

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΥΓΕΙΑΣ
ΤΜΗΜΑ ΒΙΟΧΗΜΕΙΑΣ ΚΑΙ ΒΙΟΤΕΧΝΟΛΟΓΙΑΣ

ΑΝΑΠΤΥΞΗ
ΒΙΟΠΛΗΡΟΦΟΡΙΚΟΥ
ΕΡΓΑΛΕΙΟΥ ΓΙΑ ΤΟΝ
ΕΝΤΟΠΙΣΜΟ ΚΑΙ ΤΗΝ
ΕΞΕΛΙΚΤΙΚΗ
ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΤΩΝ
ΕΥΚΑΡΥΩΤΙΚΩΝ
ΜΕΤΑΦΟΡΕΩΝ ΣΑΚΧΑΡΩΝ
ΤΗΣ ΥΠΕΡΟΙΚΟΓΕΝΕΙΑΣ
MFS

ΝΙΚΟΛΑΙΔΗΣ ΜΑΡΙΟΣ

UNIVERSITY OF THESSALY
FACULTY OF LIFE SCIENCES
DEPARTMENT OF BIOCHEMISTRY AND BIOTECHNOLOGY

**DEVELOPMENT OF A
BIOINFORMATICS
TOOL FOR THE
IDENTIFICATION AND
EVOLUTIONARY
CLASSIFICATION OF
THE EUKARYOTIC MFS
SUPERFAMILY SUGAR
TRANSPORTERS**

Η παρούσα διπλωματική εργασία εκπονήθηκε στο εργαστήριο Βιοπληροφορικής, του τμήματος Βιοχημείας και Βιοτεχνολογίας, Σχολής Επιστημών Υγείας του Πανεπιστημίου Θεσσαλίας (ΠΘ).

Υπεύθυνος καθηγητής:

Αμούτζιας Γρηγόριος, Επίκουρος Καθηγητής Βιοπληροφορικής στη Γενωμική, Τμήμα Βιοχημείας και Βιοτεχνολογίας, Π.Θ.

Τριμελής επιτροπή:

- Αμούτζιας Γρηγόριος, Επίκουρος Καθηγητής Βιοπληροφορικής στη Γενωμική, Τμήμα Βιοχημείας και Βιοτεχνολογίας, Π.Θ.
- Φλεμετάκης Εμμανουήλ, Αναπληρωτής Καθηγητής Γεωπονικής Βιοτεχνολογίας, Γεωπονικό Πανεπιστήμιο Αθηνών
- Φριλίγγος Ευστάθιος, Καθηγητής Βιολογικής Χημείας, Τμήμα Ιατρικής, Πανεπιστήμιο Ιωαννίνων

ΕΥΧΑΡΙΣΤΙΕΣ

Αρχικά, θα ήθελα να ευχαριστήσω ιδιαίτέρως τον υπεύθυνο καθηγητή μου τον κ. Αμούτζια Γρηγόριο για όλη τη βοήθεια που μου παρείχε, για όλα τα αντικείμενα που με δίδαξε και όλες τις συμβουλές του καθ'όλη τη διάρκεια της διπλωματικής εργασίας. Επίσης θα ήθελα να ευχαριστήσω όλα τα μέλη του εργαστηρίου για τη βοήθεια τους. Τέλος, θα ήθελα να ευχαριστήσω τους κοντινούς μου ανθρώπους για την υποστήριξη τους.

Περιεχόμενα

ΠΕΡΙΛΗΨΗ	7
ABSTRACT	8
1 ΕΙΣΑΓΩΓΗ.....	9
1.1 Μεταφορά μικρών μορίων διαμέσου της κυτταρικής μεμβράνης9	
1.1.1 Παθητική διάχυση.....	10
1.1.2 Διευκολυνόμενη διάχυση.....	10
1.1.3 Ενεργητική μεταφορά.....	13
1.2 Μεταφορείς σακχάρων.....	15
1.2.1 Sodium – Glucose Transporters SGLT.....	16
1.2.2 SWEET transporters.....	18
1.2.3 MFS sugar transporters - Glucose transporters (GLUTs)19	
1.3 Δομή των μεταφορέων σακχάρων	20
1.4 Η Major Facilitator Superfamily	23
1.5 Εξελικτικές αναλύσεις των μεταφορέων σακχάρων της MFS	24
2 ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ.....	26
2.1 Λογισμικό	26
2.1.1 Linux Ubuntu	26
2.1.2 PERL 5.26	26
2.1.3 Python 3.6.....	26
2.1.4 MEME Suite	26
2.1.5 HMMER 3.1	26
2.1.6 ETE 3.....	27
2.1.7 Seaview 4.6.....	27
2.1.8 Interactive Tree of Life (iTOL)	27
2.1.9 Usearch	28
2.2 Λήψη πρωτεωμάτων.....	28
3 ΑΠΟΤΕΛΕΣΜΑΤΑ	30
3.1 Απομόνωση ευκαρυωτικών μεταφορέων σακχάρων της MFS	30
3.2 Δημιουργία MEME Motifs των ευκαρυωτικών ακολουθιών	30
3.3 Αναγνώριση των προκαρυωτικών μεταφορέων σακχάρων MFS	30
3.4 Ανίχνευση εξελικτικών ομάδων στους μεταφορείς.....	32
3.5 Κατανομή των μεταφορέων σακχάρων MFS στους οργανισμούς	32
3.6 Εξελικτικές ομάδες των μεταφορέων σακχάρων.....	33

3.7	Εργαλείο ανίχνευσης και ταξινόμησης των μεταφορέων σακχάρων	39
3.7.1	Σειρά διεργασιών (workflow) του εργαλείου	39
3.7.2	Ανάπτυξη ως διαδικτυακός server	41
4	ΣΥΖΗΤΗΣΗ	43
	ΒΙΒΛΙΟΓΡΑΦΙΑ	45

ΠΕΡΙΛΗΨΗ

Τα σάκχαρα αποτελούν την κύρια πηγή ενέργειας όλων των οργανισμών και για αυτό οι πρωτεΐνες μεταφορείς τους είναι από τις σημαντικότερες πρωτεΐνες φορείς στα κύτταρα. Οι μεταφορείς σακχάρων κατατάσσονται σε τρεις μεγάλες κατηγορίες, τους μεταφορείς σακχάρων που ανήκουν στη Major Facilitator Superfamily που είναι και η μεγαλύτερη κατηγορία εκ των τριών, τους συμμεταφορείς νατρίου-γλυκόζης SGLT και τους SWEETs. Οι μεταφορείς σακχάρων της Major Facilitator Superfamily έχουν προσελκύσει το ενδιαφέρον αρκετών ερευνητικών ομάδων λόγω της εξέχουσας σημασίας τους, ωστόσο μέχρι σήμερα δεν υπάρχει μια εκτεταμένη φυλογένεση που να συμπεριλαμβάνει αντιπροσωπευτικούς οργανισμούς από όλα τα βασίλεια. Σκοπός της εργασίας μας ήταν να πραγματοποιήσουμε φυλογενετική ανάλυση των συγκεκριμένων μεταφορέων, ώστε να διακρίνουμε την ύπαρξη εξελικτικών ομάδων και να βρούμε συντηρημένα μοτίβα στις ακολουθίες των συγκεκριμένων μεταφορέων με απώτερο σκοπό τη δημιουργία ενός βιοπληροφορικού εργαλείου το οποίο έχει την ικανότητα να ανιχνεύει και να ταξινομεί τους μεταφορείς σακχάρων της MFS.

ABSTRACT

Sugars compose the main source of energy of all organisms and therefore their transport proteins are among the most important proteins in the cells. Sugar transporters are divided in three major categories, the sugar transporters belonging to the Major Facilitator Superfamily, the largest category of the three, the SGLT sodium-glucose co-transporters and the SWEETs. Sugar transporters of the Major Facilitator Superfamily have been the focus of several research groups because of their importance, but to date there is no extensive phylogenetic analysis involving representative organisms from all kingdoms of life. The aim of our work was to perform a phylogenetic analysis of the MFS transporters, in order to identify the various evolutionary groups and to find conserved motifs in the sequences of these carriers. The final goal is to create a bioinformatics tool that has the ability to detect and classify the MFS sugar transporters.

1 ΕΙΣΑΓΩΓΗ

1.1 Μεταφορά μικρών μορίων διαμέσου της κυτταρικής μεμβράνης

Κάθε κύτταρο διαθέτει μια λιπιδιακή διπλοστιβάδα φραγμό από το εξωτερικό περιβάλλον. Μικρά μόρια και λιπόφιλες ουσίες έχουν την ικανότητα να διαχέονται ελεύθερα – παθητικά, διαμέσου της κυτταρικής μεμβράνης Figure 1. Ωστόσο, για τα υδρόφιλα μόρια, η μεταφορά συστατικών από το περιβάλλον στο εσωτερικό του κυτάρου και αντιστρόφως είναι μια άκρως ελεγχόμενη διαδικασία που καταλύεται (όχι παθητικά) από εξειδικευμένες πρωτεΐνες, που ονομάζονται πρωτεΐνες μεταφορείς Figure 1 και συγκροτούν περίπου το 20-30% των γονιδίων του ανθρώπου (Cooper, Geoffrey M., and Robert E. Hausman, 2011; Yan, 2017).

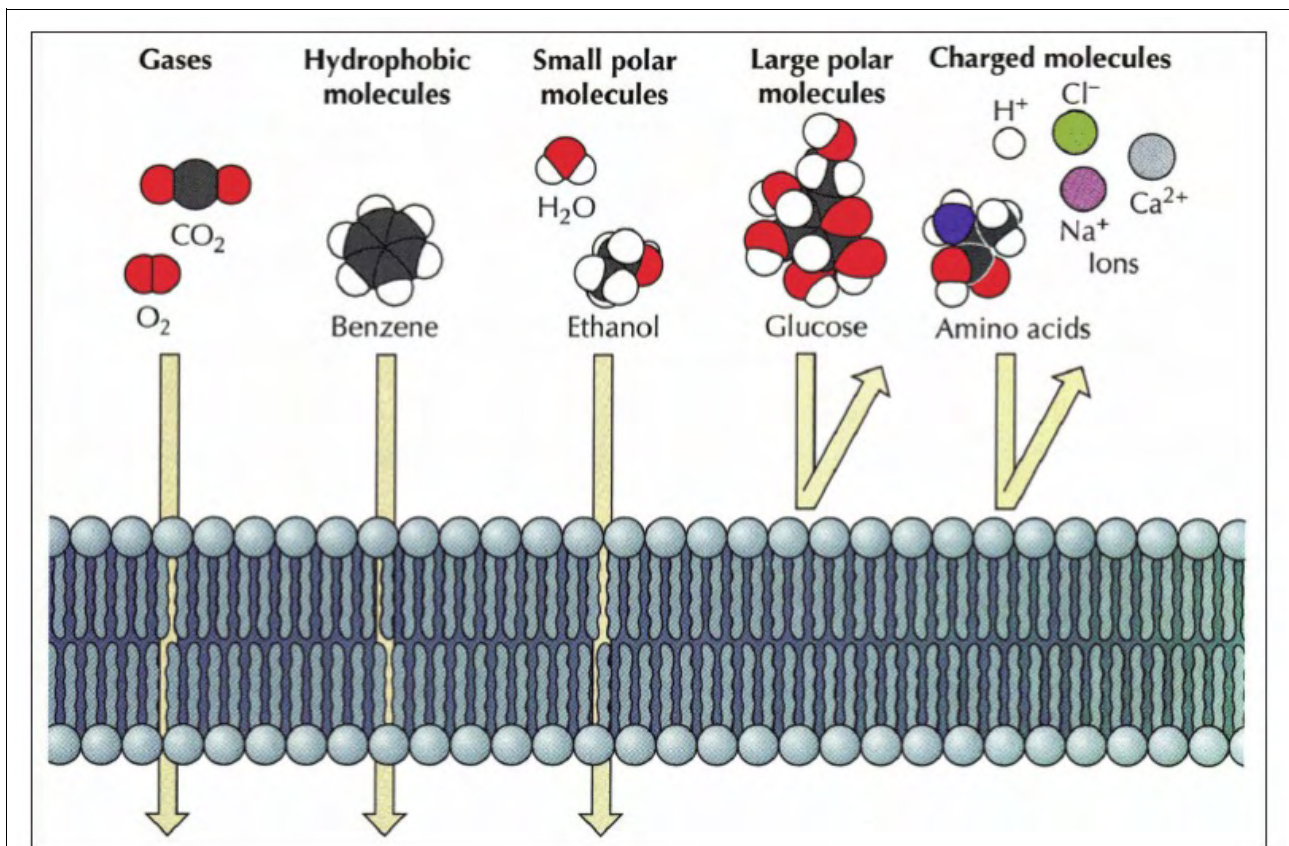


Figure 1: Η διαπερατότητα της λιπιδιακής διπλοστιβάδας από διάφορα μόρια (Chien, 2013).

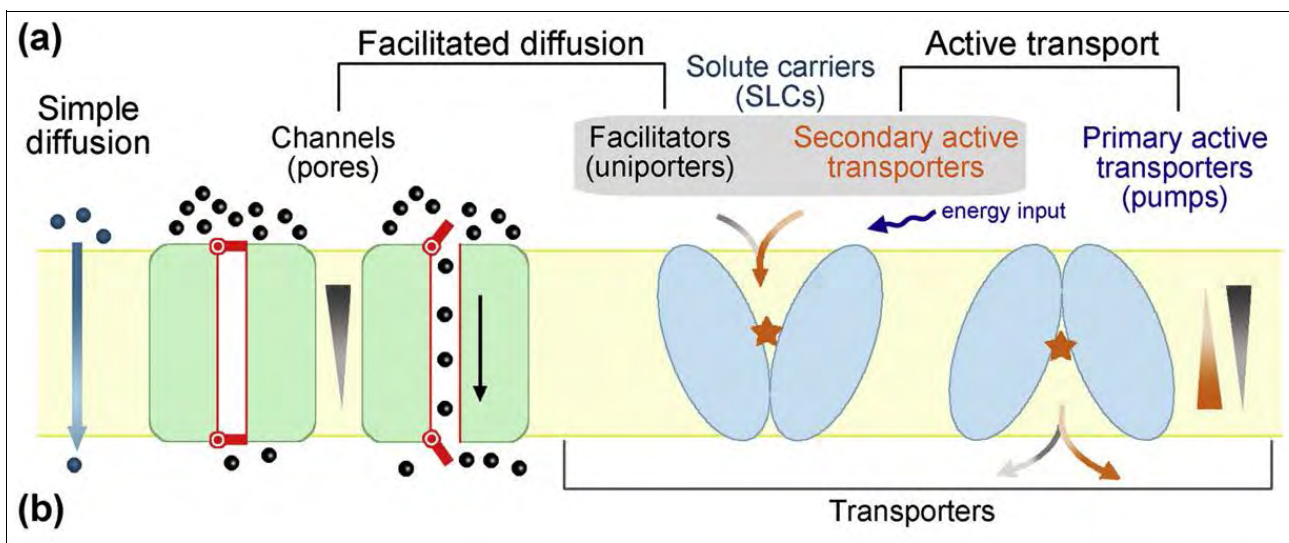


Figure 2: Οι μέθοδοι μεταφοράς μορίων δια μέσω της μεμβράνης. Η πρώτη μέθοδος αποτελεί την παθητική διάχυση, όπου μόρια διαχέονται μέσω της μεμβράνης και φθάνουν στον ενδοκυττάριο χώρο. Η δεύτερη μέθοδος αποτελεί τη διευκολυνόμενη διάχυση (facillitated diffusion) που επιτυγχάνεται είτε μέσω πρωτεϊνών καναλιών είτε μέσω πρωτεϊνών μεταφορέων. Οι παραπάνω διαδικασίες διαχωρίζονται από την ενεργητική μεταφορά καθώς η τελευταία καταλύεται από ειδικούς μεταφορείς με την ταυτόχρονη κατανάλωση ενέργειας (Yan, 2017).

