto per B1-B4 sia salito da 3,8 a 4,6 µg/g dalla prima alla seconda trasformazione, tale promotore ha confermato i suoi limiti. Ciò dimostra che il metodo di valutazione è sufficientemente solido da non essere influenzato da variabili esterne. In altre parole, questi dati consentono di validare il metodo di valutazione dei promotori utilizzato in questo lavoro.

6.5 Tessuto-specificità

Qualunque modifica alla sequenza di un promotore tessuto-specifico potrebbe, almeno in linea di principio, modificarne le caratteristiche di specificità. Nell'assemblare parti di diversi promotori endosperma-specifici è possibile che nella sequenza ottenuta manchino gli elementi responsabili della specificità. Questi elementi non sono noti con certezza, anche se l'ipotesi più accreditata è che la regolazione dell'espressione sia soprattutto positiva (Wu *et al.*, 1998; Wu *et al.*, 2000); infatti, l'unico elemento che appare coinvolto nella soppressione dell'espressione al di fuori dell'endosperma è AACA (Suzuki *et al.*, 1998).

La proteina *reporter* utilizzata in questo lavoro è la beta-glucosidasi umana, un enzima lisosomiale che degrada i glucocerebrosidi che compongono il distrato lipidico della membrana cellulare. Questo enzima, se espresso nel tessuto vegetativo, mostra un chiaro effetto fitotossico; qualora espresso oltre un certo limite nel seme di tabacco, ha inoltre dimostrato di inibire la germinazione in modo irreversibile (Reggi *et al.*, 2005). L'uso di GCasi come *reporter* per saggiare l'attività di promotori sintetici ha avuto proprio lo scopo di evidenziare eventuali perdite di

specificità. Giacché per tutti i promotori è stato possibile ottenere una popolazione di piante mature e per tutte le popolazioni analizzare la progenie di un congruo numero d'individui, è ragionevole concludere che nessuno dei promotori saggiati abbia indotto elevati livelli di espressione in tessuti diversi dall'endosperma. Ciò nonostante, il fatto che la produttività di seme non sia stata omogenea fra i vari promotori ha indotto a condurre ulteriori indagini con lo scopo di verificare se quanto rilevato fosse correlato a un effetto fitotossico.

L'osservazione di eventuali effetti fitotossici sui trasformati primari è resa complicata dal fatto che queste piante, anche in assenza di un effetto specifico, dimostrano spesso un fenotipo alterato e un rallentamento dello sviluppo, dovuti allo stress subito in fase di rigenerazione e di selezione *in vitro*.

Le piante ottenute dalla trasformazione di febbraio 2012, coltivate in serra nel periodo che va da giugno a ottobre, non hanno manifestato alcun problema di produttività di seme a differenza di quelle ottenute dalla trasformazione di novembre 2011, le quali sono state coltivate in serra fra febbraio e giugno. Dal momento che alcune trasformazioni erano state replicate a febbraio 2012, è stato possibile un confronto diretto tra popolazioni di piante diversamente allevate. Le piante trasformate con il medesimo costrutto hanno spesso mostrato comportamenti contrastanti al variare delle condizioni di crescita, dimostrando che il costrutto non è l'elemento principalmente implicato nella ridotta capacità di produrre seme. A titolo di es. le piante trasformate con B1-B4 hanno prodotto una quantità sufficiente di semi nel 42% e 93% dei casi, se allevate nel periodo vernino-primaverile e primaverile-estivo, rispettivamente. Vale la pena ricordare che il secondo dato supera quello registrato per piante trasformate con i promotori naturali GluC e GluB4 (88% e 84%, rispettivamente). Queste evidenze hanno portato a concludere che la sola osservazione della produttività di seme non è sufficiente a stabilire un nesso causale fra la ridotta produttività e l'azione fitotossica di GCasi. Inoltre, all'interno della popolazione trasformata con C-Glb-B4 è stata valutata la correlazione fra livello di espressione in endosperma e produttività di seme. Assumendo che

siano proprio i migliori espressorori a soffrire maggiormente dell'attività fitotossica della proteina *reporter*, sarebbe attesa una relazione di proporzionalità inversa fra contenuto di GCasi in seme e quantità di seme prodotto e ciò secondo una funzione iperbolica.

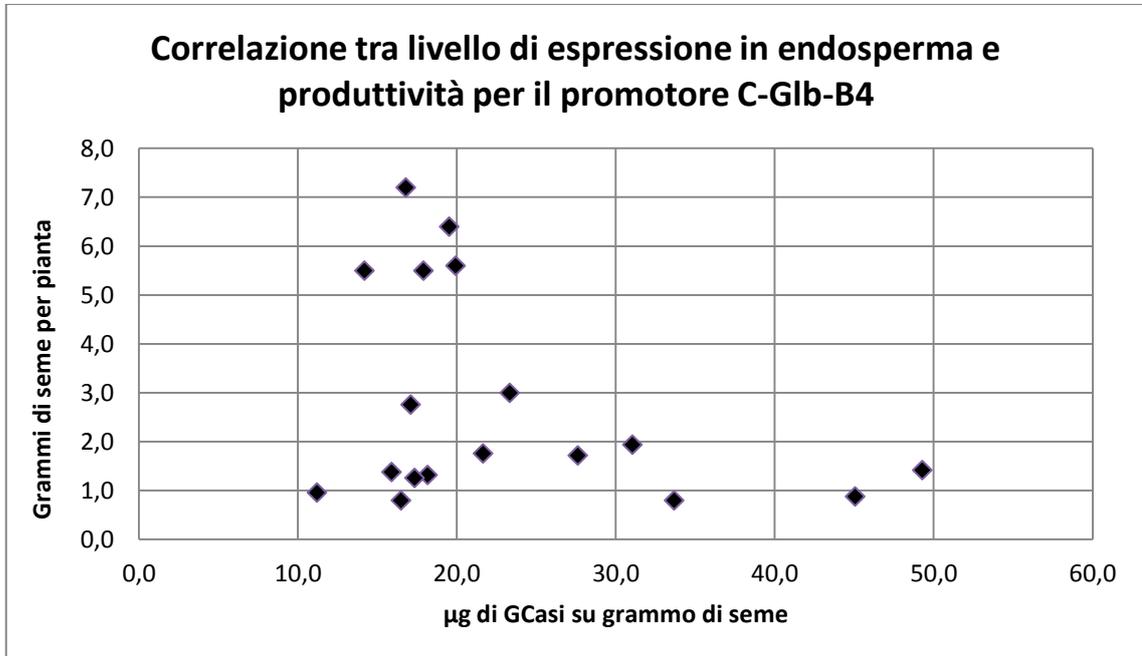


Grafico 11. La produttività di GCasi e di seme in piante appartenenti alla popolazione trasformata con il promotore C-Glb-B4.

La semplice osservazione del grafico 11 porta immediatamente a escludere che vi sia una stretta relazione tra livello di espressione e stato di sofferenza delle piante.

Per evidenziare un'eventuale minima espressione del transgene al di fuori dell'endosperma è stato deciso di procedere alla quantificazione della proteina *reporter* nel tessuto fogliare. In media la concentrazione di GCasi rilevata negli estratti di foglie di piante C-Glb-B4 è risultata 15 volte più elevata di quella rilevata in B1-B4. Va tuttavia osservato che in termini assoluti si tratta di un contenuto residuale di GCasi, tant'è vero che il valore è stato espresso in ppm sulla proteina totale estratta.

Per verificare se la concentrazione di GCasi rilevata in foglia è correlata alla concentrazione di GCasi in endosperma, e quindi sussiste effettivamente una perdita di tessuto-specificità del promotore, sono state analizzate singole piante C-Glb-B4. E' ragionevole ritenere che, in caso di perdita di tessuto-specificità, i migliori espressorori evidenzino un maggiore

accumulo di proteina *reporter* negli altri tessuti; in questi casi dovrebbe sussistere una relazione di proporzionalità diretta tra le variabili descritta da una funzione lineare.

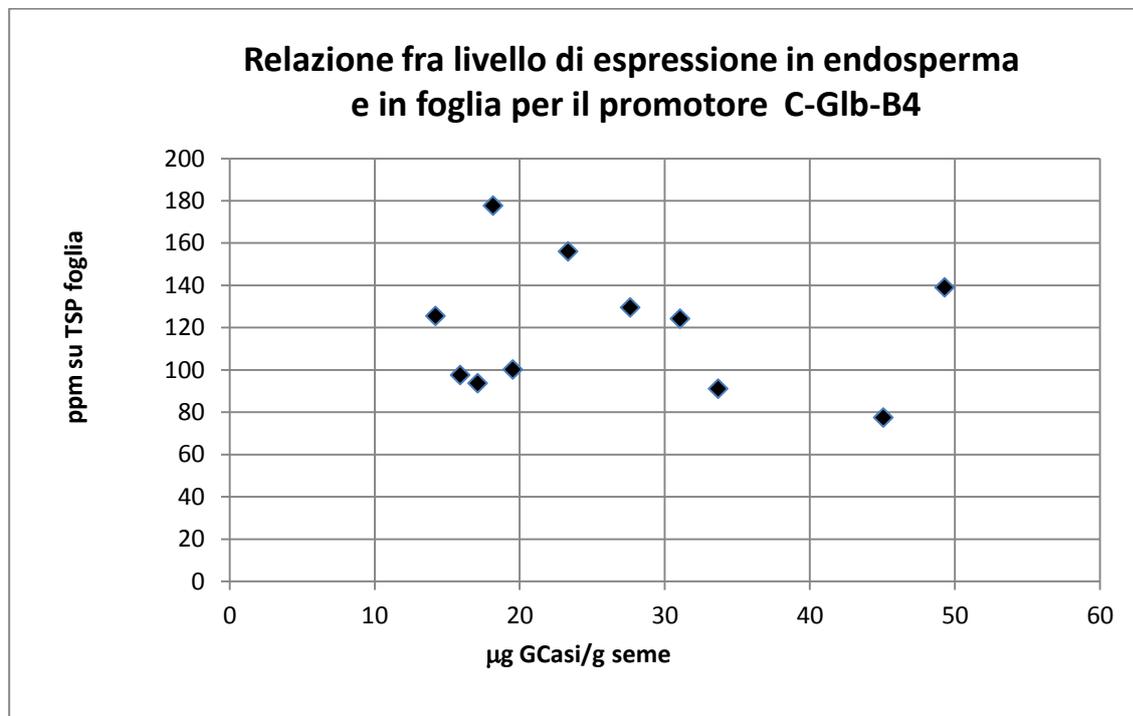


Grafico 12. Correlazione fra livello di espressione di GCasi in endosperma (ascissa) e in foglia (ordinata).

Apparentemente il livello di GCasi in foglia non è in alcun modo correlato al livello di espressione in endosperma. Questa osservazione porta ipotizzare che, a livello di cromatina, il sito d'integrazione possa favorire o meno l'espressione in un determinato tessuto. Se il transgene si integra in una zona particolarmente attiva in foglia è possibile che si verifichi un'attività trascrizionale residua al di fuori di questo tessuto. L'ipotesi, anche se affascinante, non aiuta comunque a stabilire un nesso causale fra perdita di specificità e ridotta produttività di seme. Per verificare l'esistenza di questo nesso i dati sull'espressione in foglia sono stati posti in relazione con quelli sulla produttività di seme nelle piante C-Glb-B4. Questa operazione ha avuto lo scopo di valutare quanto la presenza di GCasi nel tessuto fogliare abbia effettivamente inficiato la produzione di seme delle varie piante. Ammettendo che le ridotte concentrazioni di GCasi abbiano manifestato un effetto tossico, sarebbe nuovamente attesa una relazione di

proporzionalità inversa tra le due variabili, descritta da una funzione iperbolica.

