



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΥΓΕΙΑΣ
ΤΜΗΜΑ ΒΙΟΧΗΜΕΙΑΣ-ΒΙΟΤΕΧΝΟΛΟΓΙΑΣ

Διπλωματική εργασία
**ΠΡΟΣΔΙΟΡΙΣΜΟΣ ΤΩΝ ΡΥΘΜΙΣΤΙΚΩΝ ΣΤΟΙΧΕΙΩΝ ΤΟΥ ΓΟΝΙΔΙΟΥ
LSK1 ΣΤΟ ΦΥΤΟ *Lotus japonicus***

ΤΣΙΟΥΡΗ ΟΛΓΑ

ΛΑΡΙΣΑ 2019

IDENTIFICATION OF REGULATORY ELEMENTS OF *LSK1* GENE IN *Lotus japonicus*

Τριμελής επιτροπή:

Παπαδοπούλου Καλλιόπη: Αναπληρώτρια Καθηγήτρια Βιοτεχνολογίας Φυτών, Τμήμα Βιοχημείας και Βιοτεχνολογίας, Πανεπιστήμιο Θεσσαλίας

Γιακουντής Αντώνιος: Επίκουρος Καθηγητής Μοριακής Βιολογίας και Γονιδιωματικής, Τμήμα Βιοχημείας και Βιοτεχνολογίας, Πανεπιστήμιο Θεσσαλίας

Γκαραγκούνης Κωνσταντίνος: Ακαδημαϊκός Υπότροφος, Τμήμα Βιοχημείας και Βιοτεχνολογίας, Πανεπιστήμιο Θεσσαλίας.

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα πτυχιακή εργασία πραγματοποιήθηκε στο εργαστήριο Βιοτεχνολογίας Φυτών και Περιβάλλοντος του τμήματος Βιοχημείας και Βιοτεχνολογίας. Σε αυτό το πολύ οργανωμένο και συνάμα φιλικό περιβάλλον γνώρισα ανθρώπους που με στήριξαν και με συμβούλευαν καθ' όλη τη διάρκεια της πτυχιακής μου. Συγκεκριμένα θα ήθελα να ευχαριστήσω την επιβλέπουσα καθηγήτρια κ.Παπαδοπούλου Καλλιόπη, Αναπληρώτρια καθηγήτρια Βιοτεχνολογίας Φυτών. Σας ευχαριστώ που με επιλέξατε στο εργαστήριο σας και για τη στήριξη και τις συμβουλές που μου δώσατε. Επίσης θα ήθελα να ευχαριστήσω τους Γκαραγκούνη Κωσταντίνο και Νταντάμη Ελένη, μεταδιδακτορικούς ερευνητές στο εργαστήριο βιοτεχνολογίας φυτών. Σας ευχαριστώ για την υπομονή και τις συμβουλές που μου δώσατε. Τέλος δε θα μπορούσα να παραλείψω να ευχαριστήσω την οικογένεια μου που με στηρίζει πάντα και με έχει κάνει τον άνθρωπο που είμαι σήμερα, καθώς και τους φίλους μου για όλα τα καλά και τα άσχημα που έχουμε περάσει όλα αυτά τα χρόνια.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

| | |
|---|-----------|
| ΠΕΡΙΛΗΨΗ..... | 6 |
| ABSTRACT..... | 7 |
| 1. ΕΙΣΑΓΩΓΗ..... | 8 |
| 1.1.1 ΜΕΤΑΓΡΑΦΗ ΜΙΑ ΠΡΟΕΠΙΣΚΟΠΙΣΗ..... | 8 |
| 1.1.2 ΡΥΘΜΗΣΤΙΚΑ ΣΤΟΙΧΕΙΑ ΤΗΣ ΜΕΤΑΓΡΑΦΗΣ ΣΤΟΥΣ ΕΥΚΑΡΥΩΤΕΣ..... | 8 |
| 1.1.3 ΓΕΝΙΚΕΣ ΑΡΧΕΣ <i>in silico</i> ΕΥΡΕΣΗΣ ΡΥΘΜΙΣΤΙΚΩΝ ΣΤΟΙΧΕΙΩΝ..... | 11 |
| 1.2.1 ΔΟΜΗ ΤΩΝ ΖΩΙΚΩΝ <i>GSK3/SHAGGY LIKE</i> ΚΙΝΑΣΩΝ..... | 12 |
| 1.2.2 ΡΥΘΜΙΣΗ ΤΗΣ ΕΚΦΡΑΣΗΣ ΚΑΙ ΤΗΣ ΕΝΕΡΓΟΤΗΤΑΣ ΤΗΣ <i>GSK3</i> | 13 |
| 1.2.3 ΜΗΧΑΝΙΣΜΟΣ ΚΑΤΑΛΥΣΗΣ ΤΗΣ <i>GSK3</i> | 13 |
| 1.2.4 ΡΟΛΟΣ ΤΗΣ <i>GSK3</i> ΣΕ ΣΗΜΑΤΟΔΩΤΙΚΑ ΜΟΝΟΠΑΤΙΑ..... | 14 |
| 1.3.1 ΦΥΤΙΚΕΣ <i>GSK3/SHAGGY LIKE</i> ΚΙΝΑΣΕΣ..... | 17 |
| 1.3.2 ΡΟΛΟΣ ΤΩΝ ΦΥΤΙΚΩΝ SK ΚΙΝΑΣΩΝ ΣΕ ΣΗΜΑΤΟΔΩΤΙΚΑ ΜΟΝΟΠΑΤΙΑ.. | 17 |
| 1.4 <i>GSK3/SHAGGY LIKE</i> ΚΙΝΑΣΕΣ ΣΤΟ <i>Lotus japonicus</i> | 19 |
| 1.5 Ο ΟΡΓΑΝΙΣΜΟΣ- ΜΟΝΤΕΛΟ <i>Lotus japonicus</i> | 21 |
| 1.6 ΦΥΜΑΤΙΟΓΕΝΕΣΗ ΜΙΑ ΕΠΙΣΚΟΠΗΣΗ..... | 22 |
| 1.7 ΦΥΜΑΤΙΟΓΕΝΕΣΗ ΣΤΟ <i>Lotus japonicus</i> | 23 |
| 1.8 ΜΗΧΑΝΙΣΜΟΣ ΑΥΤΟΡΡΥΘΜΙΣΗΣ ΤΗΣ ΦΥΜΑΤΙΟΓΕΝΕΣΗΣ (ΑΟΝ)..... | 25 |
| 2. ΠΡΟΓΡΑΜΜΑΤΑ ΚΑΙ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΗΘΗΚΑΝ. | 26 |
| 2.1 ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ ΤΟΥ <i>Lotus japonicus</i> | 26 |
| 2.2 ΠΡΟΓΡΑΜΜΑΤΑ ΕΥΡΕΣΗΣ ΜΕΤΑΘΕΤΩΝ ΣΤΟΙΧΕΙΩΝ..... | 28 |
| 2.3 ΠΡΟΓΡΑΜΜΑΤΑ ΕΥΡΕΣΗΣ ΡΥΘΜΙΣΤΙΚΩΝ ΣΤΟΙΧΕΙΩΝ..... | 28 |
| 2.4 ΠΡΟΓΡΑΜΜΑΤΑ ΕΥΡΕΣΗΣ ΥΠΟΚΥΤΤΑΡΙΚΟΥ ΕΝΤΟΠΙΣΜΟΥ..... | 31 |
| ΣΤΟΧΟΣ ΤΗΣ ΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ..... | 33 |

| | |
|--|-----------|
| 3. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ..... | 33 |
| 3.1 ΑΠΟΜΟΝΩΣΗ ΓΕΝΕΤΙΚΟΥ ΥΛΙΚΟΥ ΚΑΙ ΕΝΙΣΧΥΣΗ ΤΟΥ ΥΠΟΚΙΝΗΤΗ ΤΗΣ LSK1.. | 33 |
| 3.2 ΠΡΟΕΤΟΙΜΑΣΙΑ ΤΟΥ ΦΟΡΕΑ 1391xc ΓΙΑ ΚΛΩΝΟΠΟΙΗΣΗ..... | 34 |
| 3.3 ΚΛΩΝΟΠΟΙΗΣΗ ΤΟΥ ΥΠΟΚΙΝΗΤΗ ΣΕ PGEM-T ΚΑΙ PCAMBIA 0391xc..... | 36 |
| 3.4 ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ ΑΝΑΛΥΣΗ..... | 37 |
| 4. ΑΠΟΤΕΛΕΣΜΑΤΑ..... | 40 |
| 4.1 ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΕΙΡΑΜΑΤΩΝ..... | 40 |
| 4.2 CpG ΝΗΣΙΔΕΣ, ΜΕΤΑΘΕΤΑ ΣΤΟΙΧΕΙΑ ΚΑΙ S/MARs..... | 43 |
| 4.3 TATA -box ΚΑΙ ΛΟΙΠΑ ΡΥΘΜΙΣΤΙΚΑ ΣΤΟΙΧΕΙΑ..... | 45 |
| 4.4 RNA TFs, miRNAs ΚΑΙ RIBOSWITCHES..... | 48 |
| 5. ΣΥΖΗΤΗΣΗ..... | 51 |
| 6. ΒΙΒΛΙΟΓΡΑΦΙΑ..... | 52 |

ΠΕΡΙΛΗΨΗ

Στα ψυχανθή φυτά το μονοπάτι αυτορρύθμισης της φυματιογένεσης (AON) συμβάλλει στη διατήρηση του ισοζυγίου μεταξύ των αναγκών του φυτού σε άζωτο και την σπατάλη άνθρακα καταστέλλοντας τη δημιουργία επιπλέον φυματίων. Επιπρόσθετα είναι γνωστό το γεγονός πως διάφοροι δευτερογενείς μεταβολίτες εμφανίζουν διαφορετική έκφραση μεταξύ των ριζών με φυμάτια και αυτών χωρίς. Συνεπώς η εύρεση παραγόντων που δρουν στο μονοπάτι του AON είναι υψίστης σημασίας. Στο *Lotus japonicus* μέσα από πειράματα υπερέκφρασης και υποέκφρασης έχει βρεθεί ότι μία GSK3 like κινάση παίζει κάποιο ρόλο στην ρύθμιση του αριθμού των φυματίων η LSK1 (*Lotus* SHAGGY-LIKE Kinase). Στη παρούσα διπλωματική εργασία πραγματοποιήθηκε ο προσδιορισμός των ρυθμιστικών στοιχείων αυτού το γονιδίου. Στον υποκινητή της LSK1 βρέθηκαν διάφορα πιθανά TATA-boxes καθώς και θέσεις πρόσδεσης μεταγραφικών παραγόντων των κλάσεων ERF, LBD και bZIP. Επιπλέον βρέθηκε ένα CpG island ανοδικά του TSS. Από την κλωνοποίηση 2kb από τον υποκινητή του LSK1 από 2 διαφορετικούς οικοτύπους (Gifu και MG20) και τη την σύγκριση μεταξύ τους, προέκυψε ότι οι 2 υποκινητές εμφανίζουν ελάχιστες διαφορές, με ποιο σημαντική την ύπαρξη διαφορετικού σημείου έναρξης της μεταγραφής.

ABSTRACT

In legumes the autoregulation of nodulation pathway helps in maintaining the equilibrium between the need for nitrogen and the waste of sugars by repressing the creation of new nodules. Moreover, it is known that the genes of various secondary metabolites have different expression levels in roots with and without nodules. Therefore, it is very important to find the components of this pathway. In *Lotus japonicus* a GSK3-like kinase named as LSK1 (*Lotus* SHAGGY-LIKE Kinase) seems to play a role in controlling the number of nodules according to overexpression and silencing experiments. In this Bsc thesis the regulatory elements that play a role in the expression of this gene were found. In the LSK1 promoter there were possible tata-boxes, ERF, LBD & bZIP families of transcription factor binding sites and a CpG island was found upstream of the TSS. From the cloning of the 2kb promoter fragment of LSK1 and the comparison between different ecotypes (Gifu and MG20) it was discovered that the main difference between the two was the location of the transcription starting site.

1.ΕΙΣΑΓΩΓΗ

1.1.1 ΜΕΤΑΓΡΑΦΗ- ΜΙΑ ΕΠΙΣΚΟΠΙΣΗ

Η Μεταγραφή αποτελεί το πρώτο βήμα της γονιδιακής έκφρασης και αποτελεί τη διαδικασία εκείνη που από το δίκλωνο DNA χρησιμοποιώντας ως εκμαγείο μία από τις δημιουργείται μονόκλωνο RNA με τη βοήθεια πρωτεϊνών. Οι πρωτεΐνες αυτές χωρίζονται σε 2 κατηγορίες τις RNA πολυμεράσες που καταλύουν την δημιουργία του νουκλεοτιδικού πολυμερούς και τους μεταγραφικούς παράγοντες οι οποίοι προσδένονται σε συγκεκριμένες για τον καθένα αλληλουχίες στο DNA, στρατολογώντας τις πολυμεράσες στο σημείο έναρξης της μεταγραφής (TSS συνήθως έχει αδενίνη). Σε αυτό το σημείο, σε περιοχές ανοδικά αυτού από αυτό το σημείο και σε κάποιες δεκάδες βάσεις καθοδικά οροθετείται μια περιοχή που περιέχει αλληλουχίες-μοτίβα πρόσδεσης μεταγραφικών παραγόντων (μήκος 100nt-1000nt) (Tropp B.E 2012).

1.1.2 ΡΥΘΜΙΣΤΙΚΑ ΣΤΟΙΧΕΙΑ ΤΗΣ ΜΕΤΑΓΡΑΦΗΣ ΣΤΟΥΣ ΕΥΚΑΡΥΩΤΕΣ

Τα ρυθμιστικά στοιχεία της μεταγραφής που θα αναλυθούν στη συνέχεια αποτελούν: τα μοτίβα πρόσδεσης μεταγραφικών παραγόντων, τα CpG νησίδια (CpG islands), τα μεταθετά στοιχεία, τα μικρά RNAs (miRNAs), οι πρωτεΐνες που προσδένονται στο RNA και οι ριβοδιακόπτες (riboswitches).

ΜΟΤΙΒΑ ΠΡΟΣΔΕΣΗΣ ΜΕΤΑΓΡΑΦΙΚΩΝ ΠΑΡΑΓΟΝΤΩΝ

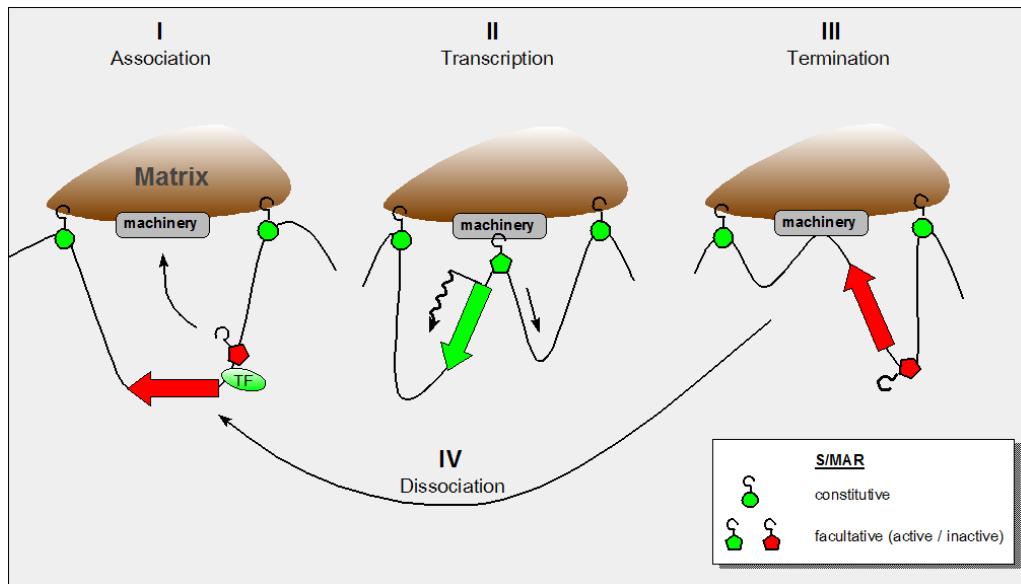
Τα μοτίβα πρόσδεσης μεταγραφικών παραγόντων διακρίνονται σε 3 κατηγορίες ανάλογα την απόσταση τους από το TSS σε μοτίβα που κεντρικού υποκινητή, σε μοτίβα του ρυθμιστικού υποκινητή και στα απομακρυσμένα μοτίβα (Tropp B.E 2012). Ο κεντρικός υποκινητής βρίσκεται -40 έως +40 (Kutach and Kadonaga, 2002) (+καθοδικά, – ανοδικά) βάσεις σχετικά με το TSS και αποτελείται από τα εξής ρυθμιστικά στοιχεία:

1. TATA box, που είναι A/T rich αλληλουχία αποτελούμενη από 8 νουκλεοτίδια (Porto et al.,2014) [στα φυτά αποτελεί την αλληλουχία TATAWAW (W = A/T) (Loganantharaj, 2006)] που βρίσκεται 25-35 νουκλεοτίδια ανοδικά του TSS (Burley and Roeder, 1996). Στο μοτίβο αυτό προσδένεται η πρωτεΐνη TBP που αποτελεί υπομονάδα του μεταγραφικού παράγοντα TFIID. Ο TFIID στρατολογεί τους TFIIA και TFIIIB οι οποίοι στρατολογούν το σύμπλοκο TFIIF-RNA πολυμεράση (Porto et al., 2014). Το TBP προσδένεται μόνο στην κωδική αλυσίδα του DNA (Iraní et al., 2010). Το TATA- box δεν υπάρχει σε όλους τους υποκινητές γονιδίων (Suzuki et al., 2001) και η θέση που περιβάλλει αυτή την αλληλουχία είναι GC-rich στον άνθρωπο και στον καπνό (Horikoshi et al., 1992, Kiran et al., 2006). Είναι υπεύθυνο για την εξειδικευμένη μεταγραφή γονιδίων (Kiran et al., 2006) στα φυτά.

2. BRE (basal. regulatory element), το οποίο βρίσκεται ανοδικά του TSS (5'-G/C-G/C-G/A-C-G-C-C-3' ακολουθούμενο από το T του TATA box) (Lagrange et al., 1998) ωστόσο BPEs δεν έχουνε βρεθεί σε υποκινητές φυτικών οργανισμών και ζυμομυκήτων (Smale, & Kadonaga, 2003). Στον άνθρωπο σε έλλειψη του TATA box το στοιχείο αυτό βρίσκεται 81 με 75 βάσεις ανοδικά του TSS.

3. Το Inr το οποίο περιλαμβάνει ουσιαστικά το TSS (-2 έως +5) και έχει διαφορετική αλληλουχία στον άνθρωπο (Py-Py-A₊₁-N-(T/A)-Py-Py, Smale et al., 1998), στη *Drosophila* (Py-Py-A₊₁-N-(T/A)-Py-Py Arkhipova, 1995, Kutach and Kadonaga, 2000; Thomas and Chiang, 2006) και στα φυτά (Py-T-C-A-N-T-Py-Py, Nakamura et al., 2002 όπου Py πυριμιδίνη). Απουσία του TATA box μπορεί να επάγει την μεταγραφή καθώς προσδέεται σε αυτό ο TFIID (Porto et al., 2014).

4. DPE (downstream promoter element) το οποίο τοποθετείται +28 με +32 (Porto et al., 2014). Η ενδιαμέση περιοχή του με το Inr καθορίζει την πρόσδεση του TFIID, καθώς με αλλαγή 1 nt η μεταγραφική δραστηριότητα μειώνεται (Kutach & Kadonaga, 2000). Στα φυτά το DPE απαντάται σε υποκινητές που αποκρίνονται σε διάφορες συνθήκες όπου εμφανίζεται σε διάφορες θέσεις σε πολλά αντίγραφα (Sawant et al., 2001). Αυτό το στοιχείο μπορεί να δράσει όταν δεν υπάρχει TATA box ωστόσο γίνεται τα TATA box, DPE, Inr να συνυπάρξουν (Juven-Gershon & Kadonaga, 2010). Τα μοτίβα του ρυθμιστικού υποκινητή αποτελούν : τα μοτίβα CCAAT στο -100 (στα φυτά GGCCAATCT -80), CG box(-90 έως 200, Tropp B.E, 2012) και τα φυτικά CAT Box (GCCAAC -49), G Box (TGACGGTGT -66) (Porto et al., 2014). Στο CG προσδέεται ο μεταγραφικός παράγοντας Sp1, ο οποίος, ωστόσο, δεν εμφανίζεται στα φυτά (Yamamoto et al., 2007). Στη συνέχεια ακολουθούν θέσεις πρόσδεσης μεταγραφικών παραγόντων που αποκρίνονται σε συγκεκριμένες συνθήκες και ποικίλλουν από οργανισμό σε οργανισμό επιδρώντας είτε θετικά είτε αρνητικά στο ρυθμό της μεταγραφής (Tropp B.E 2012). Στους απομακρυσμένους υποκινητές επάγονται τα μοτίβα ενισχυτών και τα Scaffold/matrix attachment regions (S/MARs). Στα στοιχεία ενισχυτών ανήκουν διάφορα μοτίβα που επηρεάζουν την μεταγραφή από πολύ μεγάλη απόσταση, μπορούν να δρουν ενισχύοντας την ήδη επαγόμενη έκφραση ανοδικά και καθοδικά του γονιδίου, με οποιαδήποτε κατεύθυνση προς αυτό (Tropp B.E. 2012). Στοιχεία ενισχυτών αποτελούν τα προαναφερθέντα CCAAT και CG τα οποία εμφανίζονται ως ενισχυτές τόσο στα ζώα όσο στα φυτά (Tropp B.E., 2012, Kumar et al., 2000). Τα S/MARs αποτελούν θέσεις πρόσδεσης των πυρηνικών μικροσωληνίσκων που βοηθούν στην οργάνωση του και είναι κυρίως AT-rich περιοχές (Mirkovitch J et al., 1984). Τα S/MARs διακρίνονται σε ιδιόσυστατα και επαγόμενα όπου επηρεάζουν την μεταγραφή όπως φαίνεται στην εικόνα :



Εικόνα 1. Ρόλος των S/MARs στη μεταγραφή. Με πράσινο τα ιδιοσυστάτα S/MAR με κόκκινο το επαγόμενο.

Επίσης τα επαγόμενα S/MARs μπορούν να βρίσκονται και σε περιοχές ενισχυτών επιρεάζοντας τη δράση τους. Παρόλου που τα στοιχεία του κεντρικού υποκινητή εμφανίζουν συγκεκριμένο προσανατολισμό (5'-3' προς την έναρξη της μεταγραφής) για τα υπόλοιπα μοτίβα δεν ισχύει το ίδιο στο *Arabidopsis* (Lis and Walther, 2016). Επιπλέον υπάρχουν πρωτεΐνες που δε δρουν άμεσα στο DNA αλλά δρουν μαζί με άλλους μεταγραφικούς παράγοντες επάγοντας ή καταστέλλοντας.

CpG ΝΗΣΙΔΕΣ ΚΑΙ ΜΕΤΑΘΕΤΑ ΣΤΟΙΧΕΙΑ

Οι CpG νησίδες αποτελούν περιοχές του γονιδιώματος οι οποίες φέρουν πολλές επαναλήψεις του δινουκλεοτιδίου κυτοσίνης γουανίνης. Αυτές οι περιοχές πρέπει να έχουν τουλάχιστον 200 bp μήκος το περιεχόμενο σε CG να είναι πάνω από 50% το CG/GC ratio να είναι πάνω από 0,6 και το παρατηρούμενο/εκτιμητέο ratio να είναι πάνω από 0,6. Τα νησίδια αποτελούν περιοχές με "ανοιχτή" διαμόρφωση χρωματίνης (Deaton and Bird, 2011 και Medvedeva et al., 2010) τα οποία κυρίως βρίσκονται κοντά στο TSS αποτελώντας σημεία πρόσδεσης μεταγραφικών παραγόντων διαφορετικών από τα παραπάνω (Deaton and Bird, 2011, Medvedeva et al., 2010). Οι νησίδες αυτές μπορούν να βρίσκονται σε περιοχές εκτός και εντός γονιδίων επηρεάζοντας την μεταγραφή (Deaton and Bird, 2011 και Medvedeva et al., 2010) και μπορούν να μεθυλιώνονται και να απομεθυλιώνονται επιδρώντας στην επαγωγή ή όχι της έκφρασης γονιδίων. Τα μεταθετά τα οποία αποτελούν τμήματα DNA που σε ορισμένες συνθήκες μετακινούνται στο γονιδίωμα, στους ευκαρυώτες αποτελούνται από 2 κατηγορίες τα DNA μεταθετά και τα ρετρομεταθετά. Τα DNA μεταθετά αποτελούνται από επάκριες αντίστροφες επαναλήψεις και στο εσωτερικό τους περιέχουν μία τρανσποζάση ή ένζυμο που καταλύει τη μεταφορά τους με το μηχανισμό αποκοπής επικόλλησης (Tropp B.E.,

2012). Αυτά τα στοιχεία διακρίνονται σε αυτόνομα και μη αυτόνομα από το αν διαθέτουν το γονίδιο της τρανσποζάσης ή όχι αντίστοιχα. Τα ρετρομεταθετά διαφέρουν από τα άλλα καθώς στο κύκλο μεταφοράς τους δημιουργείται ένα μονόκλωνο RNA το οποίο μετατρέπεται σε cDNA και επανεισάγεται με μηχανισμό αντιγραφής επικόλλησης. Τα LTR ρετροτρανσποζόνια διαθέτουν στα άκρα τους μακριές τερματικές επαναλήψεις (LTRs) μήκους 300-500 bp. Στο εσωτερικό τους φέρουν το γονίδιο gag που κωδικοποιεί για μία πρωτεΐνη που αλληλεπιδρά με το DNA σχηματίζοντας ένα ιώδες σωματίο. Επίσης περιέχουν το γονίδιο pol που κωδικοποιεί για ένα πολυπεπτιδίο από το οποίο θα προκύψουν μια μεταγραφάση, μια ιντεργκράση και μία πρωτεάση. Ακόμη υπάρχουν και ρετρομεταθετά χωρίς LTR (Tropp B.E., 2012). Η προσθήκη μεταθετών στοιχείων μπορεί να ωθήσει στη δημιουργία νέων ρυθμιστικών μοτίβων καθώς και στην αλλαγή της απόστασης μεταξύ αυτών (Testori et al., 2012).

miRNAs, riboswitches και πρωτεΐνες που προσδένονται στο RNA

Τα miRNAs αποτελούν μία οικογένεια μη κωδικών RNA τα οποία παράγονται ενδογενώς από ένα οργανισμό και επιδρούν στην καταστολή της έκφρασης γονιδίων τόσο σε επίπεδο μεταγραφής όσο και σε μετα-μεταγραφικό επίπεδο σε ζώα και φυτά (Catalanotto, Cogoni and Zardo, 2016, Samad et al., 2017). Τα riboswitches αποτελούν μια περιοχή του αγγελιοφόρου RNA της οποίας η δευτεροταγής δομή μεταβάλλεται με την πρόσδεση ενός μικρού μορίου, επιδρώντας στη γονιδιακή έκφραση (Nudler E., Mironov A.S., 2004, Vitreschak A.G., Rodionov D.A., Mironov A.A., Gelfand M.S., 2004). Εκτός από τα βακτήρια μόλις μια οικογένεια riboswitch έχει ανακαλυφθεί στα αρχαία σε ορισμένους μύκητες και στα φυτά (Wachter et al., 2007 Swati, D., 2017, Mukherjee et al., 2018). Οι πρωτεΐνες πρόσδεσης στο RNA είναι πρωτεΐνες που προσδένονται (<https://meshb.nlm.nih.gov/record/ui?name=RNA-Binding%20Proteins>) σε μονόκλινα ή δίκλινα μόρια RNA με κυριότερους ρόλους την επαγωγή ή καταστολή του ματίσματος σε μία συγκεκριμένη περιοχή, την πολυαδενυλίωση και την καταστολή της μετάφρασης (Glisovic et al., 2008, Fu and Ares, 2014).