1.1.1 Παθητική διάχυση

Η παθητική διάχυση Figure 2 αποτελεί μια διαδικασία που δεν απαιτεί την κατανάλωση ενέργειας και την συμμετοχή πρωτεϊνών, καθώς τα λιπόφιλα μόρια λόγω των φυσικοχημικών τους ιδιοτήτων περνούν διαμέσου της κυτταροπλασματικής μεμβράνης, από τον εξωκυττάριο στον ενδοκυττάριο χώρο και αντίστροφα. Τα μόρια διαχέονται περνώντας από την διπλοστιβάδα φωσφολιπιδίων, καταλήγοντας στον ενδοκυττάριο ή εξωκυττάριο χώρο αντίστοιχα με την κατεύθυνση της πορείας τους. Ο παράγοντας που καθορίζει την κατεύθυνση των μορίων είναι η διαβάθμιση συγκέντρωσης ανάμεσα στις δύο πλευρές της κυτταρικής μεμβράνης. Μικρά, μη πολικά και υδρόφοβα μόρια έχουν την ικανότητα να μεταφέρονται μέσω παθητικής διάχυσης. Αντιθέτως, τα μεγαλύτερα πολικά και αφόρτιστα μόρια, όπως η γλυκόζη, καθώς και τα φορτισμένα μόρια οποιουδήποτε μεγέθους αδυνατούν να περάσουν παθητικά κατά μήκος της μεμβράνης, με αποτέλεσμα να είναι απαραίτητη η μεταφορά μέσω της διαδικασίας της διευκολυνόμενης διάχυσης (Cooper, Geoffrey M., and Robert E. Hausman, 2011).

1.1.2 Διευκολυνόμενη διάχυση

Αναφορικά με την διευκολυνόμενη διάχυση Figure 2, ισχύει κι εδώ ότι η διαβάθμιση των συγκεντρώσεων μεταξύ των δύο πλευρών της κυτταροπλασματικής μεμβράνης καθορίζει την πορεία μεταφοράς των μορίων, δίχως την κατανάλωση ενέργειας. Η επαφή με τα υδρόφοβα λιπίδια της διπλοστιβάδας, που αποτελεί και τον περιοριστικό παράγοντα

για τη μεταφορά των φορτισμένων και μεγάλων μορίων, αποφεύγεται με την βοήθεια κυτταροπλασματικών πρωτεϊνών, επιτυγχάνοντας έτσι την μεταφορά των μορίων αυτών. Οι πρωτεΐνες που καταλύουν τη συγκεκριμένη διεργασία διακρίνονται σε πρωτεΐνες-φορείς και πρωτεΐνες-διαύλους. Οι πρωτεΐνες-φορείς (carrier proteins) δεσμεύουν το μόριο προς μεταφορά σε ένα σημείο και η δέσμευση επάγει την αλλαγή της στερεοδιαμόρφωσης της πρωτεΐνης συνεπάγοντας την αποδέσμευση και μεταφορά του μορίου στην άλλη πλευρά. Αντίθετα, οι πρωτεΐνες-διάυλοι (channel proteins) δεν αλλάζουν την διαμόρφωση τους αλλά παραμένουν ανοιχτοί πόροι μέσω των οποίων περνούν ελεύθερα τα μόρια ανάλογα με το μέγεθος και το φορτίο τους (Cooper, Geoffrey M., and Robert E. Hausman, 2011).

1.1.2.1 Διευκολυνόμενη διάχυση μέσω πρωτεϊνών μεταφορέων

Οι πρωτεΐνες φορείς υπάρχουν σε όλους τους οργανισμούς της βίωσης και κατέχουν εξέχοντα ρόλο σε πολλαπλές φυσιολογικές διεργασίες. Οι πρωτεΐνες μεταφορείς όπως προαναφέρθηκε περνούν από μια σειρά αλλαγών στη διαμόρφωση τους προκειμένου να καταστεί επιτυχής η μεταφορά του μορίου υποστρώματος κατά τη φορά της διαβάθμισής του. Μια εν τω βάθει ματιά στο συγκεκριμένο μηχανισμό φαίνεται στην εικόνα Figure 3, όπου παρουσιάζονται οι αλλαγές στις διαμορφώσεις των μεταφορέων GLUT. Για να πραγματοποιηθεί η μεταφορά ενός μορίου υποστρώματος απαιτείται ένας πλήρης κύκλος μεταβολών της διαμόρφωσης του μεταφορέα μεταξύ των καταστάσεων i) έξω στραμμένος (outward-facing) με το “άνοιγμα” του μεταφορέα να βρίσκεται στραμμένο προς την εξωκυτταρική πλευρά, ii) φραγμένος λόγω της ύπαρξης του υποστρώματος στη θέση σύνδεσης (occluded) και iii) έσω στραμμένος (inward-facing) όπου ο μεταφορέας “ανοίγει” προς το κυτταρόπλασμα. Ο συγκεκριμένος κύκλος εναλλαγών ονομάζεται κύκλος της εναλλασσόμενης πρόσβασης (alternating access cycle) (Yan, 2017).

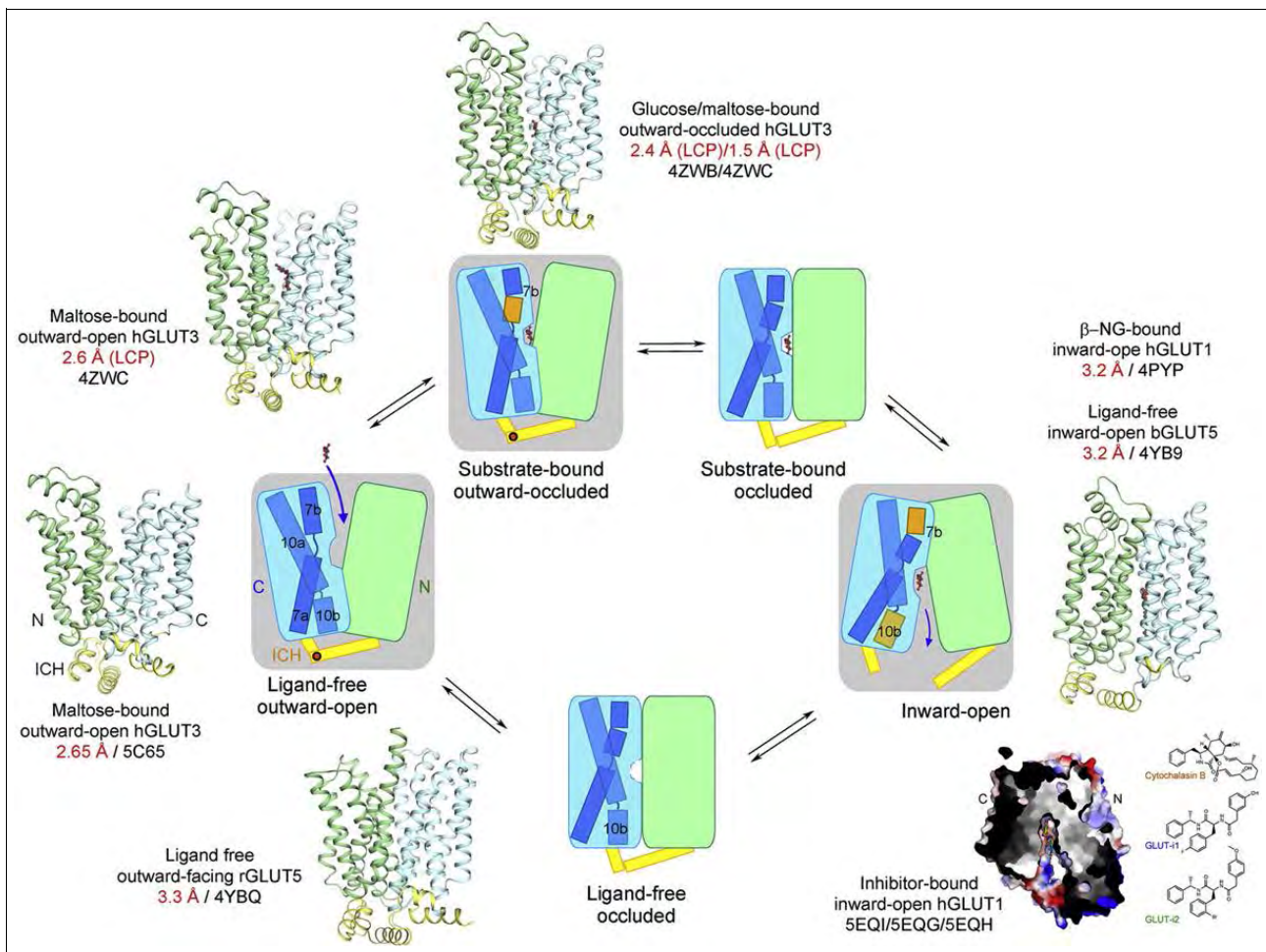


Figure 3: Ο κύκλος αλλαγών στη διαμόρφωση των μεταφορέων GLUT κατά την μεταφορά ενός μορίου υποστρώματος (Yan, 2017).

1.1.2.2 Διευκολυνόμενη διάχυση μέσω πρωτεϊνών διαύλων

Οι καλύτερα μελετημένες πρωτεΐνες διαύλοι είναι οι διαύλοι ιόντων που ο ρόλος τους είναι η μεταφορά ιόντων κατά μήκος των μεμβρανών (Cooper, Geoffrey M., and Robert E. Hausman, 2011). Οι διαύλοι ιόντων απαντώνται σε όλους τους οργανισμούς και έχουν κεντρίσει το ενδιαφέρον πολλαπλών ερευνητικών ομάδων λόγω του κρίσιμου ρόλου που επιτελούν σε κυτταρικές διεργασίες. Τα κανάλια ιόντων σχηματίζουν υδρόφιλους επιλεκτικούς πόρους που επιτρέπουν παθητική μεταφορά των ανόργανων ιόντων υποστρωμάτων. Ιόντα κατάλληλου μεγέθους και φορτίου (κυρίως Na^+ , K^+ , Ca^{2+} , Cl^-) διασχίζουν τη μεμβράνη σύμφωνα με την ηλεκτροχημική τους κλίση με ρυθμό που υπολογίζεται έως και 1000 φορές υψηλότερος από εκείνον που μπορεί να επιτευχθεί από άλλες πρωτεΐνες που καταλύουν μεταφορά διαμέσου της μεμβράνης. Η επιλεκτικότητα των καναλιών βασίζεται τόσο στο μέγεθος του πόρου όσο και στο φορτίο του εσωτερικού του, η επιλεκτικότητα στα ανιόντα επιτυγχάνεται με την ύπαρξη θετικά φορτισμένων αμινοξέων ενώ για τα κατιόντα αρνητικά φορτισμένα. Το δεύτερο σημαντικό χαρακτηριστικό των

διαύλων ιόντων είναι ότι εμφανίζουν ένα είδος φραγμού αποτρέποντας τη συνεχή διάχυση των ιόντων κατά μήκος της μεμβράνης. Για να μεταβούν στην ανοιχτή κατάσταση οι διάυλοι πρέπει να δεχτούν ένα ερέθισμα, εξειδικευμένο για τον τύπο τους, και ανάλογα με τον τύπο του ερεθίσματος ταξινομούνται σε ορισμένες κατηγορίες όπως φαίνεται στην εικόνα **Figure 4** (Kulbacka et al., 2017).

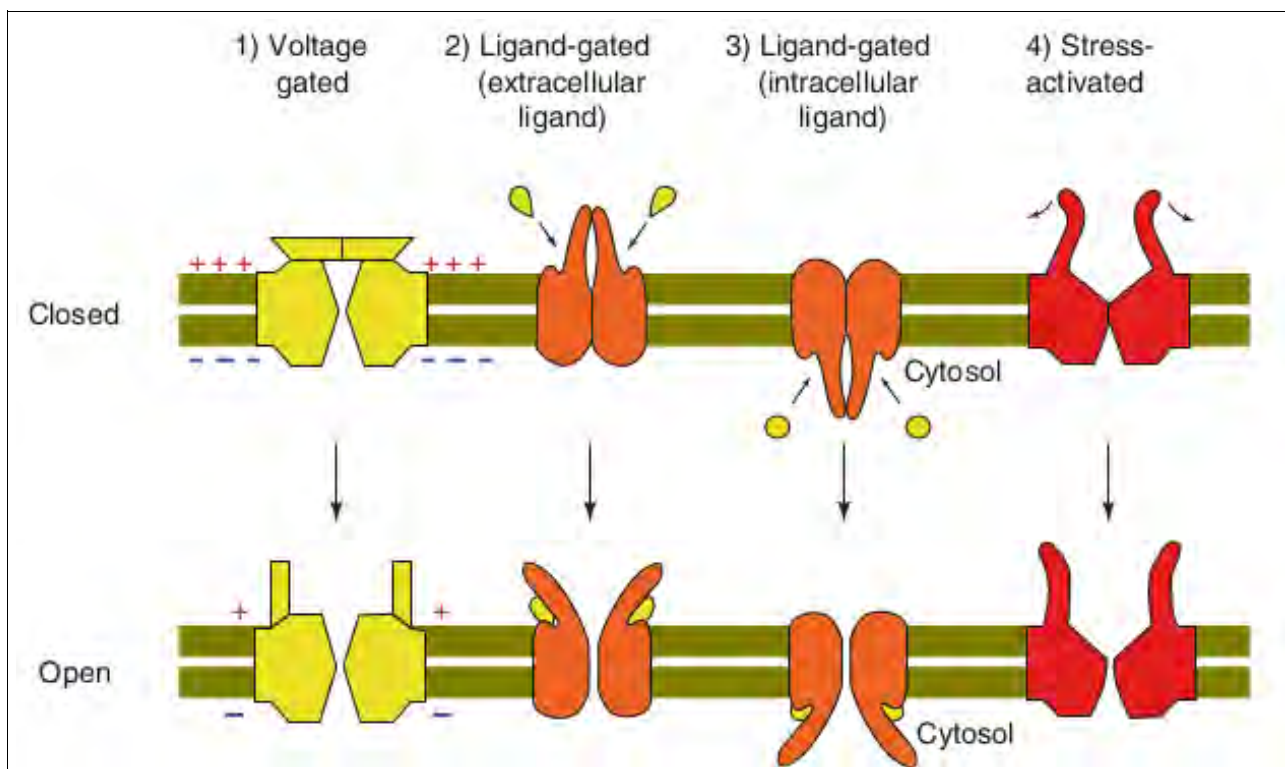


Figure 4: Οι τύποι των καναλιών ανάλογα με το ερέθισμα που τα οδηγεί στην ανοιχτή κατάσταση (Kulbacka et al., 2017).