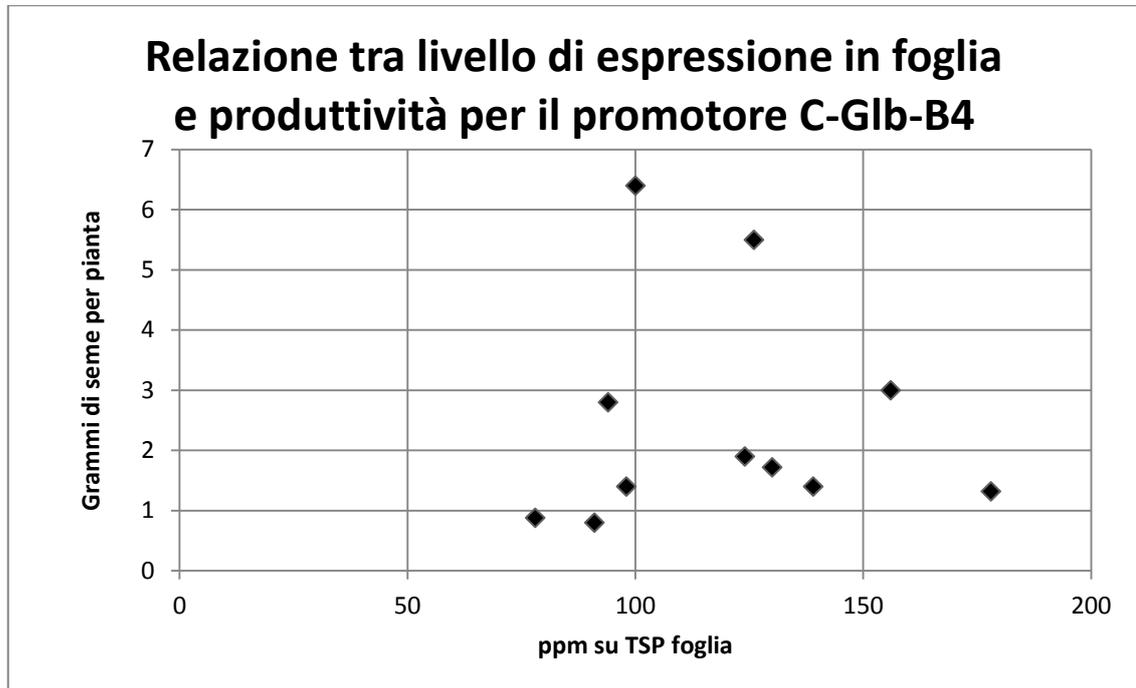


Grafico 13. Produttività di seme e livello di espressione di GCasi in foglie di piante appartenenti alla popolazione trasformata con il promotore C-Glb-B4.

L'andamento della produttività di seme rispetto all'espressione in foglia difficilmente porta a concludere che l'effetto tossico di GCasi sia la causa della ridotta produzione di seme da parte della popolazione trasformata con C-Glb-B4.

Infine, la quantificazione del livello di espressione di proteina *reporter* è stata condotta nelle foglie dei migliori espressoristi di ciascuna delle popolazioni di trasformati primari (§ Tabella 38). Dai risultati di questo esperimento è certo che nelle piante trasformate con C-Glb-B4 ha luogo un'espressione fogliare di GCasi superiore a tutte le altre piante. Quando questo risultato viene tuttavia rapportato al livello di espressione in endosperma, ottenendo l'indice di *leakage* (§ 5.6.1), la differenza tra promotori permane anche se ridimensionata.

Questa evidenza porta a concludere che il rischio di compromissione della tessuto-specificità di un promotore non è remoto. Nel caso di C-Glb-B4 la presenza di proteina *reporter* al di fuori del tessuto bersaglio sembra essere dovuta più alla forza del promotore che non alla perdita di elementi

di controllo negativo, dal momento che l'indice di *leakage* supera di poco quello dei migliori promotori. Osservando la sequenza di C-Glb-B4 non è possibile identificare alcuna ragione che dia conto di un profilo di espressione diverso da GluB4. Probabilmente una sequenza particolarmente ricca di motivi di riconoscimento per fattori di trascrizione può divenire parzialmente attiva dal punto di vista trascrizionale anche a causa di un riconoscimento aspecifico da parte di fattori trans-attivanti presenti in alte concentrazioni in tessuti diversi da quello bersaglio. La verifica di una tale ipotesi esula però dagli scopi di questo dottorato.

Tutti i promotori sintetici realizzati nel secondo ciclo hanno come base il promotore della globulina oppure il promotore della glutelina B4. Qu e Takaiwa (2004) hanno pubblicato i risultati di un esperimento che evidenziava la distribuzione della proteina *reporter* fra i vari tessuti del seme in caso di utilizzo dell'uno e dell'altro promotore. In tale esperimento, gli autori si sono serviti dell'attività enzimatica della proteina *reporter* utilizzata (GUS) per condurre un esame istochimico. I risultati di questo esperimento evidenziano in maniera intuitiva il profilo spaziale di espressione dei promotori, ma hanno soltanto valore qualitativo. Nel presente studio, per dimostrare il sito prioritario di accumulo della proteina *reporter*, si è deciso di seguire una via diversa. I tessuti costituenti il seme sono stati separati meccanicamente mediante una processo industrialmente consolidato. Il vantaggio di ricorrere alla sbiancatura del seme per valutare la distribuzione della proteina *reporter* consiste nel fatto che questo metodo viene correntemente utilizzato nell'ambito delle normali operazioni di processamento del seme di riso. I risultati che si sono ottenuti anticipano dunque un'informazione importante per l'applicazione dei promotori sintetici nell'espressione di proteine di interesse farmaceutico.

Confrontando i dati relativi alle percentuali di recupero e di perdita di proteina ricombinante per i vari promotori sintetici, non è stato ravvisato alcun caso di alterazione del rapporto di distribuzione tra seme sbiancato e farinaccio. Per questo motivo si può concludere che la manipolazione delle sequenze non ha modificato il profilo di espressione spaziale nell'endosperma di riso tipico dei promotori utilizzati come base di sintesi.

6.6 In conclusione

I risultati ottenuti alla fine del secondo ciclo sperimentale raggiungono tutti gli scopi fissati all'inizio di questo dottorato e forniscono informazioni importanti e nuovi spunti per l'attività futura.

E' stato messo a punto un sistema integrato di sviluppo di promotori sintetici tenendo conto dei dati bibliografici esistenti a proposito di elementi *cis*-regolatori contenuti in promotori naturali endosperma-specifici e di un'analisi fine della loro sequenza.

Sono stati prodotti 7 promotori sintetici ed altrettanti vettori di espressione utilizzati in esperimenti di trasformazione di riso in presenza di opportuni controlli (promotori naturali). Sono state realizzate popolazioni di piante transgeniche analizzate ai fini della valutazione della forza e tessuto-specificità dei promotori in prova nonché della validità del metodo di design adottato.

Alcune interpretazioni fornite in letteratura relativamente all'effetto di motivi regolatori specifici non hanno trovato conferma negli esperimenti condotti nell'ambito del presente dottorato. Altre volte invece le ipotesi basate sui dati pubblicati sono state confermate dai risultati. In linea di massima le previsioni basate sull'analisi di sequenza si sono sempre dimostrate fondate.

Tre promotori sintetici hanno superato il miglior promotore naturale, in un caso è stato raggiunto un livello di espressione medio 2,5 volte più elevato. Ciò dimostra che ingegnerizzare la sequenza del promotore può elevare significativamente il livello di espressione di un transgene in endosperma di riso senza comprometterne la tessuto-specificità.

Il metodo di progettazione combinatorio si è dimostrato efficace. L'integrazione dell'attività di *design* con la pianificazione delle vie sintetiche mediante *Clone Manager 5* ha permesso di ottenere in un tempo relativamente breve (2 mesi ca.) i costrutti necessari per la trasformazione. Alcuni promotori sono stati realizzati applicando su piccola scala tecniche di sintesi parallela, che hanno ulteriormente abbattuto i tempi di lavoro. L'utilizzo di questa modalità di lavoro potrà essere in futuro applicata su

scala più vasta in cicli sperimentali volti all'ulteriore miglioramento dei promotori sintetici.

La ripetizione di tre trasformazioni ha fornito i dati necessari per dimostrare che il metodo di valutazione dei promotori sintetici è sufficientemente solido e che, entro certi limiti, le variabili ambientali non sono in grado di compromettere seriamente i risultati.

7 BIBLIOGRAFIA

Andrianantoandro E, Basu S, Karig DK, Weiss R, 2006. Synthetic biology: new engineering rules for an emerging discipline. *Mol. Systems Biol.* 2, 2006 .0028.

Bailey TL, Elkan C, 1995. Unsupervised learning of multiple motifs in biopolymers using expectation maximization. *Mach. Learn.* 21, 51–80.

Barrett CL, Kim TY, Kim HU, Palsson BØ, Lee SY, 2006. Systems biology as a foundation for genomescale synthetic biology. *Curr. Opin. Biotechnol.* 17, 488–492.

Beer MA, Tavazoie S, 2004. Predicting gene expression from sequence. *Cell* 117, 185–198.

Comai L, Moran P, Maslyar D, 1990. Novel and useful properties of a chimeric plant promoter combining CaMV 35S and MAS elements. *Plant Mol. Biol.* 15, 373-381.

Daniell H, Streatfield SJ, Wycoff K, 2001. Medical molecular farming: production of antibodies, biopharmaceuticals and edible vaccines in plants. *Trends Plant Sci.* 6, 219-226.

Datta S, Datta S, 2006. Evaluation of clustering algorithms for gene expression data. *BMC Bioinformatics* 125, 3048-3052.

Depicker A, Stachel S, Dhaese P, Zambryski P, Goodman HM, 1982. Nopaline synthase: transcript mapping and DNA sequence. *J. Mol. Appl. Genet.* 1, 561-573.

Doyle JJ, Doyle JL, 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19, 11-15.

Egelkroun E, Rajan V, Howard JA, 2012. Overproduction of recombinant proteins in plants. *Plant Sci.* 184, 83-101.

Evangelista RL, Kusnadi AR, Howard JA, Nikolov ZL, 1998. Process and economic evaluation of the extraction and purification of recombinant beta-glucuronidase from transgenic corn. *Biotechnol Prog.* 14, 607-614.

Fessele S, Maier H, Zischek C, Nelson PJ, Werner T, 2002. Regulatory context is a crucial part of gene function. *Trends Genet.* 18, 60-63.

Fischer R, Stoger E, Schillberg S, Christou P, Twyman RM, 2004. Plant-based production of biopharmaceuticals. *Curr. Opin. Plant Biol.* 7, 152-158.

Fischer R, Emans N, 2000. Molecular farming of pharmaceutical proteins. *Transgenic Research* 9, 279-299.

Giddings G, Allison G, Brooks D, Carter A, 2000. Transgenic plants as factories for biopharmaceuticals. *Nature Biotechnology* 18, 1151-1155.

Guet CC, Elowitz MB, Hsing WH, Leibler S, 2002. Combinatorial synthesis of genetic networks. *Science* 296, 1466-1470.