1.1.3 ΓΕΝΙΚΕΣ ΑΡΧΕΣ *in silico* ΕΥΡΕΣΗΣ ΡΥΘΜΙΣΤΙΚΩΝ ΣΤΟΙΧΕΙΩΝ

Με τη βοήθεια πειραματικών μεθόδων (De Sa' et al., 2018, Bartlett et al., 2017) έγινε εφικτή η κατασκευή βάσεων δεδομένων που περιέχουν ρυθμιστικά μοτίβα. Για την εύρεση *in silico*, μοτίβων στους υποκινητές χρησιμοποιούνται προγράμματα τα οποία ομαδοποιούνται σε 3 είδη ανάλογα με την αρχή της μεθόδου τους:

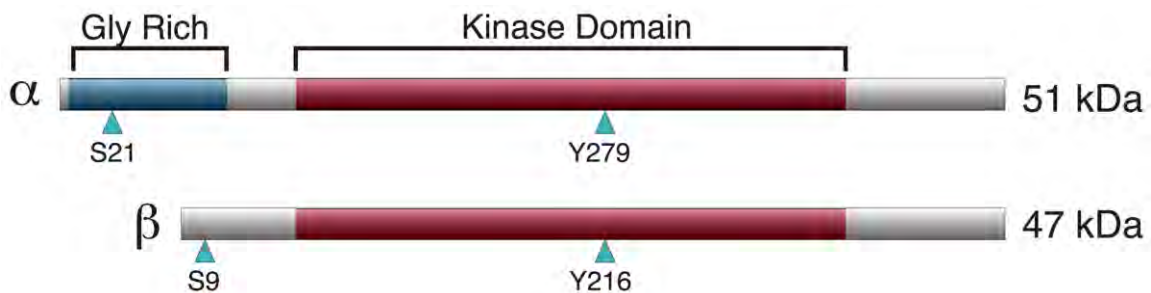
1. Word based algorithms. Σε αυτό τον τύπο για κάθε νουκλεοτιδική αλληλουχία ορισμένου μήκους γίνεται υπολογισμός των επαναλήψεων της στην αλληλουχία
2. Probabilistic models. Συγκεκριμένα αυτά τα προγράμματα χρησιμοποιούν position weight matrixes διαθέτοντας ένα matrix που δηλώνει την πιθανότητα

κάθε νουκλεοτιδίου να βρίσκεται σε αυτή τη θέση στο μοτίβο. Τα καλύτερα matrices επιλέγονται σύμφωνα με το information content για κάθε matrix.

3. Phylogenetic Footprinting. Η εύρεση κοινών ρυθμιστικών στοιχείων με φυλογενετική ανάλυση σε ομόλογους υποκινητές (Das and Dai, 2007).

1.2.1 ΔΟΜΗ ΤΩΝ ΖΩΙΚΩΝ GSK3/SHAGGY LIKE ΚΙΝΑΣΩΝ

Μία από τις πιο καλά μελετημένες κινάσες στα ζώα τα τελευταία 20 χρόνια είναι η GSK3/SHAGGY LIKE (glucogen synthase kinase 3). Πρόκειται για ένα αρκετά ενδιαφέρον ένζυμο καθώς είναι ιδιοσυστάτα ενεργό της οποίας η δράση καταστέλλεται αντί να ενεργοποιείται από ανοδικά σηματοδοτικά μονοπάτια (reviewed in Beurel et al., 2015). Ομόλογες πρωτεΐνες με αυτό το ένζυμο έχουν βρεθεί σε όλους τους ευκαρυώτες, το οποίο δρα φωσφορυλιώνοντας τα υποστρώματα του στα κατάλοιπα Ser/Thr με τον πιο γνωστό ρόλο της τη φωσφορυλίωση και απενεργοποίηση της συνθάσης του γλυκογόνου (Cohen et al., 1982). Η GSK3 των θηλαστικών συναντάται σε 2 ισομορφές την GSK-3α και την GSK-3β (Woodgett, 1990) όπου τα γονίδια που τις κωδικοποιούν βρίσκονται στον άνθρωπο στο χρωμόσωμα 19 και 3 αντίστοιχα. Η GSK-3β και άλφα εκφράζονται σε όλους τους ευκαρυώτες με 90% ομοιότητα με το ενεργό κέντρο (reviewed in Ali et al., 2001) πλην των πτηνών στα οποία δεν υπάρχει η GSK3α (Alon et al., 2011). Οι δύο ισομορφές παρουσιάζουν διαφορές στο μοριακό τους βάρος (GSK-3α: 51 kDa, GSK-3β: 47 kDa) παρά το γεγονός ότι στην περιοχή με ενεργότητα κινάσης υπάρχει 98% ομολογία μεταξύ των 2 (Woodgett, 1990). Οι 2 πρωτεΐνες διαφέρουν στα 76 τελευταία κατάλοιπα της C-τελικής περιοχής (36% ομοιότητα), με μεγαλύτερη την διαφορά στη N-τελική περιοχή όπου περιλαμβάνει περισσότερα αμινοξέα στην ισομορφή α (περιοχή πλούσια σε γλυκίνη) όπως φαίνεται και στην εικόνα 1 (Doble et al., 2003).



Εικόνα 2. Σχηματική αναπαράσταση των α και β ισομορφών της GSK3. Με βέλη φαίνονται οι θέσεις φωσφορυλίωσης στις 2 πρωτεΐνες (Doble et al., 2003).

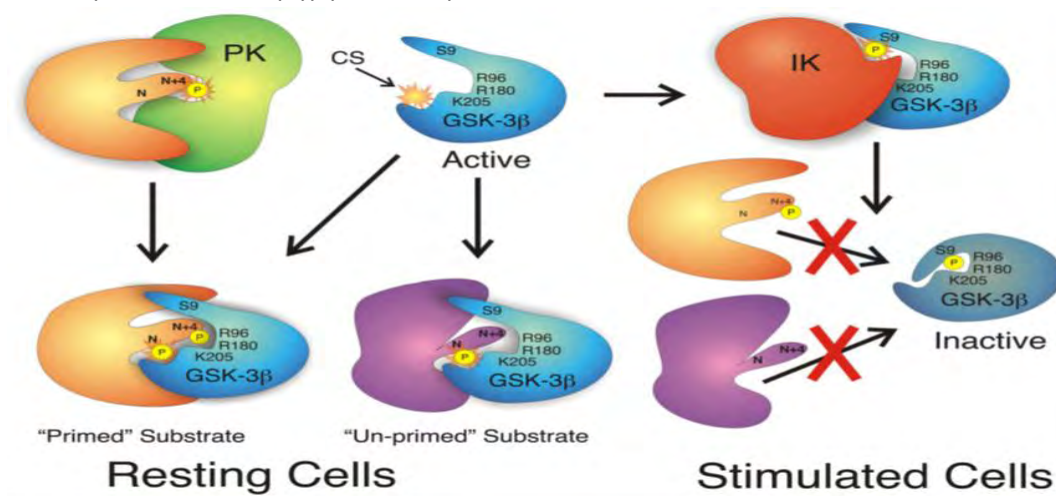
1.2.2 ΡΥΘΜΙΣΗ ΤΗΣ ΕΚΦΡΑΣΗΣ ΚΑΙ ΤΗΣ ΕΝΕΡΓΟΤΗΤΑΣ ΤΗΣ GSK3

Η GSK-3β έχει παρόμοια επίπεδα έκφρασης στα διάφορα όργανα των ζώων με εξαίρεση τα Th17 κύτταρα όπου εμφανίζει 10πλάσια έκφραση από ότι στους υπόλοιπους υποτύπους των T-λεμφοκυττάρων στον άνθρωπο. Επιπλέον τα επίπεδα έκφρασης και τον 2 ισόμορφών είναι 3-4 μεγαλύτερα κατά την ανάπτυξη του εγκεφάλου σε ποντίκια. Από τους Fai Lau et al., (1999) προσδιορίστηκε η θέση του σημείου έναρξης της μεταγραφής (TSS) το οποίο βρίσκεται 952 βάσεις ανοδικά του ATG. Ο υποκινητής της GSK-3β είναι κατηγορίας 0 καθώς δεν περιέχει ούτε TATA box ούτε Inr (CTCAMTCT ή YAATCYYY, M = A ή C, Y = C ή T). Η αλληλουχία η οποία επάγει την μεταγραφή βρέθηκε 427 με 384 νουκλεοτίδια ανοδικά του TSS ενώ αυτή που καταστέλλει 1421 με 1363 nt ανοδικά (Fai Lau et al., 1999). Στο TSS υπάρχει μια σειρά νουκλεοτιδίων που επιτρέπουν στο μεταγραφικό παράγοντα CP-2 να προσδεθεί, ο οποίος έχει βρεθεί να συνδέεται με την Fe65 η οποία αλληλεπιδρά, με μια πρώιμη αμυλοειδής πρωτεΐνη υπεύθυνη για τη νόσο του αλτσχάιμερ (Fai Lau et al., 1999). Επίσης ο υποκινητής διαθέτει CCAAT boxes (& ATTG) το οποίο αποτελεί ρυθμιστικό στοιχείο μεταβολικών ενζύμων καθώς και θέσεις πρόσδεσης άλλων μεταγραφικών παραγόντων όπως : SP-1, c-Myb, AP2 και CRE. (Fai Lau et al., 1999). Κατά τη διάρκεια της μεταγραφής του γονιδίου της GSK-3β μπορεί να προκύψουν εναλλακτικά μετάγραφα και κατά συνέπεια διαφορετικές πρωτεΐνες από εναλλακτικό μάτισμα. Παράδειγμα αποτελεί η GSK-3β2 πρωτεΐνη η οποία βρίσκεται σε αφθονία στον εγκέφαλο των θηλαστικών (Mukai et al., 2002) συμμετέχοντας στο σχηματισμό του άξονα των νευρικών κυττάρων (Wood-Kaczmar et al., 2009, Castaño et al., 2010). Επιπρόσθετα πρέπει να σημειωθεί ότι η GSK-3β2 εμφανίζει μικρότερη δραστηριότητα απέναντι σε αρκετά από τα υποστρώματα της σε σχέση με τη GSK-3β (Mukai et al., 2002, Soutar et al., 2010). Η ενεργότητα του συγκεκριμένου ενζύμου επηρεάζεται από τις μετά-μεταφραστικές τροποποιήσεις καθώς φωσφορυλίωση σε κατάλοιπα σερίνης (εικόνα 1) οδηγεί στη δημιουργία ψευδουποστρώματος που δρα σαν ανταγωνιστικός αναστολέας (Doble et al., 2003). Άλλες μετά-μεταφραστικές τροποποιήσεις περιλαμβάνουν την πέψη από μεταλλοπρωτεϊνάσες (Kandasamy and Schulz, 2009) σε ενεργά τμήματα και την παρεμπόδιση της δράσης της GSK3 με τη ADP-ριβοζυλίωση (Feijs et al., 2013, Rosenthal et al., 2013).

1.2.3 ΜΗΧΑΝΙΣΜΟΣ ΚΑΤΑΛΥΣΗΣ ΤΗΣ GSK3

Η GSK3 κινάση για να μπορέσει να φωσφορυλιώσει τα υποστρώματα της απαιτεί να είναι έχει προηγηθεί φωσφορυλίωση από άλλη κινάση (priming kinase) (Doble et al., 2003). Συγκεκριμένα η 1^η φωσφορυλίωση θα γίνει στο κατάλοιπο S/T που βρίσκεται 3 αμινοξέα μακριά (προς το c-τελικό άκρο) από το κατάλοιπο που θα φωσφορυλιώσει η

GSK3 (S/T-X-X-X-S/T(P)) (εικόνα 3).



Εικόνα 3. Μηχανισμός δράσης του ενζύμου και ο ρόλος των μετα-μεταφραστικών τροποποιήσεων (PK:priming kinase, IK:inactivating kinase) (Doble et al., 2003).

1.2.4 ΡΟΛΟΣ ΤΗΣ GSK3 ΣΕ ΣΗΜΑΤΟΔΟΤΙΚΑ ΜΟΝΟΠΑΤΙΑ

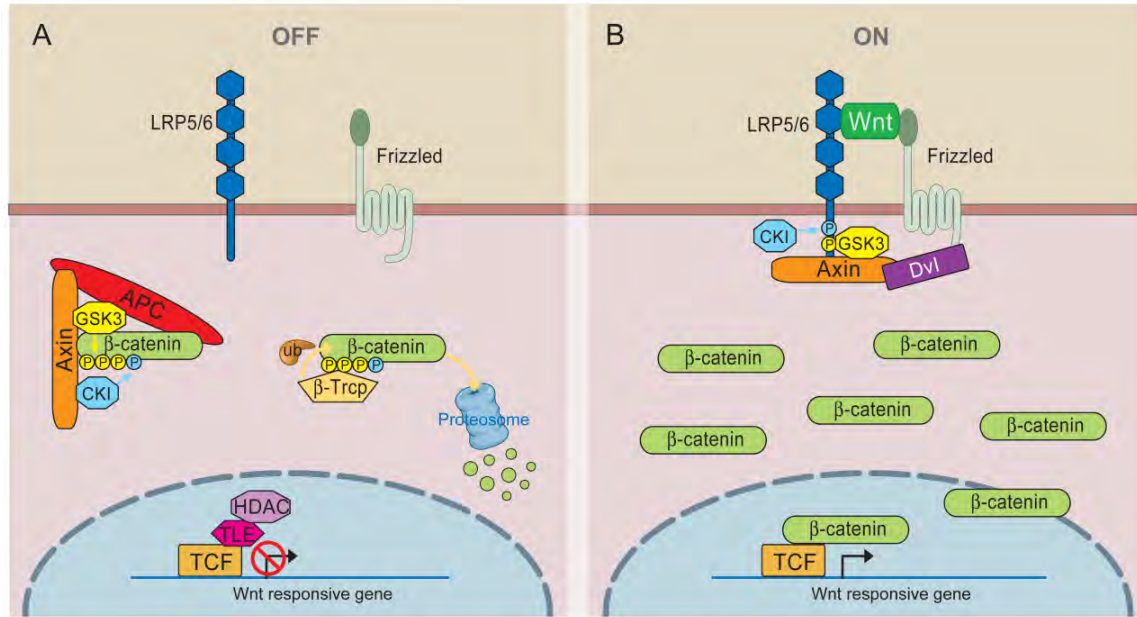
Όπως έχει ήδη αναφερθεί ο πρώτος ρόλος της GSK3 που ανακαλύφθηκε αφορούσε τη συμμετοχή της στο μονοπάτι της ινσουλίνης. Υπάρχουν ωστόσο και άλλα σημαντικά μονοπάτια μεταγωγής σήματος τόσο σε θηλαστικά όσο και σε έντομα τα οποία θα αναλυθούν παρακάτω.

Το μονοπάτι Wnt/β-catenin

Οι Wnt πρωτεΐνες αποτελούν μία οικογένεια εκκρινόμενων γλυκοπρωτεϊνών που μέσω του μονοπατιού σηματοδότηση τους παίζουν ρόλο στον κυτταρικό πολλαπλασιασμό, την ομοίωση σε επίπεδο ιστού καθώς και την κυτταρική διαφοροποίηση σε εμβρυικά στάδια (Logan and Nusse, 2004). Οι γλυκοπρωτεΐνες καθώς και το μονοπάτι αυτό έχει ανακαλυφθεί στον άνθρωπο, στα ποντίκια, στον *Xenopus*, στην *Drosophila*, (MacDonald B et al., 2009) ακόμα και στον *Caenorhabditis elegans* (Mizumoto and Sawa, 2007). Μετά από 20 χρόνια έχει εξεταστεί με βιοχημικές μελέτες πλήρως το μονοπάτι αυτό το οποίο ρυθμίζεται ως εξής:

1. Απουσία του Wnt η β-catenin (μεταγραφικός συνενεργοποιητής) καθίσταται ανενεργή καθώς φωσφορυλιώνεται από την δράση του συμπλόκου της πρωτεΐνης Axin (η οποία φέρνει κοντά στην β-catenin τις κινάσες CKI-κινάση της καζείνης- και GSK3) που την οδηγεί στα πρωτεοσώματα. Το σύμπλοκο μεταγραφικών παραγόντων TCF-TLE καθώς και η αποακετυλάση των ιστονών εμποδίζουν την μεταγραφή των γονιδίων απόκρισης στο Wnt.

2. Παρουσία του Wnt αυτό προσδένεται στον μεμβρανικό υποδοχέα του Frizzled (Fz) και στον συνυποδοχέα LRP6 ή LRP5 (υποδοχέας των λιποπρωτεϊνών χαμηλής πυκνότητας). Αυτό οδηγεί στην στρατολόγηση της πρωτεΐνης Dishevelled (Dvl), η οποία μαζί με το σύμπλοκο πρόσδεσης εμποδίζει την Axin (η οποία στρατολογείται από τις παραπάνω κινάσες οι οποίες φώσφορυλιώνουν των συνυποδοχέα) να δράσει απενεργοποιώντας την β -catenin. Έτσι ο μεταγραφικός αυτός παράγοντας είναι ελεύθερος να δράσει μαζί με τον TCF επάγοντας την έκφραση των γονιδίων απόκρισης στο Wnt, (MacDonald et al., 2009).

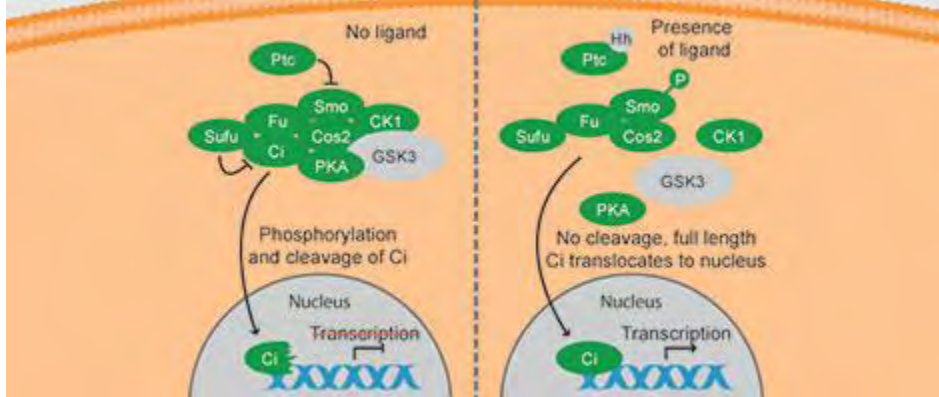


Εικόνα 4. Η μεταγωγή σήματος στο μονοπάτι Wnt/ β -catenin απουσία και παρουσία του Wnt (MacDonald B et al., 2009).

Το μονοπάτι hedgehog

Το μονοπάτι αυτό, που συμβάλλει στην κυτταρική διαφοροποίηση κατά την εμβρυογένεση ανακαλύφθηκε πρώτη φορά στη δροσόφιλα από τους Nusslein-Volhard et al., (1980), στο οποίο η σηματοδότηση ξεκινάει από την πρόσδεση της πρωτεΐνης Hedgehog στον υποδοχέα της Patched (Ptc) ένα διαμεμβρανικό υποδοχέα (Hooper and Scott 2005, Østerlund and Kogerman, 2006). Η Ptc δρά ως καταστολέας της πρωτεΐνης Smo ενός διαμεμβρανικού υποδοχέα με 7 α -έλικες. Καθοδικά της smo υπάρχει ένα σύμπλοκο πρωτεϊνών (HSC) το οποίο περιέχει εκτός των άλλων τις κινάσες PKA, CK1 και GSK3 καθώς και το μεταγραφικό παράγοντα Cubitus interruptus (Ci). Απουσία της Hedgehog η Ptc καταστέλλει τη δράση της smo, μη επιτρέποντας την απελευθέρωση της Ci από το σύμπλοκο η οποία οδηγείται στα πρωτεασώματα. Έτσι δημιουργείται μια μορφή Ci που δρα κατασταλτικά στην έκφραση γονιδίων. Παρουσία της Hedgehog η Smo είναι ελεύθερη να δράσει και οι κινάσες απομακρύνονται από το σύμπλοκο κι έτσι ο Ci δεν οδηγείται στα πρωτεασώματα και

επάγει την έκφραση των γονιδίων απόκρισης στην Hedgehog. Το μονοπάτι αυτό εμφανίζεται και στα θηλαστικά όπου υπάρχουν 3 πρωτεΐνες Hedgehog Sonic, Indian και Desert 2 Ptc και 3 Ci (Gli1, Gli2 που δρουν ως επαγωγείς και ο Gli3 που δρα ως καταστολέας) (Hooper and Scott 2005).

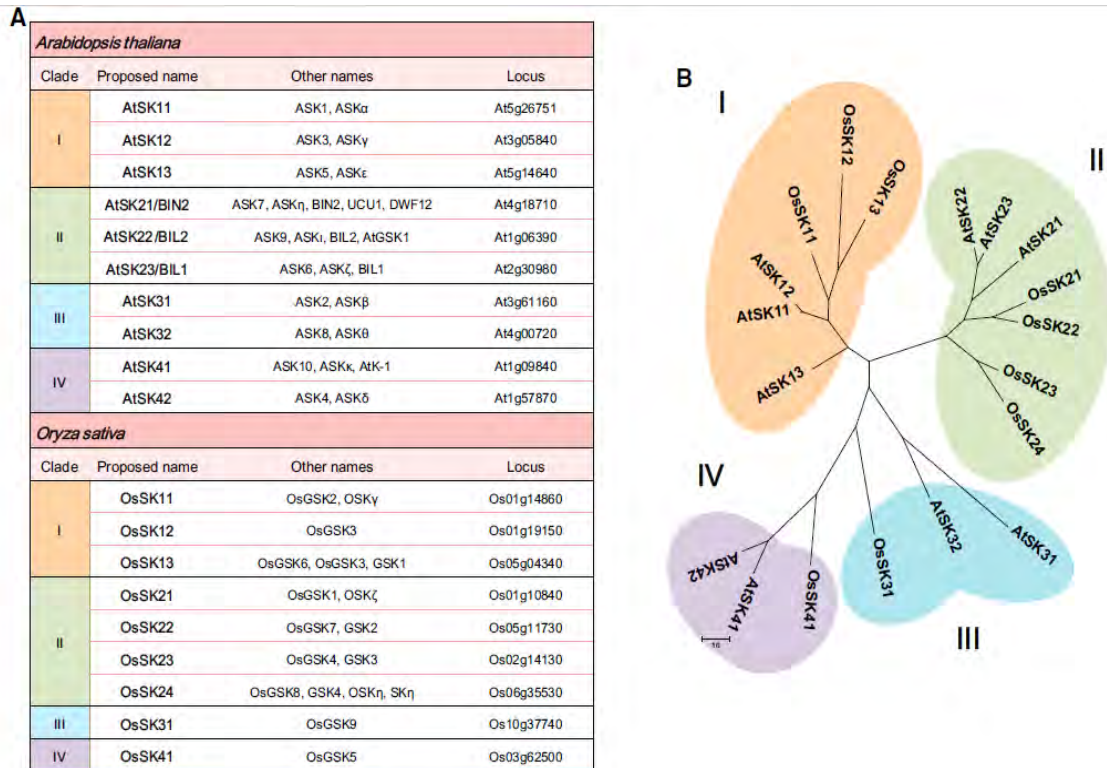


Εικόνα 5. Η μεταγωγή σήματος στο μονοπάτι Hedgehog απουσία και παρουσία της πρωτεΐνης (https://www.abcam.com/neuroscience/the-role-of-gsk3-in-cell-signal.ing).

1.3.1 ΦΥΤΙΚΕΣ GSK3/SHAGGY LIKE ΚΙΝΑΣΕΣ

Οι φυτικές GSK3/SHAGGY LIKE κινάσες αποτελούν μια ευρύτερη ομάδα σε σχέση με αυτή των ζώων αφού κάθε οργανισμός διαθέτει πολλές GSK3 κινάσες οι οποίες έχουν ανακαλυφθεί πρόσφατα. Αυτού του τύπου οι κινάσες δρουν στα φυτά συμμετέχοντας σε σημαντικές για τις επιβίωση τους διαδικασίες όπως :στην κυτταρική αύξηση, στην ανάπτυξη των ριζών, των άνθων, των στομάτων, στη διαφοροποίηση του ξυλώματος καθώς και σε αποκρίσεις στο φως και σε βιοτικούς και αβιοτικούς παράγοντες (Youn and Kim, 2015). Μία σημαντική λειτουργική διαφορά ανάμεσα σε ζωικές και φυτικές κινάσες αυτής της κατηγορίας είναι ότι στις δεύτερες δεν απαιτείται να έχει προηγηθεί φωσφορυλίωση του υποστρώματος τους (He et al., 2002, Zhao et al., 2002, Jonak and Hirt 2002). Επιπλέον η αμινοτελική περιοχή των φυτικών SKs διαφέρει αρκετά από αυτή των ζωικών γεγονός, που υποδηλώνει ότι στις φυτικές δεν υπάρχει ως μεταμεταφραστική τροποποίηση η φωσφορυλίωση στο κατάλοιπο σερίνης (Youn and Kim, 2015). Η αμινοξική περιοχή των ενζύμων φαίνεται να είναι σημαντική για την υποκυτταρική τους τοποθέτηση ενώ η καρβοξυτελική για την αλληλεπίδραση τους με τα υποστρώματα (Youn and Kim, 2015). Υποκυτταρικά η φυτικές SKs έχουν βρεθεί στον πυρήνα, (Vert and Chory, 2006, Ryu et al., 2010, Youn et al., 2013, Kim et al., 2009) στην πλασματική μεμβράνη, (Bayer et al., 2012), στα πλαστίδια (Kempra et al., 2007) και στο κυτταρόπλασμα, (Youn et al., 2013). Ακόμη ενδιαφέρον παρουσιάζει το γεγονός ότι οι φυτικές SKs μπορούν και να αυξάνουν και να μειώνουν με τη δράση τους την ενεργότητα των υποστρωμάτων τους (και να έχουν και θετικές και αρνητικές αποκρίσεις) σε σχέση με τις ζωικές όπου το αποτέλεσμα είναι κατασταλτικό. Στο φυτό μοντέλο *Arabidopsis Thaliana* έχουν βρεθεί 10 SKs (Yoo et al., 2006), με ρόλους σε σηματοδοτικά μονοπάτια

που αφορούν :το μονοπάτι μεταγωγής σήματος των βρασσινοστεροειδών, το οσμοτικό στρες και την ανάπτυξη των άνθων (Saidi, Hearn and Coates, 2012). Σε άλλα φυτά όπως στο ρύζι υπάρχουν 9 τέτοια ένζυμα ενώ στο *Medicago Truncatula* υπάρχουν 2 κινάσες οι MsK1 και MsK4 που παίζουν ρόλο στο βιοτικό στρες και στο μεταβολισμό των υδατανθράκων αντίστοιχα (Saidi, Hearn and Coates, 2012).



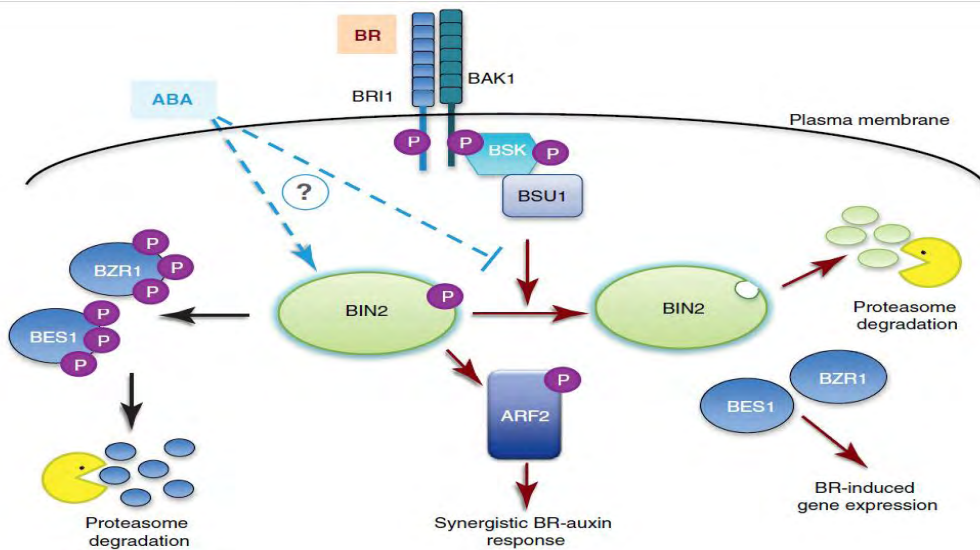
Εικόνα 6. α) Οι SKs στο *Arabidopsis* και στο ρύζι και ο φυλογενετικός τους διαχωρισμός β) (Youn and Kim, 2015).