1.1.3 Ενεργητική μεταφορά

Όλες οι περιπτώσεις διάχυσης, διευκολυνόμενης ή παθητικής, έχουν ένα κοινό γνώρισμα, την μετακίνηση των μορίων σύμφωνα με την διαβάθμιση της συγκέντρωσής τους. Η ενεργητική μεταφορά πραγματοποιεί την αντίστροφη διαδικασία μετακινώντας μόρια αντίθετα στη συγκεκριμένη διαβάθμιση Figure 2, διεργασία που είναι μη ευνοούμενη και απαιτεί την κατανάλωση ενέργειας είτε άμεσα είτε έμμεσα (Cooper, Geoffrey M., and Robert E. Hausman, 2011).

1.1.3.1 Ενεργητική μεταφορά μέσω υδρόλυσης του ATP

Η ενεργητική μεταφορά κατά την οποία υδρολύεται το ATP αποτελεί τη περίπτωση της άμεσης κατανάλωσης ενέργειας. Ο μηχανισμός μεταφοράς με τη χρήση ATP καταλύεται από τις πρωτεΐνες μεταφορείς που ανήκουν στην υπεροικογένεια ABC (ATP-binding cassette) και είναι παρούσες σε όλα τα βασίλεια των οργανισμών. Ο μηχανισμός μεταφοράς των μεταφορέων ABC έχει μελετηθεί εκτεταμένα σε διάφορους αντιπροσώπους

με χαρακτηριστικά παραδείγματα να αποτελούν οι μεταφορείς ιόντων, που είναι επίσης γνωστοί ως αντλίες ιόντων, αλλά και ο μεταφορέας μαλτόζης. Ο μηχανισμός που χρησιμοποιείται υπακούει στο μοντέλο *alternating access*, που προαναφέρθηκε στη παθητική διάχυση μέσω πρωτεϊνικών μεταφορέων, αν και αυτό δεν αποτελεί καθολικό κανόνα, όπως φαίνεται και στην εικόνα Figure 5 (Hopfner, 2016).

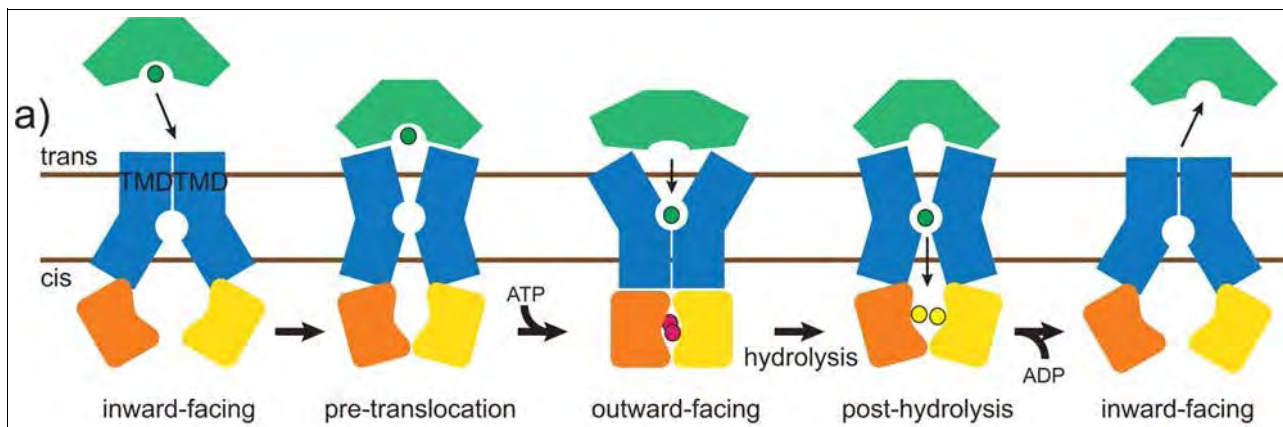
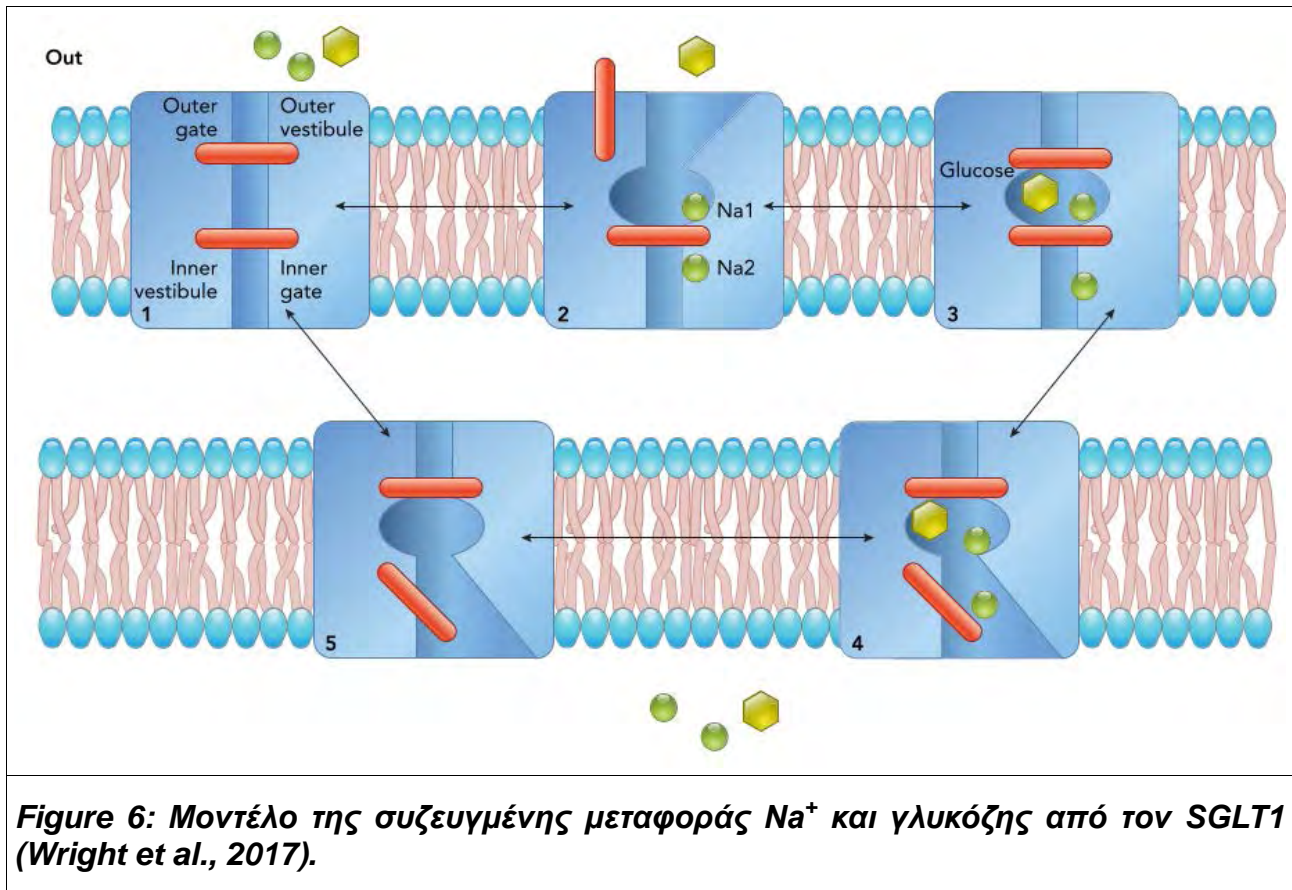


Figure 5: Ο μηχανισμός μεταφοράς μέσω ABC πρωτεϊνών. Η πρωτεΐνη δέσμευσης μαλτόζης (πράσινη) δεσμεύει τη μαλτόζη (πράσινη σφαίρα) και το σύμπλοκο δεσμεύεται στο μεταφορέα ABC προκαλώντας μια αλλαγή στη διαμόρφωση (*pre-translocation*). Στη συνέχεια πραγματοποιείται μετάβαση του μεταφορέα στην “εξω στραμμένη” (*outward-facing*) διαμόρφωση που συμβαίνει κατά τη δέσμευση της ATP στη *cis*, η οποία κατέχει περιοχές δέσμευσης νουκλεοτιδίων (*Nucleotide Binding Domain NDB*). Η μαλτόζη απελευθερώνεται από τη πρωτεΐνη δέσμευσης της και δεσμεύεται εντός του μεταφορέα στις διαμεμβρανικές περιοχές (*Transmembrane Domain*). Η υδρόλυση ATP επάγει μια τελευταία αλλαγή της διαμόρφωσης του μεταφορέα η οποία οδηγεί στην απελευθέρωση μαλτόζης στο εσωτερικό του κυττάρου (Hopfner, 2016).

1.1.3.2 Ενεργητική μεταφορά μέσω συµμεταφορέων

Όπως προαναφέρθηκε η μεταφορά ενός μορίου αντίθετα στη διαβάθμιση συγκέντρωσης του είναι μια διαδικασία που χρειάζεται την κατανάλωση ενέργειας. Πέρα από την άμεση χρήση ενέργειας υπό τη μορφή τριφωσφορικών νουκλεοτιδίων, τα κύτταρα έχουν αναπτύξει ένα ενδιαφέρον μηχανισμό σύμφωνα με τον οποίο συζευγνύουν την μη ευνοούμενη μεταφορά του μορίου ενδιαφέροντος με την ευνοούμενη μεταφορά ενός άλλου μορίου (συνήθως ιόντος) καλύπτοντας έτσι τις ενεργειακές απαιτήσεις της διεργασίας (Cooper, Geoffrey M., and Robert E. Hausman, 2011). Ένα χαρακτηριστικό παράδειγμα αποτελεί ο συµμεταφορέας SGLT1 που καταλύει τη μεταφορά της γλυκόζης, συζευγμένη με μεταφορά δύο μορίων νατρίου, από τον εντερικό αυλό στα κύτταρα του εντερικού επιθηλίου. Αρχικά ο συµμεταφορέας είναι κλειστός, στη συνέχεια εξωτερικό Na^+ συνδέεται πρώτα στις θέσεις Na_1 και Na_2 για να ανοίξει την εξωτερική πύλη και να επιτρέψει τη γλυκόζη να δεσμευτεί στο εσωτερικό Figure 6. Αυτό το εξωτερικό φράγμα κλείνει για να εγκλωβίσει το σάκχαρο και να μην του επιτρέψει να διαφύγει στο εξωτερικό υδατικό διάλυμα. Με το άνοιγμα του εσωτερικού φράγματος, το Na^+ και η γλυκόζη διαχέονται για να

φθάσουν στο κυτταρόπλασμα. Ο κύκλος ολοκληρώνεται με την επιστροφή της πρωτεΐνης στην αρχική θέση (Wright et al., 2017).



1.2 Μεταφορείς σακχάρων

Στους πολυκύτταρους οργανισμούς τα διαλυτά σάκχαρα χρησιμοποιούνται ως πηγές ανθρακικών σκελετών για τη βιοσύνθεση πολλών άλλων κυτταρικών ενώσεων, οσμολύτες, μόρια-σινιάλα, παροδικές αποθήκες ενέργειας και ως μόρια-μεταφορείς. Μεταφορείς σακχάρων υπάρχουν σε όλους τους οργανισμούς και αποτελούν υψίστης σημασίας πρωτεΐνες, καθώς καθιστούν δυνατή τη μεταφορά των μορίων σακχάρων, για τα οποία εξειδικεύονται, εντός και εκτός του κυττάρου με ελεγχόμενο τρόπο. Μέχρι σήμερα έχουν αναγνωριστεί τρεις μεγάλες κατηγορίες μεταφορέων σακχάρων, οι μεταφορείς σακχάρων που ανήκουν στην υπεροικογένεια μεταφορέων MFS (Major Facilitator Superfamily), οι μεταφορείς SGLT (Sodium Glucose symporters) και οι SWEETs (Chen et al., 2015). Οι τρεις κατηγορίες μεταφορέων έχουν ξεχωριστούς φυσιολογικούς ρόλους και συνοψίζονται στο μοντέλο των (Deng and Yan, 2016) Figure 7. Η κατηγορία των μεταφορέων σακχάρων που ανήκουν στην υπεροικογένεια MFS είναι η μεγαλύτερη (Pao et al., 1998).