Guido NJ, Wang X, Adalsteinsson D, McMillen D, Hasty J, Cantor CR, Elston TC, Collins JJ, 2006. A bottom-up approach to gene regulation. *Nature* 439, 856–860.

Guilfoyle TJ, 1997. The structure of plant gene promoters. *Genetic Engin.* 19, 15–47.

Hiei Y, Ohta S, Komari T, Kumashiro T, 1994. Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J.* 6, 271-282.

Holt KE, Millar AH, Whelan J, 2006. ModuleFinder and CoReg: alternative tools for linking gene expression modules with promoter sequences motifs to uncover gene regulation mechanisms in plants. *Plant Methods* 2, 2-8.

Hood EE, Woodard SL, Horn ME, 2002. Monoclonal antibody manufacturing in transgenic plants--myths and realities. *Curr. Opin. Biotechnol.* 13, 630-635.

Hwang YS, McCullar C, Huang N, 2001. Evaluation of expression cassettes in developing rice endosperm using a transient expression assay. *Plant Science* 161, 1107-1116.

Hwang YS, Yang D, McCullar C, Wu L, Chen L, Pham P, Nandi S, Huang N, 2002. Analysis of the rice endosperm-specific globulin promoter in transformed rice cells. *Plant Cell Rep.* 20, 842–847.

Izawa T, Foster R, Chua NH, 1993. Plant bZip protein DNA binding specificity. *J. Mol. Biol.* 230, 1131–1144.

Kinkhabwala A, Guet CC, 2008. Uncovering cis Regulatory Codes Using Synthetic Promoter Shuffling. *PLoS ONE* 3(4): e2030. doi:10.1371/journal.pone.0002030

Krishnan HB, Okita TW, 1986. Structural relationship among the rice glutelin polypeptides. *Plant Physiol* 81, 748-753.

Last DI, Brettell RIS, Chamberlain DA, Chaudhury AM, Larkin PJ, Marsh EL, Peacock WJ, Dennis ES, 1991. pEmu: an improved promoter for gene expression in cereal cells. *Theor. Appl. Genet.* 81, 581-588.

Lawrence CE, Altschul SF, Boguski MS, Liu JS, Neuwald AF, Wootton JC, 1993. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science* 262, 208–214.

Lin JJ, 1995. Electrotransformation of *Agrobacterium*. *Methods Mol. Biol.* 47:171-178.

Liu L, White MJ, MacRae TH, 1999. Transcription factors and their genes in higher plants functional domains, evolution and regulation. *Eur. J. Biochem.* 262, 247-57.

Liu WX, Liu HL, Chai ZJ, Xu XP, Song YR, Qu le Q, 2010. Evaluation of seed storage-protein gene 5' untranslated regions in enhancing gene expression in transgenic rice seed. *Theor. Appl. Genet.* 121, 1267-1274.

Matys V, Fricke E, Geffers R, Gösling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Münch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E, 2003. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* 31, 374–378.

Miles JS, Guest JR, 1984. Nucleotide sequence and transcriptional start point of the phosphomannose isomerase gene (*manA*) of *Escherichia coli*. *Gene* 32, 41-48.

Mitsuhara I, Ugaki M, Hirochika H, Ohshima M, Murakami T, Gotoh Y, Katayose Y, Nakamura S, Honkura R, Nishimiya S, Ueno K, Mochizuki A, Tanimoto H, Tsugawa H, Otsuki Y, Ohashi Y, 1996. Efficient promoter cassettes for enhanced expression of foreign genes in dicotyledonous and monocotyledonous plants. *Plant Cell Physiol.* 37, 49–59.

Ni M, Cui D, Einstein J, Narasimhulu S, Vergara CE, Gelvin SB, 1995. Strength and tissue specificity of chimeric promoter derived from octopine and manopine synthase gene. *Plant J.* 7, 661–676.

Okita TW, Hwang YS, Huilo J, Kim WT, Aryan, AP, LarsenR, Krishnan HB, 1989. Structure and expression of the rice glutelin multigene family. *J. Biol. Chem.* 264, 12573–12581.

Onodera Y, Suzuki A, Wu CY, Washida H, Takaiwa F, 2001. A rice functional transcriptional activator, RISBZ1, responsible for endosperm-specific expression of storage protein genes through GCN4 motif. *J. Biol. Chem.* 276, 14139-52.

Oszvald M, Kang TJ, Tomoskozi S, Jenes B, Kim TG, Cha YS, Tamas L, Yang MS, 2008. Expression of cholera toxin B subunit in transgenic rice endosperm. *Mol. Biotechnol.* 40, 261-268.

Pan SJ, Reeck GR, 1988. Isolation and characterization of rice alpha-globulin. *Cereal Chemistry* 65, 316-319.

Patti T, Bembi B, Cristin P, Mazzarol F, Secco E, Pappalardo C, Musetti R, Martinuzzi M, Versolatto S, Cariati R, Dardis A, Marchetti S, 2012. Endosperm-specific expression of human acid beta-glucosidase in a waxy rice. *Rice* 2012, 5:34.

Pietrzak M, Burri M, Herrero JJ, Mosbach K, 1989. Transcriptional activity is inducible in the cauliflower mosaic virus 35S promoter engineered with the heat shock consensus sequence. *FEBS Lett.* 249, 311-315.

Pilpel Y, Sudarsanam P, Church GM, 2001. Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat. Genet.* 29, 153-159.

Qu LQ, Takaiwa F, 2004. Evaluation of tissue specificity and expression strength of rice seed component gene promoters in transgenic rice. *Plant Biotechnol. J.* 2, 113-125.

Qu LQ, Xing YP, Liu WX, Xu XP, Song YR, 2008. Expression pattern and activity of six glutelin gene promoters in transgenic rice. *J. Exp. Bot.* 59, 2417-2424.

Rombauts S, Florquin K, Lescot M, Marchal K, Rouzé P, van de Peer Y, 2003. Computational approaches to identify promoters and *cis*-regulatory elements in plant genomes. *Plant Physiol.* 132, 1162-1176.

Rushton PJ, Reinstädler A, Lipka V, Lippok B, Somssich IE, 2002. Synthetic plant promoters containing defined regulatory elements provide novel insights into pathogen and wound-induced signaling. *Plant Cell* 14, 749–762.

Sambrook J, Fritsch EF, Maniatis T, 1989. *Molecular cloning: a laboratory manual 2nd edition*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schmidt RJ, Ketudat M, Aukerman MJ, Hoschek G, 1992. Opaque-2 is a transcriptional activator that recognizes a specific target site in 22-kD zein genes. *Plant Cell* 4, 689-700.

Schwechheimer C, Bevan M, 1998. The regulation of transcription factor activity in plants. *Trends Plant. Sci.* 10, 378–383.

Sharma A, Sharma R, Imamura M, Yamakawa M, Machii H, 2000. Transgenic expression of cecropin B, an antibacterial peptide from *Bombyx mori*, confers enhanced resistance to bacterial leaf blight in rice. *FEBS Lett.* 484, 7-11.

Shorrosh BS, Wen L, Zen KC, Huang JK, Pan JS, Hermodson MA, Tanaka K, Muthukrishnan S, Reeck GR, 1992. A novel cereal storage protein: molecular genetics of the 19 kDa globulin of rice. *Plant Mol. Biol.* 18, 151-154.

Singh KB, 1998. Transcriptional regulation in plants: the importance of combinatorial control. *Plant Physiol.* 118, 1111–1120.

Stoger E, Ma JK, Fischer R, Christou P, 2005. Sowing the seeds of success: pharmaceutical proteins from plants. *Curr. Opin. Biotechnol.* 16, 167-173.

Streatfield SJ, Jilka JM, Hood EE, Turner DD, Bailey MR, Mayor JM, Woodard SL, Beifuss KK, Horn ME, Delaney DE, Tizard IR, Howard JA, 2001. Plant-based vaccines: unique advantages. *Vaccine* 19, 2742-2748.

Steffens NO, Galuschka C, Schindler M, Bülow L, Hehl R, 2004. AthaMap: an online resource for *in silico* transcription factor binding sites in the *Arabidopsis thaliana* genome. *Nucleic Acids Res.* 32, D368–D372.

Suzuki A, Wu CY, Washida H, Takaiwa F, 1998. Rice MYB protein OSMYB5 specifically binds to the AACA motif conserved among promoters of genes for storage protein glutelin. *Plant Cell Physiol.* 39, 555-559.

Takagi H, Hiroi T, Yang L, Tada Y, Yuki Y, Takamura K, Ishimitsu R, Kawauchi H, Kiyono H, Takaiwa F, 2005. A rice-based edible vaccine expressing multiple T cell epitopes induces oral tolerance for inhibition of Th2-mediated IgE responses. *Proc. Natl. Acad. Sci. USA* 102:17525-30.

Takaiwa F, Kikuchi S, Oono K, 1987. A rice glutelin family: major type of glutelin mRNAs can be divided into two classes. *Mol. Gen. Genet.* 208, 15–22.

Takaiwa F, Oono K, 1990. Genomic DNA sequences of two new genes for new storage protein glutelin in rice. *Jpn. J. Genet.* 66, 161–171.

Takaiwa F, Oono K, Wing D, Kato A, 1991. Sequence of three members and expression of a new major subfamily of glutelin genes from rice. *Plant Mol. Biol.* 17, 875–885.

Takaiwa F, Sakuta C, Choi SK, Tada Y, Motoyama T, Utsumi S, 2008. Co-expression of soybean glycinins A1aB1b and A3B4 enhances their accumulation levels in transgenic rice seed. *Plant Cell Physiol.* 49, 1589-99.

Takaiwa F, Takagi H, Hirose S, Wakasa Y, 2007. Endosperm tissue is good production platform for artificial recombinant proteins in transgenic rice. *Plant Biotechnol. J.* 5, 84-92.

Takaiwa F, Yamanouchi U, Yoshihara T, Washida H, Tanabe F, Kato A, Yamada K, 1996. Characterization of common *cis*-regulatory elements responsible for the endosperm-specific expression of members of the rice glutelin multigene family. *Plant Mol. Biol.* 30, 1207-1221.

Tanaka K, Sugimoto T, Ogawa M, Kasai Z, 1980. Isolation and characterization of two types of protein bodies in rice endosperm. *Agric. Biol. Chem.* 44, 1633–1639.

Tompa M, Li N, Bailey TL, Church GM, De Moor B, Eskin E, Favorov AV, Frith MC, Fu Y, Kent WJ, Makeev VJ, Mironov AA, Noble WS, Pavesi G, Pesole G, Régnier M, Simonis N, Sinha S, Thijs G, van Helden J, Vandenbogaert M, Weng Z, Workman C, Ye C, Zhu Z, 2005. Assessing computational tools for the discovery of transcription factor binding sites. *Nat. Biotechnol.* 23(1), 137-144.

Twyman RM, Stoger E, Schillberg S, Christou P, Fischer R, 2003. Molecular farming in plants: host systems and expression technology. *Trends Biotechnol.* 21, 570-578.

Van Helden J, 2003. Regulatory sequence analysis tools. *Nucleic Acids Res.* 31, 3593–3596.

Venter M, 2007. Synthetic promoters: genetic control through *cis* engineering. *Trends Plant Sci.* 12, 118-124.

Vilar JMG, 2006. Modularizing gene regulation. *Mol. Systems Biol.* 2, 2006.0016.