1.3.2 ΡΟΛΟΣ ΤΩΝ ΦΥΤΙΚΩΝ SK ΚΙΝΑΣΩΝ ΣΕ ΣΗΜΑΤΟΔΩΤΙΚΑ ΜΟΝΟΠΑΤΙΑ

ΜΟΝΟΠΑΤΙ ΣΗΜΑΤΟΔΟΤΗΣΗΣ ΤΩΝ ΒΡΑΣΣΙΝΟΣΤΕΡΟΕΙΔΩΝ(BR)

Η πρώτη φυτική GSK3 κινάση που βρέθηκε με γενετικό έλεγχο ήταν η BR-INSENSITIVE 2 (BIN2) που συμμετέχει στο μονοπάτι απόκρισης στα βρασσινοστεροειδή (Li et al., 2001). Πρόκειται για ένα αρκετά συντηρημένο μονοπάτι διότι εμφανίζεται σε φυτά, μήκητες και ζώα (Li et al, 1997) και είναι υπεύθυνο για τις αναπαραγωγικές αποκρίσεις στα αγγειόσπερμα (Kim, and Wang, 2010, Clouse, 2011, Yang, C.J. et al., 2011, Gudesblat and Russinova, 2011). Στο φυτό *Arabidopsis thaliana* απουσία ορμόνης η BIN2 η οποία βρίσκεται σε φωσφορυλιωμένη μορφή φωσφορυλιώνει τους μεταγραφικούς

παράγοντες BRASSINAZOLE RESISTANT 1, BZR1 and bri1 EMS SUPPRESSOR1, BES1/BZR2 (επάγουν τη έκφραση γονιδίων απόκρισης στα βρασσινοστεροειδή) και τους οδηγεί στα πρωτεασώματα (Kim, and Wang, 2010, Clouse et.al., 2011, Yang, C.J. et al., 2011). Παρουσία των βρασσινοστεροειδών οι ορμόνες αυτές προσδένονται σε υποδοχείς όπως BRI1 και BAK1 (BRI1-associated receptor kinase) που δρα ως συνυποδοχέας και προκαλεί τη φωσφορυλίωση του BAK1 και την ενεργοποίηση του BRI1. Η BRI1 φωσφορυλιώνει BSKs που είναι κινάσες απόκρισης στο BR οι οποίες με τη σειρά τους φωσφορυλιώνουν και ενεργοποιούν την φωσφατάση BSU1. Τέλος η φωσφατάση απενεργοποιεί την BIN2 και κατά συνέπεια την οδηγεί στα πρωτεασώματα, επιτρέποντας έτσι την επαγωγή των γονιδίων απόκρισης στα BRs (Saidi, Hearn and Coates, 2012).



Εικόνα 6.α) Μονοπάτι σηματοδότησης των μπρασσινοστεροειδών στο φυτό *Arabidopsis thaliana* (Saidi Y., Hearn T.J. and Coates J.C, 2012).

ΜΟΝΟΠΑΤΙΑ ΑΠΟΚΡΙΣΗΣ ΣΤΟ ΑΒΙΟΤΙΚΟ ΚΑΙ ΒΙΟΤΙΚΟ ΣΤΡΕΣ

Οι SKs τόσο στο *Arabidopsis* όσο και σε άλλα φυτά φαίνεται να κατέχουν σημαντικό ρόλο σε αβιοτικές, βιοτικές καταπονήσεις. Συγκεκριμένα με αύξηση της αλατότητας παρατηρείται αύξηση των δραστικών ριζών οξυγόνου και συνεπώς αύξηση των μεταγράφων ορισμένων SKs. Μια κινάση της οποίας έχει χαρακτηριστεί ο ρόλος σε αυτό το μονοπάτι είναι της *AtSK11* (Del Santo et al., 2012). Ο ρόλος του συγκεκριμένου καταλύτη είναι η αύξηση του αναγωγικού δυναμικού στο κύτταρο (NADPH) για την εξουδετέρωση των ριζών, πράγμα το οποίο πετυχαίνει μέσω της φωσφορυλίωσης της αφυδρογονάσης της 6-φωσφορικής γλυκόζης (G6PD). Στα σηματοδοτικά μονοπάτια που αφορούν την απόκριση σε βιοτικές καταπονήσεις οι SKs φαίνεται να συμμετέχουν στο μονοπάτι των MARK κινάσων. Αναλυτικότερα στο φυτό *Medicago Truncatula* έχουν παρατηρηθεί μειωμένα επίπεδα μεταγράφου και πρωτεΐνης (στο γονίδιο *MsK1*) όταν το φυτό εκτεθεί σε παράγοντες που επάγουν τους μηχανισμούς άμυνας των φυτών όπως το βακτηριακό ένζυμο cellulase (Wrzaczek et al., 2007). Υπερέκφραση αυτού του ενζύμου

στο *Arabidopsis* έδωσε ως φαινότυπο αυξημένη ευαισθησία στο παθογόνο *Pseudomonas syringae* και καταστολή του μονοπατιού κινασών.

1.4 GSK3/SHAGGY LIKE ΚΙΝΑΣΕΣ ΣΤΟ *Lotus japonicus*

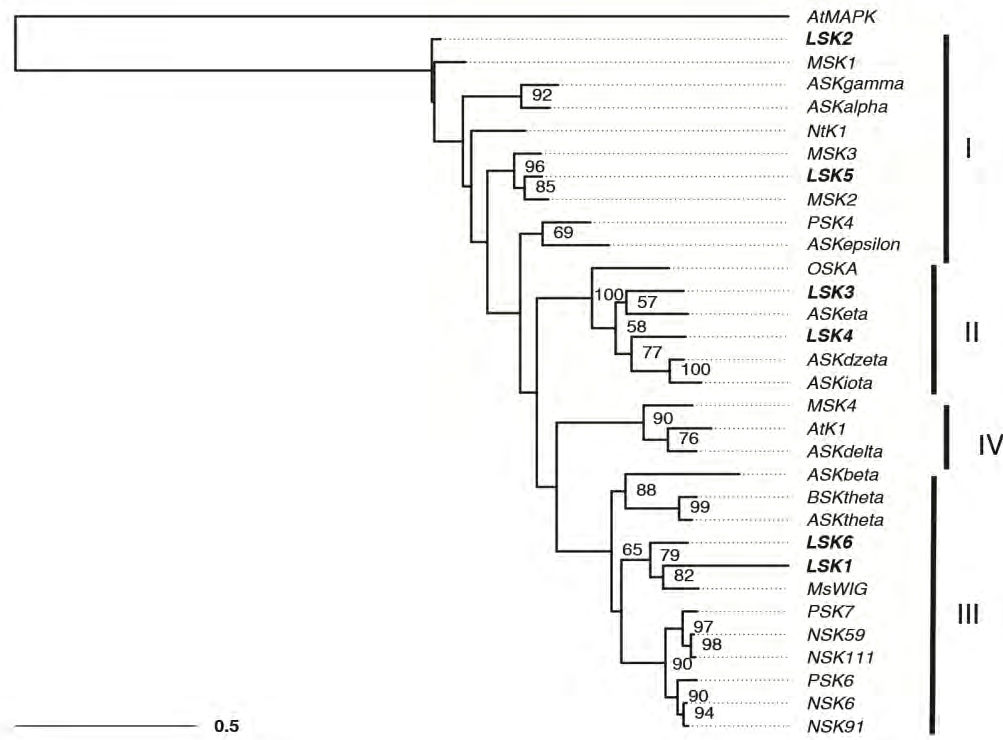
Οι μόνες SKs που έχουνε βρεθεί στο *Lotus* προέρχονται από την πρωτεωμική ανάλυση (western blot με αντίσωμα που προσδένεται σε συντηρημένη περιοχή κινασών) μιας cDNA βιβλιοθήκη ρίζας φυματίου (Kameshita et al., 2004). Με την αλληλούχηση των κλώνων καθώς και τη φυλογενετική ανάλυση των πρωτεϊνών προέκυψαν 164 κλώνοι που σχετίζονται με μετάγραφα των οποίων η μετάφραση τους οδηγεί στην παραγωγή κινασών Ser/Thr. Τα ένζυμα αυτά διαχωρίζονται σε 5 οικογένειες: την SnRK, την GSK-3, την Ndr, την Ark και την οικογένεια των κινασών που δρουν ως υποδοχείς. Από την ανάλυση των προβλεπόμενων πιθανών αλληλουχιών προέκυψε η αντιστοίχιση των cDNA κλώνων με τους cDNA κλώνους αυτής της μελέτης 1 γονίδιο σε κάθε κλώνο 15 γονίδια) όπως φαίνεται στον παρακάτω πίνακα:

| Clone | Number of clones ^a | Accession No. ^b |
|--------|-------------------------------|----------------------------|
| LNZ001 | 1 | AB115547 |
| LNZ002 | 23 | AB113570 |
| LNZ003 | 5 | AB113571 |
| LNZ016 | 17 | AB113572 |
| LNZ019 | 29 | AB115548 |
| LNZ020 | 23 | AB113573 |
| LNZ021 | 7 | AB115549 |
| LNZ024 | 8 | AB115550 |
| LNZ031 | 33 | AB113574 |
| LNZ109 | 1 | AB184970 |
| LNZ149 | 1 | AB184971 |
| LNZ215 | 12 | AB184972 |
| LNZ293 | 1 | AB184973 |
| LNZ344 | 1 | AB184974 |
| LNZ404 | 1 | AB184975 |

Πίνακας 1. Αντιστοίχιση των κλώνων της cDNA βιβλιοθήκης (a) με τους γονιδιωματικούς .b είναι τα accession numbers που προέκυψαν σε αυτή τη μελέτη (Kameshita et al., 2004) .

Από τη φυλογενετική ανάλυση των γονιδιωματικών αυτών κλώνων προέκυψε ότι στην οικογένεια GSK-3 ανήκουν οι LNZ020 και το LNZ031. Ο κλώνος LNZ020 περιλαμβάνει ένα γονίδιο που κωδικοποιεί μία κινάση, με ομολογία 76% με την AtSKθ του *A. thaliana* και κωδικοποιεί για μια πρωτεΐνη με 467 αα. Ο κλώνος LNZ031 αντιστοιχεί στην αλληλουχία AB113574.1 (GI:62857013) και παρουσιάζει 89% ομοιότητα με την AB113573.1 και 78% ομοιότητα με την AtSKα κινάση του *A. thaliana*. Στις πρωτεΐνες αυτές δόθηκαν τα ονόματα: Lotus SHAGGY-Like 1 *LSK1* (LNZ020) και *LSK2* (LNZ031). Η LSK1 φαίνεται να παίζει σημαντικό ρόλο στην ρύθμιση του αριθμού των φυματίων όπως φαίνεται από πειράματα σίγησης (Plitsi, MSc Thesis, 2016) και υπερέκφρασης της (Abramidou, BSc Thesis, 2017). Πρόκειται σίγουρα για μία πρωτεΐνη με ενεργότητα κινάσης η δράση της οποίας διαπιστώθηκε με πειράματα αυτοφωσφορύλιωσης (Psarrakou, MSc Thesis, 2014). Επιπρόσθετα μέσω ημιποσοτικής PCR πραγματικού χρόνου βρέθηκε ότι τα επίπεδα του μεταγράφου LSK1 αυξάνονται κατά τα πολύ αρχικά στάδια της

φυματιογένεσης (15 λεπτά μετά την μόλυνση) και μειώνονται στις 48 ώρες μετά την μόλυνση με το ριζόβιο (Psarrakou, MSc Thesis, 2014). Αξίζει να σημειωθεί ότι η προσθήκη του τριτερπενοειδούς λουπεόλη στο σύστημα προκαλεί τόσο την μείωση των επιπέδων έκφρασης (48 ώρες μετά τη μόλυνση) όσο και την μείωση της ενεργότητας της (Psarrakou, MSc Thesis, 2014). Μέσω BLAST βρέθηκαν και άλλα γονίδια (LSK2-LSK6) από τα οποία η LSK6 πιθανόν να παίζει κάποιο ρόλο στα αρχικά στάδια της φυματιογένεσης (Garagounis et al., 2018).



Εικόνα 7. Φυλογενετική ανάλυση των πιθανών κινασών (με bold) στο φυτό *Lotus japonicus* (Garagounis et al., 2018).

1.5 Ο ΟΡΓΑΝΙΣΜΟΣ- ΜΟΝΤΕΛΟ *Lotus japonicus*

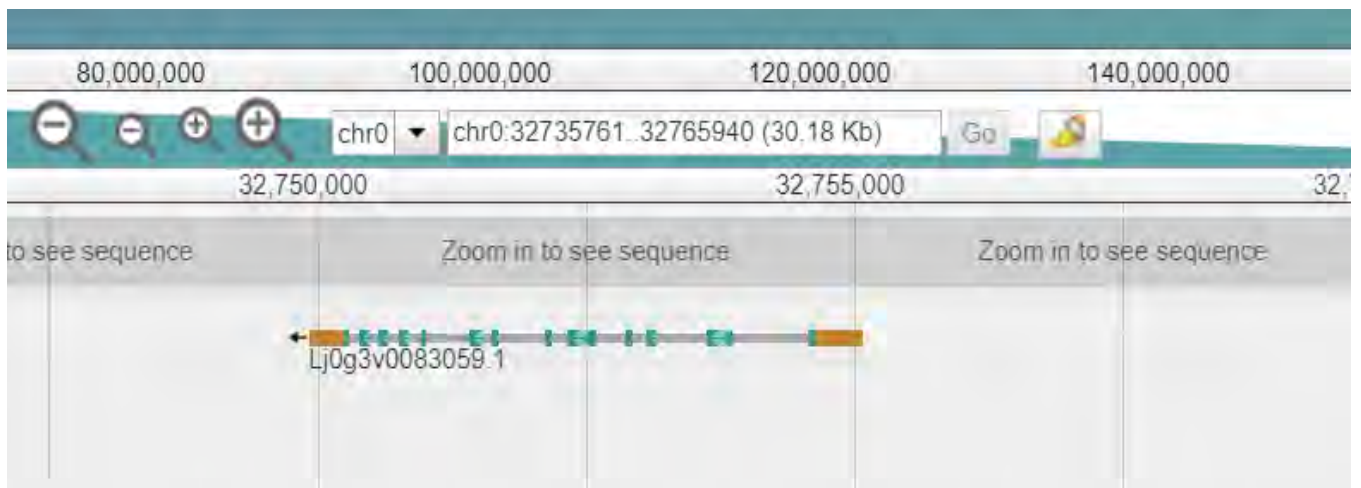
Πρόκειται για ένα φυτό με κύκλο ζωής 3 μήνες το οποίο ανήκει στην τάξη των ψυχανθών και στην οικογένεια Leguminosae (Fabaceae) μια από τις μεγαλύτερες οικογένειες στα αγγειόσπερμα. Ανήκει στους διπλοειδείς οργανισμούς, μπορεί να αυτογονιμοποιείται και διαθέτει μικρό γονιδίωμα (472MB MG20, 442.8 MB Gifu, Ito M. et al., 2000 Sato et al., 2008) το οποίο κάνει τις μελέτες σε επίπεδο γονιδιώματος πιο εύκολες (Stougaard, 2005). Οι παραπάνω λόγοι μαζί με το γεγονός ότι πρόκειται για φυτό μικρού μεγέθους με μεγάλα άνθη, μεγάλη παραγωγή σπόρων, εύκολα καλλιεργήσιμο και δεκτικό σε πρωτόκολλα μετασχηματισμού (Stougaard, 2005) το κάνουν ιδανικό

οργανισμό μοντέλο. Συγκεκριμένα μέσω αυτού το οργανισμού έχει γίνει και γίνεται , μελέτη των συμβιωτικών σχέσεων με βακτήρια κατά τη φυματιογένεση (Niwa et al., 2001, Kawaguchi et al., 2002), των συντενικών σχέσεων με άλλους οργανισμούς καθώς και του μοριακού μονοπατιού της φυματιογένεσης. Το φυσικό περιβάλλον ανάπτυξης του βρίσκεται στην Κίνα στην Ιαπωνία στην Κορέα και σε κάποιο βαθμό μέχρι το Αβγανιστάν. Η πιο βέλτιστες εργαστηριακές συνθήκες για το *Lotus* είναι: 16h/8h (μεγαλύτερη ημέρα), υγρασία 70% και θερμοκρασίες 18-22° C, ενώ το υπερβολικό πότισμα αντιπροτείνεται (Stougaard, 2005). Το είδος αυτό διαθέτει πολλούς οικότυπους με ποιο καλά μελετημένους και χρησιμοποιημένους τους MG20 (Miyakojima) και Gifu. Ο πρώτος οικότυπος εμφανίζεται στην Ιαπωνία χρησιμοποιείται για την αλληλούχηση του *Lotus* καθώς περιέχει περισσότερους πολυμορφισμούς από κάθε άλλο οικότυπο ,καθώς και ιδιαίτερα χαρακτηριστικά όπως μειωμένα επίπεδα ανθοκυανινών στα πέταλα, μεγάλα μαύρα σπέρματα και λίγα τριχίδια. Ο δεύτερος οικότυπος ανακαλύφθηκε το 1950 σε ένα ποτάμι στην πόλη Gifu της Ιαπωνίας το οποίο οδήγησε στην καθιέρωση του ως ερευνητικό εργαλείο μεταλλάξεων στο *Lotus*.



Εικόνα 8. Το φυτό μοντέλο *Lotus japonicus* .

Η πρώτη απόπειρα αλληλούχησης του γονιδιώματος έγινε το 2008 από τους Sato et al. όπου το ποσοστό επικάλυψη του γονιδιώματος ήταν 67% με γονιδιακή επικάλυψη στο 91,3%. Από το 2016 έχει δημοσιευτεί το τρίτο draft του γονιδιώματος που φέρει 98% επικάλυψη και εκτός από τα 6 πυρηνικά χρωμοσώματα διαθέτει και ένα τεχνητό χρωμόσωμα (0) που αποτελείται από contigs τα οποία δεν μπορούν να αντιστοιχηθούν κάπου στο γονιδίωμα και ενώνονται μεταξύ τους με εκφυλισμένες νουκλεοτιδικές επαναλύψεις (NNs) (Mun et al., 2016). Η LSK1 ανακαλύφθηκε μέσω blast ότι βρίσκεται στο χρωμόσωμα 0 και είναι το μοναδικό γονίδιο σε αυτό τον κλώνο. Τα γονίδια των υπόλοιπων SK κινασών του *Lotus* (LSK2-LSK3) είναι τοποθετημένα στα χρωμοσώματα 5, 6, 3, 1, 1 αντίστοιχα (blast με το γονιδίωμα δεδομένων από Garagounis et al., 2018).



Εικόνα 9. Το gene id για την LSK1 και η χρωμοσωμική της θέση στο 3^ο draft του *Lotus japonicus* (data derived from a genome browser in <https://lotus.au.dk/>).

1.6 ΦΥΜΑΤΙΟΓΕΝΕΣΗ- ΜΙΑ ΕΠΙΣΚΟΠΗΣΗ

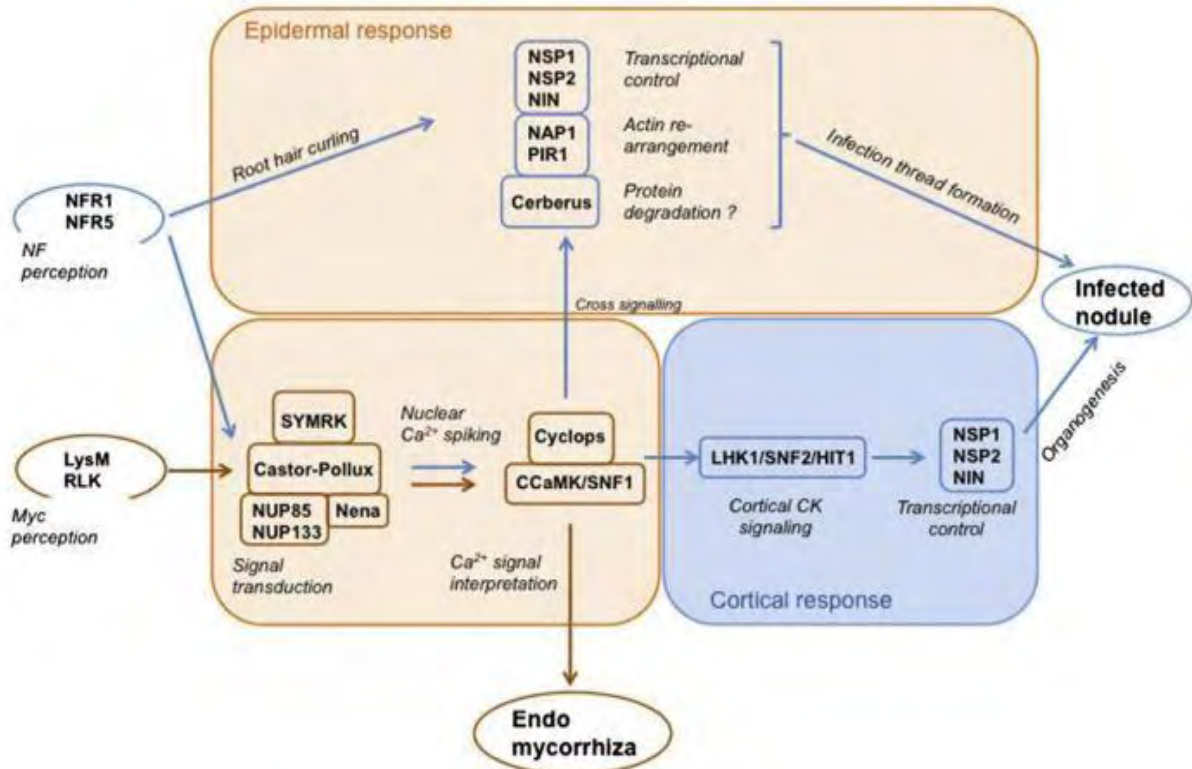
Για λόγους άγνωστους δεν υπάρχει ευκαρυωτικό ένζυμο που να διασπά τον τριπλό δεσμό του αζώτου. Εξάιρεση αποτελούν μία οικογένεια φυτών που ανήκουν στην οικογένεια των Leguminosae (Fabaceae) τα οποία μπορούν να δεσμεύουν και να ανάγουν το ατμοσφαιρικό άζωτο σε αμμώνιο (Downie, 2015), κάνοντας τα ικανά να ζουν σε περιβάλλον φτωχό σε άζωτο και δίνοντας τους μεγάλη διατροφική αξία (<https://nutritiondata.self.com/facts/legumes-and-legume-products/4326/2>). Αυτό το επιτυγχάνουν συνάπτοντας συμβιωτικές σχέσεις με ένα συγκεκριμένο στέλεχος βακτηρίων τα ριζόβια [γένη *Azorhizobium*, *Bradyrhizobium*, *Mesorhizobium*, *Rhizobium*, and *Sinorhizobium* (Romero & Mellado, 1996) (Qi & Shengming, 2017)]. Για να μπορέσει να επιτευχθεί ο αποικισμός των ριζοβίων απαιτείται από το φυτό ο σχηματισμός δομών που ονομάζονται φυμάτια. Αυτές οι δομές έχουν την ικανότητα να εγκολπώνουν στο εσωτερικό τους το καθένα έως και 10^9 ριζόβια (Downie A.J., 2015) δρώντας σαν εργοστάσια παραγωγής αμμωνίας καθώς εκμεταλλεύονται το πλήρως τη δράση του βακτηριακού ενζύμου νιτρογενάση που μετατρέπει το N_2 σε αμμώνιο (Debrosses & Stougaard, 2011). Το φυτό ως αντάλλαγμα παρέχει σε αυτά τα βακτηροειδή (δεν είναι αυτόνομα γιατί δεν έχουν όλα τα απαραίτητα ένζυμα για την επιβίωση) όλα τα απαραίτητα για τη διατήρηση της ζωής συστατικά όπως: υδατάνθρακες, πρωτεΐνες και ένα περιβάλλον με χαμηλή συγκέντρωση οξυγόνου, ώστε να μην αναστέλλεται η δράση της νιτρογενάσης των βακτηροειδών (Debrosses & Stougaard, 2011). Τα φυμάτια μπορούν να διακριθούν σε 2 κατηγορίες :

1. Τα καθορισμένα φυμάτια τα οποία έχουν βρεθεί στα γένη *Glycine*, *Phaseolus* και *Lotus*. Σε αυτά τα φυμάτια χάνεται η μεριστωματική τους ικανότητα μετά τη μόλυνση (την εγκόλπωση των ριζοβίων)

2. Τα μη καθορισμένα φυμάτια τα οποία έχουν βρεθεί στα γένη *Pisum* και *Medicago* όπου το μερίστωμα υφίσταται σε όλη τη διάρκεια ζωής του φυματίου (Newcomb, 1981). Και οι δύο τύποι φυματίων χαρακτηρίζονται από περιφερειακά αγωγά στοιχεία και ένα κεντρικό μερίστωμα που περιλαμβάνει μολυσμένα και μη κύτταρα (Mylona et al., 1995).

1.7 ΦΥΜΑΤΙΟΓΕΝΕΣΗ ΣΤΟ *Lotus japonicus*

Το φυτό *Lotus japonicus* συνάπτει συμβιωτικές σχέσεις με το βακτηροειδές *Mesorhizobium loti*. Σε συνθήκες χαμηλού αζώτου στο έδαφος επάγονται ουσίες που προσελκύουν τα ριζόβια με κυριότερες τα φλαβονοειδή. Τα βακτήρια σε απόκριση παράγουν λιποχιτινο-ολιγοσακχαρίτες που ονομάζονται Nod παράγοντες (Nod factors, NFs) και αναγνωρίζονται από επιδερμικά κύτταρα στο *L. japonicus*. Αυτοί οι ολιγοσακχαρίτες προσδένονται σε μία πρωτεΐνη που μοιάζει με υποδοχέα και δρα ως ετεροδιμερής κινάση σερίνης/θρεονίνης, που περιέχουν έναν τομέα LysM στην εξωκυττάρια περιοχή τους (LysM RLKs, Radutoiu et al., 2003) οι NFR5/1. Ταυτόχρονα φαίνεται να υπάρχει ένας δευτερογενές σήμα οφειλόμενο στην αναγνώριση των παραγόντων αυτών από τον υποδοχέα LjSYMRK, ο οποίος αποτελεί μία κινάση με επικράτειες που διαθέτουν μεγάλο αριθμό καταλοίπων λυσίνης (Orlloyd & Downie, 2004). Το μονοπάτι μεταγωγής σήματος που ακολουθεί είναι κοινό τόσο στη συμβίωση με ριζόβια όσο και στην όσο και στη συμβίωση με μυκορριζικούς μύκητες (Guinel, 2015). Συγκεκριμένα μέσω ενός μονοπατιού το οποίο ανιχνεύει και αλληλεπιδρά με τα επίπεδα ασβεστίου, όπου εμπλέκονται κανάλια κατιόντων [CASTOR στο *L. japonicus* (Imaizumi-Anraku et al., 2004)] και νουκλεοπορίνες [NUP133, NUP85 (Kanamori et al., Saito et al., 2007)] προκαλείται η αύξηση του πυρηνικού ασβεστίου (Calcium spiking). Το Calcium spiking έχει υπολογιστεί ότι ξεκινά 10 λεπτά περίπου μετά τη μόλυνση με ριζόβια (Serna-Sanz et al., 2011). Ακολουθεί η ενεργοποίηση ενός κεντρικού ρυθμιστή, μιας κινάσης εξαρτώμενης από ασβέστιο/καλμοδουλίνη, CCaMK στο *L. japonicus* (Levy et al., 2004) και μιας πρωτεΐνης που περιέχει μια δομή coiled coil που φωσφορυλιώνεται από την CCaMK (CYCLOPS, Yano et al., 2008). Στη συνέχεια ακολουθεί η ενεργοποίηση των GRAS μεταγραφικών παραγόντων NSP1 and NSP2 (Heckmann et al., 2006, Hirsch et al., 2009), καθώς και του γονιδίου *NIN* που λειτουργεί ως μεταγραφικός ρυθμιστής (Schauser et al., 1999), μέλη της οικογένειας μεταγραφικών παραγόντων ERF και NF-Y. Επιπλέον από τους CCaMK/CYCLOPS επάγεται η παραγωγή ενός μεταγραφικού παράγοντα ERN1 ο οποίος δρα ανεξάρτητα από το *NIN* (Cerrí, 2017, Kawaharada, 2017). Έτσι πραγματοποιείται η παραγωγή πρωτεϊνών η δράση των οποίων οδηγεί στην βακτηριακή επιμόλυνση και την φυματιογένεση (madsen et al., 2010). Τα πρώτα φυμάτια εμφανίζονται στις ρίζες μετά από 5 με 7 μέρες ενώ η καθήλωση αζώτου ξεκινάει 10 μέρες μετά τη μόλυνση και τα ώριμα φυμάτια εμφανίζονται 21 μέρες μετά (Dazzo et al., 1998).