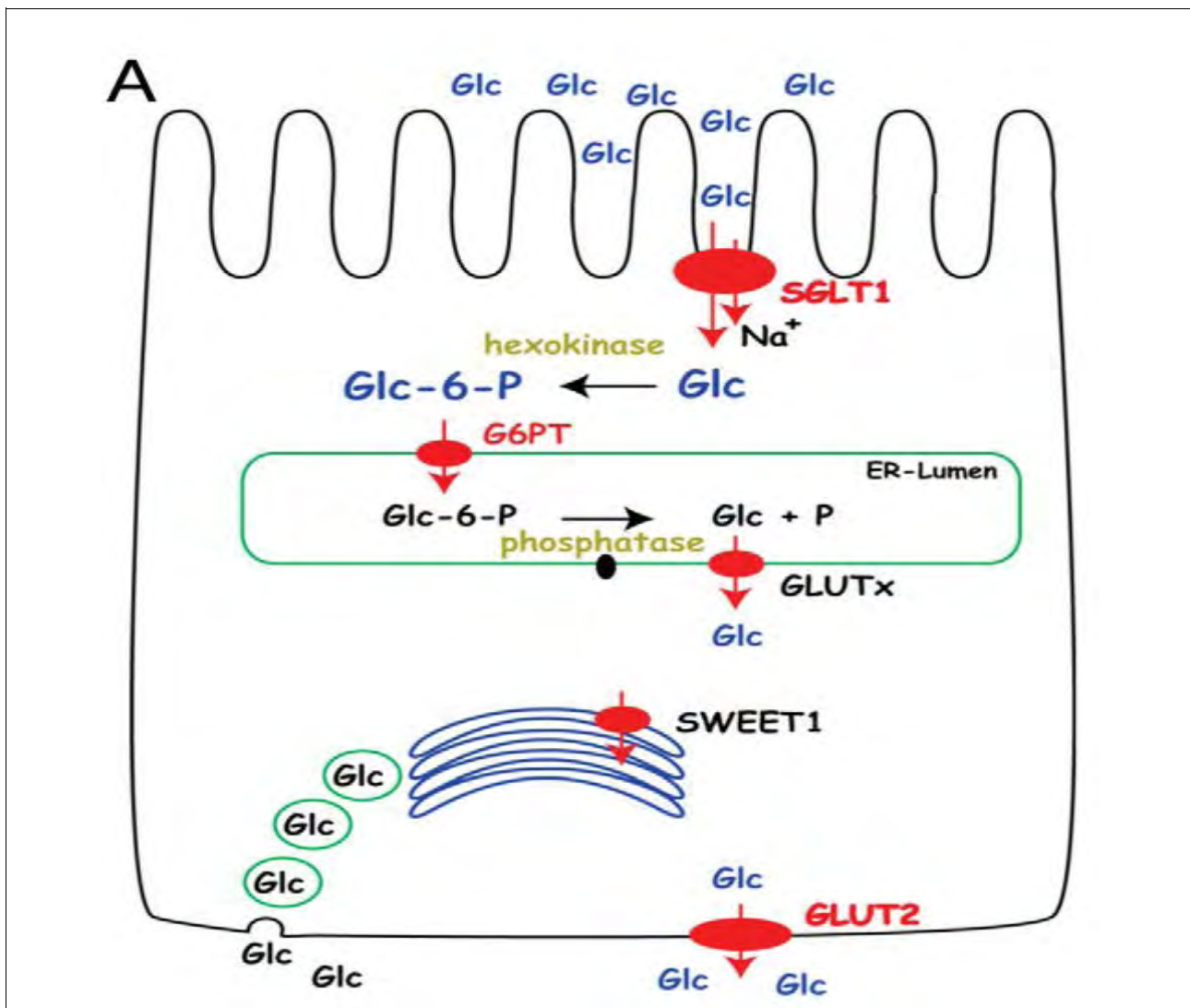


Figure 7: Απλοποιημένο μοντέλο της τοποθέτησης και της λειτουργίας των τριών μεγάλων κατηγοριών μεταφορέων γλυκόζης στα επιθηλιακά κύτταρα του εντέρου του ανθρώπου (Deng and Yan, 2016).

1.2.1 Sodium – Glucose Transporters SGLT

Ο πρώτος μεταφορέας SGLT, ο SGLT1 από το εντερικό επιθήλιο, κλωνοποιήθηκε πριν από 30 χρόνια και αποτέλεσε βάση για μεγάλες εξελίξεις στην ιατρική έρευνα. Η συγκεκριμένη οικογένεια μεταφορέων κωδικοποιείται από την οικογένεια των γονιδίων SLC5 και περιέχει 11 συμμεταφορείς και έναν μεταφορέα, τον SGLT3, που αναφέρεται ως “αισθητήρας γλυκόζης” και εκφράζεται στους νευρώνες, την ηπατική πυλαία φλέβα και το λεπτό έντερο παίζοντας άμεσο ρόλο στη σηματοδότηση που διεγείρει την πρόσληψη τροφής. Ο μηχανισμός εισαγωγής γλυκόζης του SGLT1 περιλαμβάνει δύο διαμορφώσεις, την “outward-facing conformation” η οποία είναι και η διαμόρφωση που δέχεται τα δύο μόρια νατρίου, με τη γλυκόζη και μεταπηδά στην “inward-facing conformation” που απελευθερώνει τα μόρια στο εσωτερικό του κυττάρου. Ο SGLT1 έχει επίσης ενοχοποιηθεί

για την απορρόφηση νερού από το εντερικό επιθήλιο σε συνδυασμό με τη μεταφορά της γλυκόζης (Wright et al., 2017). Φυλογενετικές αναλύσεις υποδεικνύουν ότι οι SGLTs αποτελούν μια οικογένεια ονομαζόμενη SSF (sodium:solute symporter family) που βρίσκεται υπό την ομπρέλα της υπεροικογένειας APC (amino acid/polyamine/organocation superfamily) που στο σύνολο της περιέχει 14 οικογένειες Figure 8 (Hoglund et al., 2011).

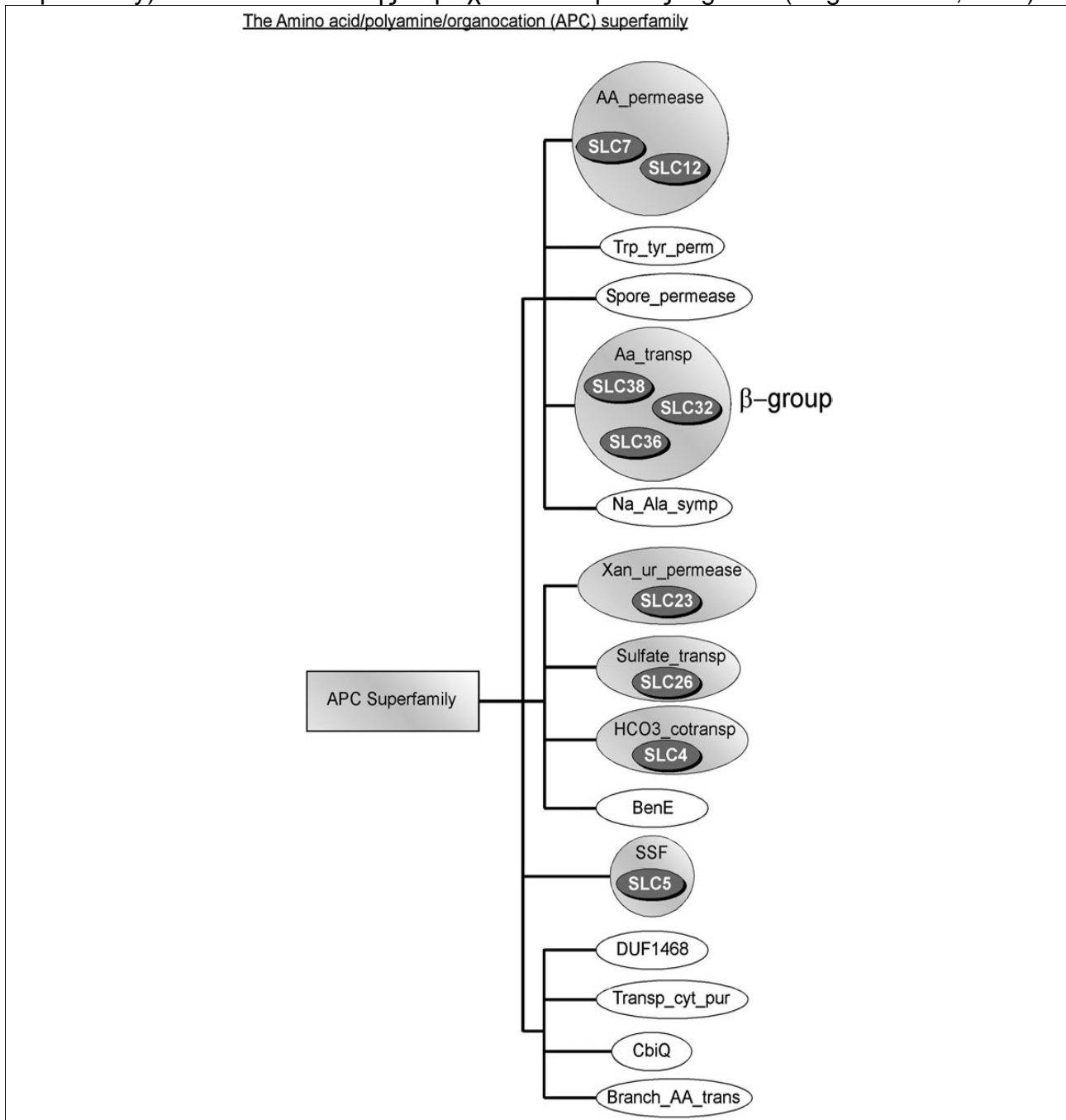


Figure 8: Η υπεροικογένεια APC. Οι μεταφορείς SGLT ανήκουν στην οικογένεια SSF (Hoglund et al., 2011).

1.2.2 SWEET transporters

Η οικογένεια μεταφορέων SWEET είναι η πιο πρόσφατα ανακαλυφθείσα οικογένεια. Σύμφωνα με τη βιβλιογραφία η συγκεκριμένη οικογένεια εμφανίζει ομόλογα τόσο σε προκαρυώτες όσο και στα βασίλεια των ευκαρυωτών. Στα φυτά έχει βρεθεί πως μεταφέρουν γλυκόζη ή σουκρόζη, ενώ μελέτες σε διάφορα ζώα όπως ο *C.elegans* (CeSWEET1) και ο άνθρωπος (HsSWEET1) έχουν αποσαφηνίσει ότι λειτουργία τους είναι η μεταφορά γλυκόζης. Φυλογενετικές μελέτες στα αγγειόσπερμα έχουν αναδείξει την ύπαρξη τεσσάρων κατηγοριών στους συγκεκριμένους μεταφορείς όπως φαίνεται και στην εικόνα Figure 9 (Eom et al., 2015). Γενικότερα τα ομόλογα των SWEET και τα βακτηριακά ομόλογα SemiSWEETs έχουν ταξινομηθεί ως μέλη της MtN3 saliva family. Εκτεταμένες μελέτες των συγκεκριμένων μεταφορέων έχουν πραγματοποιηθεί κυρίως στα φυτά και έχουν αναδείξει ότι εμπλέκονται σε διεργασίες κυτταρικής εκροής σακχάρων και μεταφοράς από τα κενοτόπια (Chen et al., 2015).



Figure 9: Φυλογενετικό δένδρο των μεταφορέων SWEET από δεκαπέντε αντιπροσώπους των αγγειόσπερμων (Eom et al., 2015).

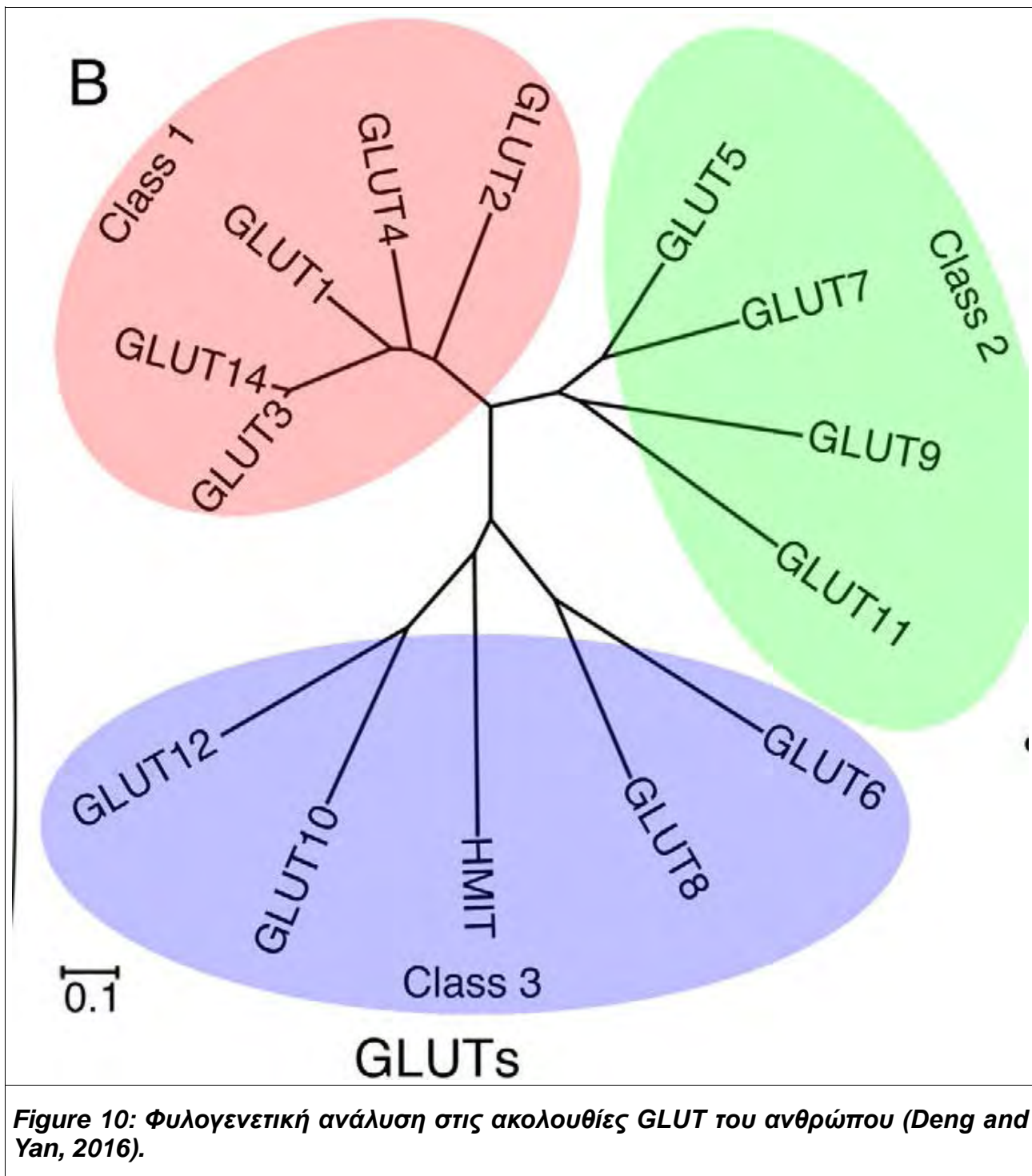
1.2.3 MFS sugar transporters - Glucose transporters (GLUTs)

Οι μεταφορείς GLUT ανήκουν στην οικογένεια Sugar Porter family που συγκαταλέγεται στην υπεροικογένεια Major Facilitator Superfamily (MFS) και μετά από πολυετή έρευνα έχουν ανακλυφθεί ποικίλα ομόλογα, ακόμα και στους προκαρυώτες. Ο πρώτος GLUT που κλωνοποιήθηκε και κρυσταλλώθηκε ήταν ο GLUT1 (Chen et al., 2015). Στον άνθρωπο έχουν ανακαλυφθεί 14 μεταφορείς GLUT που σύμφωνα με πρόσφατα reviews διαχωρίζονται σε 3 κλάσεις (Deng and Yan, 2016) , όπως φαίνεται και στην εικόνα Figure 10 .