Wakasa Y, Yasuda H, Oono Y, Kawakatsu T, Hirose S, Takahashi H, Hayashi S, Yang L, Takaiwa F, 2001. Expression of ER quality control-related genes in response to changes in BiP1 levels in developing rice endosperm. *Plant J.* 65, 675-689.

Wang G and Zhang W, 2006. A steganalysis-based approach to comprehensive identification and characterization of functional regulatory elements. *Genome Biology* 2006, 7:R49 doi:10.

Washida H, Wu CY, Suzuki A, Yamanouchi U, Akihama T, Harada K, Takaiwa F, 1999. Identification of *cis*-regulatory elements required for endosperm expression of the rice storage protein glutelin gene GluB-1. *Plant Mol. Biol.* 40, 1-12.

Werner T, Fessele S, Maier H, Nelson PJ, 2003. Computer modelling of promoter organization as a tool to study transcriptional coregulation. *FASEB J.* 17, 1228–1237.

Wu CY, Suzuki A, Washida H, Takaiwa F, 1998. The GCN4 motif in a rice glutelin gene is essential for endosperm-specific gene expression and is activated by opaque-2 in transgenic rice plants. *Plant J.* 14, 673-683.

Wu C, Washida H, Onodera Y, Harada K, Takaiwa F, 2000. Quantitative nature of the Prolamin-box, ACGT and AACA motifs in a rice glutelin gene promoter: minimal *cis*-element requirements for endosperm-specific gene expression. *Plant J.* 23, 415-421.

Yamagata H, Sugimoto T, Tanaka K, Kasai Z, 1982. Biosynthesis of storage proteins in developing rice seeds. *Plant Physiol.* 70, 1094-1100.

Yamamoto MP, Onodera Y, Touno SM, Takaiwa F, 2006. Synergism between RPB1 Dof and RISBZ1 bZIP activators in the regulation of rice seed expression genes. *Plant Physiol.* 141, 1694-1707.

Yang D, Wu L, Hwang YS, Chen L, Huang N, 2001. Expression of the REB transcriptional activator in rice grains improves the yield of recombinant proteins whose genes are controlled by a Reb-responsive promoter. *Proc. Natl. Acad. Sci. USA* 98, 11438-43.

Yang J, Zhang J, Huang Z, Wang Z, Zhu Q, Liu L, 2002. Correlation of cytokinin levels in the endosperms and roots with cell number and cell division activity during endosperm development in rice. *Annals of Botany* 90, 369-377.

Yang L, Wakasa Y, Kawakatsu T, Takaiwa F, 2009. The 3'-untranslated region of rice glutelin GluB-1 affects accumulation of heterologous protein in transgenic rice. *Biotechnol. Lett.* 31, 1625-1631.

Yoshihara T, Takaiwa F, 1996. *Cis*-regulatory elements responsible for quantitative regulation of the rice seed storage protein glutelin GluA-3 gene. *Plant Cell Physiol.* 37, 107-111.

Yoshihara T, Washida H, Takaiwa F, 1996. A 45-bp proximal region containing AACAA and GCN4 motif is sufficient to confer endosperm-specific expression of the rice storage protein glutelin gene, GluA-3. *FEBS Lett.* 383, 213-218.

8 SEQUENZE DEI PROMOTORI SINTETICI

Di seguito sono riportate le sequenze dei promotori sintetici che sono stati costruiti nell'ambito di questo dottorato. Per i promotori costruiti sulla base di GluB4 la sequenza termina in 3' con il sito di inizio trascrizione, mentre per tutti gli altri la sequenza termina in 3' con il sito *XhoI* (GAGCTC) che serve ad ancorare il promotore sintetico alla TATA-box di GluB4.

Le sette sequenze sono disposte seguendol'ordine alfabetico dei nomi dei relativi promotori sintetici.

16-Glb

```
1 TTAATTAGCG ATGGTGCAGG TCACGTACCC AAATTTCTTC ACTGTTGGAT
51 CAACTAGAGT AGTTAAACGA GGGCATGTGA TGAAGGCTAG CTATTTGAAA
101 TTTTCCAATT ATCCCTGCAT AAGTCAGGCT ACAATAGCAC CTGGACTACA
151 TGCAGGGATT ACAAATAGG TGGTAACCAC ATTTACCGCG TTAACCCTAT
201 CAAATTCAAA TAAATTTTAA AAGTAATTTG ATTTTTTTAA TAAATTTTGT
251 ATGGTTTCTC AAGCTTTTAT TTGGTTACCG TGCTTACTGC CGGAGGCAAT
301 GGGAAACCCT CACTAGAAGT TGCACCTGTT CTTGTCTGTG CACCATATCA
351 TGTTGAATCA TGTGCGTTGT GTCCTTTCGG AAGAACCGAT TTAATACATG
401 ACTCATCAAT TCCACTTTAC GTATCAAAAG GTTTGTATATG GGGGCAATGC
451 TTTTGTGAAA TTAAATTTTT ATTTTTCGTC ACGTTGTATC TAGTTAAACA
501 CTACCTACCT ACCATTACAA AACCTCATTC CACAAAACGA TGCATCTAGT
551 GATTCCTGGA GGGAGGAGAG GGGAGAGATG GTGAGAGAGG AGGAAGAAGA
601 GGAGGGGTGA CAATGATATG TGGGGCCATG TGGGCCCCAC CTTTTTTTAA
651 TTCATTCTTT TGTGAAACT GACATGTGGG TCCCATGAGA TTTATTATTT
701 TTCGGATCGA ATTGCCACGT AAGCGCTACG TCAATGCTAC GTCAGATGAA
751 GACCGAGTCA AATTAGCCAC GTAAGCGCCA CGTCAGCCAA AACCACCATC
801 CAAACCGCCG AGGGACCTCA TCTGCACTGG TTTTGATAGT TGAGGGACCC
851 GTTGTATCTG GTTTTTCGAT TGAAGGACGA AAATCAAATT TGTTGACAAG
901 TTAAGGGACC TTAAATGAAC TTATTCCATT TCAAATATT CTGTGAGCCA
951 TATATACCGT GGGCTTCCAA TCCTCCTCAA ATTAAAGGGC CTTTTTAAAA
1001 TAGATAATTG CCTTCTTTCA GTCACCCATA AAAGTACAAA ACTACTACCA
1051 ACAAGCAACA TGCGCAGTTA CACACATTTT CTGCACATTT CCACCACGTC
1101 ACAAAGAGCT AAGAGTTATC CCTAGGACAA TCTCATTAGT GTAGATACAT
1151 CCATTAATCT TTTATCAGAG GCAAACGTAA AGCCGCTCTT TATGACAAAA
1201 ATAGGTGACA CAAAAGTGTT ATCTGCCACA TACATAACTT CAGAAATTAC
1251 CCAACACCAA GAGAAAAATA AAAAAAATC TTTTTCGCAAG CTCCAAATCT
1301 TGGAAACCTT TTTCACTCTT TGCAGCATTG TACTCTTGCT CTTTTTCCAA
```

1351 CCGATCCATG TCACCCTCAA GCTTCTACTT GATCTACACG AAGCTCACCG
1401 TGCACACAAC CATGGCCACC TCGAG

B1-B4

1 TTCTTGCTAC CAACAAC TTC ACAAAGTAGT AGTCAACCAA AACTATGCTA
51 AGGAATCACC TCACTTCCGC CCATGACCGT GAGCACGACT GTTCAAACAG
101 TTTGTTAATC TCTACAAAGA AGGTACTACTT TACCTACACA ACGCCACTAA
151 CCTGAGTTAC CCAGCCCATG CAAAATAGCC ACGTCTTGTG ACTTAAGGGA
201 TTTTCGCGACA AGGCATTTTCG AAAGCCCACA CAAGGACACC TTATGAAAAC
251 TGGAGGGGTC CCACAGACCA ACAACAAGTT AGGTCCCAAA CCATGTTGTG
301 CCAGGAAAAA TCCAAGGGGT CCTCCCCAAC ACCACCCCGA CAAATCCACT
351 TGTCCATTGG CATCAAGATT TGCCTGACCT AGCTAATTAC TCAGCCAGGC
401 ATGTCACAAT TCACCCATGT GGTCACACAT GTTAGGTTGG ATGAAATTCT
451 AAAGGAATCG GTCCATATGA GCAAGAACGA GAAAACCATA CCACCAGTAC
501 TTCTACCGAA ATACGAGTTT AGTAAACTCA TTTGTTTTCA AGGCACCGAC
551 CCAGGTGTGT CGGGTTTTTC AGGGATTTTG TAAACCCAAG TTTTACCCAT
601 AGTTGATCAT TCAAATTTTG AGGAGGGTCA TTGGTATCCG TACCTGAGGG
651 CACGAATACT GAGACCTAGC ATTGTAGTCG ACCAAGGAGG TTAATGCAGC
701 AATTGTAGGT GGGCCTGTT GGTTATATTG CAAACTGCGG CCAACATTTT
751 ATGTGTAATT TAGAGATGTG CATTTTGAGA AATGAAATAC TTAGTTTCAA
801 ATTATGGGCT CAAATAATCA AAGGTGACCT ACCTTGCTTG ATATCTTGAG
851 CTTCTTCTC GTATTCCGCG CACTAGGACT CTTCTGGCTC CGAAGCTACA
901 CGCCCAATGG CACTCACTGG CCTACCTCAC CGGTCAAGGT CTATCCTTTA
951 GGGTCACCAA CACAGATGAC CAAACGCTCG TCTTCACCAA CGTCGTGCCA
1001 CCAGGATGGA AGTTTGGCCA GACATTTGCA AGCAAGCTGC AGTTCAAGTG
1051 AGAGGAGAAG CCTGAATTGA TACCGGAGCG TTTCTTTTGG GAGTAACATC
1101 TCTGGTTGCC TAGCAAACAT ATGATTGTAT ATAAGTTTCG TTGTGCGTTT
1151 ATTCTTTCGG TGTGTAAAAT AACATACATG CTTTCCTGAT ATTTTCTTGT
1201 ATATATGTAC ACACACACGA CAAATCCTTC CATTTCTATT ATTATTGAAC
1251 AATTTAATTG CGAGGGCGAG TACTTGTCTG TTTACCTTTT TTTTTCAGAT
1301 GGCATTTTAT AGTTTAACTT TTCATGGACC GGCAGTAGTT CTAACCATGA
1351 ATGAAAAGAA ATCATAGTCC ACACCACGCA GGGACATTGT GGTCAATTTA
1401 GACAAGACGA TTTGATTAAT GTCTTGTATG ATATGGTCGA CAGTGAGGAC
1451 TAACAAACAT ATGGCATATT TTATTACCGG CGAGTTAAAT AAATTTATGT
1501 CACAGTAATA AACTGCCTAA TAAATGCACG CCAGAAAATA TAATGATAAA
1551 AAAAAGAAAA GATACATAAG TCCATTGCTT CTACTTTTTT AAAAATTTAA
1601 TCCAACATTT TCTATTTTTT GGTATAAACT TGGAAGTACT AGTTGGATAT
1651 GCAAAATCAT CTAACCTCCA TATATTTTAT CAATTTGTTT ACTTTACATA