Εικόνα 10. Μονοπάτι μεταγωγής σήματος της φυματογένεσης (madsen et al. 2010).

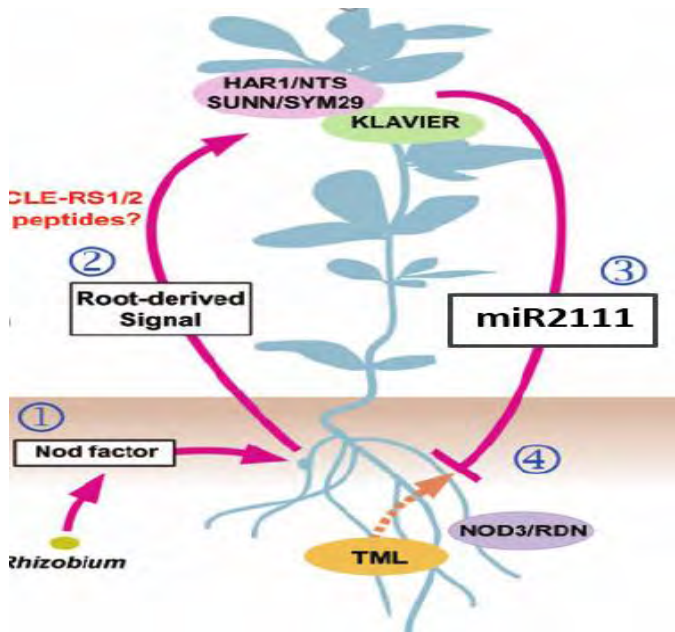


Εικόνα 11. Ώριμα φυμάτια στο *Lotus japonicus* (Credit: Masayoshi Kawaguchi).

1.8 ΜΗΧΑΝΙΣΜΟΣ ΑΥΤΟΡΡΥΘΜΙΣΗΣ ΤΗΣ ΦΥΜΑΤΙΟΓΕΝΕΣΗΣ (ΑΟΝ)

Πρόκειται για ένα μηχανισμό αυτορρύθμισης του αριθμού των φυματίων (Autoregulation Of Nodulation, AON) ο οποίος καθορίζεται από τη συνισταμένη των ενεργειακών αναγκών του φυτού και όπου τα παλαιότερα φυμάτια καταστέλλουν τη δημιουργία νέων φυματίων (Caetano-Anolles & Gresshoff, 1991) και φαίνεται να ξεκινάει περίπου 5 ημέρες μετά τη μόλυνση (Suzuki et al., 2008). Κάποιες από τις πρωτεΐνες που συμμετέχουν σε αυτό το μηχανισμό έχουν ανακαλυφθεί. Αναλυτικότερα εικάζεται ότι σήματα από τη ρίζα (με ισχυρότερους υποψηφίους τα CLE-RS1/2 πεπτίδια)

μεταφέρονται στο βλαστό και ενεργοποιούν την HAR1 κινάσης σερίνης/θρεονίνης με μια περιοχή πλούσια σε κατάλοιπα λευκίνης (LRR domain). Η HAR1 μεταφέρεται στη ρίζα όπου με την TML (TOO MUCH LOVE) μία F box πρωτεΐνη καταστέλλουν τη φυματιογένεση (Kouchi et al., 2010). Λόγω προηγούμενων ευρημάτων φαίνεται και η LSK1 να παίζει ρόλο στο AON χωρίς να είναι γνωστό πως καθώς τα υποστρώματα της δεν έχουν βρεθεί (Garagounis et al., 2018). Πρόσφατα έχει δειχθεί ότι ένα miRNA το miR2111 παίζει ρόλο στις μετα-μεταφραστικές τροποποιήσεις του TML (Tsikou et al., 2018). Οι ορμόνες που φέρεται να συμμετέχουν στη μείωση του αριθμού των φυματίων στο *Lotus* είναι το αμψισικό οξύ (ABA) (Tomimaga et al., 2009) και το γιασμονικό (JA) (Nakagawa, 2006) το οποίο έχει βρεθεί ωστόσο ότι παίζει και θετικό ρόλο στη φυματιογένεση στο MG20 (Suzuki A. et al., 2011), το σαλικυλικό οξύ SA (Stacey et al., 2006), ενώ δεν είναι ξεκάθαρος ο ρόλος των βρασσινοστεροειδών (BR) στη ρύθμιση της φυματιογένεσης.



Εικόνα 11. Το AON μονοπάτι στο *Lotus japonicus* (Kouchi et al., 2010, Tsikou et al., 2018) (τροποποιημένη).

2.ΠΡΟΓΡΑΜΜΑΤΑ ΚΑΙ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΗΘΗΚΑΝ

2.1 ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ ΤΟΥ *Lotus japonicus*

Την πιο σύγχρονη βάση δεδομένων του *Lotus japonicus* αποτελεί το Lotus base (<https://lotus.au.dk/>) στο οποίο περιέχονται πληροφορίες σχετικά με το γονιδίωμα, το μεταγράμμα, το πρωτείνωμα, τα μεταθετά στοιχεία και τη συνέκφραση γονιδίων (Mun et al., 2016). Συγκεκριμένα περιέχει ένα genome browser (JBrowse v1.12.0) ο οποίος συμβάλλει στην οπτικοποίηση της αλληλουχίας των 3.0 και 2.5 version του *Lotus japonicus* (MG20)(γενωμικό μιτοχονδιακό και χλωροπλαστικό) καθώς και των επιβεβαιωμένων μεταγράφων (σε FASTA format) και των πιθανών πρωτεϊνών ,μεταθετών στοιχείων ,κενών στην αλληλούχηση (GFF3 format) και επικάλυψη του γονιδιώματος από RNA-seq του MG20 και Gifu (bedGraph format) (Buels et al., 2016) (Mun et al. 2016).Κάθε γονίδιο που εισάγεται στη μηχανή αναζήτησης του Lotus base πρέπει να έχει την παρακάτω μορφή: LjXg3vYYYYYYY.Z

- όπου το Xg δηλώνει το χρωμόσωμα το οποίο βρίσκεται το γονίδιο (1-6 και το ψευδοχρωμόσωμα 0, mito, chloro)
- το 3v το 3^ο draft του *Lotus japonicus*
- α YYYYYYY αποτελούν ψηφία μοναδικά για κάθε γονίδιο και
- το Z δείχνει τον αριθμό (version) του μεταγράφου (πολλά μετάγραφα από εναλλακτικό μάτισμα)

(<https://lotus.au.dk/tools/trex>),(www.kazusa.or.jp/lotus/summary3.0.html)π.χ Lj0g3v0083059.1 (LSK1).

Επιπλέον περιέχει έναν expression atlas για την εύρεση της έκφραση ενός γονιδίου η τη διαφορική έκφραση μεταξύ γονιδίων (Verdier et al., 2013), το οποίο περιέχει δεδομένα από RNA-seq τα οποία διαρκώς ανανεώνονται και γίνεται η αντιστοίχιση τους με την v3.0 του γονιδιώματος (Mun et al., 2016). Το περιβάλλον εργασίας του expression atlas έχει μια γραμμή προσθήκης των id των μεταγράφων, μια καρτέλα επιλογής των συνθηκών στις οποίες θα γίνει ο προσδιορισμός των μεταγράφων και 3 επιλογές εξαγωγής δεδομένων raw data, normalized data, standardized data. Τα raw data δίνουν ως output τις απλές τιμές έκφρασης κάθε γονιδίου (ο αριθμός των τιμών για κάθε γονίδιο εξαρτάται από τον αριθμό των βιολογικών και τεχνικών επαναλήψεων), με την μέση τιμή και την τυπική απόκλιση των μετρήσεων. Τα κανονικοποιημένα και τυποποιημένα αποτελέσματα προκύπτουν από τις εξισώσεις στις παρακάτω εικόνα :

$$x'_s = \frac{(\log_{10}x_s) - (\log_{10}x)_{min}}{(\log_{10}x)_{max} - (\log_{10}x)_{min}} \quad (1)$$

$$x'_s = \frac{x_s - \mu}{\sigma} \quad (2)$$

Εικόνα 12. 1) κανονικοποίηση αποτελεσμάτων για κάθε γονίδιο (x_s απόλυτη τιμή έκφρασης δείγματος x_{min} η μικρότερη τιμή έκφρασης του γονιδίου στις επιλεγμένες συνθήκες x_{max} η μεγαλύτερη τιμή) 2) τυποποίηση αποτελεσμάτων με η μέση τιμή έκφρασης του γονιδίου στις επιλεγμένες συνθήκες σ η τυπική απόκλιση σε αυτές (Mun et al., 2016).

Ουσιαστικά η κανονικοποίηση μετατρέπει όλα τις τιμές έκφρασης σε τιμές μεταξύ 0 και 1. Αντίθετα στην τυποποίηση οι μεταβλητές τροποποιούνται με τέτοιο τρόπο έτσι ώστε να ακολουθεί την κανονική κατανομή με $N(\mu, \sigma^2)$ (0,1) (Mun et al., 2016).

Enter accession number or GI here

Separate each accession number of GI with a comma, space or tab.
Unsure what to do? [Try a sample data](#) from Mun et al., 2016.

Entire LjGEA dataset by gene ID

If no columns are selected by the user, all columns will be queried, returning a full dataset (default behavior).

Enter a keyword to filter conditions. Each space-separated keyword is treated as an AND, i.e. the condition has to match all listed keywords:

Filter conditions by entering a keyword...

| <input type="checkbox"/> | COLUMN | PLANT SPECIES | PLANT ECOTYPE | PLANT GENOTYPE | STANDARD | EXPERIMENTAL FACTOR | AGE (DAY) |
|--------------------------|-------------------|------------------------|---------------|----------------|----------|---------------------|-----------|
| <input type="checkbox"/> | WT_control1 | <i>Lotus japonicus</i> | Gifu | Wildtype | - | Draught Response | 35 |
| <input type="checkbox"/> | WT_Drought1 | <i>Lotus japonicus</i> | Gifu | Wildtype | - | Draught Response | 39 |
| <input type="checkbox"/> | Ljgln2_2_Control1 | <i>Lotus japonicus</i> | Gifu | Wildtype | - | Draught Response | 35 |
| <input type="checkbox"/> | Ljgln2_2_Drought1 | <i>Lotus japonicus</i> | Gifu | Wildtype | - | Draught Response | 39 |

Εικόνα 13. Το περιβάλλον εργασίας του expression atlas.

Ακόμη είναι δυνατόν να γίνει προσδιορισμός της συνέκφρασης μεταξύ γονιδίων από το πρόγραμμα CORGI. Αυτό το πρόγραμμα χρησιμοποιεί τον δείκτη συσχέτισης Pearson (Pearson's Correlation Coefficient (PCC)) μία στατιστική μέθοδο που υπολογίζει τη γραμμική συσχέτιση μεταξύ 2 μεταβλητών τα οποία έχουν προκύψει μετά από ιεραρχική ομαδοποίηση (hierarchical clustering) (Mun et al., 2016). Το περιβάλλον εργασίας είναι παρόμοιο με αυτό του expression atlas και ως αποτέλεσμα εμφανίζεται μία λίστα με αποτελέσματα με $PCC=R^2$ αριθμούς κατά φθίνουσα σειρά και από δίπλα τα γονίδια στα οποία αντιστοιχεί αυτός ο αριθμός συσχέτισης για το γονίδιο-query (μόνο ένα γονίδιο μπορεί να αναζητηθεί τη φορά).

Τέλος αυτή η βάση δεδομένων διαθέτει BLAST (νουκλεοτίδια για v2.5, 3.0 πρωτεΐνες για v.3.0) με δυνατότητα σύνδεσης με όλα τα παραπάνω.

2.2 ΠΡΟΓΡΑΜΜΑΤΑ ΕΥΡΕΣΗΣ ΜΕΤΑΘΕΤΩΝ ΣΤΟΙΧΕΙΩΝ

Το πρόγραμμα εύρεσης μεταθετών στοιχείων τα οποία έχουν χρησιμοποιηθεί στο *Lotus japonicus* είναι το **repeat masker** (<http://www.repeatmasker.org/>) και το **plant MITE database** (<http://pmite.hzau.edu.cn/>) (Holligan, Zhang, Jiang, Pritham, & Wessler, 2006)(Chen, Hu, Zhang, Lu, & Kuang, 2014). Το repeat masker σκανάρει την αλληλουχία για μεταθετά στοιχεία και άλλες επαναλήψεις που έχουν βρεθεί στο συγκεκριμένο γονιδίωμα. Τα μεταθετά στοιχεία του repeat masker έχουν προκύψει από το πρόγραμμα **RECON** το οποίο σκανάρει όλο το γονιδίωμα για μεταθετά στοιχεία (DNA και ρετροστοιχεία) αφού πρώτα γίνει blast με τον εαυτό του (Tabata and Stougaard, 2016). Αντίστοιχα τα προγράμματα **RBSB** και **MITEhunter** λειτουργούν με τον ίδιο τρόπο και τα δεδομένα που προκύπτουν μεταφέρονται στο P MITE database. Έπειτα με την βοήθεια του repeat masker (όπου διαθέτει προγράμματα που κάνουν blast) και της μηχανής blast στο plant mite database είναι δυνατή η εύρεση μεταθετών στοιχείων και mites. Το repeat masker παίρνει πληροφορίες από την βάση δεδομένων για μεταθετά στοιχεία **repbase** (<https://www.girinst.org/>) (Bao et al., 2015).

2.3 ΠΡΟΓΡΑΜΜΑΤΑ ΕΥΡΕΣΗΣ ΡΥΘΜΙΣΤΙΚΩΝ ΣΤΟΙΧΕΙΩΝ

Τα προγράμματα που χρησιμοποιήθηκαν για την εύρεση ρυθμιστικών στοιχείων είναι τα: PlantTFDB 4.0, PlantPAN 3.0, Nsite,nsiteM-PL, ScanWM-PL και CpGFinder.

PlantTFDB 4.0: Πρόκειται για μία βάση δεδομένων των μεταγραφικών παραγόντων και των νουκλεοτιδικών αλληλουχιών που αυτοί προσδέονται (αλληλουχίες πρόσδεσης υψηλής ποιότητας). Περιέχει ουσιαστικά πίνακες (Position specific scoring matrices-PSSM ή position weight matrices-PWM), όπου προσδιορίζουν την πιθανότητα κάθε νουκλεοτιδίου να βρίσκεται σε κάθε θέση.



Εικόνα 14: Εικονική αναπαράσταση ενός πίνακα όπου δείχνει την πιθανότητα ύπαρξης ενός νουκλεοτιδίου σε κάθε θέση(στην εικόνα το matrix του μεταγραφικού παράγοντα LjRAP2.4).

Παρόλο που τα ρυθμιστικά στοιχεία που είναι κυρίως διαθέσιμα αφορούν μονάχα το *arabidopsis* στο PlantTFDB 4.0 γίνεται αντιστοίχιση αυτών με μοτίβα πρόσδεσης μεταγραφικών παραγόντων για κάθε φυτό (<http://planttfdb.cbi.pku.edu.cn/>)(Jin et al.,2017). Συγκεκριμένα πραγματοποιώντας reciprocal blast hits (RBHs) ανάμεσα σε μεταγραφικούς παράγοντες του *arabidopsis* (η όποιου άλλου μεταγραφικού παράγοντα

είναι γνωστή η θέση πρόσδεσης) και καθενός από 165 είδη φυτών γίνεται μετάφορά των ρυθμιστικών στοιχείων, και των gene ontologies σε κάθε οργανισμό (Jin et al., 2017). Επίσης είναι διαθέσιμος ένας binding site prediction server για την εύρεση μοτίβων για κάθε οργανισμό. Αυτός ο server χρησιμοποιεί το πρόγραμμα FIMO (Find Individual Motif Occurrences) (<http://meme-suite.org/doc/fimo.html>) (Grant, 2011) το οποίο σκανάρει την αλληλουχία, συγκρίνει το score κάθε PSSM με αυτό ενός ολιγονουκλεοτιδίου σε αυτήν και εξάγει τα ρυθμιστικά στοιχεία των οποίων τα score μετατρέπονται, ένα p-value μικρότερο από αυτό που έχει επιλεγεί από το χρήστη και το οποίο υπολογίζεται με δυναμικό προγραμματισμό (<http://meme-suite.org/doc/fimo.html>). Επίσης για κάθε ένα μοτίβο υπολογίζεται και το q-value το οποίο αποτελεί την πιθανότητα το μοτίβο αυτό να έχει βρεθεί τυχαία. Επιπλέον στο PlantTFDB 4.0 διατίθεται μια συλλογή των genome browsers για κάθε είδος (PlantRegMap) το οποίο διαθέτει πιθανά ρυθμιστικά στοιχεία που προέρχονται από πολλαπλή στοίχιση των γωνιδωμάτων και από τα μοτίβα που βρέθηκαν με το FIMO (p-value=10⁻⁴).

PlantPAN 3.0: Αποτελεί μια βάση δεδομένων για ρυθμιστικά στοιχεία μεγαλύτερη από τις μέχρι τώρα υπάρχουσες όπου αντιστοιχεί τα ρυθμιστικά στοιχεία όπως και η προιγούμενη όμως δεν χρησιμοποιεί FIMO (<http://plantpan.itps.ncku.edu.tw/>) (Chow et al.,2019). Ακόμα εκτός από ρυθμιστικά στοιχεία συμβάλλει και στην εύρεση CpG islands στην αλληλουχία (50% CG content, 0.6 exp/obsrd ration, 200nt AT Skew, CG Skew, η πιθανότητα να βρίσκεται στο TSS και η πιθανότητα να βρίσκεται στη συγκεκριμένη αλησίδα του DNA).

Nsite: Αυτό το πρόγραμμα σκανάρει για ρυθμιστικά στοιχεία (μόνο consensus sequence χωρίς να δηλαδή νουκλεοτίδια που βρίσκονται για πάντα σε αυτή τη θέση για το μεταγραφικό παράγοντα – consensus sequence) χρησιμοποιώντας τις βάσεις δεδομένων RegsiteAN DB, RegsitePL DB και ooTFD DB για ζώα, φυτά και ζώα αντίστοιχα. Αναλυτικότερα ως πιθανά ρυθμιστικά στοιχεία ορίζονται αυτά που βρίσκονται σε ένα διάστημα εμπιστοσύνης ,ενώ θα πρέπει να βρίσκονται κάτω από ένα κατόφλι και ένα κατώφλι για ψευδώς θετικά (<http://www.softberry.com/cgi-bin/programs/promoter/nsite>) (Shahmuradov and Solovyeu, 2015). Επιπλέον διαθέτει το nsiteM-PL το οποίο συντελεί στο προσδιορισμό των ρυθμιστικών στοιχείων που έχουν βρεθεί μεταξύ 2 αλληλουχιών.

| | |
|--|--------------------------|
| change minimal similarity level in percents | 80 |
| change maximal expected by chance number of motifs | 0.1 |
| Write query sequences to output file. | <input type="checkbox"/> |
| Positions of motifs found are given in relation to right boundaries of upstream sequences, if query sequences include upstream sequences of genes by themselves or as part of gene sequence. If this option disabled, positions of motifs found are given as in query sequences. | <input type="checkbox"/> |
| Mean expected number (real, ≥ 0 .) | 0.01 |
| Statistical significance level (real, > 0 ., ≤ 100 .) | 0.95 |
| Allow some variation of distance between two blocks of RE. | <input type="checkbox"/> |
| Put here other options, if any. | |

Εικόνα 15: Πίνακας των παραμέτρων που χρησιμοποιεί το πρόγραμμα Nsite(default τιμές)

ScanWM-P: Πρόκειται για ένα πρόγραμμα το οποίο βρίσκει weight matrices (PSSM) σε αλληλουχίες χρησιμοποιώντας τη βάση δεδομένων RegsitePL (μοτίβα που έχουν βρεθεί πειραματικά για διάφορα είδη φυτών. Για να δώσει ως αποτέλεσμα ένα μοτίβο το πρόγραμμα αυτό χρησιμοποιεί ένα κατώφλι πάνω από το οποίο το μοτίβο αυτό θεωρείται στατιστικά σημαντικό (εικόνα 16) (<http://www.softberry.com/cgi-bin/programs/promoter/scanwmp.pl>). Όπως και σε όλα τα παραπάνω προγράμματα το σκανάρισμα πραγματοποιείται και στις 2 αλυσίδες.

Direct chain

Inverse chain

Advanced options [see example of value in (.):]

| | |
|--|-----|
| Threshold type - way of entering "cut-off values" for weight matrixes. When "threshold type" is 1 the "cut-off value" = "average" - "threshold value" * "std_dev", where: "average" is an average deviation; "std_dev" is a root-mean-square deviation. | |
| When "threshold type" is 2 the "cut-off value" = "wm_min_value" + "threshold value" * ("wm_max_value"-"wm_min_value"), where "wm_min_value", "wm_max_value" is minimal and maximal weights which the weight matrix can give. "Threshold value" here should belong to an interval (0; 1). | 2 |
| Threshold value | 0.9 |
| Put here other options, if any. | |

Εικόνα 16: Πίνακας των παραμέτρων που χρησιμοποιεί το πρόγραμμα ScanWP-PL(default τιμές) . Το wm_min_val.ue είναι το μικρότερο δυνατό score από ένα matrix .Οι τιμές που βρίσκονται στα τετραγωνάκια καθορίζονται από το χρήστη.

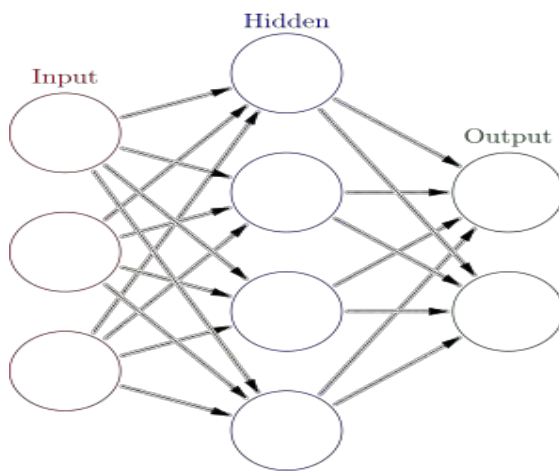
CpGFinder: Με αυτό γίνεται έλεγχος για το αν υπάρχουν CpG νησίδες χρησιμοποιώντας τις εξής παραμέτρους οι τιμές των οποίων καθορίζονται από το χρήστη :

1. %CG περιεχόμενο
2. P(CpG)/exp το παρατηρούμενο/ αναμενόμενο αποτέλεσμα
3. Ο λόγος CG/GC
4. Τα ελάχιστα CpG σε μία αλληλουχία
5. Το ελάχιστο μήκος του νησιδίου

Το CpGFinder (<http://www.softberry.com/cgi-bin/programs/promoter/cpgfinder.pl>) έχει χρησιμοποιηθεί για την εύρεση νησιδίων για πολλούς οργανισμούς μεταξύ αυτών και φυτά (<http://www.softberry.com> στην καρτέλα the program's usage in scientific publications).

2.4 ΠΡΟΓΡΑΜΜΑΤΑ ΕΥΡΕΣΗΣ ΥΠΟΚΥΤΤΑΡΙΚΟΥ ΕΝΤΟΠΙΣΜΟΥ

Τα προγράμματα που χρησιμοποιήθηκαν για την εύρεση του υποκυτταρικού εντοπισμού αποτελούν τα BUSCA και locTree 3.0 τα οποία χρησιμοποιούν τεχνητή νοημοσύνη για να βρουν τον υποκυτταρικό εντοπισμό πρωτεϊνών. Συγκεκριμένα το BUSCA βασίζεται σε μία μορφή των τεχνητών νευρωνικών δικτύων ,που αποτελούν μαθηματικό ανάλογο των βιολογικών νευρωνικών δικτύων. Κάθε νευρώνας παίρνει τιμές από 0 έως 1 (1 ενεργοποίηση 0 απενεργοποίηση) και για να μπορέσει να ενεργοποιήσει κάθε νευρώνα πρέπει η τιμή του να ξεπεράσει ένα "κατώφλι". Συνεπώς σε αυτά τα προγράμματα η πληροφορία (σε pixels) μετατρέπεται σε κόμβους καθένας από τους οποίους διαθέτει τις παραπάνω τιμές. Εάν κάποιος από αυτούς διαπεράσουν ένα threshold τότε δίνουν input στην επόμενη σειρά από κόμβους και αυτή με τη σειρά της στην επόμενη σειρά μετά από αυτήν φτάνοντας σε ένα τελικό μοναδικό αποτέλεσμα (υποκυτταρικός εντοπισμός).

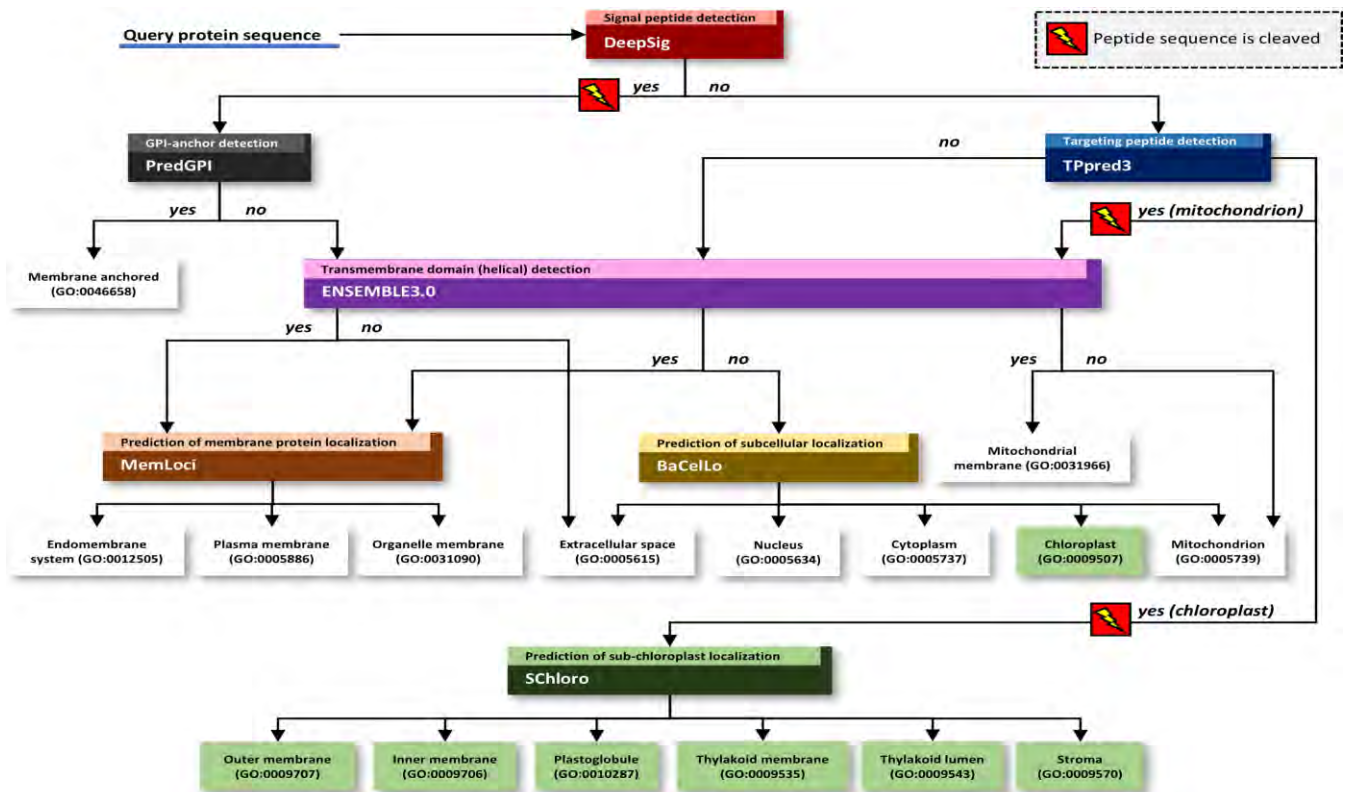


Εικόνα 17: Σχηματική απεικόνιση ενός τεχνητού νευρωνικού δικτύου. Input layer :οι τιμές από τα pixels της εικόνας ,hidden layer τα ενδιάμεσα στάδια και output layer οι τελικές τιμές του δικτύου . ο αριθμός τους είναι ενδεικτικός και καθορίζεται από το χρήστη έτσι ώστε να δίνουν το επιθυμητό αποτέλεσμα

Συνεπώς γίνεται training του δικτύου με πρωτεΐνες όπου είναι γνωστή η υποκυτταρική θέση έτσι ώστε να βρεθούν οι κατάλληλοι αριθμοί κόμβων σε κάθε σειρά που θα δώσουν το σωστό αποτέλεσμα (Savojarido et al., 2018).