Η πρώτη κλάση (Class I) εμπεριέχει τους μεταφορείς οι GLUT1-4 και το GLUT14, στη δεύτερη κλάση (Class II) ανήκουν οι GLUT5, 7, 9 και 11, ενώ στην τρίτη κλάση (Class III) συναντάμε τους GLUT6, 8, 10, 12 και τον HMIT που καταλύει τη μεταφορά της μυοϊνοσιτόλης και μπορεί να συναντηθεί με τον όρο GLUT 13.

Η κλάση 1 έχει τραβήξει το ενδιαφέρον της επιστημονικής κοινότητας περισσότερο από κάθε άλλη κατηγορία, καθώς τα μέλη της αποτελούν και τους πρώτους σε σειρά μεταφορείς που ανακαλύφθηκαν. Οι φυσιολογικοί ρόλοι των μεταφορέων GLUT 1-4 έχουν μελετηθεί διεξοδικά. Ο GLUT1 είναι ένας μεταφορέας γλυκόζης που εντοπίζεται σε όλους τους ιστούς, αλλά κύριοι ρόλο έχει στα ερυθροκύτταρα και τον αιματοεγκεφαλικό φραγμό. Ο GLUT2 είναι ο κύριος μεταφορέας γλυκόζης στα ηπατικά κύτταρα και τα κύτταρα του εντέρου και παρουσιάζει μια εξαιρετικά υψηλή τιμή Km για τη γλυκόζη, που εξυπηρετεί το ρόλο του ως αισθητήρα της συγκέντρωσης της γλυκόζης. Επίσης συγκριτικά με τους GLUT 1, 3 & 4 αποτελεί το μόνο GLUT που καταλύει την αμφίδρομη ροή γλυκόζης ανάλογα με τη συγκέντρωση της γλυκόζης στο αίμα, αν δηλαδή ο οργανισμός βρίσκεται σε κατάσταση μεταγευματική ή κατάσταση νηστείας. Ο GLUT3 εκφράζεται κυρίως σε νευρικά κύτταρα και αναφέρεται επίσης ως “νευρωνικός μεταφορέας γλυκόζης” και εμφανίζει τη χαμηλότερη τιμή Km, όπως είναι λογικό καθώς η γλυκόζη αποτελεί την κύρια πηγή τροφής για τα νευρικά κύτταρα, σε συνδυασμό με το υψηλότερο ποσοστό ανακύκλωσης μεταξύ των GLUT1-4. Ο GLUT4 έχει υψηλά επίπεδα έκφρασης στο λιπώδη ιστό και τους σκελετικούς μύες. Είναι γνωστό ότι ανταποκρίνεται στην ινσουλίνη καθώς η σηματοδότηση της επάγει τη μετεγκατάσταση του από ενδοκυτταρικά κυστίδια στην πλασματική μεμβράνη.

Οι μεταφορείς της κλάσης 2 δεν έχουν αυστηρή εξειδίκευση για τη μεταφορά γλυκόζης. Ο GLUT 5 έχει ως βασικό υπόστρωμα τη φρουκτόζη και αποτελεί το βασικό τρόπο πρόσληψης της στα εντερικό επιθήλιο. Οι GLUT 7 και 11 καταλύουν τη μεταφορά τόσο γλυκόζης όσο και φρουκτόζης. Ο GLUT 9 μεταφέρει ουρικό οξύ. Οι μεταφορείς της κλάσης 3 αποτελούν τους λιγότερο μελετημένους μεταφορείς (Deng and Yan, 2016).



1.3 Δομή των μεταφορέων σακχάρων

Κάθε κατηγορία από τις προαναφερθείσες έχει μια χαρακτηριστική τυπική δομή, που παίρνει μια χαρακτηριστική τρισδιάστατη διαμόρφωση στο χώρο και συναντάται σε όλα τα μέλη της. Οι δομές των MFS, SGLT και SWEET είναι διαφορετικές, ωστόσο

εμφανίζουν και αξιοσημείωτες ομοιότητες. Οι θέσεις πρόσδεσης είναι τοποθετημένες στο κέντρο της πλασματικής μεμβράνης και πιο συγκεκριμένα στη διεπιφάνεια μεταξύ των δυο λιπιδιακών στιβάδων, επίσης υπάρχουν κατάλοιπα αρωματικών αμινοξέων που έχουν κρίσιμο ρόλο είτε στην επιλεκτική δέσμευση του υποστρώματος στο μεταφορέα και η περαιτέρω κατάλυση της μεταφοράς είτε τη δημιουργία φυσικού φραγμού ώστε να μην μπορεί το υπόστρωμα να διαφύγει (Chen et al., 2015).

Η τυπική δομή της υπεροικογένειας MFS, αποτελείται από 12 διαμεμβρανικές περιοχές (12 TM) οργανωμένες σε δύο ψευδοσυμμετρικά μισά και τη ενδομεμβρανική ελικοειδή περιοχή ICH που αποτελείται από τέσσερις α-έλικες IC 1-4, όπως φαίνεται και στην εικόνα Figure 11a και θα αναλυθεί παρακάτω (Chen et al., 2015).

Οι SGLT, όπως προαναφέρθηκε, ανήκουν στην υπεροικογένεια APC και εμφανίζουν τη χαρακτηριστική “LeuT” αναδίπλωση, που συγκροτείται από 10 διαμεμβρανικές περιοχές (10 TM) χωρισμένες ανά 5 με τη πρώτη πεντάδα να αναφέρεται ως το αμινοτελικό domain και τη δεύτερη ως το καρβοξυτελικό domain, με τα άκρα και των δύο να τοποθετούνται προς την κυτταροπλασματική πλευρά της μεμβράνης. Η δομή που παρουσιάζεται στις εικόνες αποτελεί την τυπική και κρυσταλλική δομή του vSGLT, που προέρχεται από τον οργανισμό *Vibrio parahaemolyticus* και έχει κάποιες διαφορές από την τυπική APC δομή. Στη τυπική δομή του vSGLT, που βλέπουμε στα αριστερά της Figure 11b, οι διαμεμβρανικές περιοχές TM2-6 και οι TM7-11 αποτελούν τη LeuT αναδίπλωση (Deng and Yan, 2016).

Οι ευκαρυωτικοί SWEET εμφανίζουν μια προβλεπόμενη τοπολογία η οποία περιλαμβάνει δύο επαναλήψεις τριών διαμεμβρανικών περιοχών (3 TM) που διαχωρίζονται από μια διαμεμβρανική περιοχή, όπως φαίνεται και στο δεξί τμήμα της Figure 11c, ενώ τα βακτηριακά ομόλογα τους διαθέτουν μόλις 3 διαμεμβρανικές περιοχές (3 TM) και ονομάζονται SemiSWEETs (Feng and Frommer, 2015). Στην εικόνα Figure 12 παρατίθενται αντιπροσωπευτικές τρισδιάστατες δομές μεταφορέων, μια για κάθε υπεροικογένεια, όπως αυτές φαίνονται σε πρόσοψη και σε κάτοψη (Chen et al., 2015; Deng and Yan, 2016).

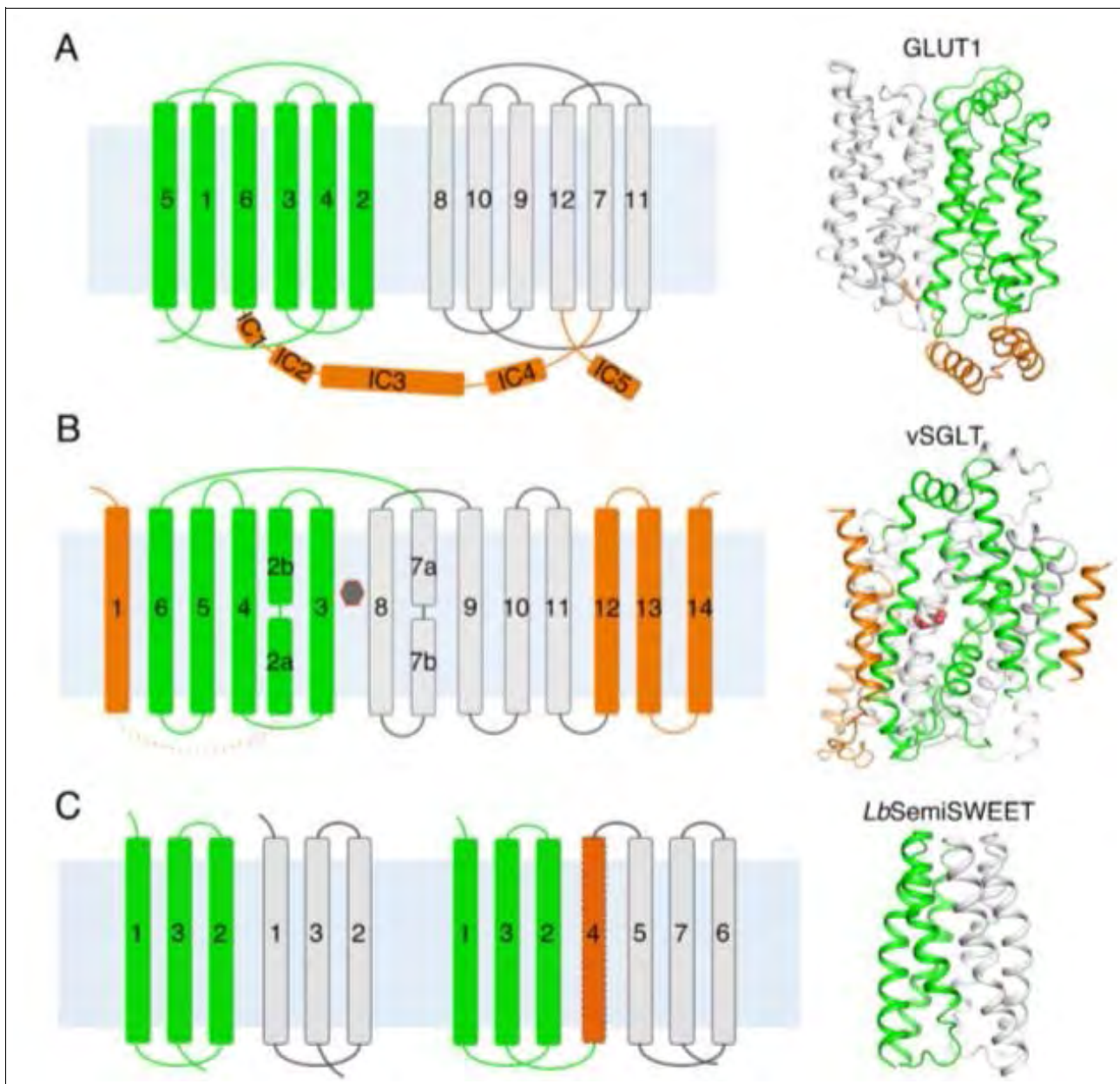


Figure 11: Οι τυπικές δομές των μεταφορέων σακχάρων.(A) Δομή των GLUTs. Διαθέτουν 12 διαμεμβρανικές έλικες, η οποίες μοιράζονται σε δύο domains, το N-τελικό (πράσινο) και το C-τελικό (λευκό) και μια ενδοκυτταρική έλικα που ονομάζεται ICH (πορτοκαλί). Δεξιά φαίνεται η δομή του ανθρώπινου GLUT1 (PDB entry: 4PYP). (B) Δομή των SGLTs. Οι διαμεμβρανικές περιοχές TM2-11 συνιστούν «5+5» ανάστροφες επαναλήψεις (πράσινο και λευκό) της LeuT αναδίπλωσης του vSGLT. Το μαύρο εξαγώνο υποδηλώνει το υπόστρωμα του μεταφορέα. Δεξιά φαίνεται η κλειστή διαμόρφωση του vSGLT, όπως υπάρχει δημοσιευμένη στην PDB (PDB entry: 3DH4). (Γ) Δομή των SWEETs. Στα αριστερά φαίνεται η τοπολογία των SemiSWEETs και η προβλεπόμενη τοπολογία του ανθρώπινου SWEET1. Και οι δύο είναι διατεταγμένες σε μοτίβο 1 -3-2 και χρωματισμένες με πράσινο και λευκό. Στα δεξιά φαίνεται η δομή του LbSemiSWEET (PDB entry: 4QNC) (Deng and Yan, 2016).