1701 TGGGAGAGGA TAGTATGTCA AAGAAAATGA CAACAAGCTT ACAAGTTTCT
 1751 TATTTTAAAA GTTCCGCTAA CTTATCAAGC ATAGTGTGCC ACGCAAAACT
 1801 GACAACAAAC CAACAAATTT AAGGAGCGCC TAACTTATCA TCTATGACAT
 1851 ACCGCACAAA ATGATAACAT ACTAGAGAAA CTTTATTGCA CAAAAGGAAA
 1901 TTTATCCATA AGGCAAAGGA ACATCTTAAG GCTTTGGATA TACATTTACC
 1951 AACAAGCATT GTTTGTATTA CCCCTAAAGC GCAAGACATG TCATCCATGA
 2001 GTCATAGTGT GTATATCTCA ACATTGCAAA GCTACCTTTT TTCTATTATA
 2051 CTTTTCGCAT TATAGGCTAG ATATTATCTA TACATGTCAA CAAACTCTAT
 2101 CCCTACGTCA TATCTGAAGA TTCTTTTCTT CACTATATAA GTTGGCTTCC
 2151 CTGTCATTGA ACTCACATC

B1-G1b

1 TCTTGCTACC AACAACTTCA CAAAGTAGTA GTCAACCAAA ACTATGCTAA
 51 GGAATCACCT CACTTCCGCC CATGACCGTG AGCACGACTG TTCAAACAGT
 101 TTGTTAATCT CTACAAAGAA GGTACACTTT ACCTACACAA CGCCACTAAC
 151 CTGAGTTACC CAGCCCATGC AAAATAGCCA CGTCTTGTGA CTTAAGGGAT
 201 TTCGCGACAA GGCATTTCGA AAGCCACAC AAGGACACCT TATGAAAAC
 251 GGAGGGGTCC CACAGACCAA CAACAAGTTA GGTCCCAAAC CATGTTGTGC
 301 CAGGAAAAAT CCAAGGGGTC CTCCCAACA CCACCCGAC AAATCCACTT
 351 GTCCATTGGC ATCAAGATTT GCCTGACCTA GCTAATTACT CAGCCAGGCA
 401 TGTCACAATT CACCCATGTG GTCACACATG TTAGGTTGGA TGAAATTCTA
 451 AAGGAATCGG TCCATATGAG CAAGAACGAG AAAACCATAC CACCAGTACT
 501 TCTACCGAAA TACGAGTTTA GTAAACTCAT TTGTTTTCAA GGCACCGACC
 551 CAGGTGTGTC GGGTTTTCCA GGGATTTTGT AAACCCAAGT TTTACCCATA
 601 GTTGATCATT CAAATTTTGA GGAGGGTCAT TGGTATCCGT ACCTGAGGGC
 651 ACGAATACTG AGACCTAGCA TTGTAGTCGA CCAAGGAGGT TAATGCAGCA
 701 ATTGTAGGTG GGGCCTGTTG GTTATATTGC AAAGTGCAGC CAACATTTCA
 751 TGTGTAATTT AGAGATGTGC ATTTTGTAGAA ATGAAATACT TAGTTTCAA
 801 TTATGGGCTC AAATAATCAA AGGTGACCTA CCTTGCTTGA TATCTTGAGC
 851 TTCTTCCTCG TATTCCGCGC ACTAGGACTC TTCTGGCTCC GAAGCTACAC
 901 GTGGAACGAG ATAACTCAAC AAAACGACCA AGGAAAAGCT CGTATTAGTG
 951 AGTACTAAGT GTGCCACTGA ATAGATCTCG ATTTTTGAGG AATTTTAGAA
 1001 GTTGAACAGA GTCAATCGAA CAGACAGTTG AAGAGATATG GATTTTCTAA
 1051 GATTAATTGA TTCTCTGTCT AAAGAAAAAA AGTATTATTG AATTAATGG
 1101 AAAAAGAAAA AGGAAAAAGG GGATGGCTTC TGCCTTTTTG GGCTGAAGGC
 1151 GGCGTGTGGC CCAGCGTGCT GCGTGGGGAC AGCCGAGCGA ACACACGACG
 1201 GAGCAGCTAC GACGAACGGG GGACCGAGTG GACCGGACGA GGATGTGGCC
 1251 TAGTGATTCC TGGAGGGAGG AGAGGGGAGA GATGGTGAGA GAGGAGGAAG

1301 AAGAGGAGGG GTGACAATGA TATGTGGGGC CATGTGGGCC CCACCATTTT
 1351 TTAATTCATT CTTTTGTTGA AACTGACATG TGGGTCCCAT GAGATTTATT
 1401 ATTTTTCGGA TCGAATTGCC ACGTAAGCGC TACGTCAATG CTACGTCAGA
 1451 TGAAGACCGA GTCAAATTAG CCACGTAAGC GCCACGTCAG CCAAAAACCAC
 1501 CATCCAAACC GCCGAGGGAC CTCATCTGCA CTGGTTTTGA TAGTTGAGGG
 1551 ACCCGTTGTA TCTGGTTTTT CGATTGAAGG ACGAAAATCA AATTTGTTGA
 1601 CAAGTTAAGG GACCTTAAAT GAACTTATTC CATTTCAAAA TATTCTGTGA
 1651 GCCATATATA CCGTGGGCTT CCAATCCTCC TCAAATTAAA GGGCCTTTTT
 1701 AAAATAGATA ATTGCCTTCT TTCAGTCACC CATAAAAGTA CAAAACACT
 1751 ACCAACAAGC AACATGCGCA GTTACACACA TTTTCTGCAC ATTTCCACCA
 1801 CGTCACAAAG AGCTAAGAGT TATCCCTAGG ACAATCTCAT TAGTGTAGAT
 1851 ACATCCATTA ATCTTTTATC AGAGGCAAAC GTAAAGCCGC TCTTTATGAC
 1901 AAAAATAGGT GACACAAAAG TGTTATCTGC CACATACATA ACTTCAGAAA
 1951 TTACCCAACA CCAAGAGAAA AATAAAAAAA AATCTTTTTG CAAGCTCCAA
 2001 ATCTTGAAA CCTTTTTTAC TCTTTGCAGC ATTGTACTCT TGCTCTTTTT
 2051 CCAACCGATC CATGTCACCC TCAAGCTTCT ACTTGATCTA CACGAAGCTC
 2101 ACCGTGCACA CAACCATGGC CACCTCGAG

C-B4

1 GCGTTCAAGA TTTATTTTTG GTATTTAATT TACTTGCTTA AGTCAGATAT
 51 ATTCCCATCG TTGCAGGTTT GTCACTTAGT ATTATTATTA AGCGCTCTAG
 101 CACTAGGACT CTGGATAAAT AAGAAAGTTT ATTCACGAGG CTAGAGTAGT
 151 AATCAATAAC ATAAGCGTGG TGTCTAGGTC AGCGGTTATC TTCATATGTA
 201 GTGTGCTCCA TGGAAAGTGA GGTAGGAGGA AGGTGGTGAC AGTCCCGTCC
 251 GTCCTTTGTA TCCCTCCATG TTCGGGTATA TCATAGAGCT ACAGGCTAGA
 301 CTTAGCTTGG CAGACTAGGG GAGAGCCGGT GCTCGAAGCA ATCCATGAGG
 351 CTTTACATTT AACATAAGTT AGTAAATTAA CCCATAGGAA TCATCTCTAG
 401 ACTGAACCTA CCAGTAGTTG TGCTTGGATA TAATTATATT CCTACATATA
 451 CATAACGTT CCCTGCGATT AGATACCCTT GGAATACTCT AAGGTGAAGT
 501 GCTACAGCGG TATCCGTGCG CTTGCGGATT TATCTGTGAC CGTATCAAAT
 551 ACCAACAGGT AGATACAAGG AATCATCTCT CCTATCCATT GGTTTATCAT
 601 CTTTTAAAAT TATCTCTTGC TCTCCTATTG CCTCTGCAAC TGCGGATAGG
 651 TGTTTCTCAA CAATGAAGGT TGTGAAGAAT GCTTTGTGCA ACAAGATGGA
 701 TGACAAGTAT CTCAGCCATA GCCTCATTTG CTTTGTAGAA AAGGATATGT
 751 CGGACACAAT CACTAAGTAT CACCGTGGAA AGGATGCACT GTATGCCCTA
 801 TCTATATTTA CCATTTAGTA ATATTTATAT GGCTTGTGCT AACTTTATGT
 851 TGTCTTTACA GGCAATAACA TTATTTGGAA GGCATATCTA TATATTAATA
 901 TTTAAGATAA TGTAATATCT CAAAGTTTTT ATAAGCTGCA ATGAGGTGAG

951 TTTCACTTAG CTTTCTAACT TGTTATGAGT TATAGATGCA TGCCACCAGT
 1001 CATTFFFFFFAT CTTGCATCAG CCCCTGCCTG TTAGAATATG TTTCTTTGTC
 1051 TGGGAGTCCA TGTCAACTAG CCAATTTCCA AATATATGAA CAAAACATATG
 1101 TGGCCTTTGT AACCCAAATG AGATAAAGAC TACTCTCCAT AGAAATTCTA
 1151 CAGGGTTCCT TGCCTGAAGA AGGGTGGCCT GCGGTTCCAC ATTAACGGTC
 1201 ACGACTACTT CCAGCTAGTA CTGGTGACCA ACGTCGCGGC GGCAGGGTCA
 1251 ATCAAGTCCA TGGAGGTTAT GGGTTCCAAC ACAGCGGATT GGATGCCGAT
 1301 GGCACGTAAC TGGGGCGCCC AATGGCACTC ACTGGCCTAC CTCACCGGTC
 1351 AAGGTCTATC CTTTAGGGTC ACCAACACAG ATGACCAAAC GCTCGTCTTC
 1401 ACCAACGTCG TGCCACCAGG ATGGAAGTTT GGCCAGACAT TTGCAAGCAA
 1451 GCTGCAGTTC AAGTGAGAGG AGAAGCCTGA ATTGATACCG GAGCGTTTCT
 1501 TTTGGGAGTA ACATCTCTGG TTGCCTAGCA AACATATGAT TGTATATAAG
 1551 TTTTCGTTGTG CGTTTATTCT TTCGGTGTGT AAAATAACAT ACATGCTTTC
 1601 CTGATATTTT CTTGTATATA TGTACACACA CACGACAAAT CCTTCCATTT
 1651 CTATTATTAT TGAACAATTT AATTGCGAGG GCGAGTACTT GTCTGTTTAC
 1701 CTTTTTTTTT CAGATGGCAT TTTATAGTTT AACCTTTCAT GGACCGGCAG
 1751 TAGTTCCTAAC CATGAATGAA AAGAAATCAT AGTCCACACC ACGCAGGGAC
 1801 ATTGTGGTCA TTTTAGACAA GACGATTTGA TTAATGTCTT GTATGATATG
 1851 GTCGACAGTG AGGACTAACA AACATATGGC ATATTTTATT ACCGGCGAGT
 1901 TAAATAAATT TATGTCACAG TAATAAACTG CCTAATAAAT GCACGCCAGA
 1951 AAATATAATG ATAAAAAAAA GAAAAGATAC ATAAGTCCAT TGCTTCTACT
 2001 TTTTTAAAAA TTAAATCCAA CTTTTTCTAT TTTTTGGTAT AAACCTGGAA
 2051 GTAAGTAGTTG GATATGCAAA ATCATCTAAC CTCCATATAT TTCATCAATT
 2101 TGTTTACTTT ACATATGGGA GAGGATAGTA TGTCAAAGAA AATGACAACA
 2151 AGCTTACAAG TTTCTTATTT TAAAAGTTCC GCTAACTTAT CAAGCATAGT
 2201 GTGCCACGCA AAAGTACAAA CAAACCAACA AATTTAAGGA GCGCCTAACT
 2251 TATCATCTAT GACATACCGC ACAAATGAT AACATACTAG AGAAACTTTA
 2301 TTGCACAAAA GGAAATTTAT CCATAAGGCA AAGGAACATC TTAAGGCTTT
 2351 GGATATACAT TTACCAACAA GCATTGTTTG TATTACCCCT AAAGCGCAAG
 2401 ACATGTCATC CATGAGTCAT AGTGTGTATA TCTCAACATT GCAAAGCTAC
 2451 CTTTTTTCTA TTATACTTTT CGCATTATAG GCTAGATATT ATCTATACAT
 2501 GTCAACAAAC TCTATCCCTA CGTCATATCT GAAGATTCTT TTCTTCACTA
 2551 TATAAGTTGG CTTCCCTGTC ATTGAACTCA CATC