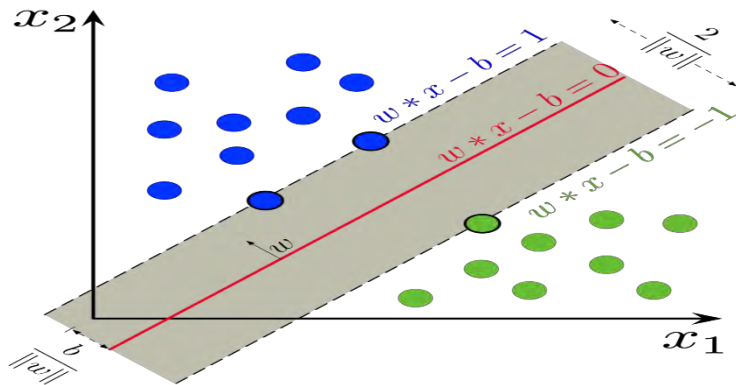
BUSCA: Το πρόγραμμα αυτό χρησιμοποιεί μία παραλλαγή των τεχνητών νευρωνικών δικτύων τα συγκλινόμενα νευρωνικά δίκτυα (CNNs) τα οποία παρουσιάζουν μεγαλύτερη ικανότητα αναγνώρισης μοτίβων και χρησιμοποιούνται ευρέως στον υποκυτταρικό

εντοπισμό πρωτεϊνών (Savojardo et al., 2018) (<http://busca.biocomp.unibo.it/>). Το πρόγραμμα δέχεται ως input αλληλουχίες πρωτεϊνών και ως output δίνει τον υποκυτταρικό εντοπισμό τους με το gene ontology term για κάθε ένα από αυτά. Για να το επιτύχει αυτό χρησιμοποιεί ένα σύνολο προγραμμάτων για να κατηγοριοποιήσει το πεπτικό σήμα που βρήκε από το DeepSig (CNNs, Savojardo et al., 2018) όπως φαίνονται στην παρακάτω εικόνα:



Εικόνα 18: Σύνοψη των προγραμμάτων που χρησιμοποιούνται για την εύρεση του υποκυτταρικού εντοπισμού. (C Savojardo , 2018).

locTree 3.0: Βασίζεται στα support vector machines (SVM) τα οποία επιτρέπουν τη διάκριση μεταξύ δεδομένων επιτρέποντας την προσομοίωση της μεταφοράς πρωτεϊνών στο κύτταρο (Goldberg et al., 2014). Τα SVM αποτελούν διανύσματα σχεδιασμένα έτσι ώστε να διακρίνουν τα δεδομένα όπως φαίνεται στην παρακάτω εικόνα:



Εικόνα 18: Αναπαράσταση της λειτουργίας ενός support vector machine. Για να είναι καλό το διάνυσμα πρέπει η συντεταγμένες του (matrix) να απέχουν από τα δεδομένα μία τιμή b .

Όταν η διάκριση των δεδομένων δεν μπορεί να γίνει γραμμικά (υποκυτταρικός εντοπισμός τότε χρησιμοποιείται ένα μη γραμμικό kernel matrix (kernel: ο χώρος μεταξύ 2 διανυσμάτων). Μεταξύ του εσωτερικού και του εξωτερικού χώρου γίνεται η διάκριση των δεδομένων (Goldberg et al., 2014).

ΣΤΟΧΟΣ ΤΗΣ ΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Όπως έχει προκύψει από πειράματα υπερέκφρασης και υποέκφρασης η LSK1 παίζει σημαντικό ρόλο στην αυτορρύθμιση του μηχανισμού σχηματισμού φυματίου. Δεδομένου ότι η δέσμευση αζώτου επιδρά στη βιομάζα ενός φυτού και ότι η φυματιογένεση επηρεάζει τα επίπεδα έκφρασης μεταβολιτών φαρμακευτικής σημασίας γίνεται κατανοητό ότι η εύρεση του μονοπατιού στο οποίο δρά η LSK1, με την εύρεση παραγόντων που ρυθμίζουν την έκφραση της, αποκτά επιστημονικό ενδιαφέρον. Συνεπώς στόχος της παρούσας πτυχιακής εργασίας είναι η εύρεση των πιθανών παραγόντων που είτε επάγουν είτε καταστέλλουν την έκφραση της LSK1. Πραγματοποιήθηκε *in silico* ανάλυση των ρυθμιστικών στοιχείων με τη βοήθεια των προαναφερθέντων προγραμμάτων αλλά και μοτίβων που είναι διαθέσιμα από τη βιβλιογραφία. Χρησιμοποιώντας τη βάση δεδομένων για το MG20 έγινε κλωνοποίηση των 2 kb ανοδικά από το ATG της LSK1 από το Gifu σε ένα pCambia 0391 xc φορέα. Τέλος πραγματοποιήθηκε σύγκριση των αποτελεσμάτων σε MG20 και Gifu. Από τα αποτελέσματα βρέθηκαν τόσο πιθανά TATA -boxes όσο και μεταγραφικοί παράγοντες πιο ανοδικά σε σχέση με το TSS. Ακόμη βρέθηκε ότι οι 2 υποκινητές εμφανίζουν μεγάλη ομολογία και τα σημεία που κυρίως διαφέρουν αφορούν το πόσο μακριά από το ATG βρίσκεται το TSS.

3. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ

3.1 ΑΠΟΜΟΝΩΣΗ ΓΕΝΕΤΙΚΟΥ ΥΛΙΚΟΥ ΚΑΙ ΕΝΙΣΧΥΣΗ ΤΟΥ ΥΠΟΚΙΝΗΤΗ ΤΗΣ LSK1

Σε 2 Eppendorf 1,5 ml συλλέχθηκαν 6+6 φύλλα από 3 διαφορετικά WT φυτά *Lotus japonicus* Gifu (2 φύλλα από κάθε φυτό). Τα δείγματα λειοτριβήθηκαν σε υγρό άζωτο και στη συνέχεια ακολούθησε η απομόνωση του γενωμικού DNA με τη μέθοδο CTAB. Συγκεκριμένα 2μl μερκαπτοαιθανόλης και 200 μl CTAB(100mM Tris-Cl pH=8.0, 20mM EDTA pH=8, 1,4M NaCl, 2% w/v CTAB, 1% w/v PVP 40000) (mix) που είχαν προθερμανθεί στους 65°C (10 λεπτά στο υδατόλουτρο) προστέθηκαν σε κάθε δείγμα (101 μl) και ακολούθησε επώαση τους στους 65°C για 15 min. Έπειτα ακολούθησε η προσθήκη 150 μl διαλύματος 1:25 ισοαμυλική αλκοόλη /χλωροφόρμιο σε κάθε δείγμα, vortex για 30 min και φυγοκέντρηση στα 13.000 rpm για 5 λεπτά. Έπειτα η υδατική φάση κάθε δείγματος μεταφέρθηκε σε καινούργιο Eppendorf όπου προστέθηκε ισοπροπανόλη 0,7 X του δειγματικού όγκου, επώαση στον πάγκο για 10 min και φυγοκέντρηση στα 13.000 rpm για 15 λεπτά. Από τα δείγματα αφαιρέθηκε το υπερκείμενο, προστέθηκαν 500 μl αιθανόλης στο καθένα και ακολούθησε φυγοκέντρηση στα 13.000 rpm για 5 λεπτά. Στη συνέχεια ακολούθησε αφαίρεση του υπερκειμένου και επώαση στους 37 °C (heat block). Τέλος ακολούθησε επαναδιαλυτοποίηση σε 20 μl υπερκάθαρου αποστειρωμένου H₂O με RNase (1μl RNase σε 50μl ddH₂O). Το κόματι 2 kb ανωδικά του γωνιδίου LSK1 ενισχύθηκε με τους εκκινητές : LSK1Pr₂₀₉₀F: GCAGGATCCGGGAATTTGTGGAGGCTCTG και LSK1Pr_{uni}R: CGACCCGGGTTTCATGGTTGCTGATTCCACTCA οι οποίοι διαθέτουν παλίνδρομες αλληλουχίες για την BamHI και SmaI αντίστοιχα. Η pcr πραγματοποιήθηκε με την Q5 πολυμεράση με τις παρακάτω ποσότητες αντιδραστηρίων : Q5 2x mm 25 μl, 10 μM For 1.25 μl, 10 μM Rev 1.25 μl, template 0.8 μl (αραίωση 1:10) , H₂O έως τα 50 μl. Οι κύκλοι της PCR ρυθμίστηκαν ως εξής : 30'' στους 98 °C , 10'' στους 98 °C, 30'' στους 67 °C ,1' στους 72 °C (επανάληψη των τριών σταδίων για 35 κύκλους) και τέλος 2' στους 72 °C και στους 4°C για ∞.

3.2 ΠΡΟΕΤΟΙΜΑΣΙΑ ΤΟΥ ΦΟΡΕΑ 1391xc ΓΙΑ ΚΛΩΝΟΠΟΙΗΣΗ

Ο φορέας 1391xc διαθέτει ένα γονίδιο GUS, ένα καναμυκίνης, ένα υγρομυκίνης και έναν 35S υποκινητή για το γονίδιο της υγρομυκίνης όπως φαίνεται στην παρακάτω εικόνα:

°C και πέψη με *Ascl* για να καταστραφούν οι φορείς στους οποίους δεν έχει πραγματοποιηθεί σωστά η αντίδραση λιγάσης (20 μl template, 1,5 μl *Ascl*, 2,5 μl *cutsmart* buffer, 0,5 μl H₂O) και επώαση 2h στους 37. Πραγματοποιήθηκε μετασχηματισμός σε κύτταρα DH5a με 7 μl δείγματος και αλκαλική λύση. Στα δείγματα τα οποία μετά από πέψη είχαν μοτίβο ίδιο με το 0391xc (1391xc χωρίς 35S και υδρομυκίνη) και σε άκοπο control πραγματοποιήθηκε PCR με εκκινητές για τον 35S (0,4 μl dNTPs, 0,4 μl For 35S, 0,4 μl Rev 35S, 0,08 μl *Kapa Taq*, 2 μl 10x buffer, 16,72 μl H₂O) και το μηχάνημα της PCR ρυθμίστηκε ως εξής: 95°C για 5', 95°C για 30", 60°C για 30", 72°C για 40" (και τα 3 στάδια για 30 κύκλους), 72°C για 5' και στους 4°C για ∞. Τα δείγματα φορτώθηκαν σε 1,2% gel αγαρόζης (1XTAE) μαζί με το άκοπο.

3.3 ΚΛΩΝΟΠΟΙΗΣΗ ΤΟΥ ΥΠΟΚΙΝΗΤΗ ΣΕ pGEM-T ΚΑΙ pCAMBIA 0391xc

Επειδή εμφανίστηκαν προβλήματα κατά την κλωνοποίηση του υποκινητή σε αυτούς τους φορείς (γιατί η *Bam*HI και η *Sma*I έκοβαν στο ίδιο σημείο) αρχικά έγινε κλωνοποίηση του υποκινητή σε pGEM-T φορέα και μετά κόψιμο και μεταφορά στον pCAMBIA 0391xc. Η εισαγωγή στον pGEM-T έγινε με TA cloning συνεπώς ο ενισχυμένος υποκινητής υποβλήθηκε σε κατεργασία από την *kapa taq* πολυμεράση (0,5 ul template, 1ul dATP, 5 ul buffer B *kapa Taq*, 0,1 ul *Kapataq*, 43,42 ul H₂O) με επώαση στους 72°C for 20'. Στη συνέχεια ακολούθησε η αντίδραση λιγάσης: 0,5 ul pGEM- T, 3,5 ul insert, 5 ul buffer 2x, 1ul ligase και αναμονή 1h στους 25°C. Αφού ακολούθησε ο μετασχηματισμός βακτηρίων (plates με 50ul Xgal. 2%, 10 ul IPTG). Ο μετασχηματισμός έγινε με δεκτικά κύτταρα *e coli* στα οποία προστέθηκε ο pGEM- T και ακολούθησε heat shock (επώαση στον πάγο για 30 λεπτά, μεταφορά στους 42°C για 50 δευτερόλεπτα και πάλι στον πάγο). Όταν το δείγμα βρισκόταν στον πάγο τοποθετήθηκαν σε αυτό 920μl LB και ακολούθησε επώαση για 45' στους 37 °C σύντομη φυγοκέντρηση για 30 δευτερόλεπτα, αφαίρεση 800μl και επίστρωση σε τριβλία. Στη συνέχεια έγινε αλκαλική λύση και ακολούθησε πέψη με *Sma*I και *Pst*I με το φορέα (οι φορείς του PGEM-T στάλθηκαν για αλληλούχιση με εκκινητές για T7 και SP6 και στις πέψεις χρησιμοποιήθηκαν αυτοί που ήταν antisense σε σχέση με τον T7 υποκινητή του PGEM).

Για το 0391xc οι τιμές έχουν ως εξής για τη *Sma*I:

| | |
|-------------------------------|-------------------------------|
| 4,2 ul DNA vector | 5,5 ul DNA insert |
| 1,5 ul buffer <i>cutsmart</i> | 1,5 ul buffer <i>cutsmart</i> |
| 1,5 ul <i>Sma</i> I | 1,5 ul <i>Sma</i> I |
| 7,8 ul H ₂ O | 1 ul RNase 10x |
| | 5,5 ul H ₂ O |

Πίνακας 2: Σύνοψη των αντιδραστηρίων που χρησιμοποιήθηκαν για την πέψη του 0391xc και υποκινητή με την *Sma*I.

Και για την PstI:

| | |
|------------------------|------------------------|
| 15 ul DNA vector | 15 ul DNA insert |
| 1,5 ul buffer cutsmart | 1,5 ul buffer cutsmart |
| 1,5 ul PstI | 1,5 ul PstI |
| 12 ul H ₂ O | 12 ul H ₂ O |

Πίνακας 3: Σύνοψη των αντιδραστηρίων που χρησιμοποιήθηκαν για την πέψη του 0391xc και υποκινητή με την PstI .

Ο χρόνος επώασης για την SmaI ήταν για 2 h στους 25°C και για 10' στους 65 °C ενώ της PstI ήταν 2 h στο 37 °C. Μετά από ηλεκτροφόρηση σε gel αγαρόζης και gel extraction έγινε η αντίδραση με λιγάση (NEB) με συγκέντρωση φορέα 4,0 ul και insert 13 ul. Για να ελεγχθεί αν έχει γίνει η αντίδραση θα γίνει πέψη με SpeI η οποία κόβει εκατέρωθεν του σημείου ένθεσης (1% gel αγαρόζης). Στη συνέχεια 8 primers στάλθηκαν για sequencing από τη lab supplies:

| |
|------------------------------|
| P1: CACATTGCGGACGTTTTTAAT |
| P2: GCGGAGGTGTCACGGTGTCAAC |
| P3: GACCTCTGTTCTTTGCTCTTCG |
| P4: CAGTTGTCTTCACACGCACAC |
| P5: GCAATTGCACTATTATCTC |
| P6: CAAGTTAAAACCTAAAACACG |
| P7: GAGACAAATTATAGAGAGAG |
| Gus5'R: CAGACTGAATGCCACAGGCC |

Πίνακας 3: primers που χρησιμοποιήθηκαν για τη αλληλούχιση του υποκινητή στον φορέα 0391xc-LSK1₂₀₉₀

Έλεγχος των αποτελεσμάτων της αλληλούχισης έγινε με το πρόγραμμα ugene και το assembly των αλληλουχιμένων κομματιών με το Combine FASTA (https://www.genscript.com/sms2/combine_fasta.html).

3.4 ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ ΑΝΑΛΥΣΗ

Τόσο ο υποκινητής της LSK1 του gifu όσο και αυτός του MG20 (2 και 10kb) ελέγχθηκαν για ρυθμιστικά στοιχεία με τα προγράμματα που θα αναλυθούν παρακάτω. Συγκεκριμένα 10 kb από τη γενωμική περιοχή ανοδικά του ATG συλλέχθηκαν από το genome browser του Lotus base (Chr0:32754629...32764629). Μέ τη βοήθεια του reverse complement (<http://reverse-complement.com/>) έγινε διαθέσιμη η κωδική αλυσίδα του DNA κάνοντας τη αλληλουχία έτοιμη προς ανάλυση. Και οι 2 αλληλουχίες ελέγχθηκαν για μοτίβα χρησιμοποιώντας το PlantTFDB με p value 10^{-5} και 4.78×10^{-5} αντίστοιχα για το 10 και 2 kb κομμάτι. Ο λόγος για τον οποίον γίνεται αυτό είναι ότι για να μπορούν να βρεθούν σωστά μοτίβα το p-value θα πρέπει να αυξάνεται όσο αυξάνονται και οι κιλοβάσεις του γονιδίου (<http://meme-suite.org/doc/fimo-tutorial.html>). Επίσης για να βεβαιωθούμε ότι η αλληλουχία που αποτελεί σημείο πρόσδεσης μεταγραφικών παραγόντων δεν βρέθηκε τυχαία, δεχόμαστε ως μοτίβα αυτά που έχουν q value < 0,05

(Noble, 2009) για το 2kb κομμάτι. Επιπλέον έγινε σκανάρισμα των αλληλουχιών με το Nsite θέτοντας ως παράμετρο τα μοτίβα να έχουν 100% ομοιότητα με τις αντίστοιχες αλληλουχίες του υποκινητή ως default και στο ScanW-PL (default parameters). Επιπλέον οι αλληλουχίες ελέγχθηκαν από το PlantPAN 3.0 για όλους τους φυτικούς οργανισμούς και για CpG νησίδια μαζί με το CpG finder. Για να βρεθεί εάν η εύρεση αυτών το νησιδίων είναι σημαντική έγινε σύγκριση της με τη σύσταση της υπόλοιπης αλληλουχίας χρησιμοποιώντας το πρόγραμμα ugene (<http://ugene.net/>) (Okonechnikov et al., 2012). Δεδομένο του ότι το γονίδιο επάγεται νωρίς κατά τη φυματιογένεση τότε πιθανοί μεταγραφικοί παράγοντες που παίζουν ρόλο στην έκφραση του αποτελούν οι προαναφερθέντες μεταγραφικοί παράγοντες που δρουν νωρίς στην φυματιογένεση και πριν από την έναρξη της (Miyake et al., 2004) καθώς και οι πρωτεΐνες του πίνακα 2. Επιπλέον λόγω της συμμετοχής του γωνιδίου στο AON εξετάστηκε μέσω αυτών των προγραμμάτων η ύπαρξη μοτίβων από μεταγραφικούς παράγοντες που συμμετέχουν στο μονοπάτι των ABA και JA καθώς και διάφορα TATA boxes από τη βιβλιογραφία (Miyake et al., 2004). Επιπλέον πραγματοποιήθηκε coexpression ανάλυση για την LSK1 με το GORGI. Σύμφωνα με μελέτες που έχουν γίνει στο *Arabidopsis* για την εύρεση συνεκφραζόμενων γονιδίων έχει προταθεί το Highest reciprocal ranking για την εύρεση των πιο συνεκφρασμένων γονιδίων (Obayashi and Kinoshita, 2009) (Liesecke et al., 2018). Δηλαδή εάν για το A γονίδιο το B αποτελεί το πιο συνεκφρασμένο ισχύει το ίδιο εάν ερευνηθούν τα PCCs για το B; Συνεπώς έγινε εύρεση των PCC για την LSK1 και για το κάθε πιθανό συνεκφραζόμενο γονίδιο ξεχωριστά το PCC της κινάσης. Το γονίδιο που εμφανιζόταν πιο ψηλά στην κατάταξη των PCCs και εμφάνιζε πιο ψηλά την κινάση θεωρήθηκε το ποιο συνεκφραζόμενο. Χρησιμοποιήθηκαν ως βάσεις δεδομένων για γονιδιακή έκφραση τα LjGEA και Kellys et al., (2017). Στα heat maps που προέκυψαν έγινε κανονικοποίηση έτσι ώστε να μπορούν να διακριθούν οι μεταβολές της έκφρασης ανεξάρτητα από την απόλυτη τιμή έκφρασης. Επίσης τόσο ο υποκινητής του συνεκφραζόμενου γονιδίου όσο και οι υποκινητές των μεταγραφικών παραγόντων που επάγονται νωρίς κατά τη φυματιογένεση ελέγχθηκαν για κοινά ρυθμιστικά στοιχεία με το Nsite-PL ψάχνοντας για επικαλυπτόμενα ρυθμιστικά στοιχεία με το PlantTFDB. Με τη βοήθεια του expression atlas προσδιορίστηκαν τα επίπεδα έκφρασης των μεταγραφικών παραγόντων σε ρίζες φυτών 3 εβδομάδων 1,3,7,14,21 μέρες μετά τη μόλυνση (Høglund N et al., 2009) καθώς και σε μη εμβολιασμένα με ή χωρίς άζωτο (Gifu). Επιπρόσθετα από το Kelly et al., (2017) καταγράφηκαν τα επίπεδα έκφρασης σε ρίζες από φυτά 14 ημερών 1 και 3 ημέρες από εμβολιασμό με H₂O, 1 και 3 ημέρες από εμβολιασμό με το m.loti, ρίζες 7 και 21 μέρες αντίστοιχα από φυτά 12 και 26 ημερών (R7A και Gifu σε όλα τα παραπάνω RNA-seq). Ο υποκυτταρικός εντοπισμός των TFs προσδιορίστηκε με τα προγράμματα BUSCA (πιο πρόσφατο) και LocTree 3.0 (πιο ακριβές για των πυρηνικό εντοπισμό σε φυτά και μύκητες).

| Gene Name | Accession No. | Transcription Factor Family | 3 h | 24 h | 2N | 4N | 7N | 12N |
|----------------------------|--------------------------|-----------------------------|-----|------|-----|-----|-----|-----|
| Group I (initial response) | | | | | | | | |
| <i>LjERF1</i> | AB378626 | AP2-EREBP | 1.8 | 1.7 | 1.2 | 0.1 | 0.4 | 0.5 |
| <i>HDZ-M48</i> | AB378627 | Homeobox | 1.8 | 1 | 0.7 | 0.1 | 0.1 | 0.1 |
| <i>CBF-A22</i> | AB378628 | CCAAT | 1.6 | 1.3 | 1.1 | 0.3 | 0.5 | 0.2 |
| <i>bZIP-R91</i> | AB378629 | bZIP | 1.9 | 1.1 | 0.9 | 0.2 | 0.3 | 0.9 |
| <i>LjRAP2.4</i> | AB378630 | AP2-EREBP | 1.9 | 1.1 | 1.4 | 0.6 | 1.6 | 0.5 |
| <i>ZF-M39</i> | AB378631 | C2H2 | 1.8 | 1.1 | 0.8 | 1 | 1.2 | 0.8 |
| <i>LjERF2</i> | AB378632 | AP2-EREBP | 1.7 | 1.3 | 0.6 | 0.8 | 1.4 | 1.6 |

Πίνακας 4: Λίστα των μεταγραφικών παραγόντων που δρουν νωρίς κατά τη φυματιογένεση (Asamizu et al.,2008).

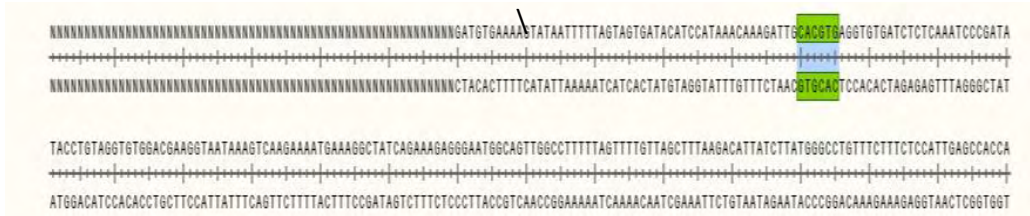
Τα μοτίβα των οποίων η ύπαρξη δεν μπορούσε να προσδιοριστεί με τη βοήθεια κάποιου προγράμματος (TATA-boxes ή μοτίβα από τη βιβλιογραφία), εντοπίστηκαν manually με τη βοήθεια του προγράμματος snargene viewer (<https://www.snargene.com/snargene-viewer/>) όπως φαίνεται στην εικόνα 21. Οι 2 υποκινητές ελέγχθηκαν για την ύπαρξη S/MARs με τη βοήθεια του MAR-Wiz το οποίο δίνει ως output S/MARs για τις περιοχές οι οποίες πληρούν τα παρακάτω κριτήρια:

The image shows a screenshot of the MAR-Wiz software interface. It is divided into two main sections: 'Core MAR Rules' and 'New MAR Rules'. Under 'Core MAR Rules', there are six checkboxes, all of which are checked: 'Origin of Replication Rule', 'TG-Richness Rule', 'Curved DNA Rule', 'Kinked DNA Rule', 'Topoisomerase II Recognition Rule', and 'AT-Richness Rule'. Under 'New MAR Rules', there are two checkboxes: 'Consensus Motif' and 'ATC Rule', both of which are unchecked.

Εικόνα 20: Σύνολο τόσο των γενικών όσο και ειδικών κριτηρίων (για φυτά New MAR Rules) για τον εντοπισμό MARs σε μία αλληλουχία.

Στο πρώιμο και ώριμο μετάγραφο του MG20 έγινε έλεγχος για τον εντοπισμό miRNAs χρησιμοποιώντας τον blast server από τη βάση δεδομένων miRbase (<http://www.mirbase.org/index.shtml>) (Kozomara et al., 2018) και riboswitches με τη βοήθεια του Riboswitch Scanner (<http://service.iiserkol.ac.in/~riboscan/application.html>)

Με τη βοήθεια των gene ontology consortium, Lotus base, PlantTFDB και PlantPAN 3.0 αναζητήθηκαν RNA binding πρωτεΐνες. Τέλος με τη βοήθεια της βάσης δεδομένων Repbase και του προγράμματος SCAN2, καθώς επιτρέπει τη σύγκριση μεταξύ αλληλουχιών πολλών κλοβάσεων, η ύπαρξη και η ομοιότητα μεταθετών στοιχείων που έχουν βρεθεί στο Lotus japonicus.