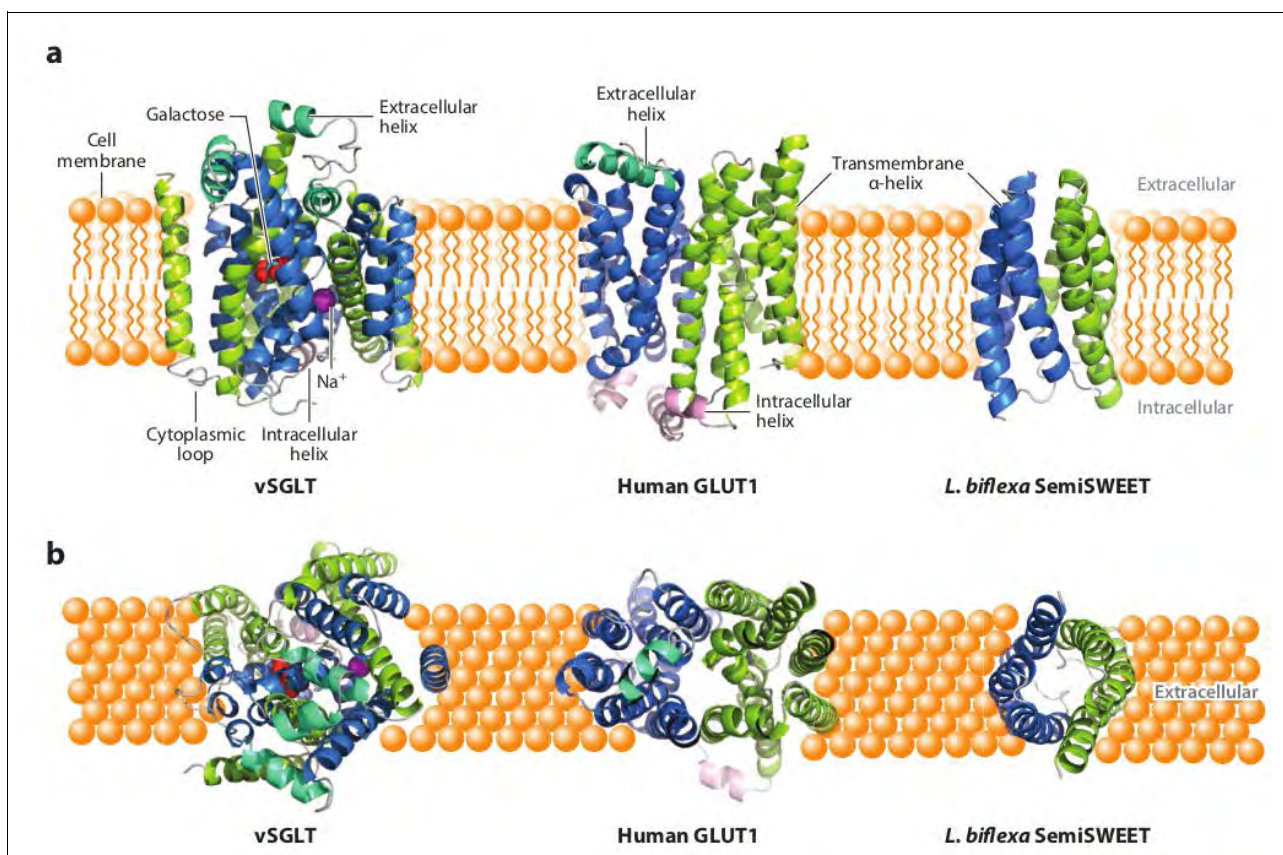


Figure 12: Τρισδιάστατες δομές των τριών κατηγοριών μεταφορέων σακχάρων στους ευκαρυώτες. α) Πρόσωση του μοντέλου vSGLT από το *Vibrio parahaemolyticus* που αποτελεί ομόλογο των μεταφορέων SGLT του ανθρώπου (PDB : 3DH4), του ανθρώπινου GLUT1 (PDB : 4PYP) και το βακτηριακού semiSWEET από το *Leptospira biflexa* που αποτελεί ομόλογο των μεταφορέων SWEET (PDB : 4QNC) . Κάτοψη των ίδιων δομών όπου με μπλε και ανοιχτό πράσινο αντικατοπτρίζονται τα αμινοτελικά και καρβοξυτελικά μέρη των πρωτεϊνών αντίστοιχα. Η εξωτερική έλικα, η ενδοκυτταρική έλικα και ο κυτταροπλασματικός βρόγχος είναι χρωματισμένοι πράσινο, ανοιχτό ροζ και γκρι αντίστοιχα.

1.4 Η Major Facilitator Superfamily

Η Major Facilitator Superfamily (MFS) είναι η μεγαλύτερη υπεροικογένεια δευτερογενών μεταφορέων που εντοπίζεται στην βιόσφαιρα και εντός της υπεροικογένειας εντοπίζονται 12 οικογένειες μεταφορέων με μια να ονομάζεται MFS, η οικογένεια 2.A.1. Οι υποοικογένειες μέλη της 2.A.1 (89 στον αριθμό με 5 μη χαρακτηρισμένες) μεταφέρουν ένα σύνολο παρεμφερών ενώσεων όπως απλοί μονοσακχαρίτες, ολιγοσακχαρίτες, αμινοξέα, πεπτιδία, βιταμίνες και συμπαραγοντες ενζύμων (Perland and Fredriksson, 2017). Η πρώτη υποοικογένεια (2.A.1.1) είναι η Sugar Porter subfamily που θα μελετήσουμε για τη

φυλογενετική ανάλυση και τη δημιουργία του εργαλείου μας. Πολλές από τις υποοικογένειες έχουν προσδιοριστεί φυλογενετικά και δεν περιέχουν κανένα λειτουργικά χαρακτηρισμένο μέλος και ονομάζονται UMFs (Unknown Major Facilitators)(Reddy et al., 2012). Οι μεταφορείς αυτής της υπεροικογένειας φαίνεται να προήλθαν από μια μονή hairpin δομή με δύο διαμεμβρανικές περιοχές, η οποία τριπλασιάστηκε σε μια δομή με 6 διαμεμβρανικές περιοχές. Η μονάδα με τις έξι διαμεμβρανικές περιοχές στη συνέχεια διπλασιάστηκε και το αποτέλεσμα ήταν μια δομή με 12 διαμεμβρανικές περιοχές. Αν και η τοπολογία των 12 διαμεμβρανικών περιοχών είναι η βασική δομή των μεταφορέων της MFS, υπάρχουν και οικογένειες που εμφανίζουν 14 διαμεμβρανικές περιοχές, από τις οποίες οι δύο επιπλέον έλικες εντοπίζονται κεντρικά, ανάμεσα στις δυο επαναλαμβανόμενες μονάδες των 6 διαμεμβρανικών περιοχών και ίσως έχουν προέλθει από έναν ενδογενή (intragenic) διπλασιασμό μιας δομής φουρκέτας (hairpin). Έχει αποδειχθεί ότι υπάρχουν ομόλογα με 13, 14, 15 ή 16 διαμεμβρανικές περιοχές, αλλά και με 24 διαμεμβρανικές περιοχές. Τα τελευταία ομόλογα προήλθαν από πλήρη διπλασιασμό και παρουσιάζουν διπλές λειτουργίες.

Δομικές αναλύσεις (Deng and Yan, 2016; Yan, 2015) έχουν φανερώσει ότι όλοι οι μεταφορείς της υπεροικογένειας μοιράζονται χαρακτηριστική αναδίπλωση στον πυρήνα, που ονομάζεται MFS fold Figure 11a. Η αναδίπλωση MFS περιλαμβάνει δύο domains, που αποτελούνται από 6 διαδοχικές διαμεμβρανικές περιοχές (1-6 & 7-12) και ονομάζονται N και C domains. Οι δυο αυτές περιοχές παρουσιάζουν διπλή ψευδοσυμμετρία που σχετίζεται με τον άξονα ψ. Σε κάθε domain, οι 6 διαμεμβρανικές περιοχές οργανώνονται σε ζευγάρι «3+3» ανεστραμμένων επαναλήψεων με τις διαμεμβρανικές περιοχές της κάθε επανάληψης να παίζουν σημαντικό ρόλο στη λειτουργία και τη δομή της πρωτεΐνης. Συγκεκριμένα, κάθε πρώτη έλικα κάθε δεματίου τριών ελίκων (TMs 1,4,7 και 10) τοποθετείται στο κέντρο του μεταφορέα και όλες μαζί δομούν τον πόρο μεταφοράς. Η πλειοψηφία των καταλοίπων εκεί έχουν ως ρόλο τη σύνδεση του υποστρώματος. Εξωτερικά του πυρήνα βρίσκονται οι TMs 2, 5, 8 και 11, οι οποίες μεσολαβούν την επαφή των δύο domain, ενώ κάποια κατάλοιπα του ίσως συμβάλλουν στη δέσμευση του υποστρώματος και στη μεταφορά του. Εξωτερικά των TMs 1, 4, 7 και 10 τοποθετούνται οι TMs 3, 6, 9 και 12 συμβάλλοντας στην δομική σταθερότητα του μεταφορέα. Τέλος, παρατηρώντας τη δομή σαν σύνολο, είναι ενδιαφέρον ότι οι αντίστοιχες TMs κάθε ενός από τις δύο επαναλήψεις τριών TMs, πάντα τοποθετούνται η μία δίπλα στην άλλη και έχουν αντίθετες κατευθύνσεις.

1.5 Εξελικτικές αναλύσεις των μεταφορέων σακχάρων της MFS

Η φυλογενετική ανάλυση των (Wilson-O'Brien et al., 2010) ήταν από τις πρώτες εργασίες που συμπεριέλαβαν τόσο τους GLUT του ανθρώπου όσο και ορθόλογα από άλλες εξελικτικές ομάδες (συνολικά 20 οργανισμοί εκτός του ανθρώπου). Στα αποτελέσματά τους αναφέρουν ότι οι GLUT των τάξεων 1 & 2 εμφανίζονται ως καλά υποστηριζόμενοι ξεχωριστοί κλάδοι. Ο κλάδος που περιέχει τις πρωτεΐνες της πρώτης τάξης εμφανίζει κοντινά ομόλογα στην *Drosophila melanogaster* υποδεικνύοντας την ύπαρξη της συγκεκριμένης τάξης και στα ασπόνδυλα, ένα συμπέρασμα που ισχύει και για τους GLUT 6 & 8. Αντίθετα η δεύτερη τάξη εμφανίζει ορθόλογα μόνο σε ακολουθίες σπονδυλωτών. Τέλος παρατήρησαν ότι οι μεταφορείς ινοσιτόλος/μυοϊνοσιτόλης από τα

σπονδυλωτά, τα ασπόνδυλα, τους μύκητες και τα φυτά που χρησιμοποίησαν διαμορφώνουν ένα ξεχωριστό κλάδο. Λίγα χρόνια αργότερα δημοσιεύθηκε μια φυλογενετική ανάλυση όπου συμπεριλαμβάνονταν 13 εκδυσόζωα (12 έντομα & *C.elegans*), 6 μετάζωα που ανήκουν σε διαφορετικές εξελικτικές γραμμές και ορισμένοι φυτικοί οργανισμοί ως εξωομάδα. Τα αποτελέσματα της συγκεκριμένης ανάλυσης έρχονται σε συμφωνία με των (Wilson-O'Brien et al., 2010) καθώς αλληλουχίες εντόμων αποτελούν ορθόλογα των GLUT τάξης 1 των θηλαστικών. Επίσης οι GLUT 6 & 8 εμφάνισαν μεγάλο αριθμό ομόλογων ακολουθιών στα ασπόνδυλα και πιο συγκεκριμένα τους μεταφορείς Tret (Trehalose transporters) των εντόμων (Stock et al., 2013).

Η ανάλυση των (Jia et al., 2019) αποτελεί την πιο πρόσφατη και εκτεταμένη ανάλυση των μεταφορέων σακχάρων των μεταζώων. Στη μελέτη τους οι GLUT ταξινομούνται σε 3 ομάδες, όπως ήταν μέχρι τώρα γνωστό, αλλά σε κάθε ομάδα κατανέμονταν διαφορετικά ορθόλογα από τα καθιερωμένα (Deng and Yan, 2016; Wilson-O'Brien et al., 2010). Επίσης αναφέρεται ότι τα έντομα κατέχουν μεγάλο αριθμό μεταφορέων.

Τέλος, στα φυτά έχει δειχθεί η ύπαρξη επτά υποοικογενειών που έχουν ονομαστεί ανάλογα με το υποστρώματα που μεταφέρουν, το μεταλλαγμένο φαινότυπο που δημιουργούν και την κατάσταση stress που επάγει την έκφραση τους (Doidy et al., 2012). Αναλύσεις σε διάφορους φυτικούς οργανισμούς, όπως το *Arabidopsis thaliana* και το ρύζι (Johnson et al., 2006), η ντομάτα (Reuscher et al., 2014) το σταφύλι (Afoufa-Bastien et al., 2010) και το αχλάδι (Li et al., 2015) έχουν ανιχνεύσει και κατατάξει τους μεταφορείς σακχάρων τους στις 7 προαναφερθείσες υποοικογένειες.

Γενικά έχουν πραγματοποιηθεί διάφορες έρευνες για την εύρεση, το χαρακτηρισμό και την εξελικτική κατηγοριοποίηση των MFS μεταφορέων σακχάρων, ωστόσο κάθε ανάλυση από τις παραπάνω είναι περιορισμένη σε ένα εύρος εξελικτικών ομάδων ή ακόμα και σε λίγους οργανισμούς. Επίσης στη βιβλιογραφία δεν υπάρχει κάποιο εργαλείο που να μπορεί να ανιχνεύει και να ταξινομεί τους μεταφορείς σακχάρων.

2 ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ

2.1 Λογισμικό

Για την πραγμάτωση των αναλύσεων της διπλωματικής εργασίας ήταν απαραίτητη η χρήση εξειδικευμένου λογισμικού και γλωσσών προγραμματισμού τα οποία και παρατίθενται.

2.1.1 Linux Ubuntu

Τα Linux Ubuntu αποτελούν ένα λειτουργικό σύστημα ανοιχτού λογισμικού για υπολογιστές που διατίθεται δωρεάν για τους χρήστες. Το βασικό περιβάλλον τερματικού των Ubuntu χρησιμοποιεί το BASH και ενδείκνυται για τη διαχείριση μεγάλου όγκου δεδομένων (Canonical, n.d.).

2.1.2 PERL 5.26

Η PERL (Practical Extraction and Report Language) είναι μια γλώσσα προγραμματισμού που εξειδικεύεται στο χειρισμό αρχείων κειμένου και την εξαγωγή ή εισαγωγή πληροφορίας σε αυτά. Ενδείκνυται για την ανάλυση βιοπληροφορικών δεδομένων (“About Perl - www.perl.org,” n.d.).

2.1.3 Python 3.6

Η Python είναι μια ευρέως διαδομένη αντικειμενοστρεφής (object – oriented) γλώσσα προγραμματισμού παρόμοια με τη Ruby (scripting languages). Διαθέτει πλήθος βιβλιοθηκών που την καθιστούν μια γλώσσα προγραμματισμού με τεράστιο εύρος εφαρμογών (“Welcome to Python.org,” n.d.). Η συγκεκριμένη γλώσσα φέρει βιβλιοθήκες για επιστημονική έρευνα και δη βιολογική – βιοπληροφορική έρευνα όπως για παράδειγμα τη biopython.

2.1.4 MEME Suite

Πρόκειται για έναν διαδικτυακό server που μπορεί να εγκατασταθεί και τοπικά και που μπορεί να χρησιμοποιηθεί για την εύρεση motifs (μοτίβων) σε ακολουθίες που μπορεί να έχουν κάποια βιολογική σημασία. Ο αλγόριθμος MEME που βρίσκεται ενσωματωμένος στο εργαλείο ανακαλύπτει νέα, χωρίς κενά μοτίβα (επαναλαμβανόμενα και σταθερού μήκους) των αλληλουχιών εισόδου και δημιουργεί τα motif logos. Ο αλγόριθμος MAST (Motif Alignment Search Tool) στοιχίζει στις ακολουθίες επερώτησης τα MEME motifs που βρέθηκαν επιτρέποντας έτσι τη σύγκριση ακολουθιών (Bailey et al., 2009).