C-Glb

1 GCGTTCAAGA TTTATTTTTG GTATTTAATT TACTTGCTTA AGTCAGATAT
 51 ATTCCCATCG TTGCAGGTTT GTCACCTAGT ATTATTATTA AGCGCTCTAG
 101 CACTAGGACT CTGGATAAAT AAGAAAGTTT ATTCACGAGG CTAGAGTAGT

151 AATCAATAAC ATAAGCGTGG TGTCTAGGTC AGCGGTTATC TTCATATGTA
 201 GTGTGCTCCA TGGAAAGTGA GGTAGGAGGA AGGTGGTGAC AGTCCCGTCC
 251 GTCCTTTGTA TCCCTCCATG TTCGGGTATA TCATAGAGCT ACAGGCTAGA
 301 CTTAGCTTGG CAGACTAGGG GAGAGCCGGT GCTCGAAGCA ATCCATGAGG
 351 CTTTACATTT AACATAAGTT AGTAAATTAA CCCATAGGAA TCATCTCTAG
 401 TGATTCCTGG AGGGAGGAGA GGGGAGAGAT GGTGAGAGAG GAGGAAGAAG
 451 AGGAGGGGTG ACAATGATAT GTGGGGCCAT GTGGGCCCCA CCATTTTTTTA
 501 ATTCATTCTT TTGTTGAAAC TGACATGTGG GTCCCATGAG ATTTATTATT
 551 TTTCGGATCG AATTGCCACG TAAGCGCTAC GTCAATGCTA CGTCAGATGA
 601 AGACCGAGTC AAATTAGCCA CGTAAGCGCC ACGTCAGCCA AAACCACCAT
 651 CCAAACCGCC GAGGGACCTC ATCTGCACTG GTTTTGATAG TTGAGGGACC
 701 CGTTGTATCT GGTTTTTCGA TTGAAGGACG AAAATCAAAT TTGTTGACAA
 751 GTTAAGGGAC CTTAAATGAA CTTATTCCAT TTCAAAATAT TCTGTGAGCC
 801 ATATATAACG TGGGCTTCCA ATCCTCCTCA AATTAAAGGG CCTTTTTAAA
 851 ATAGATAATT GCCTTCTTTC AGTCACCCAT AAAAGTACAA AACTACTACC
 901 AACAAGCAAC ATGCGCAGTT ACACACATTT TCTGCACATT TCCACCACGT
 951 CACAAAGAGC TAAGAGTTAT CCCTAGGACA ATCTCATTAG TGTAGATACA
 1001 TCCATTAATC TTTTATCAGA GGCAAACGTA AAGCCGCTCT TTATGACAAA
 1051 AATAGGTGAC ACAAAGTGT TATCTGCCAC ATACATAACT TCAGAAATTA
 1101 CCCAACACCA AGAGAAAAAT AAAAAAAAAAT CTTTTTGCAA GCTCCAAATC
 1151 TTGGAAACCT TTTTCACTCT TTGCAGCATT GTACTCTTGC TCTTTTTCCA
 1201 ACCGATCCAT GTCACCCTCA AGCTTCTACT TGATCTACAC GAAGCTCACC
 1251 GTGCACACAA CCATGGCCAC CTCGAG

C-Glb-B4

1 GCGTTCAAGA TTTATTTTTG GTATTTAATT TACTTGCTTA AGTCAGATAT
 51 ATTCCCATCG TTGCAGGTTT GTCACCTAGT ATTATTATTA AGCGCTCTAG
 101 CACTAGGACT CTGGATAAAT AAGAAAGTTT ATTCACGAGG CTAGAGTAGT
 151 AATCAATAAC ATAAGCGTGG TGTCTAGGTC AGCGGTTATC TTCATATGTA
 201 GTGTGCTCCA TGGAAAGTGA GGTAGGAGGA AGGTGGTGAC AGTCCCGTCC
 251 GTCCTTTGTA TCCCTCCATG TTCGGGTATA TCATAGAGCT ACAGGCTAGA
 301 CTTAGCTTGG CAGACTAGGG GAGAGCCGGT GCTCGAAGCA ATCCATGAGG
 351 CTTTACATTT AACATAAGTT AGTAAATTAA CCCATAGGAA TCATCTCTAG
 401 TGATTCCTGG AGGGAGGAGA GGGGAGAGAT GGTGAGAGAG GAGGAAGAAG
 451 AGGAGGGGTG ACAATGATAT GTGGGGCCAT GTGGGCCCCA CCATTTTTTTA
 501 ATTCATTCTT TTGTTGAAAC TGACATGTGG GTCCCATGAG ATTTATTATT
 551 TTTCGGATCG AATTGCCACG TAAGCGCTAC GTCAATGCTA CGTCAGATGA
 601 AGACCGAGTC AAATTAGCCA CGTAAGCGCC ACGTCAGCCA AAACCACCAT

651 CCAAACCGCC GAGGGACCTC ATCTGCACTG GTTTTGATAG TTGAGGGACC
 701 CGTTGTATCT GGTTTTTCGA TTGAAGGACG AAAATCAAAT TCTACAGGGT
 751 TCCTTGCGTG AAGAAGGGTG GCCTGCGGTT CACCATTAAC GGTCACGACT
 801 ACTTCCAGCT AGTACTGGTG ACCAACGTCG CGGCGGCAGG GTCAATCAAG
 851 TCCATGGAGG TTATGGGTTC CAACACAGCG GATTGGATGC CGATGGCACG
 901 TAACTGGGGC GCCCAATGGC ACTCACTGGC CTACCTCACC GGTCAAGGTC
 951 TATCCTTTAG GGTCACCAAC ACAGATGACC AAACGCTCGT CTTACCAAC
 1001 GTCGTGCCAC CAGGATGGAA GTTTGGCCAG ACATTTGCAA GCAAGCTGCA
 1051 GTTCAAGTGA GAGGAGAAGC CTGAATTGAT ACCGGAGCGT TTCTTTTGGG
 1101 AGTAACATCT CTGGTTGCCT AGCAAACATA TGATTGTATA TAAGTTTCGT
 1151 TGTGCGTTTA TTCTTTCGGT GTGTAAAATA ACATACATGC TTTCTGATA
 1201 TTTTCTTGTA TATATGTACA CACACACGAC AAATCCTTCC ATTTCTATTA
 1251 TTATTGAACA ATTTAATTGC GAGGGCGAGT ACTTGTCTGT TTACCTTTTT
 1301 TTTTTCAGAT GGCATTTTAT AGTTTAACTT TTCATGGACC GGCAGTAGTT
 1351 CTAACCATGA ATGAAAAGAA ATCATAGTCC ACACCACGCA GGGACATTGT
 1401 GGTCATTTTA GACAAGACGA TTTGATTAAT GTCTTGTATG ATATGGTCGA
 1451 CAGTGAGGAC TAACAAACAT ATGGCATATT TTATTACCGG CGAGTTAAAT
 1501 AAATTTATGT CACAGTAATA AACTGCCTAA TAAATGCACG CCAGAAAATA
 1551 TAATGATAAA AAAAAGAAAA GATACATAAG TCCATTGCTT CTACTTTTTT
 1601 AAAAATTAATA TCCAACATTT TCTATTTTTT GGTATAAACT TGGAAGTACT
 1651 AGTTGGATAT GCAAATCAT CTAACCTCCA TATATTTTAT CAATTTGTTT
 1701 ACTTTACATA TGGGAGAGGA TAGTATGTCA AAGAAAATGA CAACAAGCTT
 1751 ACAAGTTTCT TATTTTAAAA GTTCCGCTAA CTTATCAAGC ATAGTGTGCC
 1801 ACGCAAACT GACAACAAAC CAACAAATTT AAGGAGCGCC TAACCTATCA
 1851 TCTATGACAT ACCGCACAAA ATGATAACAT ACTAGAGAAA CTTTATTGCA
 1901 CAAAAGGAAA TTTATCCATA AGGCAAAGGA ACATCTTAAG GCTTTGGATA
 1951 TACATTTACC AACAAGCATT GTTTGTATTA CCCCTAAAGC GCAAGACATG
 2001 TCATCCATGA GTCATAGTGT GTATATCTCA ACATTGCAAA GCTACCTTTT
 2051 TTCTATTATA CTTTTCGCAT TATAGGCTAG ATATTATCTA TACATGTCAA
 2101 CAACTCTAT CCCTACGTCA TATCTGAAGA TTCTTTTCTT CACTATATAA
 2151 GTTGGCTTCC CTGTCATTGA ACTCACATC

Glb-B4

1 AATTTGTTGA CAAGTTAAGG GACCTTAAAT GAACTTATTC CATTTCAAAA
 51 TATTCTGTGA GCCATATATC CGTGGGCTTC CAATCCTCCT CAAATTAAG
 101 GGCCTTTTTA AAATAGATAA TTGCCTTCTT TCAGTCACCC ATAAAAGTAC
 151 AAACTACTA CCAACAAGCA ACATGCGCAG TTACACACAT TTTCTGCACA
 201 TTTCCACCAC GTCACAAAGA GCTAAGAGTT ATCCCTAGGA CAATCTCATT

251 AGTGTAGATA CATCCATTAA TCTTTTATCA GAGGCAAACG TAAAGCCGCT
301 CTTTATGACA AAAATAGGTG ACACAAAAGT GTTATCTGCC ACATACATAA
351 C TTCAGAAAT TACCCAACAC CAAGAGAAAA ATAAAAAAAA ATCTTTTTGC
401 AAGCTCCAAA TCTTGGAAC CTTTTTCACT CTTTGCAGCA TTGTACTCTT
451 GCTCTTTTTT CAACCGATCC ATGTCACCCT CAAGCTTACA AGTTTCTTAT
501 TTTAAAAGTT CCGCTAACTT ATCAAGCATA GTGTGCCACG CAAAACGAC
551 AACAAACCAA CAAATTTAAG GAGCGCCTAA CTTATCATCT ATGACATACC
601 GCACAAAATG ATAACATACT AGAGAAACTT TATTGCACAA AAGGAAATTT
651 ATCCATAAGG CAAAGGAACA TCTTAAGGCT TTGGATATAC ATTTACCAAC
701 AAGCATTGTT TGTATTACCC CTAAAGCGCA AGACATGTCA TCCATGAGTC
751 ATAGTGTGTA TATCTCAACA TTGCAAAGCT ACCTTTTTTC TATTATACTT
801 TTCGCATTAT AGGCTAGATA TTATCTATAC ATGTCAACAA ACTCTATCCC
851 TACGTCATAT CTGAAGATTC TTTTCTTCAC TATATAAGTT GGCTTCCCTG
901 TCATTGAACT CACATC