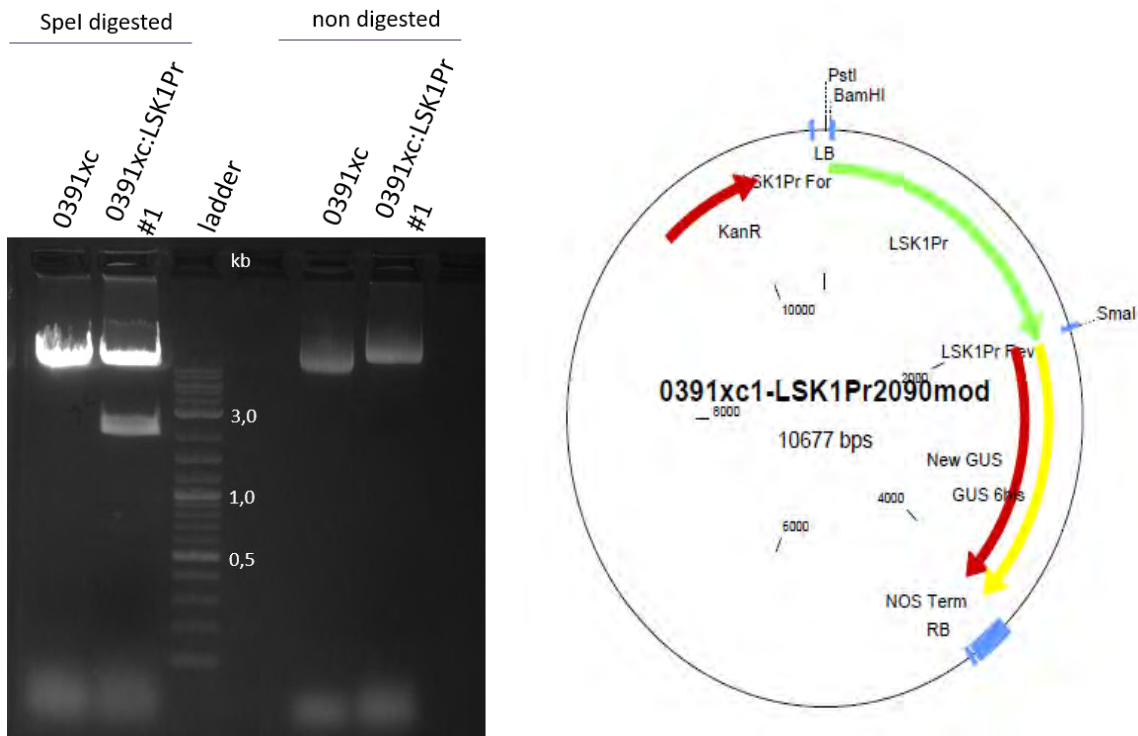


Εικόνα 21: Προσδιορισμός της ύπαρξης ενός G box με τη βοήθεια του snargene viewer (πράσινο χρώμα)

4. ΑΠΟΤΕΛΕΣΜΑΤΑ

4.1 ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΕΙΡΑΜΑΤΩΝ

Μετά την κλωνοποίηση του LSK1 υποκινητή με τον P_{cambia} 0391xc ακολούθησε πέψη με SpeI η οποία κόβει μόνο το πλασμίδιο με το ένθεμα. Όπως φαίνεται και στην παρακάτω εικόνα στο δείγμα με το όνομα 0391xc-LSK1_{pr} εμφανίζονται 2 ζώνες υποδηλώνοντας ότι έχει μπει σωστά το ένθεμα. Το τελικό πλασμίδιο αποτελείται από τον υποκινητή, ένα γονίδιο GUS καθοδικά αυτού και ένα γονίδιο ανθεκτικότητας στην καναμυκίνη εκτός των 2 συνοριακών.

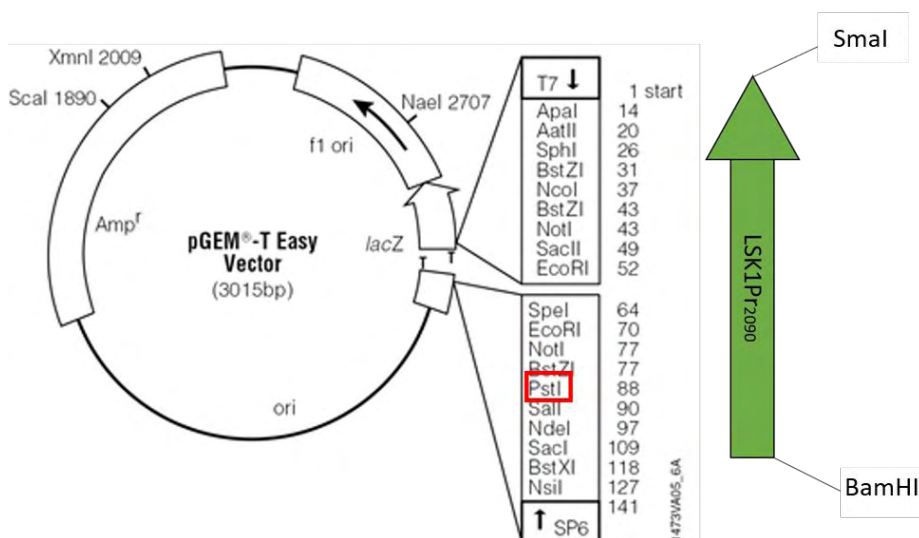


A.

B.

Εικόνα 22: Α) Τα προϊόντα πέψης των πλασμιδίων που έχουν επωαστεί με SpeI. Β) Ο τελικός φορέας p_{cambia} 0391xc-LSK1Pr2090.

Προηγουμένως είχε επιτευχθεί η ένθεση του LSK1_{Pr2090} στον pGEM-T φορέα όπως φαίνεται παρακάτω:



Εικόνα 23:Εισαγωγή του υποκινητή αντιπαράλληλα από τον υποκινητή T7 στο pGEM-T φορέα.

Από την αλληλούχηση τόσο του pGEM-T με τον υποκινητή όσο και του 0391xc-LSK1_{pr} και από assembly των αλληλουχιμένων αλληλουχιών προέκυψε η αλληλουχία του LSK1 promoter Gifu όπως φαίνεται και παρακάτω. Αν και δεν είναι ξεκάθαρο το είδος της βάσης σε κάθε θέση (ambiguous nucleotides) η αλληλουχία εμφανίζεται να έχει 97% ομοιότητα με αυτή του Lotus base με τις περισσότερες διαφορές (μη εκφυλισμένα νουκλεοτίδια) να εμφανίζονται στα 2kb ανοδικά του ATG. Γενικά φαίνεται ότι οι 2kb ανοδικά του ATG εμφανίζουν μεγάλη ομοιότητα ωστόσο δε είναι γνωστή η αλληλουχία ανοδικά των 2 kb στο Gifu οπότε δεν μπορεί να βγει συμπέρασμα για τα υπόλοιπα 8 kb. Επιπλέον σύμφωνα με το 3^o draft του Lotus τόσο από RNA-seq του MG20 όσο και του Gifu δεν προκύπτει η ύπαρξη κάποιου γονιδίου. Ωστόσο ενδιαφέρον παρουσιάζει το γεγονός πως τα σημεία έναρξης της μεταγραφής της κινάσης διαφέρουν στους 2 οργανισμούς (MG20 και Gifu RNA-seq track data στο genome browser του Lotus base). Συγκεκριμένα τα 2 TSS διαφέρουν κατά 214 nt με το TSS του Gifu να βρίσκεται πιο κοντά στο κωδικόνιο έναρξης.

```

>LSK1 Promoter Gifu
GGGAATTTGTGGAGGCTCTGCTGAGGAAATTCGAACCAGAATTGGAACCGTACATGCCAG
AACCAGTCCAAGATTCAGAGGAGGAAGAATCCCTAGGAAGCAAGAAGTCTTGGAGGCAG
AACGAAGAAGCTGTGTGGTGAAGCAAACCTGAGGCCAATTCAGTGTGTGGCGAAAT
CAGAGATCAAGCCATATTTGCCGGAGAATTTGAGATGGAAGCGTCGCGGAAAGACTCGG
AGTTTGCAATGGCGGTGGAGCAAAACACGGAGGTGTACGGGGTCAACCGGTACCGTGG
TTGATTACATCGATTGCACGTTGACGGCGAAAGGGGAGAAGATCGACGCCGAAGGGGAAA
TCAAAATCAAGCGGTACAGGAAGGTTCTTCAATGGAGAAATTCAGATTTGCAGCATTC
ATGGGAGGAAGAAGGTGCAACCAATGTTTGGCCCAAATAAGTGATTTAGGGTTCGATG
GGAGAGACACCGCCGCGGAAACCGCCGGATTCATGTCTCTGCACGGCGGTAACCGCCAA
ACGACTGCCGAACACGTTCTTTTGGTGGACGGAATTCATTTCAGACCTCTGTTCTTTG
CTCTTCGATGGTTGCCATTCGACTGGAGATAAGCCACCAGACCGCAGTGGTGGCGGTTCA
CGCTTCATCTTGGCGATTCTTGAGATCAGACTTGGTGACCAACCGTCTCCAGCTAAS
CCATCGGATTTGGCTGAKACGGCGCTGGTGKAATATTTGGCCCTCRAGTTAKTTCKCACG
GAGGAKAGTAAAGGGCGCTCGCTCATTGCAAGTCGTAACCGTTCACGGCCACCGGGCG
AAGCCATTTAGGCTTTGCAAGTTACRGCAGGGGACGCTCGTCTCTGTCGTTGCTGGTTGGT
GTGGTCCCTCTKTGTCTWTTACGCACACGTTTYYKTARGCRAAGCATGCTGGCTCAGC
TAGATTTTTCAGAAATGGGTCAAAGTTTGACCACAAGGTCAAAGCCATAATGCTAAACA
GAAGGAACGGGCCAAGAAGAAGTCTTTCTTCTCAATCATCTACAGACCAGTTCATTCCC
TTACTCTAATGGGCCAGGCCCAATATGTGTATGTAATGAGACTTGAACTTTTCGATT
AAGAAATAAAATAAATTCATCCTCTTTCGGAAGTCAAGGTGTTTGTCAAACACGAAGCTA
TTCCCAATTTCCAAAGATTCCAAGCAATTCGACTATTATCTCTAATTAACCTGACAAAC
ACTTATAAATTAAGATTGAACATTTTAAATTTGCTAATTTTAAATTTCAAATTAATCTCA
TAATTTACTTTATAGTAGATTTCATAGTTCATAGATTTCCGTTCCATAAwwWAAATAGTCA
AGATTTCCGGACAACKAAATTTTCWGTGTTTTTAATTTACCTTTTCTGGTTATTGACAAA
ATGTTTCCCTTAMCTGCATTTATCTCTAWTTCCTAACCTAATTAATTAATYATGCTAT
TCAACATTTTCATCAAATAAATCCSAAATTTGATTAKTCAAATTTACMAGTTAAACTTAWAA
TACSGTTTACGCRATAGGATATTCGTTGACGCGCCGCATCTATCCCTTTTCCGTAATAA
AATCATGAAACGAGGGCATTCCCAACAACCAACCAACCGCAACCCACCAACCTAT
GCAAGAAACGAGAAGATCTGGACCGTCCATTTCAATCCATGACCATCCAACGGCCAAAGA
TCACGAGTTTGACTGATGCACAGGTTTCAAAGTAGAATGATTAGGTTTGACCGTAGAAAA
TTCAGATTTTTTTCTCATAAAAAACGACCACCTGAAAAATCCTGTCTCCGAGCAACAG
AGACAAATATAGAGAGAGAAAAACAAGAGGGGTTTTGTTGGTGGGATTTTCAAATTTTC
CAACTGTCTGCTTAGTTTTTGGATTTGGTGTGGGTACTTGGTGGGTGAGAGAAAG
TTTGGGACTTTGTTATCTGTGGGAAAGTTAATTTTTTTTAGTTGGGTTGAGGGGATT
TAGGGTTGGATTGGTGGGTTGTGAGTGAATCAGCAACCATGAA

```

Εικόνα 24: Η αλληλουχία του υποκινητή της LSK1 στο φυτό Gifu(R=A ή G,W=A ή T,K=G ή T).

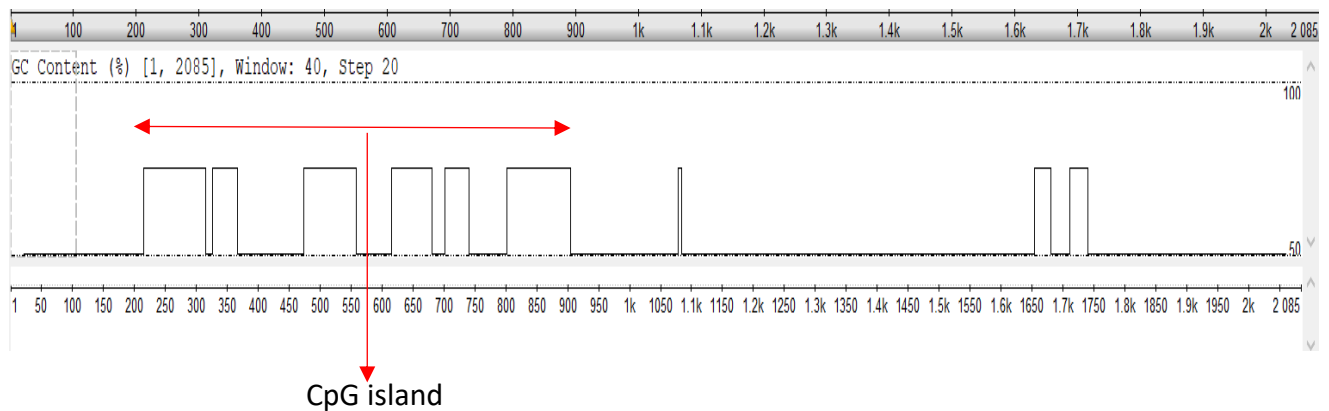
| Score | Expect | Identities | Gaps | Strand |
|-----------------|--|----------------|------------|-----------|
| 3528 bits(3912) | 0.0 | 2029/2090(97%) | 5/2090(0%) | Plus/Plus |
| Query 1 | GGGAATTTGTGGAGGCTCTGCTGAGGAAATTCGAACCAGAATTGGAACCGTACATGCCAG | 60 | | |
| Sbjct 7293 | GGGAATTTGTGGAGGCTCTGTTGAGGAAATTTGAACCAGAATTGGAACCGTACATGCCCG | 7352 | | |
| Query 61 | AACCAGTCCAAGATTCAGAGGAGGAAGAATCCCTAGGAAGCAAGAAGTCTTGGAGGCAG | 120 | | |
| Sbjct 7353 | AACCAGTCCAAGATTCAGAGGAGGAAGAATCCCTGGGAAGCAAGAAGTCTTGGAGGCAG | 7412 | | |
| Query 121 | AACGAAGAAGCTGTGTGGTGAAGCCAAACCTGAGGCCAATTCAGTGTGTGGCGAAAT | 180 | | |
| Sbjct 7413 | AACGAAGAAGCTGTGTGGTGAAGCCAAACCTGAGGCCAATTCAGTGTGTGGCGAAAT | 7472 | | |
| Query 181 | CAGAGATCAAGCCATATTTGCCGGAGAATTTGAGATGGAAGCGTCGCGGAAAGACTCGG | 240 | | |
| Sbjct 7473 | CAGAGATCAAGCCATATTTGCCGGAGAATTTGAGATGGAAGCGTCGCGGAAAGACTCGG | 7532 | | |
| Query 241 | AGTTTGCAATGGCGGTGGAGCAAAACCGGAGGTGTACGGGGTCAACCGGTACCGTGG | 300 | | |
| Sbjct 7533 | AGTTTGCAATGGTGGCGGAGCAAAACCGGAGGTGTACGGGTGTCAACCGGTACCGTGG | 7592 | | |
| Query 301 | TTGATTACATCGATTGCACGTTGACGGCGAAAGGGGAGAAGATCGACGCCGAAGGGGAAA | 360 | | |
| Sbjct 7593 | TTGATTACATCGATTGCACGTTGACGGCGAAAGGGGAGAAGATCGACGCCGAAGGGGAAA | 7652 | | |
| Query 361 | TCAAAATCAAGCGGTACAGGAAGGTTCTTCAATGGAGAAATTCAGATTTGCAGCATTC | 420 | | |
| Sbjct 7653 | TCAAAATCAAGCGGTACAGGAAGGTTCTTCAATGGAGAAATTCAGATTTGCAGCATTC | 7712 | | |

Εικόνα 25: Σύγκριση των αλληλουχιών των 2 υποκινητών (MG20 &Gifu).

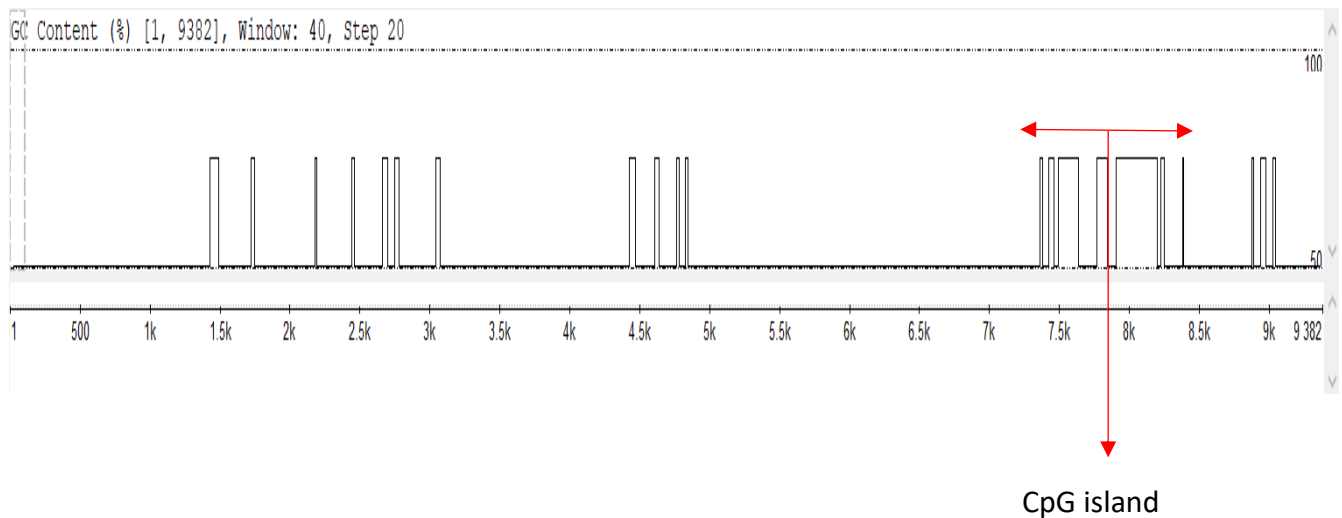
4.2 CpG ΝΗΣΙΔΕΣ ,ΜΕΤΑΘΕΤΑ ΣΤΟΙΧΕΙΑ ΚΑΙ S/MARs

Από τα δεδομένα του Repbase που εξετάστηκαν με το Scan2 προέκυψε ότι δεν έχει ομοιότητα με κανένα μεταθετό στοιχείο ομοίως και με blast ανάλυση στην Plant-Mite database. Τα MARs εμφανίζονται 2 φορές στην 10kb αλληλουχία ανοδικά του 2 kb κομματιού ενώ υπάρχει ένα κομμάτι μέσα στα 2 kb το οποίο αποτελεί CpG island και διαφέρει σημαντικά από την υπόλοιπη αλληλουχία σε MG20 και Gifu όπως φαίνεται και παρακάτω:

A)

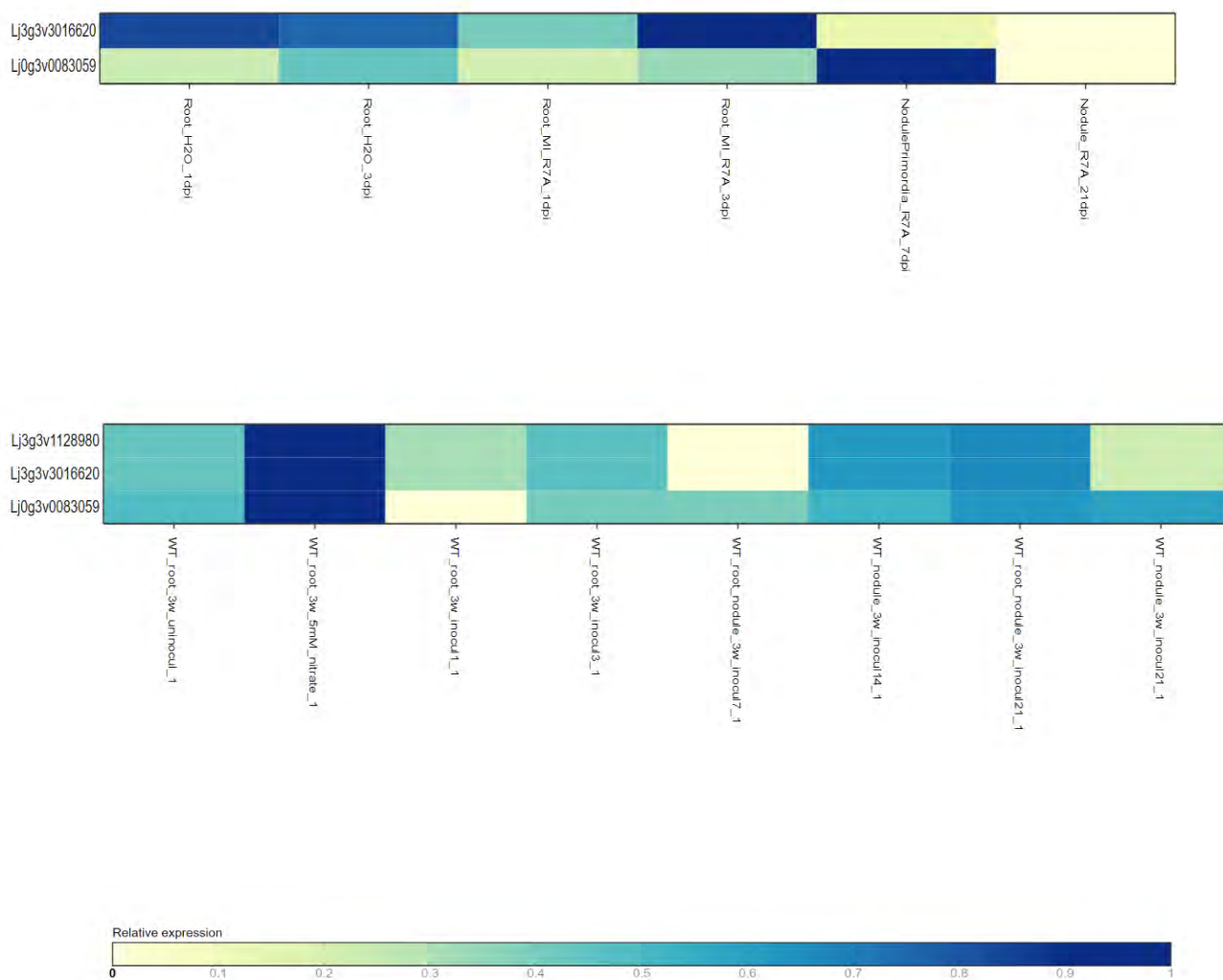


B)



Εικόνα 26: Διαφορά στο GC content μεταξύ των γενωμικών νησιδίων και των υπόλοιπων περιοχών στο 2kb υποκινητή του Gifu A) και στο 10 kb υποκινητή του MG20 B) (min cutoff value 50).

Στα S/MARs εμφανίζονται πολλά AT-hook μοτίβα (Fujimoto et al.,2004). Υπάρχουν 2 AT-hook binding πρωτεΐνες στο *Lotus* που είναι ομόλογα με του *Theobroma cacao*. Τα γονίδια αυτά παρουσιάζουν κοινά επίπεδα έκφρασης με την LSK1 σε φυτά τριών εβδομάδων όπως φαίνεται και στην εικόνα:



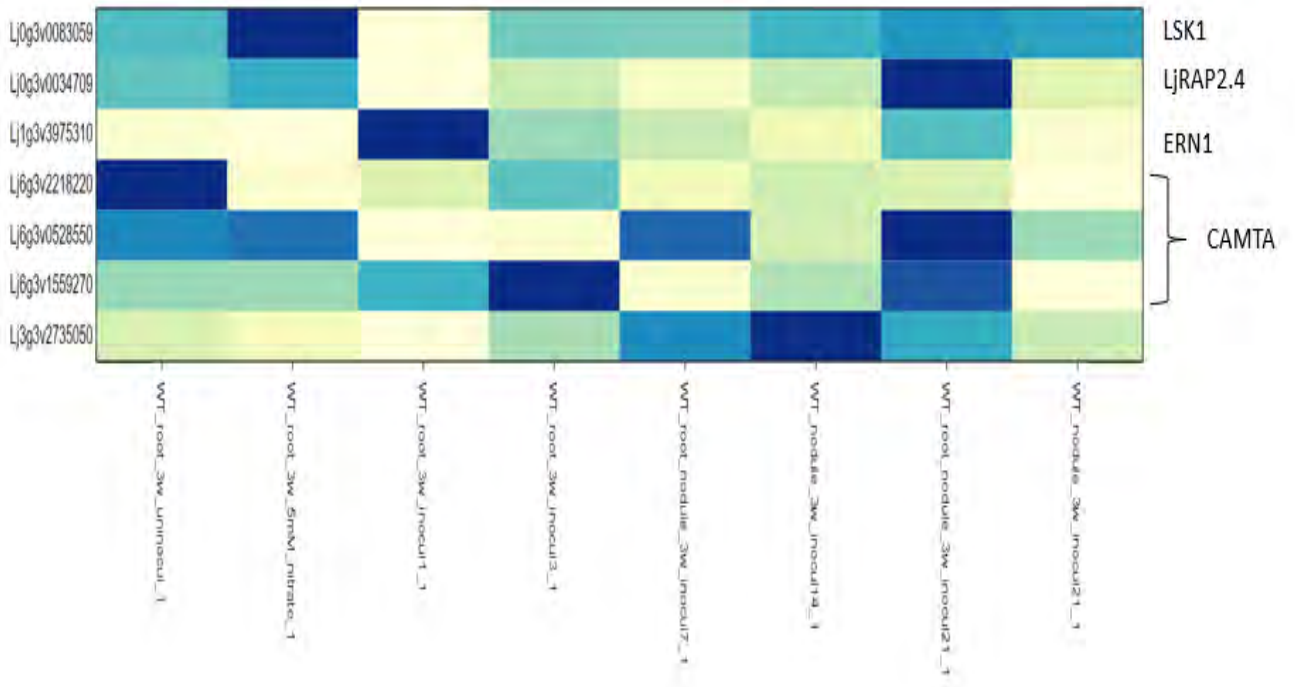
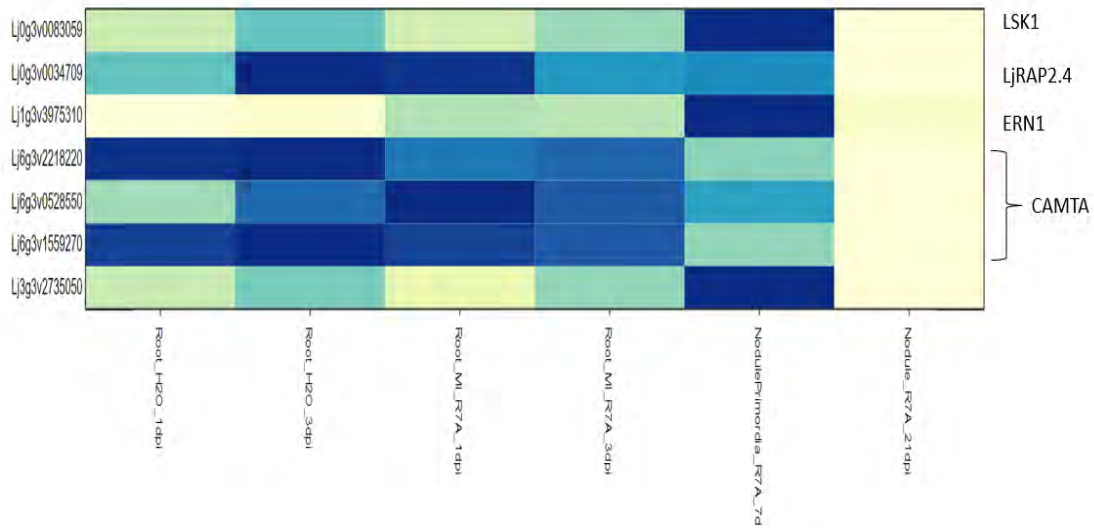
Εικόνα 27: Επίπεδα έκφρασης των AT-hook binding proteins σε RNA-seq δεδομένα από Kelly et al., 2017 (πάνω) και microarrays (Høgslund *et al.*, 2009) (κάτω). Η LSK1 είναι το Lj0g3v0083059 (κανονικοποιημένα αποτελέσματα).