2.1.5 HMMER 3.1

Το λογισμικό HMMER επιτρέπει την ανάλυση αλληλουχιών μέσω profile Hidden Markov Models (profile HMMs). Τα προφίλ HMMs μετατρέπουν μια στοίχιση πολλαπλών

ακολουθιών σε ένα μοντέλο HMM για αναζήτηση απομακρυσμένων ομόλογων ακολουθιών σε βάσεις δεδομένων (Eddy, 1998). Το HMMER αποτελεί ένα ευαίσθητο και ταχύτατο εργαλείο για την εξερεύνηση μακρινών ομόλογων (“HMMER,” n.d.). Οι αλγόριθμοι που χρησιμοποιήθηκαν για την παρούσα διπλωματική εργασία είναι ο `hmmbuild` που δημιουργεί ένα `profile hidden markov model` από μια πολλαπλή στοίχιση, ο `hmmsearch` που μετατρέπει ένα HMM σε κατάλληλη μορφή για να χρησιμοποιηθεί ως βάση δεδομένων για την εύρεση των ομόλογων, ο `hmmalign` που πραγματοποιεί την εξερεύνηση πρωτεϊνικών αλληλουχιών έναντι των προφίλ HMM που έχουν δημιουργηθεί και ο `hmmalign` που δημιουργεί πολλαπλή στοίχιση ακολουθιών στοιχίζοντας την καθεμία ξεχωριστά σε ένα `profile Hidden Markov Model`.

2.1.6 ETE 3

Το λογισμικό ETE 3 (Environment for Tree Exploration) με το πακέτο εντολών `EteToolkit` χρησιμοποιήθηκαν για την απεικόνιση, ανάλυση, επεξεργασία και χειρισμό φυλογενετικών δένδρων. Το ETE διαθέτει μια βιβλιοθήκη της Python που χρησιμοποιήθηκε για τον αυτοματισμό της επεξεργασίας των φυλογενετικών δεδομένων. Μερικά από τα εργαλεία που διαθέτει είναι ο χαρακτηρισμός των κλαδιών, η αυτόματη ανίχνευση ορθολόγων και παραλόγων και η ανεξάρτητη επεξεργασία διαφορετικών τμημάτων του ίδιου φυλογενετικού δένδρου (Huerta-Cepas et al., 2016).

2.1.7 Seaview 4.6

Το λογισμικό Seaview μπορεί να χρησιμοποιηθεί σε όλα τα λειτουργικά συστήματα (Linux, Windows, MacOS) και έχει σχεδιαστεί για να πραγματοποιεί πολλαπλές στοιχίσεις ακολουθιών DNA ή πρωτεϊνών και να δημιουργεί φυλογενετικά. Το συγκεκριμένο λογισμικό υποστηρίζει τόσο γραφικό περιβάλλον (Graphical User Interface – GUI) όσο και περιβάλλον τερματικού (Command Line Interface – CLI) (Gouy et al., 2010). Για τις πολλαπλές στοιχίσεις δίνεται η δυνατότητα χρήσης τόσο του αλγορίθμου Clustal Omega όσο και του MUSCLE, ο οποίος και χρησιμοποιήθηκε στη παρούσα εργασία. Ο υπολογισμός των φυλογενετικών δένδρων μπορεί να γίνει είτε με μεθόδους αποστάσεων και πιο συγκεκριμένα με Neighbor Joining είτε με μεθόδους χαρακτήρων, όπως η μέγιστη φειδωλότητα (Maximum Parsimony) ή η μέγιστη πιθανοφάνεια (Maximum Likelihood – Αλγόριθμος PhyML).

2.1.8 Interactive Tree of Life (iTOL)

Το interactive Tree of Life αποτελεί ένα διαδικτυακό πρόγραμμα, ανοιχτό προς το κοινό, με σκοπό την ευκολία στην απεικόνιση, το χειρισμό και το σχολιασμό φυλογενετικών δένδρων και την αποθήκευση των αποτελεσμάτων μέσω προσωπικού κωδικού χρήστη. Σαν αρχεία εισόδου υποστηρίζει όλα τα αρχεία φυλογενετικών δένδρων κοινού τύπου όπως Newick, Nexus και PhyloXML (Letunic and Bork, 2019). Βασική χρήση του στην εργασία ήταν ο χρωματισμός και η πραγματοποίηση της διεργασίας collapse (κατάρρευση) κλάδων με τιμή bootstrap χαμηλότερη από ένα όριο.

2.1.9 Usearch

Στη σημερινή εποχή όπου η βιολογική πληροφορία συσσωρεύεται συνεχώς με ταχείς ρυθμούς είναι απαραίτητη η χρήση αλγορίθμων που μπορούν ταχύτατα να πραγματοποιούν αναζητήσεις και ομαδοποιήσεις μέσα στον τεράστιο αυτό όγκο πληροφορίας. Το Usearch μέσω του αλγορίθμου Uclust έχει την ικανότητα σε πολύ σύντομο χρόνο να ομαδοποιεί πρωτεϊνικές ή νουκλεοτιδικές αλληλουχίες με βάση την ομοιότητά τους (Edgar, 2010).

2.2 Λήψη πρωτεωμάτων

Για την εργασία χρησιμοποιήθηκαν πρωτεώματα από πλήθος αντιπροσώπων που προέρχονται από όλα τα βασίλεια, δηλαδή των φυτών, των ζώων, των μυκήτων, των πρωτίστων και των προκαρυωτών. Τα πρωτεώματα των οργανισμών που χρησιμοποιήθηκαν, εκτός των φυτικών και των προκαρυωτικών πρωτεωμάτων μαζί με τις πηγές τους αναγράφονται στην εικόνα Figure 13. Στην περίπτωση των προκαρυωτικών πρωτεωμάτων δεν ήταν δυνατή η απεικόνιση, καθώς έγινε η λήψη 8226 από βακτήρια και αρχαία ποικίλλων ταξινομικών ομάδων. Τα πρωτεώματα των φυτικών οργανισμών παρελήφθησαν από τη διπλωματική εργασία της κ. Διακογεωργίου που εκπονήθηκε στο εργαστήριο μας (“ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΕΞΕΛΙΚΤΙΚΗ ΑΝΑΛΥΣΗ ΤΩΝ ΜΕΤΑΦΟΡΕΩΝ ΣΑΚΧΑΡΩΝ ΤΗΣ MAJOR FACILITATOR SUPERFAMILY,” 2018). Η λήψη από τις βάσεις δεδομένων έγινε μέσω bash script.

Όργανισμός	Μεγάλη ταξινόμηση ομάδα	Σύνολο πρωτεϊνών	Πηγή
<i>Anphimedon queenslandica</i>	Basal animals	43615	http://oct2017-metazoa.ensembl.org/species.html
<i>Mnemiopsis leidyi</i>	Basal animals	16058	http://oct2017-metazoa.ensembl.org/species.html
<i>Nematostella vectensis</i>	Basal animals	24780	http://oct2017-metazoa.ensembl.org/species.html
<i>Schistosoma mansoni</i>	Basal animals	11774	http://oct2017-metazoa.ensembl.org/species.html
<i>Trichoplax adhaerens</i>	Basal animals	11520	http://oct2017-metazoa.ensembl.org/species.html
<i>Ciona intestinalis</i>	Deuterostomes	17302	https://www.ensembl.org/info/about/species.html
<i>Eptatretus burgeri</i>	Deuterostomes	27960	https://www.ensembl.org/info/about/species.html
<i>Gallus gallus</i>	Deuterostomes	30252	https://www.ensembl.org/info/about/species.html
<i>Gorilla gorilla</i>	Deuterostomes	45194	https://www.ensembl.org/info/about/species.html
<i>Homo sapiens</i>	Deuterostomes	107844	https://www.ensembl.org/info/about/species.html
<i>Latimeria chalumnae</i>	Deuterostomes	23601	https://www.ensembl.org/info/about/species.html
<i>Mus musculus</i>	Deuterostomes	65896	https://www.ensembl.org/info/about/species.html
<i>Ornithorhynchus anatinus</i>	Deuterostomes	23584	https://www.ensembl.org/info/about/species.html
<i>Pan troglodytes</i>	Deuterostomes	49949	https://www.ensembl.org/info/about/species.html
<i>Petromyzon marinus</i>	Deuterostomes	11442	https://www.ensembl.org/info/about/species.html
<i>Pongo abelii</i>	Deuterostomes	21414	https://www.ensembl.org/info/about/species.html
<i>Strongylocentrotus purpuratus</i>	Deuterostomes	28886	http://oct2017-metazoa.ensembl.org/species.html
<i>Takifugu rubripes</i>	Deuterostomes	47841	https://www.ensembl.org/info/about/species.html
<i>Tetraodon nigroviridis</i>	Deuterostomes	23118	https://www.ensembl.org/info/about/species.html
<i>Xenopus tropicalis</i>	Deuterostomes	22718	https://www.ensembl.org/info/about/species.html
<i>Arthrobotrys oligospora</i>	Fungi	11479	https://fungi.ensembl.org/info/website/ftp/index.html
<i>Aspergillus nidulans</i>	Fungi	10534	https://fungi.ensembl.org/info/website/ftp/index.html
<i>Candida albicans</i>	Fungi	6392	https://fungi.ensembl.org/info/website/ftp/index.html
<i>Debaryomyces hansenii</i>	Fungi	6272	https://fungi.ensembl.org/info/website/ftp/index.html
<i>Eremothecium gossypii</i>	Fungi	4774	https://fungi.ensembl.org/info/website/ftp/index.html
<i>Fusarium graminearum</i>	Fungi	14145	https://fungi.ensembl.org/info/website/ftp/index.html
<i>Kluyveromyces fragilis</i>	Fungi	5074	https://fungi.ensembl.org/info/website/ftp/index.html
<i>Saccharomyces cerevisiae</i>	Fungi	5917	https://fungi.ensembl.org/info/website/ftp/index.html
<i>Schizosaccharomyces pombe</i>	Fungi	5146	https://fungi.ensembl.org/info/website/ftp/index.html
<i>Sclerotinia sclerotiorum</i>	Fungi	10175	https://fungi.ensembl.org/info/website/ftp/index.html
<i>Xylaria heveae</i>	Fungi	8201	https://fungi.ensembl.org/info/website/ftp/index.html
<i>Yarrowia lipolytica</i>	Fungi	6448	https://fungi.ensembl.org/info/website/ftp/index.html
<i>Zygosaccharomyces rouxii</i>	Fungi	4991	https://fungi.ensembl.org/info/website/ftp/index.html
<i>Bigeloviella natans</i>	Protists	21706	http://protists.ensembl.org/species.html
<i>Cryptosporidium parvum</i>	Protists	473	http://protists.ensembl.org/species.html
<i>Dictyostelium discoideum</i>	Protists	13267	http://protists.ensembl.org/species.html
<i>Emiliania huxleyi</i>	Protists	38554	http://protists.ensembl.org/species.html
<i>Eutima histolytica</i>	Protists	8113	http://protists.ensembl.org/species.html
<i>Giardia intestinalis</i>	Protists	6098	http://protists.ensembl.org/species.html
<i>Guillardia theta</i>	Protists	24969	http://protists.ensembl.org/species.html
<i>Leishmania major</i>	Protists	8315	http://protists.ensembl.org/species.html
<i>Monosiga brevicollis</i> MX1	Protists	9172	http://protists.ensembl.org/species.html
<i>Paramecium tetraurelia</i>	Protists	39642	http://protists.ensembl.org/species.html
<i>Phaeodactylum tricostatum</i>	Protists	12178	http://protists.ensembl.org/species.html
<i>Phytophthora infestans</i>	Protists	17787	http://protists.ensembl.org/species.html
<i>Plasmodium falciparum</i> 7G8	Protists	6314	http://protists.ensembl.org/species.html
<i>Thalassiosira pseudonana</i>	Protists	11673	http://protists.ensembl.org/species.html
<i>Toxoplasma gondii</i> AFR	Protists	9958	http://protists.ensembl.org/species.html
<i>Trypanosoma brucei</i>	Protists	8747	http://protists.ensembl.org/species.html
<i>Adineta vaga</i>	Protostomes	49300	http://oct2017-metazoa.ensembl.org/species.html
<i>Anopheles gambiae</i>	Protostomes	14870	http://oct2017-metazoa.ensembl.org/species.html
<i>Apis mellifera</i>	Protostomes	15314	http://oct2017-metazoa.ensembl.org/species.html
<i>Bombyx mori</i>	Protostomes	14623	http://oct2017-metazoa.ensembl.org/species.html
<i>Caenorhabditis elegans</i>	Protostomes	31574	http://oct2017-metazoa.ensembl.org/species.html
<i>Capitella teleta</i>	Protostomes	32175	http://oct2017-metazoa.ensembl.org/species.html
<i>Crassostrea gigas</i>	Protostomes	26089	http://oct2017-metazoa.ensembl.org/species.html
<i>Daphnia pulex</i>	Protostomes	30611	http://oct2017-metazoa.ensembl.org/species.html
<i>Drosophila melanogaster</i>	Protostomes	30362	http://oct2017-metazoa.ensembl.org/species.html
<i>Ixodes scapularis</i>	Protostomes	20486	http://oct2017-metazoa.ensembl.org/species.html
<i>Lingula anatina</i>	Protostomes	34105	http://oct2017-metazoa.ensembl.org/species.html
<i>Pediculus humanus</i>	Protostomes	10788	http://oct2017-metazoa.ensembl.org/species.html
<i>Rhodnius prolixus</i>	Protostomes	15456	http://oct2017-metazoa.ensembl.org/species.html
<i>Strigamia maritima</i>	Protostomes	15008	http://oct2017-metazoa.ensembl.org/species.html
<i>Zootermopsis nevadensis</i>	Protostomes	14610	http://oct2017-metazoa.ensembl.org/species.html
Prokaryotes	Prokaryotes		ftp.ncbi.nlm.nih.gov

Figure 13: Τα πρωτεώματα που χρησιμοποιήθηκαν σε συνδυασμό με τις πηγές τους. Από τη συγκεκριμένη εικόνα απουσιάζουν τόσο τα φυτικά πρωτεώματα, καθώς έχουν παραληφθεί από άλλη εργασία, όσο και τα προκαρυωτικά πρωτεώματα, λόγω πλήθους.