9 INDICE DEI GRAFICI

- Grafico 1. I tre parametri per definire l'efficienza di un promotore sono la media di tutto il campione, la media del primo quarto di classifica e il livello del miglior espressore.102
- Grafico 2. Grafico riportante i valori assunti dai parametri di valutazione dei promotori posti a confronto nel secondo ciclo sperimentale. Le popolazioni prodotte nelle due sessioni di trasformazione sono state accorpate in un unico grafico. Per i promotori che sono stati trasformati in replicato è riportata la media dei parametri ottenuti. Nella pagina successiva è riportata una figura che illustra la struttura dei promotori ordinati secondo un livello crescente di espressione indotta.105
- Grafico 3. Livelli di espressione del gene *reporter* nelle foglie dei migliori espressori per ciascun costrutto.....108
- Grafico 4. Distribuzione della proteina reporter tra endosperma e tessuto aleuronico (farinaccio).110
- Grafico 5. Promotore GluB4: distribuzione dei livelli di espressione all'interno del campione della popolazione di trasformati primari. Le barre indicano la deviazione standard.....141
- Grafico 6. Promotore Δ GluC: distribuzione dei livelli di espressione all'interno del campione della popolazione di trasformati primari . Le barre indicano la deviazione standard.....141
- Grafico 7. Istogramma che evidenzia per ciascuno dei 9 promotori il valore assunto da media di popolazione, media del I quartile, e massimo. I promotori sono ordinati secondo un livello crescente di espressione indotta.148
- Grafico 8. Istogramma relativo alla sessione di trasformazione di febbraio 2012.150
- Grafico 9. Si riportano qui i risultati per una più agevole consultazione ...176
- Grafico 10. Dati riguardanti le trasformazioni replicate a febbraio 2012...178

Grafico 11. La produttività di GCasi e di seme in piante appartenenti alla popolazione trasformata con il promotore C-Glb-B4.....	181
Grafico 12. Correlazione fra livello di espressione di GCasi in endosperma (ascissa) e in foglia (ordinata).	182
Grafico 13. Produttività di seme e livello di espressione di GCasi in foglie di piante appartenenti alla popolazione trasformata con il promotore C-Glb-B4.....	183

10 INDICE DELLE FIGURE

Figura 1. Esperimenti condotti da Wu <i>et al.</i> nel 1998. (a) <i>Loss of function</i> ; (b) <i>gain of function</i> ; (c) livello di espressione.	21
Figura 2. Esperimenti di <i>loss of function</i> condotti da Washida <i>et al.</i> nel 1999. Promotore minimo GluB1 in tabacco.	23
Figura 3. Esperimenti condotti da Wu <i>et al.</i> nel 2000 su GluB1 in riso.	23
Figura 4. Promotori sintetici costruiti su base GluB1 e relativi livelli di espressione.	34
Figura 5. Promotori sintetici tronchi costruiti su base del promotore della globulina.	35
Figura 6. Promotori sintetici ottenuti dalla ricombinazione di porzioni appartenenti al promotore minimo GluB1. A destra sono riportati i livelli di espressione specifici (seed) e aspecifici (leaf).	36
Figura 7. Promotori sintetici costruiti su base 35S CaMV Δ [-46;-1] per ricombinazione di motivi regolatori per l'espressione in endosperma.	36
Figura 8. EMSA condotto da Onodera <i>et al.</i> (2001); qualunque mutazione perturbante il motivo TGAGTCA nel competitore appare causare l'assenza di legame con RISBZ1.	51
Figura 9. Rappresentazione schematica del vettore pUC18 con in dettaglio il MCS.	61
Figura 10. Rappresentazione schematica di pGEM-T con in dettaglio il MCS.	62
Figura 11. Struttura del vettore pCAMBIA con il marcatore selezionabile PMI. Nell'immagine è riprodotto il vettore che contiene il promotore di GluB4.	63
Figura 12. Mappa degli elementi cis-regolatori nei promotori delle prolamine.	89
Figura 13. Mappa degli elementi cis-regolatori nei promotori delle gluteline.	90

Figura 14. Mappa degli elementi cis-regolatori nel promotore della globulina.	92
Figura 15. Mappa del vettore di trasformazione utilizzato nel	94
Figura 16. Schema del modulo accettore; a monte del gene <i>reporter</i> sono evidenziati i siti necessari all'inserimento del promotore, e i 2 siti <i>EcoRI</i> laterali per la sua movimentazione nel vettore finale di espressione.....	95
Figura 17. Due esempi di come i promotori sintetici sono stati inseriti nel modulo accettore.	99
Figura 18. Schema riassuntivo rappresentante i promotori studiati in riso. Sono riportati graficamente gli elementi <i>cis</i> -regolatori individuati nel lavoro di analisi delle sequenze.....	100
Figura 19. Struttura dei promotori saggiati nel secondo ciclo sperimentale. L'ordine rispecchia quello della classifica in Figura 18.	106
Figura 20. Frazioni prodotte con il processo di sbiancatura del seme di riso.	109
Figura 21. Promotore della glutelina B1: differenze fra la sequenza in CR W3 (sopra) e la sequenza depositata in banca dati, gi 42742291 gb AY427569.1 <i>Oryza sativa</i> , var. Nipponbare (japonica cultivar-group).	112
Figura 22. Promotore della globulina: differenze fra la sequenza in CR W3 (sopra) e la sequenza depositata, gi 42742297 gb AY427575.1 <i>Oryza sativa</i> , var. Nipponbare (japonica cultivar-group).....	112
Figura 23. Promotore della prolamina 10kDa, differenze fra la sequenza in CR W3 (sopra) e la sequenza depositata, gi 42742294 gb AY427572.1 <i>Oryza sativa</i> , var. Nipponbare (japonica cultivar-group).	113
Figura 24. Promotore della prolamina 16kDa: differenze fra la sequenza in CR W3 (sopra) e la sequenza depositata, gi 42742296 gb AY427574.1 <i>Oryza sativa</i> , var. Nipponbare (japonica cultivar-group)	113

Figura 25. Modulo accettore: a monte del gene <i>reporter</i> sono evidenziati i siti necessari all'inserimento del promotore, e i 2 siti <i>EcoRI</i> laterali per la sua movimentazione nel vettore di espressione.	114
Figura 26. Inserimento della porzione di GluC a monte della TATA-box nel modulo accettore pGEM-T/TATA::LLTCK::GAA::NOS.	115
Figura 27. Inserimento della cassetta Δ -GluC-GAA nel vettore di espressione tramite sub-clonazione in pUC18. Il sito <i>HindIII</i> è stato tratto da pUC18 e sfruttato per l'inserimento in pCAMBIA.	116
Figura 28. Costruzione di C-Glb mediante digestione del vettore pGEM-T-GluC con <i>XbaI</i> ; costruzione di C-Glb_2695 mediante digestione del vettore pGEM-T-GluC con <i>NheI</i>	117
Figura 29. Gel di agarosio in cui sono stati caricati i prodotti PCR duplex ottenuti su coltura batterica con i <i>primers</i> GluC <i>for</i> , Glb <i>rev2</i> e M13 <i>rev</i> su 6 colonie di C-Glb.....	118
Figura 30. Gel di agarosio in cui sono stati caricati i prodotti PCR su coltura batterica. Corsie 1-6: colonie 1-6 con <i>primers</i> GluC <i>for3</i> e Glb <i>rev2</i> . Corsie 7-12: colonie 1-6 con <i>primers</i> GluC <i>for3</i> e M13 <i>rev</i>	118
Figura 31. Sostituzione della porzione prossimale di Prol 16 con Glb. <i>XbaI</i> e <i>SpeI</i> producono estremità compatibili.....	119
Figura 32. Costruzione del promotore ibrido B1-Glb. L'apertura del vettore avviene tramite taglio con <i>BnI</i> e <i>XhoI</i>	120
Figura 33. Modalità di inversione dell'insero clonato nel MCS di pGEM-T. Questa procedura è applicabile a qualunque insero clonato fra le 2 T, purchè privo di siti di riconoscimento per <i>EagI</i>	121
Figura 34. Costruzione di un promotore ibrido con base GluB4.....	122
Figura 35. Inserimento dell'elemento GluC per digestione sequenziale con <i>AatII</i> e <i>ApoI</i> . L'insero viene saldato immediatamente a monte di GluB4, in corrispondenza del sito <i>EcoRI</i>	123
Figura 36. Costruzione del promotore ternario C-Glb-B4.	124

Figura 37. Sostituzione della porzione distale di GluB4, sita a monte di <i>HindIII</i>	125
Figura 38. Costruzione di B1-B4. L'inserto derivato da GluB1 ha un'estremità piatta in 3', prodotta dall'enzima <i>BsaAI</i> . Il vettore derivato da pGEM-T-GluB4-LLTCK viene aperto con <i>SphI</i> ed <i>EheI</i>	126
Figura 39. Gel di agarosio in cui sono stati caricati i prodotti ottenuti mediante PCR effettuata direttamente su coltura batterica. Corsia 1-3: B1-B4, colonie 1-3 <i>primers</i> GluB1 <i>for</i> e GluB1 <i>rev4</i> . Corsia 4-6: C-B4, colonie 1-3 <i>primers</i> GluC <i>for</i> GluC <i>rev4</i> . Corsia 7-9: C-Glb-B4, colonie 1-3 <i>primers</i> Glb <i>for2</i> e Glb <i>rev2</i> . Corsia 10-12 C-Glb-B4, colonie 1-3 <i>primers</i> GluC <i>for</i> e GluC <i>rev4</i>	127
Figura 40. Corsia 1: C-Glb-B4-3087 digerito con <i>SphI</i> e <i>SnaBI</i> . Corsia 2: C-Glb-B4-3087 digerito con <i>AatII</i> e <i>MluI</i> . Corsia 3: C-Glb-B4-3087 digerito con <i>SphI</i> e <i>HindIII</i> . Corsia 4: Glb-B4 digerito con <i>AatII</i> e <i>SnaBI</i> . Corsia 5: Glb-B4 digerito con <i>SphI</i> e <i>HindIII</i> . Corsia 6: C-Glb-B4 digerito con <i>AatII</i> e <i>SnaBI</i> . Corsia 7: C-Glb-B4-3087 digerito con <i>SphI</i> e <i>HindIII</i> . Corsia 8: C-Glb-B4-3087 digerito con <i>AatII</i> e <i>SnaBI</i> . Corsia 9: B1-B4 digerito con <i>AatII</i> e <i>SnaBI</i> . Corsia 10: B1-B4 digerito con <i>SphI</i> e <i>SnaBI</i> . Corsia 11: C-B4 digerito con <i>AatII</i> e <i>SnaBI</i>	128
Figura 41. Gel da cui sono state recuperate le bande relative ai vari promotori. Corsia 1: pGEM-T/TATA::LLTCK::NOS digerito con <i>XbaI</i> e <i>SacI</i> . Corsia 2: pGEM-T/Glb digerito con <i>SphI</i> e <i>XhoI</i> . Corsia 3: pGEM-T/C-Glb digerito con <i>SphI</i> e <i>XhoI</i> . Corsia 4: pGEM-T/P16-Glb digerito con <i>SphI</i> e <i>XhoI</i> . Corsia 5: pGEM-T/B1-Glb, digerito con <i>SphI</i> e <i>XhoI</i> . Corsia 6: pGEM-T/Glb-B4 digerito con <i>SphI</i> e <i>XbaI</i> . Corsia 7: pGEM-T/C-Glb-B4 digerito con <i>SphI</i> e <i>XbaI</i> . Corsia 8: pGEM-T/B1-B4 digerito con <i>SphI</i> e <i>XbaI</i>	132
Figura 42. Analisi di restrizione delle cassette di espressione (prima parte). Corsie 1 e 2: pGEM-T/P16-Glb::GCasi, colonie 1 e 2, mp digerite con <i>XhoI</i> e <i>EcoRI</i> . Corsie 3 e 4: pGEM-T/Glb::GCasi, colonie 1 e	