Στις υπόλοιπες συνθήκες σύμφωνα με τα RNA-seq και microarray δεδομένα το 1 από τα 2 TFs δεν εκφράζεται ενώ το άλλο εκφράζεται κυρίως 1 και 3 μέρες μετά τη φυματιογένεση (μεγαλύτερο peak στις 3 μέρες που φαίνεται να ακολουθεί και η LSK1)

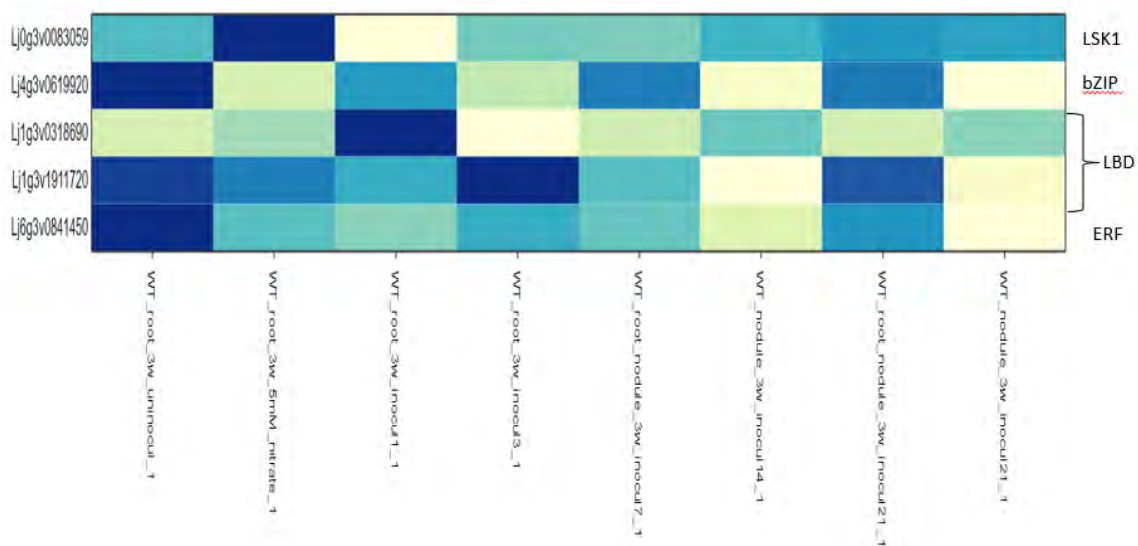
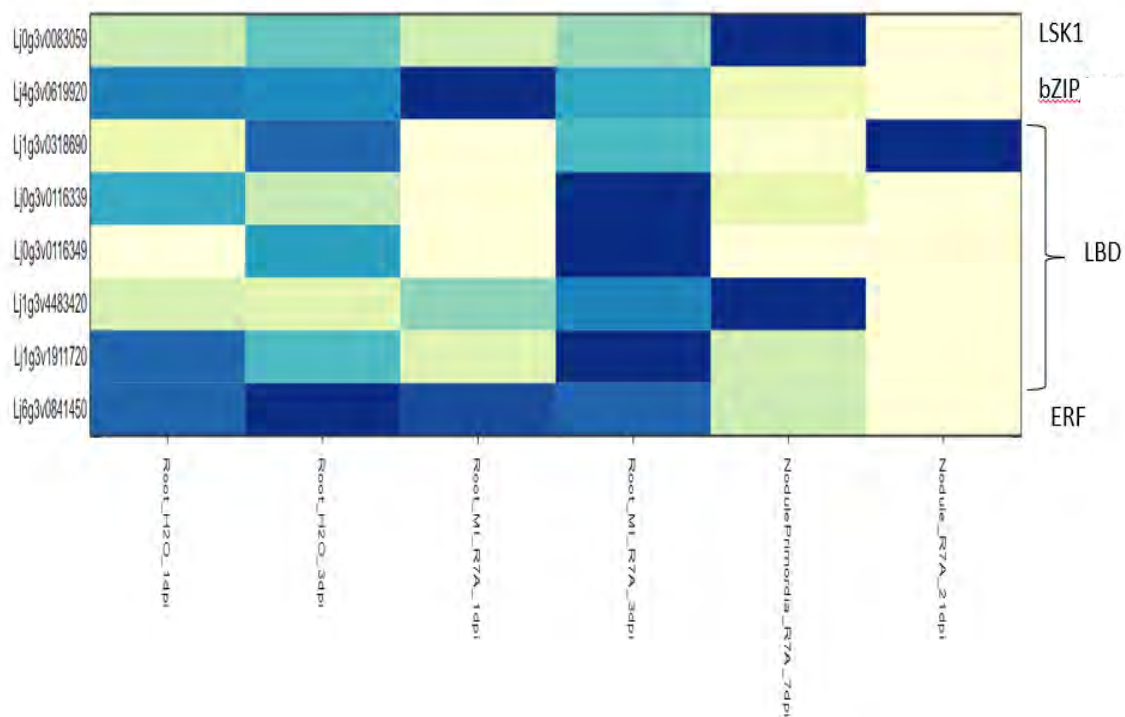
4.3 TATA -box ΚΑΙ ΛΟΙΠΑ ΡΥΘΜΙΣΤΙΚΑ ΣΤΟΙΧΕΙΑ

Όπως είναι γνωστό το TATA-box αποτελεί κεντρικό ρυθμιστικό στοιχείο με την μεγαλύτερη επίδραση στα επίπεδα μεταγραφής. Στους υποκινητές της LSK1 βρέθηκαν 2 core elements ένα TAAAAAAA box και ένα AAATAAAT box στο Gifu και MG20 αντίστοιχα. Για το 1^ο δεν είναι τόσο σίγουρο κατά πόσο είναι λειτουργικό καθώς επειδή δεν έχει κάποιο G δεν μπορεί να σχηματίσει σύμπλοκο με την πολυμεράση. (Berendzen, Kilian and Wanke, 2013) ή ότι η πρόσδεση της πολυμεράσης μπορεί να επιτευχθεί αλλά τα μεταγραφικά επίπεδα είναι χαμηλά (Flatters and Lavery, 1998). Για το 2^ο δεν αποτελεί ακριβώς TATA-box αλλά AT-hook καθώς υπάρχει μεταγραφικός παράγοντας πρόσδεσης σε AT-hook στο Arabidopsis ομόλογος με την πρωτεΐνη Lj3g3v2735050 και ίσως να μπορεί να αλληλεπιδρά με tata box binding factors (Metcalf and Wassarman, 2006) και συμμετέχει σε επιγενετικές τροποποιήσεις σε φυτά (Yun, 2012).

Σχετικά με την ύπαρξη άλλων μεταγραφικών παραγόντων στον υποκινητή βρέθηκαν σημεία πρόσδεσης των μεταγραφικών παραγόντων ERN1 και LjRAP2.4 που όπως συζητήθηκε δρουν στα αρχικά στάδια της φυματογένεσης (early nodulin genes). Ακόμη στον υποκινητή βρέθηκε μια αλληλουχία η οποία αποτελείται από 2 CG-rich περιοχές. Πιο συγκεκριμένα αποτελείται από 1 CG-rich μοτίβο το οποίο αποτελεί μοτίβο πρόσδεσης άγνωστου μεταγραφικού παράγοντα στο *Phaseolus vulgaris* και από άλλο 1 10 νουκλεοτίδια καθοδικά του προηγούμενου στο οποίο αποτελεί σημείο πρόσδεσης μεταγραφικών παραγόντων ERF στη μπανάνα *Cavendish* (Nsite-PL). Και τα 2 μοτίβα αντιστοιχούν σε CG-rich περιοχές που προσδέονται μεταγραφικοί παράγοντες στον άνθρωπο (Nsite). Γενικά αυτά τα μοτίβα είναι συντηρημένα μεταξύ οργανισμών καθώς συνήθως υπομεθυλιώνονται όπως και τα CpG islands (Babenko et al., 2017). Σε αυτή την περιοχή προσδέονται κυρίως μεταγραφικοί παράγοντες LBD (Lateral. Organ Boundaries) οι οποίοι φαίνεται να συμμετέχουν σε διάφορες διαδικασίες της φυτικής ανάπτυξης συμπεριλαμβανομένου αποκρίσεις σε νιτρώδη (Husbands et al., 2007) ενώ μπορούν να δράσουν σαν ομοδιμερή και ετεροδιμερή (Lee et al., 2017). Τέλος στην αλληλουχία προσδέεται ένας bzir μεταγραφικός παράγοντας από αυτούς που προσδέονται στα Gbox (CACGTG). Αυτός ο παράγοντας δρά και επάγεται η έκφραση του με την απόκριση στο ABA (Tomimaga et al., 2009). Οι CAMTA μεταγραφικοί παράγοντες φαίνεται να παίζουν ρόλο στη στα εναρκτήρια στάδια (Yang Y., 2015). Όλα τα παραπάνω (και παρακάτω) επίπεδα έκφρασης έχουν μελετηθεί στο Gifu (στο MG20 μόνο δεδομένα για φυμάτια στις 21 dpi). Ακόμη από τη βιβλιογραφία προέκυψε ότι υπάρχει ένας CCAAT box binding factor χωρίς να υπάρχει γνωστό matrix και ομοίως μία TGA binding αλληλουχία ως απόκριση στο SA [TGACG που βρίσκεται στη ίδια περιοχή με ERN1 (-1525) & CAMTA (-252)] (Mur et al., 2013). Παρά την ύπαρξη ενός FunTFBS σύμφωνα με τον UCSC genome browser δεν εμφανίστηκε μοτίβο σε εκείνη την περιοχή που ο μεταγραφικός του παράγοντας να πληροί τα παραπάνω κριτήρια.

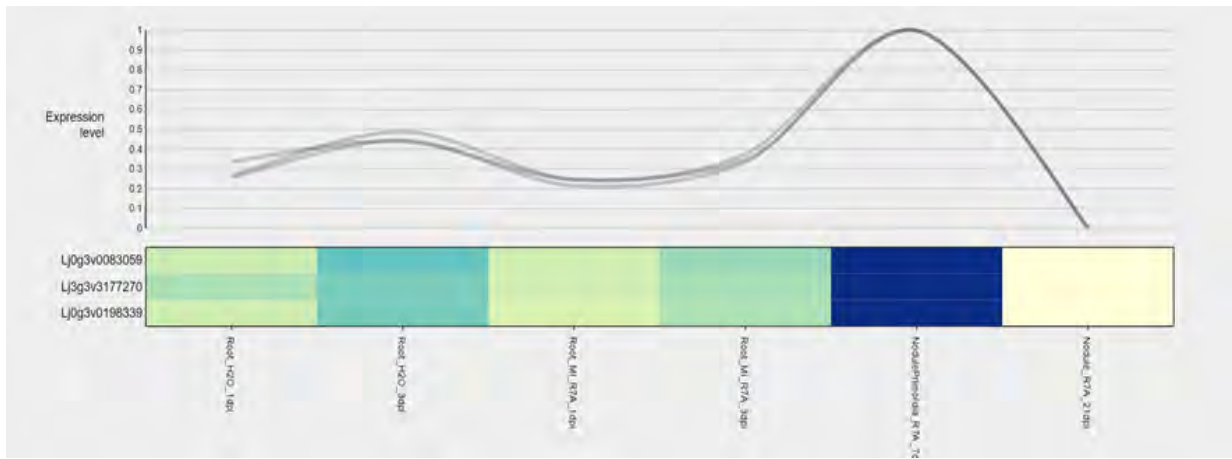


Εικόνα 28: Επίπεδα έκφρασης των παραπάνω πρωτεϊνών σε RNA-seq δεδομένα από Kelly et al., 2017 (πάνω) και microarrays Høglund N et al., 2009 (κάτω) (κανονικοποιημένα αποτελέσματα).



Εικόνα 29: Επίπεδα έκφρασης των παραπάνω πρωτεϊνών σε RNA-seq δεδομένα από Kelly et al., 2017 (πάνω) και microarrays (Høglund N et al., 2009) (κάτω) (κανονικοποιημένα αποτελέσματα).

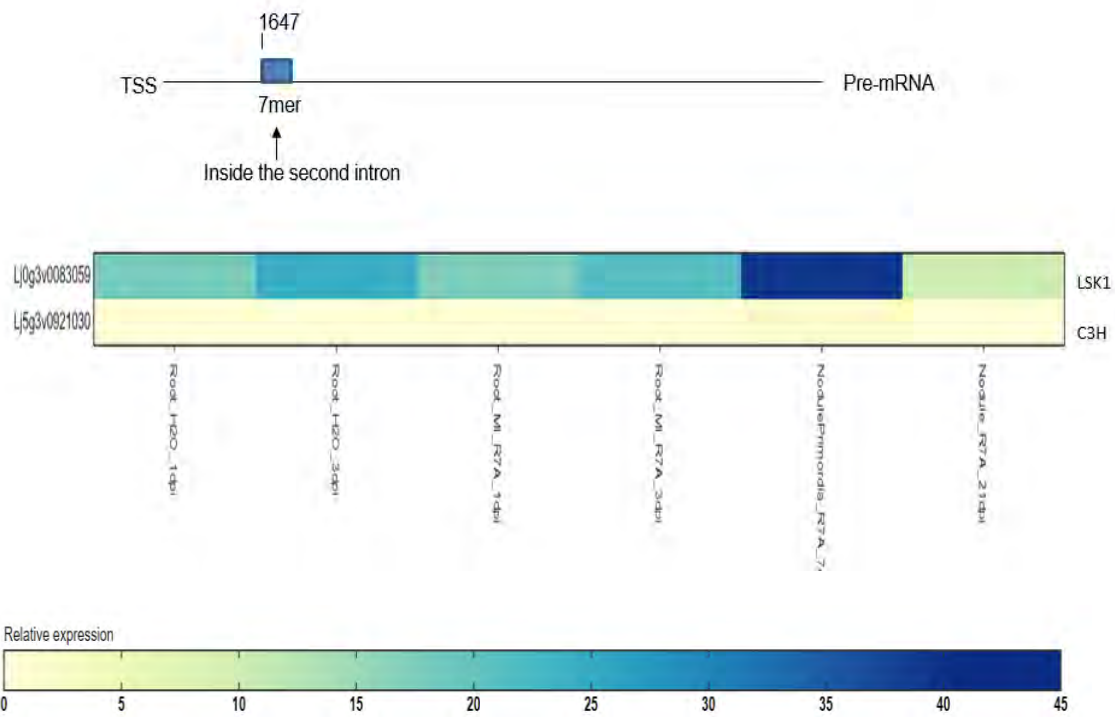
Σύμφωνα με τα παραπάνω heatmaps προκύπτει ότι το γονίδιο Lj0g3v2735050 παρουσιάζει παρόμοια μεταβολή της έκφρασης με το γονίδιο *LSK1* (RNA-seq) και παρόμοιο peak με την κινάση παρουσιάζουν οι ERN1 (RNA-seq, nodule primordia). Τέλος από την ανάλυση του πιο συνεκφρασμένου γονιδίου προέκυψε ότι μια καρβοξυπεπτιδάση της σερίνης (Lj0g3v0198339) φαίνεται να συνεκφράζεται με την *LSK1* μάλιστα από ανάλυση των υποκινητών τους προέκυψε ότι υπάρχει ένα κοινό μοτίβο GCCACCG το οποίο βρίσκεται εσωτερικά του μοτίβου πρόσδεσης για τον ERN1 παράγοντα και σύμφωνα με το Nsite αποτελεί σημείο πρόσδεσης ενός AIB4 παράγοντα σε απόκριση στο ABA στο *Arabidopsis*.



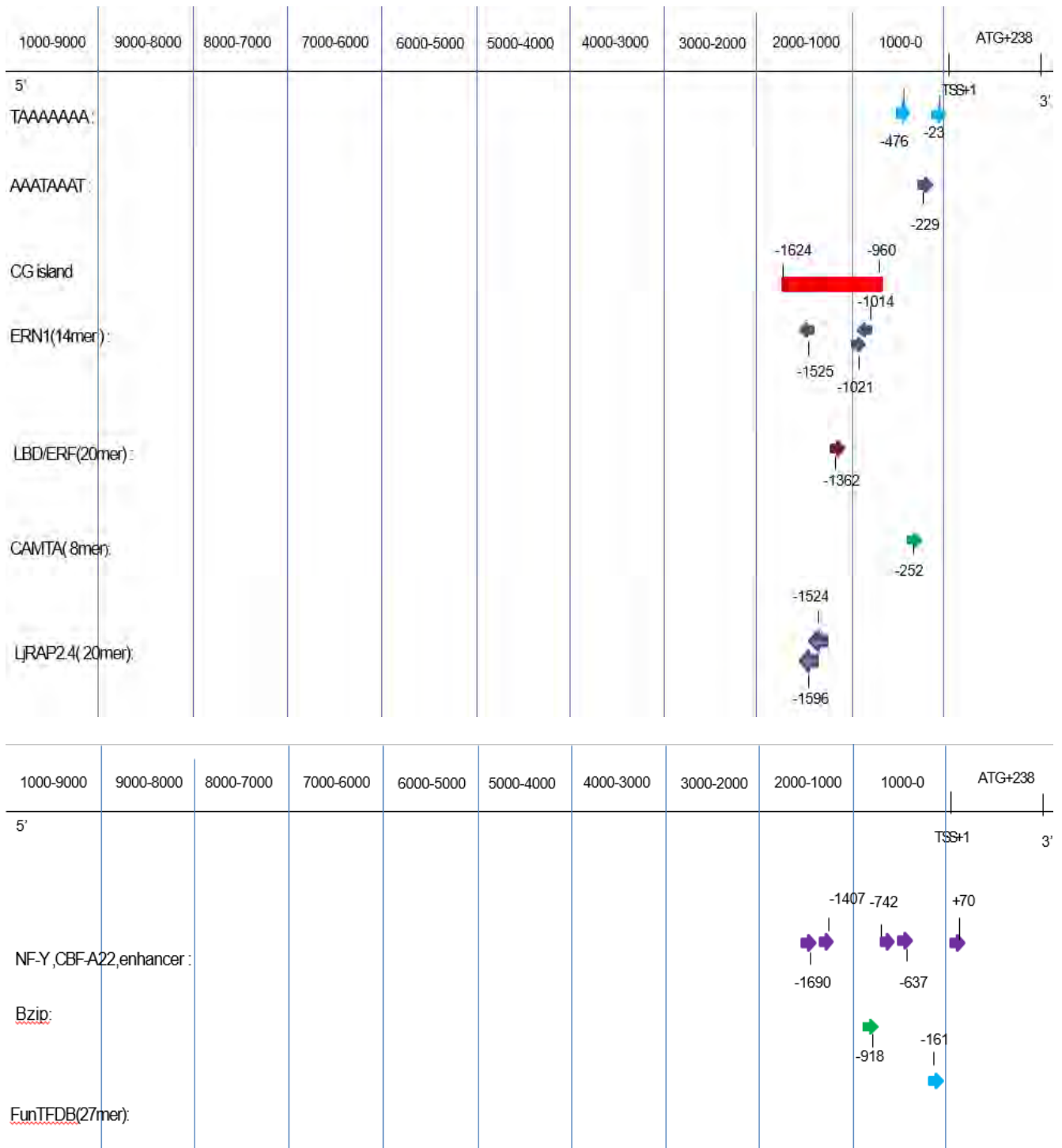
Εικόνα 30 : Επίπεδα έκφρασης των παραπάνω πρωτεϊνών σε RNA-seq δεδομένα από Kelly et al., 2017. (κανονικοποιημένα αποτελέσματα). Το 2^ο γονίδιο στην εικόνα είναι αυτό που βρέθηκε πρώτο στο CORGI.

4.4 RNA TFs, miRNAs ΚΑΙ RIBOSWITCHES

Από την αναζήτηση για miRNAs και riboswitches προέκυψε ότι δεν υπάρχει κάποιο miRNA που να αλληλεπιδρά ή η ύπαρξη ενός riboswitch στο πρώιμο και ώριμο mRNA του MG20 και σίγουρα όχι στο ώριμο μεταγράφο του Gifu (δε γνωρίζουμε την αλληλουχία του πρώιμου μεταγράφου του Gifu). Από την αναζήτηση RNA αλληλοεπιδρώντων μεταγραφικών παραγόντων σε πρώιμο και ώριμο μεταγράφο προέκυψε ότι υπάρχει 1 CH3 μεταγραφικός παράγοντας που προσδέεται στο δεύτερο ιντρόνιο της κινάσης όπως φαίνεται παρακάτω. Παρ' όλα αυτά το αποτέλεσμα δεν μπορεί να θεωρηθεί στατιστικά σημαντικό καθώς η τιμή q value που προέκυψε από το πρόγραμμα είναι απαγορευτική ($q=0,145$ ενώ πρέπει $q < 0,05$).



Εικόνα 31: Επίπεδα έκφρασης της CH3 πρωτεΐνης σε RNA-seq δεδομένα από Kelly et al., 2017 (τυποποιημένα αποτελέσματα).



Εικόνα 32: Ρυθμιστικά στοιχεία του υποκινητή της LSK1 (Gifu). Οι αριθμοί σε παρένθεση δηλώνουν το μέγεθος του μοτίβου, ο αριθμός τη θέση έναρξης του και η κατεύθυνση του βέλους την 5'-3' κατεύθυνση του. Ο enhancer παράγοντας στην εικόνα προσδένεται στην αλληλουχία CCAAT. Το CpG island προέκυψε από την αλληλοεπικάλυψη των 2 προγραμμάτων.

5.ΣΥΖΗΤΗΣΗ

Στην παρούσα πτυχιακή εργασία μελετήθηκαν τα πιθανά ρυθμιστικά στοιχεία του υποκινητή της LSK1. Συγκεκριμένα πραγματοποιήθηκε η ενίσχυση 2090 nt ανοδικά του ATG .Στη συνέχεια ακολούθησε εισαγωγή του σε PGEM-T φορέα και PCAMBIA 0391xc (μετά από κατεργασία του 1391xc). Στη συνέχεια ακολουθήθηκε βιοπληροφορική ανάλυση για την εύρεση πιθανών ρυθμιστικών στοιχείων ψάχνοντας για ρυθμιστικά στοιχεία της βιβλιογραφίας και των μεταγραφικών παραγόντων που συνδέονται στα μονοπάτια του AON και ποια από αυτά τα στοιχεία εμφανίζονται σε υποκινητή που συνεκφράζεται με την LSK1 η επάγει πιθανώς τα early nodulin genes .Από τα αποτελέσματα της ανάλυσης αυτής προέκυψε ότι στη 2 kb αλληλουχία του LSK1 Gifu εμφανίζονται θέσεις πρόσδεσης LBD/ERF παραγόντων σε CG-rich περιοχές, G box αλληλουχίες στις οποίες προσδένεται μεταγραφικός παράγοντας με απόκριση σε ABA/JA, AT-hook, ERN1, LjRAP2.4, enhancer. Ακόμη προέκυψε η ύπαρξη ενός γονιδιωματικού νησιδίου και από την RCC ανάλυση προέκυψε ότι το πιο συνεκφραζόμενο γονίδιο αποτελεί μια καρβοξυπεπτιδάση της σερίνης χωρίς να είναι γνωστό εάν οι 2 πρωτεΐνες αλληλεπιδρούν. Επίσης βρέθηκε ότι οι 2 υποκινητές αυτών των γονιδίων εμφάνιζαν ένα κοινό μοτίβο το οποίο βρίσκεται μέσα στην αλληλουχία πρόσδεσης του ERN1. Παρόλο που συνήθως τα μοτίβα πρόσδεσης μεταγραφικών παραγόντων βρίσκονται συνήθως αρκετά κοντά στο TSS (200-250 nt) ανοδικά ωστόσο δεν θεωρείται παράλογη η θέση τους καθώς όπως δείχνουν δεδομένα από το *Arabidopsis* οι TFs μπορούν να προσδένονται έως -2000 kb (Yu et al., 2016). Συνεπώς αξίζει να ερευνηθούν τυχών επιγενετικές τροποποιήσεις στο CpG island που το κατατάσσουν λειτουργικό η μη σε διάφορα στάδια της φυματογένεσης καθώς και αν στην double CG -rich περιοχή οι TFs επάγουν η καταστέλλουν την έκφραση. Επιπλέον στην αλληλουχία δε βρέθηκαν μεταγραφικοί παράγοντες που επάγονται από αυξίνες, κυτοκινίνες και γιββερελλίνες που να είναι γνωστό ότι καταστέλλουν την έκφραση (καθώς δρουν το χρονικό διάστημα πριν το σχηματισμό του μολυσματικού νηματίου-Tabata S., 2014). Εφόσον έχει κατασκευαστεί ο φορέας 0391xc με LSK1 και GUS μπορεί να γίνει εισαγωγή σε *Nicotiana benthamiana* για να ελεγχθεί εάν αυτό το τμήμα είναι ικανό να επάγει την έκφραση. Στη συνέχεια θα πρέπει να πραγματοποιηθούν deletions στον υποκινητή στις παραπάνω περιοχές για να διαπιστωθεί αν αποτελούν ρυθμιστικά στοιχεία και μπορεί να μελετηθεί ο φαινότυπος από μεταλλαγμένες σειρές LORE1. Το LORE1 αποτελεί ρετρομεταθετό στοιχείο με τυχαίες θέσεις ένθεσης στο γονιδίωμα (Fukai et al., 2012, Małolepszy et al., 2016). Στον υποκινητή της LSK1 υπάρχουν 4 LORE1 ενθέσεις που βρίσκονται στο CpG island και ανοδικά (genome browser tracks-LORE1 insertions). Η εύρεση των μοτίβων αυτών αποτελεί σημαντικό εύρημα για την εύρεση λειτουργικών διαφορών μεταξύ MG20 -Gifu, για την κατανόηση του AON καθώς και της ρύθμισης και δράσης δευτερογενών μεταβολιτών στη φυματογένεση, όπως η λουπεόλη (Delis et al., 2011) η οποία εμφανίζει αντικαρκινικές ιδιότητες (Saleem M., 2009).

6.ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Ahn, S., Huang, C.L., Ozkumur, E., Zhang, X., Chinnala, J., Yalcin, A., Bandyopadhyay, S., Russek, S., Unlü, M.S., DeLisi, C., Irani, R.J. (2012). TATA Binding Proteins Can Recognize Nontraditional DNA Sequences. *Biophysical Journal*, 103(7): 1510-1517.
2. Akhripova, I.R. (1995). Promoter elements in *Drosophila melanogaster* revealed by sequence analysis. *Genetics*, 139(3):1359-69.
3. Ali, A., Hoeflich, K. P. and Woodgett, J. R. (2001). Glycogen synthase kinase-3: properties, functions, and regulation. *Chem. Rev.* 101, 2527-2540.
4. Alon, L. T., Pietrokovski, S., Barkan, S., Avrahami, L., Kaidanovich-Beilin, O., Woodgett, J. R., Barnea, A., and Eldar-Finkelman, H. (2011). Selective loss of glycogen synthase kinase-3 α in birds reveals distinct roles for GSK-3 isozymes in tau phosphorylation. *FEBS Lett.* 585, 1158–1162.
5. Asamizu, E., Simoda, Y., Sato, S. (2008). A Positive Regulatory Role for LjERF1 in the Nodulation Process Is Revealed by Systematic Analysis of Nodule-Associated Transcription Factors of *Lotus japonicus*.
6. Babenko, V.N., Chadaeva, I.V., Orlov, Y.L. (2017). Genomic landscape of CpG rich elements in human. *BMC Evolutional Biology*, 17(Suppl 1):19.
7. Bao, W., Kojima, K.K., Kohani, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, 6:11. *Plant Physiology*, 147(7): 2030-2040.
8. Bartlett, A., O' Malley, R.C., Ecker, J.R. (2017). Mapping genome-wide transcription factor binding sites using DAP-seq. *Nature protocols*, 12(8):1659-1672.
9. Berendzen, K., Killan, J., Wanke, D. (2013). *The Analysis of Regulatory DNA: Current Developments, Knowledge and Application Uncovering Gene Regulation*. Bentham e books.
10. Beurel, E., Grieco, S.F., Jope, R.S. (2015). Glycogen synthase kinase-3 (GSK3): regulation, actions, and diseases. *Pharmacology and Therapeutics*, 148:114-131
11. Bayer, R.G., Stael, S., Rocha, A.G., Mair, A., Vothknecht, U.C., Teige, M. (2012). Chloroplast-localized protein kinases: a step forward towards a complete inventory. *J. Exp. Bot.* 63:1713-1723.
12. Butler, J.E.F., Kadonaga, J.T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes development*, 16:2583-2592.
13. Caetano-Anolles, G., Gresshoff, P. (1991) Plant genetic control of nodulation. *Annual Review of Microbiology*, 45:345–382.
14. Castaño, Z., Gordon-Weeks, P., and Kypta, R. (2010). The neuron-specific isoform of GSK-3 β is required for axon growth. *J. Neurochem.* 13:117–130.
15. Catalanotto, C., Cogoni, C., Zardo, G. (2016). MicroRNA in Control of Gene Expression: An Overview of Nuclear Functions. *International Journal of Molecular Sciences*, 17(10):1712.
16. Cerri, M.R., Wang, Q., Stolz, P., Folgmann, J., Frances, L., Katzer, K., Li X., Heckmann, A.B., Wang T.L., Downie, J.A., Klingl A., de Carvalho-Niebel, F, Xie F, Parniske, M (2017). The ERN1 transcription factor gene is a target of the CCaMK/CYCLOPS complex and controls rhizobial infection in *Lotus japonicus*. *New Phytology*, 215(1):323-337.
17. Chen, J., Hu, Q., Zhang, Y., Lu, C. and Kuang, H. (2013) P-MITE: a database for plant miniature inverted-repeat transposable elements. *Nucleic Acids Res*, doi: 10.1093/nar/gkt1000.
18. Chi-Nga, Chow, Tzong-Yi Lee, Yu-Cheng Hung, Guan-Zhen Li, Kuan-Chieh Tseng, Ya-Hsin Liu, Po-Li Kuo, Han-Qin Zheng, and Wen-Chi Chang "PlantPAN3.0: a new and updated resource for reconstructing transcriptional regulatory networks from ChIP-seq experiments in plants" *Nucleic Acids Res.* 2019.
19. Clouse, S.D. (2011). Brassinosteroids. *Arabidopsis Book*, 9:e0151.
20. Cohen, P. (1982), The role of protein phosphorylation in neural and hormonal control of cellular activity. *Nature* 296:613-620.
21. Dal Santo, S., Stampfl, H., Krasensky, J., Kempa, S., Gibon, Y, Petutschnig, E., Rozhon, W., Heuck, A., Clausen, T., and Jonak, C. (2012). Stress-induced GSK3 regulates the redox stress response by phosphorylating glucose-6-phosphate dehydrogenase in *Arabidopsis*. *Plant Cell* 24:3380–3392.

22. Das, M.K., Dai, H.K. (2007). A survey of DNA motif finding algorithms. *BMC Bioinformatics*, 8(Supp7): S21.
23. Dazzo, F.B., De Bruijn, F.L. (1998). Nodule Organogenesis and Symbiotic Mutants of the Model Legume *Lotus japonicus*. *Molecular Plant-Microbe Interactions*, DOI: 10.1094/MPMI.1998.11.7.684.
24. Deaton, A.M., Webb, S., Kerr, R.W.A., Illingworth, R.S., Guy, J., Andrews, R., Bird, A. (2011). Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome Research* 21(7): 1074–1086.
25. Delis C., Krokida, A., Georgiou, S., Peña-Rodríguez, L.M., Kavroulakis, N., Ioannou, E., Roussis, V., Osbourn, A.E., Papadopoulou, K.K. (2011). Role of lupeol synthase in *Lotus japonicus* nodule formation. *New Phytology*, 189(1):335-346.
26. Desbrosses, G.J., Stougaard, J. (2011). Root nodulation: a paradigm for how plant-microbe symbiosis influences host developmental pathways. *Cell Host and Microbe*, 10(4):348-358.
27. Doble, B.W., Woodgett, J.R. (2003). GSK-3: tricks of the trade for a multi-tasking kinase. *Journal of Cell Science*, 116:1175-1186.
28. Downie, J.A. (2015). Legume nodulation. *Current Biology*, 24(5): R184
29. Flatters, D., Lavery, R. (1998). Sequence-dependent dynamics of TATA-Box binding sites. *Biophysics Journal*, 75(1):372-381.
30. Fu, Xiang-Dong Ares Jr, Manuel (2014). Context-dependent control of alternative splicing by RNA-binding proteins. *Nature Reviews Genetics*, 15 (10):689–701.
31. Fujimoto, S., Matsunaga, S., Yonemura, M., Uchiyama, S., Azuma, T., Fukui, K. (2004). Identification of a novel plant MAR DNA binding protein localized on chromosomal surfaces. *Plant Molecular Biology*, 56(2):225-239.
32. Fukai et al. (2012) Establishment of a *Lotus japonicus* gene tagging population using the exon-targeting endogenous retrotransposon *LORE1*. *Plant J.*, 69(4). doi:10.1111/j.1365-313X.2011.04826.x.
33. Garagounis, C., Tsikou, D., Plitsi, P.K., Psarrakou, I.S., Avramidou, M., Stedel, C., Anagnostou, M., Georgopoulou ME, Papadopoulou KK (2018). *Lotus SHAGGY-like kinase 1 is required to suppress nodulation in Lotus japonicus*. *The Plant Journal*, doi: 10.1111/tpj.14207.
34. Glisovic, T., Bachorik, J.L., Yong, J., Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. *FEBS PRESS*, 582(14):1977-86.
35. Goldberg, T., Hecht, M., Burkhard, R. (2014). LocTree3 prediction of localization. *Nucleic acid research*, 42(web server issue):W350-W355.
36. Grant, C.E., Bailey, T.L., Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics*, 27(7):1017-8.
37. Gudesblat, G.E., Russinova, E. (2011). Plants grow on brassinosteroids. *Current Opinion in Plant Biology*, 14(5): 530-537.
38. Guguloth, Mahesh Kumar, Praveen Mamidala, Appa Rao, Podile (2009). Regulation of Polygalacturonase-inhibitory Proteins in Plants Is Highly Dependent on Stress and Light Responsive Elements, 2(6):238-249.
39. Guinel, F.C. (2015). Ethylene, a Hormone at the Center-Stage of Nodulation. *Frontiers in Plant Science*, 6:1121.
40. Gupta, A., Swati, D. (2016). Exploring Riboswitches in Archaeal Metagenomes. *Journal of RNAi and Gene Silencing*, 12:536-543.
41. He, J.X., Gendron, J.M., Yang, Y., Li, J., Wang, Z.Y. (2002). The GSK3-like kinase BIN2 phosphorylates and destabilizes BZR1, a positive regulator of the brassinosteroid signaling pathway in *Arabidopsis*. *PNAS*, 99(15):10185-10190.

42. Heckmann, A.B., Lombardo, F., Miwa, H., Perry, J.A., Bunnewell S, (2006) Lotus japonicus nodulation requires two GRAS domain regulators, one of which is functionally conserved in a non-legume. *Plant Physiol.* 142:1739–1750.
43. Hirsch S, Kim J, Mun A, Heckmann, A.B., Downie, J.A., Oldroyd, G.E. (2009). GRAS proteins form a DNA binding complex to induce gene expression during nodulation signaling in *Medicago truncatula*. *Plant Cell*, 21: 545–557.
44. Husbands, A., Bell, E.M., Shuai, B., Smith, H.M., Springer, P.S. (2007). LATERAL ORGAN BOUNDARIES defines a new family of DNA-binding transcription factors and can interact with specific bHLH proteins. *Nucleic acid research*, 35(19):6663-6671.
45. Hogsland, N. et al (2009). Dissection of Symbiosis and Organ Development by Integrated Transcriptome Analysis of Lotus japonicus Mutant and Wild-Type Plants. *OPEN PLOS*, 4(8):e6556.
46. Holligan, D., Zhang, X., Jiang, N., Pritham, E.J. & Wessler, S.R. (2006). The transposable element landscape of the model legume Lotus japonicus. *Genetics*, 174(4):2215-2228.
47. Hooper, J.E., Scott, M.P. (2005). Communicating with Hedgehogs. *Nat Rev Mol Cell Biol.* 6(4):306-317.
48. Horikoshi, M. et al (1992). Transcription factor TFIID induces DNA bending upon binding to the TATA element. *Proceedings of the National Academy of Sciences*, 89 (3):1060-1064.
49. Imaizumi-Anraku, H., Takeda, N., Charpentier, M., Perry, J., Miwa, H., Umehara, Y., Kouchi, H., Murakami, Y., Mulder, L., Vickers, K., Pike, J., Downie, J.A., Wang, T., Sato, S., Asamizu, E., Tabata, S., Yoshikawa, M., Murooka, Y., Wu, G.J., Kawaguchi, M., Kawasaki, S., Parniske, M., Hayashi, M. (2004). Plastid proteins crucial for symbiotic fungal and bacterial entry into plant roots. *Nature* 433(7025):527-531.
50. Jin, J.P., Tian, F., Yang, D.C., Meng, Y.Q., Kong, L., Luo, J.C. and Gao, G. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research*, 45(D1):D1040-D1045.
51. Jonak, C., Hirt, H. (2002). Glycogen synthase kinase 3/SHAGGY-like kinases in plants: an emerging family with novel functions. *Trends in Plant Science*, 7(10):457-461.
52. Juven-Gershon, T., & Kadonaga, J.T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Developmental Biology*, 15:225–229.
53. Kameshita, I., Nishida, T., Nakamura, S., Sugiyama, Y., Sueyoshi, N., Umehara, Y., Nomura, M., Tajima, S. (2005). Expression Cloning of a Variety of Novel Protein Kinases in Lotus japonicus. *J. Biochem*, 137:33–39.
54. Kanamori, N., Madsen, L.H., Radutoiu, S., Frantescu, M., Quistgaard, E.M., (2006) A nucleoporin is required for induction of Ca²⁺ spiking in legume nodule development and essential for rhizobial and fungal symbiosis. *Proc Natl Acad Sci U S A*, 103(2):359–364.
55. Kawaguchi, M., Imaizumi-Anraku, H., Koiwa, H., Niwa, S., Ikuta, A., Syono, K., Akao, S. (2002). Root, root hair, and symbiotic mutants of the model legume Lotus japonicus. *Molecular Plant Microbe Interactions*, 15(1):17-26.
56. Kawaharada, Y., James, E.K., Kelly, S., Sandal, N., Stougaard, J. (2017). The Ethylene Responsive Factor Required for Nodulation 1 (ERN1) Transcription Factor Is Required for Infection-Thread Formation in Lotus japonicus. *Molecular Plant Microbe Interactions* 30(3):194-204.
57. Kempa, S., Rozhon, W., Samaj, J., Erban, A., Baluska, F., Becker, T., Haselmayer, J., Schleiff, E., Kopka, J., Hirt, H., et al (2007). A plastid-localized glycogen synthase kinase 3 modulates stress tolerance and carbohydrate metabolism, 49:1076-1090.
58. Kelly, S., Mun, T., Andersen, S.U. (2018). Distinct Lotus japonicus Transcriptomic Responses to a Spectrum of Bacteria Ranging From Symbiotic to Pathogenic. *Frontiers in Plant Sciences*, 9:1218.
59. Kim, T.W., and Wang, Z.Y. (2010). Brassinosteroid signal transduction from receptor kinases to transcription factors. *Annu. Rev. Plant Biol.* 61:681–704.

60. Kiran, K. et al (2006). The TATA-Box Sequence in the Basal Promoter Contributes to Determining Light-Dependent Gene Expression in Plants. *Plant Physiology*, 142(1):364–376.
61. Kouchi, H., Anraku, H.I., Hayashi, M., Kawaguchi, M. (2010). How Many Peas in a Pod? Legume Genes Responsible for Mutualistic Symbioses Underground. *Plant and Cell Physiology* 51(9):1381-1397.
62. Kozomara, A., Birgaonu, M., Jones, S.G. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Research*, 47(D1):D155-D162.
63. Kutach, A.K., Kadonaga, J.T. (2000). The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. *Molecular and Cellular Biology*, 20(13):4754-4764.
64. Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., Ebright, R.H. (1998). New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes & Development*, 12 (1):34–44.
65. Lau, K.F., Miller, C.C., Anderton, B.H., Shaw, P.C. (1999). Expression analysis of glycogen synthase kinase-3 in human tissues *J. Pept. Res.*, 54:85-91.
66. Lee, H.W., Kang, N.Y., Pandey, S.K., Cho, C., Lee, S.H., Kim, J. (2017). Dimerization in LBD16 and LBD18 Transcription Factors Is Critical for Lateral Root Formation. *Plant Physiology*, 174(1):301-311.
67. Levy, J., Bres, C., Geurts, R., Chalhoub, B., Kulikova, O., Duc, G., Journet, E.P., Ane, J.M., Lauber, E., Bisseling T (2004). A putative Ca²⁺ and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. *Science*, 303:1361-1364.
68. Liesheke, F., Daudu, D., Dugé de Bernonville, T. (2018). Ranking genome-wide correlation measurements improves microarray and RNA-seq based global and targeted co-expression networks. *Scientific Reports*, 8(1):10885.
69. Lis. M., Walther, D. (2016). The orientation of transcription factor binding site motifs in gene promoter regions: does it matter? *BMC Genomics*, 17:185.
70. Logan, C.Y., Nusse, R. (2004). The Wnt signaling pathway in development and disease. *Annu. Rev. Cell Dev. Biol.*, 20:781-810.
71. Loganantharaj, R., Karim, M.E., Lakhota, A. (2006). Discriminating TATA box from putative TATA boxes in plant genome. *International Journal of Bioinformatics Research and Applications*, 2(1):36-51.
72. B.T. MacDonald, K. Tamai, X. He (2009). Wnt/ β -catenin signaling: components, mechanisms, and diseases *Dev. Cell*, 17:9-26.
73. Madsen, LH, Tirichine L, Jurkiewicz, A, Sullivan, JT, Heckmann AB, Bek AS, Ronson, CW, James, EK, and Stougaard, J (2010). The molecular network governing nodule organogenesis and infection in the model legume *Lotus japonicus*. *Nature Communication*, 1,10.
74. Małolepszy et al. (2016). The LORE1 insertion mutant resource. *Plant J.* doi:10.1111/tpj.13243.
75. Martínez-Romero, E. and Caballero-Mellado, J. (1996). *Rhizobium* phylogenies and bacterial genetic diversity. *Crit Rev Plant Sci.* 15(2):113-140.
76. Medvedeva, Y.A., Frindman, M.V., Oparina, N.J., Malko, D.B., Ermakova, E.O., Kulakovosky, I.V., Heinzl, A., Makeev, V.J. (2010). Intergenic, gene terminal, and intragenic CpG islands in the human genome *BMC Genomics*, 11:48.
77. Metcalf, C.E., Wassarman, D.A. (2006). DNA binding properties of TAF1 isoforms with two AT-hooks. *Journal Of Biological Chemistry* 281(40):30015-30023.
78. Mikrovitch, J., Mirault, M.E., Laemmli, U.K. (1984). Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell*, 39(1): 223-232.
79. Miyake, K., Ito, T., Senda, M., Ishikawa, R., Harada, T., Niizeki, M., Akada, S. (2004). R2R3-MYB transcription factors from *Lotus corniculatus* var. *japonicus* differentially regulated in roots and shoots in response to nitrogen starvation. *Lotus Newsletter*, 34: 54-59.

80. Mukai, F., Ishiguro, K., Sano, Y. and Fujita, S. C. (2002). Alternative splicing isoform of tau protein kinase I/glycogen synthase kinase 3beta. *Journal Neurochem.* 81, 1073-1083.
81. Mukherjee, S., Retwitzer, M.D., Sengupta, S. (2018). Phylogenomic and comparative analysis of the distribution and regulatory patterns of TPP riboswitches in fungi. *Scientific Reports*, 8: 5563.
82. Mukherjee, S. and Sengupta, S. (2016). Riboswitch Scanner: An efficient pHMM-based web-server to detect riboswitches in genomic sequences. *Bioinformatics* 32 (5):776-778.
83. Mun, T., Bachmann, A., Gupta, V., Stougaard, J., Andersen, S.U. (2016). *Lotus* Base: An integrated information portal for the model legume *Lotus japonicus*. *Sci. Rep.* doi:10.1038/srep39447.
84. Mur, L.A.J., Prats, E., Pierre Sandra, Hall, A.M., Hebelstrup Kim, H. (2013). Integrating nitric oxide into salicylic acid and jasmonic acid/ethylene plant defense pathways. *Frontiers in Plant Science*, 4:215.
85. Mylona, P., Pawlowski, K., Bisseling, T. (1995) Symbiotic Nitrogen Fixation. *The Plant Cell*, Vol. 7, 869-885.
86. Nakagawa, T., Kawaguchi, M. (2006). Shoot-applied MeJA suppresses root nodulation in *Lotus japonicus*. *Plant Cell Physiology*, 47(1): 176-180.
87. Nakamura, M., Tsunoda, T., Obokata, J. (2002) Photosynthesis nuclear genes generally lack TATA-boxes: a tobacco photosystem I gene responds to light through an initiator. *Plant J*, 29: 1–10.
88. Nakato, R. and Shirahige, K. (2017). Recent advances in ChIP-seq analysis: from quality management to whole-genome annotation. *Briefings in Bioinformatics*, 18(2):279-290.
89. Newcomb, W. (1981). Nodule morphogenesis and differentiation. In *Biology of the Rhizobiaceae*, K.L. Giles, and A.G. Atherly, eds (New York: Academic Press), pp. 247-297.
90. Niwa, S., Kawaguchi, M., Imazumi-Anraku, H., Chechetka, S.A., Ishizaka, M., Ikuta, A., Kouchi, H. Responses of a model legume *Lotus japonicus* to lipochitin oligosaccharide nodulation factors purified from *Mesorhizobium loti* JRL501. *Plant Molecular Microbe Interactions*, 14(7): 848-56.
91. Noble, W.S. (2009). How does multiple testing correction work? *Nature Biotechnology* 27:1135–1137.
92. Nudler, E., Mironov, A.S. (2004). The riboswitch control of bacterial metabolism. *Trends in Biochemical Sciences*, 29(1): 11-17.
93. Nüsslein-Volhard, C., Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*. *Nature*, 287(5785):795-801.
94. Obayashi, T., Kinoshita, K. (2009). Rank of Correlation Coefficient as a Comparable Measure for Biological Significance of Gene Coexpression. *DNA Research*, 16(5): 249-260.
95. Okonechnikov, K., Golosova, O., Fursov, M., the UGENE team (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, 28: 1166-1167.
96. Oldroyd, G.E.D., Downie, J.A. (2004). Calcium, kinases and nodulation signalling in legumes. *Nat. Rev. Mol. Cell Biol*, 5:566–576.
97. Pajuelo, E., Stougaard, J. (2005). *Lotus japonicus* handbook. Marquez AJ, chapter 1, pp 3-24.
98. Porto, M. S. et al (2014). *Plant Promoters: An Approach of Structure and Function*. *Molecular Biotechnology*, 56:38-49.
99. Radutoiu, S., Madsen, L.H., Madsen, E.B., Felle, H.H., Umehara, Y., Grønlund, M., Sato, S., Nakamura, Y., Tabata, S., Sandal, N., Stougaard, J. (2003). Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature*, 425(6958):585-592.
100. Roeder, R. G. (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends in Biochemical Sciences*, 21(9):327-335.
101. Saidi, Y., Hearn, T.J., Coates, J.C. (2012). Function and evolution of 'green' GSK3/Shaggy-like kinases. *Trends in Plant Science*, 17(1):39-46.
102. Saito, K., Yoshikawa, M., Yano, K., Miwa, H., Uchida, H., (2007) NUCLEOPORIN85 is required for calcium spiking, fungal and bacterial symbioses, and seed production in *Lotus japonicus*. *Plant Cell*, 19(2):610–624.

103. Samad, F.A., Sajad, M., Nazaruiddin, N., Fauzi, I.A., Murad, A.M.A., Zaimri, Z., Ismail, I. (2017). MicroRNA and Transcription Factor: Key Players in Plant Regulatory Network. *Frontiers in Plant Science*, 8:565.
104. Saleem, M. (2009). Lupeol, a novel anti-inflammatory and anti-cancer dietary triterpene. *Cancer letters* 285(2):109-115.
105. Savojardo, C., Martelli, P.L., Fariselli, P., Profiti, G., Casadio, R. (2018) BUSCA: an integrative web server to predict subcellular localization of proteins. *Nucleic Acids Research* 46(W1), W459-W466.
106. Sawant, S.V., Kiran, K., Singh, P.K., Tuli, R. (2001) Sequence architecture downstream of the initiator codon enhances gene expression and protein stability in plants. *Plant Physiol*, 126:1630–1636.
107. Schauser, L., Roussis, A., Stiller, J., Stougaard, J. (1999). A plant regulator controlling development of symbiotic root nodules. *Nature*, 402(6758):191-195.
108. Serna-Sanz, A., Parniske, M., Peck, S.C. (2011). Phosphoproteome analysis of *Lotus japonicus* roots reveals shared and distinct components of symbiosis and defense. *Molecular Plant Microbe Interactions*, 24(8):932-937.
109. Shahmuradov, I., Solovyev, V. (2015) Nsite, NsiteH and NsiteM Computer Tools for Studying Transcription Regulatory Elements. *Bioinformatics*, doi: 10.1093.
110. Smale, S.T., Jain, A., Emami, K.H., LO, K., Garraway, I.P. (1998). The Initiator Element: A Paradigm for Core Promoter Heterogeneity within Metazoan Protein-coding Genes. *Cold Spring Harbor Symposia on Quantitative Biology*, 63:21-32.
111. Smale, S.T., Kadonaga, J.T. (2003) The RNA polymerase II core promoter. *Annual Reviews Biochem*, 72:449-479.
112. Soutar, M.P., Kim, W.Y., Williamson, R., Pegg, M., Hastie, C.J., McLauchlan, H., Snider, W.D., Gordon-Wks, P.R., Sutherland C. (2010). Evidence that glycogen synthase kinase-3 isoforms have distinct substrate preference in the brain. *J Neurochem*, 115: 974–983.
113. Stacey, G., Bickley, C., Alvin, M., Soto, M.J. (2006). Effects of Endogenous Salicylic Acid on Nodulation in the Model Legumes *Lotus japonicus* and *Medicago truncatula*. *Plant Physiology* 141(4):1473-1481.
114. Suzuki, A., Hara, H., Kinoue, T., Arima, S. (2008). Split-root study of autoregulation of nodulation in the model legume *Lotus japonicus*. *Journal of Plant Research* 121(2):245-249.
115. Suzuki, A. et al (2011). *Lotus japonicus* nodulation is photomorphogenetically controlled by sensing the red/far red (R/FR) ratio through jasmonic acid (JA) signaling. *PNAS*, 108 (40):16837-16842.
116. Suzuki, Y. et al (2001). Identification and characterization of the potential promoter regions of 1031 kinds of human genes. *Genome Res*, 11:677-84.
117. Tabata, S. (2014). *The Lotus Japonicus Genome*. Springer.
118. Testori, T., Caizzi, Cutrupi, S., Friad, O., Bortoli, M.D., Cora, D., Caselle, M. (2012). *BMC Genomics*, 13:400.
119. Thomas, M.C., Chiang, C.M. (2006). The general transcription machinery and general cofactors. *Critical Reviews in Biochemistry and Molecular Biology*, 41(3):105-178.
120. Tominaga, A., Nagata, M., Futsuki, K., Abe, H., Uchiumi, T., Abe, M., Kucho, K.I., Hashiguchi, M., Akashi R, Hirsh AM, Arima S, Suzuki A (2009). Enhanced Nodulation and Nitrogen Fixation in the Abscisic Acid Low-Sensitive Mutant enhanced nitrogen fixation1 of *Lotus japonicus*. *Plant Physiology*, <https://doi.org/10.1104/pp.109.142638>.
121. Tropp, B.E. (2012). *Principles of Molecular Biology*. Jones & Bartlett Publishers.
122. Tsikou, D., Yan, Z., Holt, D.B., Abel, N.B., Reid, D.E., Madsen, L.H., Bhasin, H., Sexauer, M., Stougaard, J., Markmann, K. (2018). Systemic control of legume susceptibility to rhizobial infection by a mobile microRNA. *Science*, 362(6411):233-236.
123. Vitreschak, A.G., Rodionov, D.A., Mironov, A.A., Gelfand, M.S. (2004). Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends in Genetics*, 20(1):44-50.

124. Wachter, A., Tunc-Ozdemir, M., Grove, B.C., Green, P.J., Shintani, D.K., Breaker, R.R. (2007). Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. *Plant Cell*, 19(11):3437-3450.
125. Wang, Q.i., Yang, S., Zhu, H. (2017). Host-secreted antimicrobial peptide enforces symbiotic selectivity in *Medicago truncatula*. *PNAS*, 114(26):6854-6859.
126. Woodgett, J.R. (1990). Molecular cloning and expression of glycogen synthase kinase-3/factor A. *The EMBO Journal* 9(8): 2431–2438.
127. Wood-Kaczmar, A., Kraus, M., Ishiguro, K., Philpott, K. L., and Gordon-Weeks, P. R. (2009). An alternatively spliced form of glycogen synthase kinase-3beta is targeted to growing neurites and growth cones. *Mol. Cell. Neurosci.* 42, 184–194.
128. Wrzaczek, M., Rozhon, W., and Jonak, C. (2007). A proteasomeregulated glycogen synthase kinase-3 modulates disease response in plants. *J. Biol. Chem*, 282:5249–5255.
129. Yang, C.J., Zhang, C., Lu, Y.N., Jin, J.Q., Wang, X.L. (2011). The mechanisms of brassinosteroids' action: from signal transduction to plant development. *Molecular Plant*, 4(4):588-600.
130. Yang, Y., Sun, T., Shen, C. (2015). Genome-wide identification of CAMTA gene family members in *Medicago truncatula* and their expression during root nodule symbiosis and hormone treatments. *Frontiers in Plant Science*, 6:459.
131. Yano, K., Yoshida, S., Muller, J., Singh, S., Banba, M., Vickers, K., Markmann, K., White, C., Schuller, B., Sato, S. (2008) CYCLOPS, a mediator of symbiotic intracellular accommodation. *Proc Natl Acad Sci USA*, 105:20540-20545.
132. Yoo, M.J., Albert, V.A., Soltis, P.S., Soltis, D.E. (2006). Phylogenetic diversification of glycogen synthase kinase 3/SHAGGY-like kinase genes in plants. *BMC Plant Biology*, 6:3.
133. Youn, J.H., Kim, T.W., Kim, T.W. (2015). Structural and Functional Characterization of Arabidopsis GSK3-like Kinase AtSK12. *Molecules and Cells*, 36(6):564-570.
134. Youn, J.H., Kim, T.W. (2015). Functional Insights of Plant GSK3-like Kinases: Multi-Taskers in Diverse Cellular Signal Transduction Pathways. *Molecular Plant*, 8(4):552-565.
135. Yu, C.P., Lin, J.J., Li, W.H (2016). Positional distribution of transcription factor binding sites in *Arabidopsis thaliana*. *Scientific Reports*, 6: 25164.
136. Yuko Tsutsumi-Ishii, Takeshi Hasebe and Isao Nagaoka (2000). Role of CCAAT/Enhancer-Binding Protein Site in Transcription of Human Neutrophil Peptide-1 and -3 Defensin Genes. *The Journal of Immunology*, 164(6): 3264-3273.
137. Yun, J., Kim, Y.S., Park, C.M. (2012). The AT-hook Motif-containing Protein AHL22 Regulates Flowering Initiation by Modifying FLOWERING LOCUS T Chromatin in Arabidopsis. *Journal of Biological Chemistry*, 287(19): 15307-15316.