2, mp digerite con <i>XhoI</i> e <i>EcoRI</i> . Corsia 5: pGEM-T/Glb-B4::GCasi, colonia 1, mp digerita con <i>XhoI</i> e <i>EcoRI</i> . Corsie 6 e 7: pGEM-T/B1-B4::GCasi, colonia 5, mp digerita con <i>XbaI</i> ed <i>EcoRI</i>	133
Figura 43. Analisi di restrizione delle cassette di espressione (seconda parte). Corsie 1 e 2: pGEM-T/C-Glb::GCasi, colonie 1 e 2, mp digerite con <i>XhoI</i> e <i>EcoRI</i> . Corsie 3 e 4: pGEM-T/B1-Glb::GCasi, colonie 2 e 4, mp digerite con <i>XhoI</i> e <i>EcoRI</i> . Corsie 5 e 6: pGEM-T/GluC::GCasi, colonie 1 e 2, mp digerite con <i>XbaI</i> e <i>EcoRI</i> . Corsia 7: pGEM-T/C-B4::GCasi, colonia 1, mp digerita con <i>SnaBI</i> ed <i>EcoRI</i>	133
Figura 44. Analisi di restrizione della cassetta C-Glb-B4. Corsie 1-6: pGEM-T/C-Glb-B4::GCasi, colonie 1-6, mp digerite con <i>EcoRI</i> e <i>XbaI</i>	134
Figura 45. Tavola raffigurante le principali fasi di trasformazione del riso. A: sbramatura; B: disinfezione; C: germinazione; D: isolamento scutelli; E: callogenesi; F: selezione embrioidi; G: infezione con coltura liquida di agrobatterio; H: embrioidi su substrato CCM dopo infezione; I: embrioidi trasformati e non trasformati su substrato SMI; L: calli da embrioidi trasformati su substrato SMII; M: calli trasformati su substrato PRM; N: rigenerazione; O: differenziazione; P-Q: radicazione; R: trasferimento in torba...	135
Figura 46. Pannocchia allo stadio di botticella.	153
Figura 47. Campioni di materiale prodotto nel processo di sbiancatura.	156
Figura 48. I due promotori che comprendono la porzione 5' di GluB1. Per la legenda dei motivi regolatori vedere pagina successiva.	172
Figura 49. Porzione del promotore GluC che, in base ai risultati ottenuti appare inattiva. Notare che anche l'analisi di sequenza conferma questo dato.....	173
Figura 50. Porzione di Prol16 che è in grado di aumentare il livello di espressione di Glb. Notare come la posizione degli elementi	

regolatori sia stata traslata in direzione 5' nel promotore chimerico.....	174
Figura 51. Il promotore Glb-B4 ed i due promotori naturali che lo compongono. Si noti che in questo caso è stata utilizzata la porzione 3' di entrambi i componenti.....	174
Figura 52. Struttura di C-Glb-B4 unico ibrido ternario.	175
Figura 53. Le tre trasformazioni ripetute. Deviazioni standard rispetto alla media.....	178

11 INDICE DELLE TABELLE

Tabella 1 Esempi di sintesi di proteine farmaceutiche in pianta (a esclusione di vaccini, anticorpi, nutraceutici).	13
Tabella 2. Promotori analizzati, abbreviazioni utilizzate in questa tesi di dottorato, numeri di accessione e riferimento bibliografico da cui sono state tratte informazioni.....	47
Tabella 3. Sequenze consenso identificate in base ai dati reperiti in letteratura. Il codice di ambiguità è: R=A/G, Y=C/G, K=G/T, S=G/C, M=A/C, W=A/T, B=G/C/T, V=A/G/C, D=A/G/T, H=A/C/T, N=A/G/C/T	49
Tabella 4. Riassunto delle conclusioni tratte dalla elaborazione dei risultati di esperimenti di EMSA. R=A/G, Y=C/G, K=G/T, S=G/C, M=A/C, W=A/T, B=G/C/T, V=A/G/C, D=A/G/T, H=A/C/T, N=A/G/C/T ..	52
Tabella 5. Elenco dei promotori utilizzati con i relativi numeri di accessione.	65
Tabella 6. Elenco dei <i>primers</i> e relative caratteristiche.	66
Tabella 7. <i>Primers</i> utilizzati per il sequenziamento.....	67
Tabella 8. <i>Primers</i> per l'amplificazione del tratto contenente TATA-box e <i>leader</i>	68
Tabella 9. Coppia di oligonucleotidi utilizzati per aggiungere un sito <i>EcoRI</i> nel MCS.....	69
Tabella 10. Lista delle abbreviazioni utilizzate per costruire i nomi dei promotori sintetici composti da più sezioni di promotori naturali.	71
Tabella 11. Enzimi utilizzati per la costruzione dei derivati del promotore della globulina	71
Tabella 12. Enzimi utilizzati per la costruzione dei derivati del promotore della prolamina 10kDa.....	72
Tabella 13. Enzimi utilizzati per la i costruzione dei derivati del promotore della glutelina B4	72

Tabella 14. Siti di restrizione utilizzati per inserire i vari promotori nella cassetta di espressione	73
Tabella 15. <i>Primers</i> utilizzati per determinare l'orientamento di inserzione	73
Tabella 16. Composizione di ciascun substrato impiegato nel processo di trasformazione genetica mediato da <i>Agrobacterium tumefaciens</i> .	78
Tabella 17. Dettaglio della composizione degli ingredienti dei substrati elencati in Tabella 16.....	79
Tabella 18. <i>Primers</i> utilizzati per saggiare l'amplificabilità del DNA genomico estratto.	80
Tabella 19. <i>Primers</i> utilizzati per l'amplificazione dei geni <i>reporter</i>	80
Tabella 20. Promotori utilizzati per costruire gli ibridi e relative caratteristiche.	93
Tabella 21. Numero di individui rigenerati per promotore.	137
Tabella 22. Produttività media di seme registrata nelle diverse popolazioni di trasformati primari e nel controllo negativo (CN, var. CR W3 non trasformata).....	138
Tabella 23. Riassunto dei risultati generali ottenuti nel lavoro di trasformazione. >40 semi: numero di piante su cui è stato possibile condurre un'analisi dei livelli di espressione del gene <i>reporter</i>	139
Tabella 24. Contenuto medio di GAA e deviazione standard rilevati su 2 repliche di campionamento della progenie. ID è il codice identificativo dell'individuo	140
Tabella 25. Classifiche del livello di espressione del transgene in trasformati primari appartenenti ai due gruppi. Per ciascun promotore è stata calcolata la media dell'intero campione e la media del primo quartile, ovvero dell'insieme dei trasformati primari che occupano il primo quarto di classifica.....	142

Tabella 26. Livelli di espressione del transgene rilevati nelle piante trasformate nel mese di novembre 2011. La tabella continua nella pagina successiva.	144
Tabella 27. Livelli di espressione rilevati nelle piante trasformate nel mese di novembre 2011. La tabella continua nella pagina successiva. ..	145
Tabella 28. Livelli di espressione rilevati nelle piante trasformate nel mese di novembre 2011. La tabella continua dalle pagine precedenti. ..	146
Tabella 29. Classifica riassuntiva in cui i 9 promotori sono ordinati in base al valore medio di espressione registrato nei trasformati primari analizzati.	147
Tabella 30. Livelli di espressione del transgene nei trasformati primari ottenuti con la seconda sessione di trasformazione.	149
Tabella 31. Classifica riassuntiva dei quattro promotori trasformati a febbraio.....	150
Tabella 32. Concentrazione di proteina totale negli estratti da foglia. Valori misurati a 280nm con $e=1.7$	151
Tabella 33. Materiale raccolto da piante trasformate con Glb e con C-Glb-B4.	153
Tabella 34 Concentrazione di proteina totale negli estratti da seme immaturo e seme abortito. Valori misurati a 280nm con $e=1.7$	154
Tabella 35. Peso in grammi dei campioni di farinaccio, riso sbramato e sbiancato.	155
Tabella 36. Espressione di GCasi in foglia nella pianta B1-B4. Nella prima colonna sono elencate le concentrazioni di GCasi nell'estratto, nella seconda la concentrazione di proteina totale nell'estratto. Il valore è stato ottenuto con saggio Bradford, utilizzando BSA per la retta di taratura. Nella terza colonna l'espressione della proteina <i>reporter</i> è espressa in parti per milione sulla proteina totale. La quarta colonna riporta i dati rilevati rispetto all'espressione di GCasi in endosperma. La quinta mette in	

rapporto il valore in ppm in foglia con il livello di espressione in endosperma per definire un parametro normalizzato, chiamato leakage.....	157
Tabella 37. Espressione di GCasi in foglia per il promotore C-Glb-B4.....	157
Tabella 38. Espressione di GCasi in foglia nei migliori espressor di ciascun costruito. La quarta colonna riporta i dati rilevati rispetto all'espressione di GCasi in endosperma. La quinta mette in rapporto il valore in ppm in foglia con il livello di espressione in endosperma per definire un parametro normalizzato, chiamato leakage.....	158
Tabella 39. Espressione di GCasi in singolo seme immaturo di piante trasformate con il promotore della globulina 26kDa.....	158
Tabella 40. Espressione di GCasi in singolo seme immaturo di piante trasformate con il promotore C-Glb-B4.	159
Tabella 41. Espressione di GCasi in singolo seme abortito di piante trasformate con il promotore della globulina 26kDa e con il promotore C-Glb-B4.....	160
Tabella 42. Espressione aspecifica di GCasi nelle glume delle piante appartenenti alle popolazioni ottenute dai promotori saggiati nel secondo ciclo.....	160
Tabella 43. Distribuzione di GCasi nelle varie frazioni provenienti dal processo di sbiancatura. sgl=seme sbramato; sb= seme sbiancato; far=farinaccio (residuo di sbiancatura).	161
Tabella 44. Dati grezzi ottenuti dalle tre trasformazioni replicate.	178

12 RINGRAZIAMENTI

Un mio ringraziamento particolare va alle mie care colleghe Erika e Serena per l'immane mole di lavoro manuale che hanno svolto e che continuano pazientemente a svolgere per il bene del gruppo di ricerca.

Ringrazio poi il mio tutore, Prof. Stefano Marchetti, perché il suo duplice impegno come imprenditore e professore universitario non è mai venuto meno, nonostante i tagli generalizzati alla ricerca e l'arroganza della grande industria farmaceutica.

Ringrazio poi l'intero gruppo di ricerca per il supporto garantitomi nel corso di questi tre anni di dottorato.

Infine ringrazio la mia fidanzata Gioia per avermi innumerevoli volte pazientemente atteso con la tavola apparecchiata, quando mi attardavo fino a sera inoltrata in laboratorio.