3 ΑΠΟΤΕΛΕΣΜΑΤΑ

3.1 Απομόνωση ευκαρυωτικών μεταφορέων σακχάρων της MFS

Από τα ευκαρυωτικά πρωτεώματα που λήφθηκαν πραγματοποιήθηκε απομόνωση των πρωτεϊνών που ανήκουν στην υπεροικογένεια MFS με χρήση των Hidden Markov Models των υποοικογενειών 2.A.1.1, 2.A.1.9 και 2.A.1.19 που έχουν βρεθεί από τις πρωτεΐνες μεταφορείς σακχάρων της υπεροικογένειας MFS, όπως αναφέρεται στη διπλωματική εργασία (“ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΕΞΕΛΙΚΤΙΚΗ ΑΝΑΛΥΣΗ ΤΩΝ ΜΕΤΑΦΟΡΕΩΝ ΣΑΚΧΑΡΩΝ ΤΗΣ MAJOR FACILITATOR SUPERFAMILY,” 2018). Ως παράμετροι για την αναζήτηση επιλέχθηκε το κατώφλι E-value $1e-30$, καθώς αυτή ήταν ικανή να μπορεί να διαχωρίζει τις ομόλογες υποοικογένειες σύμφωνα με τις αναλύσεις της κ. Διακογεωργίου. Στη συνέχεια, στις απομονωμένες πλέον ακολουθίες χρησιμοποιήθηκε ο αλγόριθμος MAST, που πραγματοποιεί στοίχιση μοτίβων, με τα MEME motifs που είχαν αναπτυχθεί για την υποοικογένεια 2.A.1.1 αποσκοπώντας στο επιπρόσθετο φιλτράρισμα των ακολουθιών. Ειδικότερα, η ύπαρξη τουλάχιστον 5 εκ των 15 μοτίβων στις ακολουθίες χρησιμοποιήθηκε ως όριο. Τελικά συλλέχθηκαν 1714 πρωτεϊνικές ακολουθίες.

3.2 Δημιουργία MEME Motifs των ευκαρυωτικών ακολουθιών

Οι 1714 ευκαρυωτικές MFS ακολουθίες δόθηκαν ως αρχείο εισόδου στο Usearch και πραγματοποιήθηκε clustering με ποσοστό ομοιότητας 70% με αποτέλεσμα τον εντοπισμό 955 ακολουθιών. Με τη βοήθεια του προγράμματος MEME από το server MEME Suite ανακαλύφθηκαν 15 MEME motifs στις παραπάνω ακολουθίες με τη χρήση παραμέτρων “meme -protein -oc . -nostatus -time 18000 -mod zoops -nmotifs 15 -minw 6 -maxw 50 -objfun classic -markov_order 0”. Τα νέα αυτά ευκαρυωτικά motifs στοιχήθηκαν στις 1714 ακολουθίες με τον αλγόριθμο MAST, επιτρέποντας μας να διακρίνουμε 9 που ήταν παρόντα στις ακολουθίες σε ποσοστό άνω του 90%. Για να αυξήσουμε την ευαισθησία και τη πιστότητα της ανάλυσης η παρουσία των ευκαρυωτικών motifs χρησιμοποιήθηκε ως επόμενο κριτήριο φιλτραρίσματος, όπου για να θεωρηθεί η ακολουθία επερωτήσης μεταφορέας σακχάρων της υπεροικογένειας MFS, θα πρέπει να παρουσιάζει γενικά 11 από τα 15 μοτίβα, εκ των οποίων τα τουλάχιστον 7 να αποτελούν τα συχνότερα εμаниζόμενα, όπως προαναφέρθηκε. Μετά το συγκεκριμένο γύρο φιλτραρίσματος, 1370 ευκαρυωτικές ακολουθίες παρέμειναν και χρησιμοποιήθηκαν για την περαιτέρω ανάλυση.

3.3 Αναγνώριση των προκαρυωτικών μεταφορέων σακχάρων MFS

Για να διερευνήσουμε την ύπαρξη μεταφορέων σακχάρων στα βακτήρια και στα αρχαία, υποβάλλαμε τα πρωτεώματα τους κατά αντιστοιχία με τους ευκαρυώτες για αναζήτηση πρωτεϊνών που είναι ομόλογες της MFS 2.A.1.1 υποοικογένειας και για τον έλεγχο των συγκεκριμένων ομόλογων πρωτεϊνών που πληρούν το κριτήριο της ύπαρξης motif 7-11, καταλήγοντας τελικά σε 11342 πρωτεΐνες. Καθώς τα 8226 πρωτεώματα

αποτελούν βασικό εμπόδιο κατά την απεικόνιση των δεδομένων σε φυλογενετικά δέντρα, λόγω όγκου, επιλέχθηκαν ως αντιπροσωπευτικοί προκαρυώτες τα στελέχη που είχαν τις περισσότερες πρωτεΐνες MFS από κάθε μεγάλη ταξινομική ομάδα και παρουσιάζονται στην εικόνα Figure 14.

Organism	NCBI TaxId	Assembly	Sugar Transporters
Rhodococcus opacus PD630	543736	GCF_000599545	14
Bacillus filamentosus	1402861	GCF_000972245	11
Asaia bogorensis NBRC 16594	1231624	GCF_001547995	9
Klebsiella oxytoca	571	GCF_001022115	7
Acidobacterium capsulatum ATCC 51196	240015	GCF_000022565	5
Sphingobacterium sp. 21	743722	GCF_000192845	5
Polaribacter reichenbachii	996801	GCF_001975665	5
Sedimentisphaera salicampi	1941349	GCF_002117005	5
Coralimargarita akajimensis DSM 45221	583355	GCF_000025905	4
Pelosinus sp. UFO1	484770	GCF_000725345	4
Niabella ginsenosidivorans	1176587	GCF_001654455	4
Bacteroides fragilis	817	GCF_002849695	4
Escherichia coli K12	316385	GCF_000005845	3
Haloterrigena turkmenica DSM 5511	543526	GCF_000025325	3
Clostridium tyrobutyricum	1519	GCF_001642655	3
Thermogutta terrifontis	1331910	GCF_002277955	3
maquilingensis IC-167	397948	GCF_000018305	2
Thermoplasma acidophilum DSM 1728	273075	GCF_000195915	2
Cyclobacterium marinum DSM 745	880070	GCF_000222485	2
Burkholderia sp. Y123	1097668	GCF_000236065	2
Synechocystis sp. PCC 6803	1148	GCF_000009725	1
Salinibacter ruber DSM 13855	309807	GCF_000013045	1
Conexibacter woesei DSM 14684	469383	GCF_000025265	1
Cyanotheca sp. PCC 7822	497965	GCF_000147335	1
Methanobacterium lacus	877455	GCF_000191585	1
Haliscomenobacter hydrossis DSM 1100	760192	GCF_000212735	1
Parachlamydia acanthamoebae UV-7	765952	GCF_000253035	1
Rubrobacter radiotolerans	42256	GCF_000661895	1
Archangium gephyra	48	GCF_001027285	1
Endomicrobium proavitum	1408281	GCF_001027545	1
Campylobacter jejuni	197	GCF_002238375	1
Stanieria sp. NIES-3757	1807358	GCF_002355455	1
Nostoc flagelliforme CCNUN1	2038116	GCF_002813575	1
Akkermansia glycaniphila	1679444	GCF_900097105	1

Figure 14: Οι αντιπροσωπευτικοί προκαρυώτες.

3.4 Ανίχνευση εξελικτικών ομάδων στους μεταφορείς

Οι ευκαρυωτικές ακολουθίες χωρίστηκαν ανά βασίλειο (ζώα, φυτά, μύκητες, πρώτιστα και προκαρυώτες) και έγιναν διαδοχικά profile alignments εντός του Seaview. Η τελική στοίχιση που προέκυψε χρησιμοποιήθηκε για τον υπολογισμό φυλογενετικού δένδρου με μέθοδο αποστάσεων μέσω του αλγορίθμου BioNJ και αξιολογήθηκε με την χρήση τόσο εκατό bootstrap επαναλήψεων, όσο και μέσω εκατό επαναλήψεων της μεθόδου Transfer Bootstrap Expectation (TBE) που πραγματοποιεί μια παρεμφερή αλλά διαφορετική διαδικασία αξιολόγησης δένδρου από το κλασσικό bootstrap (Lemoine et al., 2018). Τα φυλογενετικά δένδρα, μέσω ανάλυσης με την βοήθεια των λογισμικών EteToolkit και interactive Tree of Life (iTOL) αποτέλεσαν την βάση για την ανίχνευση 37 εξελικτικών ομάδων. Οι ακολουθίες των συγκεκριμένων ομάδων στοιχήθηκαν και κατόπιν, χρησιμοποιήθηκαν για τη δημιουργία Hidden Markov Models.

3.5 Κατανομή των μεταφορέων σακχάρων MFS στους οργανισμούς

Οι οργανισμοί που χρησιμοποιήθηκαν στην ανάλυση κατηγοριοποιήθηκαν με βάση την ταξινόμηση τους και δημιουργήθηκε το γράφημα **Illustration 1**, όπου φαίνεται η κατανομή των μεταφορέων σακχάρων για τις συγκεκριμένες εξελικτικές ομάδες. Οι προκαρυωτικές εξελικτικές ομάδες εμφανίζουν ελάχιστα ομόλογα των συγκεκριμένων μεταφορέων. Στους φυτικούς οργανισμούς παρατηρούμε πολλά ομόλογα και σε συνδυασμό με τα δεδομένα της εργασίας (“ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΕΞΕΛΙΚΤΙΚΗ ΑΝΑΛΥΣΗ ΤΩΝ ΜΕΤΑΦΟΡΕΩΝ ΣΑΚΧΑΡΩΝ ΤΗΣ MAJOR FACILITATOR SUPERFAMILY,” 2018) γνωρίζουμε ότι τα περισσότερα ομόλογα εμφανίζονται στα φυτά της ξηράς με τον κοινό τους πρόγονο να υπήρξε περίπου 500 εκατομμύρια χρόνια πριν. Στα πρώτιστα παρατηρείται μια κατανομή των ομολόγων όμοια με εκείνη των κατώτερων ζωικών οργανισμών, υποδεικνύοντας ότι ο κοινός πρόγονος των μυκήτων, ζώων και πρώτιστων είχε μικρό αριθμό μεταφορέων σακχάρων της συγκεκριμένης υπεροικογένειας. Ιδιαίτερο ενδιαφέρον υπάρχει στην κατανομή των ζωϊκών ομολόγων. Αν και στον κοινό πρόγονο των ζώων φαίνεται να υπήρχαν μόνο λίγες ομόλογες πρωτεΐνες στην εξελικτική γραμμή των πρωτοστόμιων και κυρίως στα εξάποδα (έντομα) παρατηρούνται πολλοί γονιδιακοί διπλασιασμοί, γεγονός που έρχεται σε συμφωνία με την βιβλιογραφία. Στη πιο πρόσφατη εξελικτική ανάλυση των ομολόγων των GLUT στα μετάζωα οι συγγραφείς αναφέρουν πως περίπου το 38% των ακολουθιών προερχόταν από έντομα (Jia et al., 2019). Σύμφωνα τόσο με τα δικά μας, όσο και με τα δημοσιευμένα αποτελέσματα μπορούμε να θεωρήσουμε ότι στην εξελικτική πορεία των εντόμων πραγματοποιήθηκαν εκτεταμένοι γονιδιακοί διπλασιασμοί, υπόθεση που έρχεται να συμπληρώσει η περσινή δημοσίευση (Li et al., 2018) που πιστοποιούν την πραγματοποίηση πολλαπλών γονιδιακών και γενωμικών διπλασιασμών στα εξάποδα (το υπόφυλο που περιέχει τα έντομα).

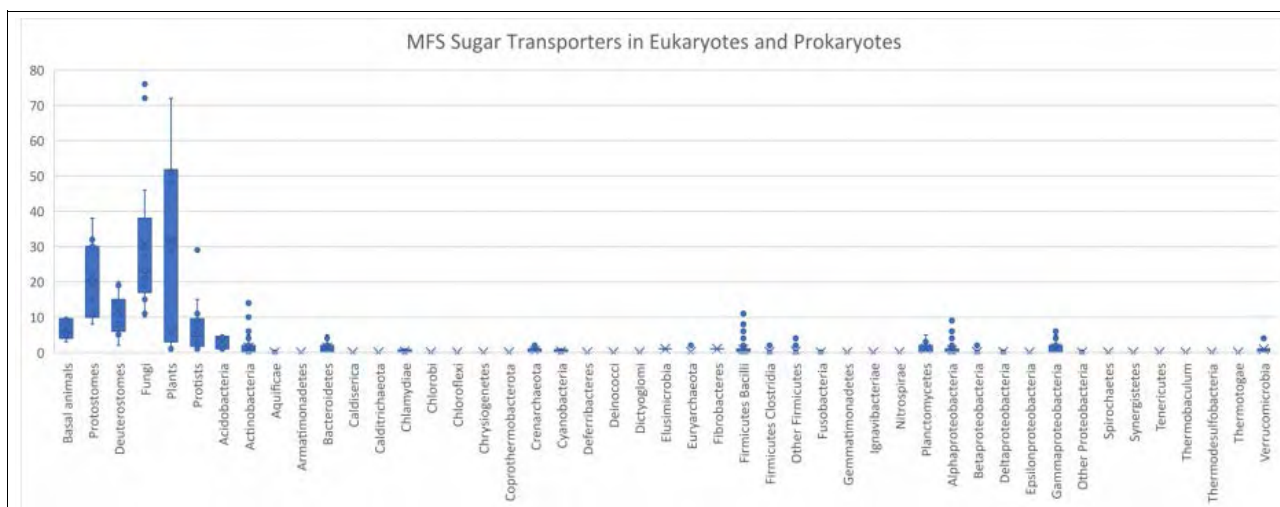


Illustration 1: Η κατανομή των μεταφορέων σακχάρων MFS στις διάφορες εξελικτικές ομάδες.

3.6 Εξελικτικές ομάδες των μεταφορέων σακχάρων

Η δημιουργία φυλογενετικού δέντρου και η αξιολόγηση του μέσω της μεθόδου bootstrap μας έδωσε τη δυνατότητα να διακρίνουμε ορισμένους κλάδους με μεγάλη αξιοπιστία (τιμή bootstrap >70) παρά τη μεγάλη εξελικτική απόσταση που υπάρχει στο σύνολο των ακολουθιών. Προκειμένου να θεωρηθεί ένας κλάδος ξεχωριστή εξελικτική ομάδα θεωρήσαμε ότι θα πρέπει να περιέχει τουλάχιστον 6 ακολουθίες από τουλάχιστον 6 οργανισμούς και τιμή bootstrap μεγαλύτερη του 70%. Σύμφωνα με το παραπάνω οδηγηθήκαμε στην αναγνώριση 37 διαφορετικών εξελικτικών ομάδων οι οποίες επεξεργάστηκαν ξεχωριστά και δημιουργήθηκαν μεμονωμένα Hidden Markov Models. Η μέθοδος αξιολόγησης δέντρου Transfer Bootstrap Expectation έχει τη δυνατότητα να αξιολογεί τους κλάδους ενός δέντρου, με μια διαφορετική μέθοδο από το bootstrap, παράγοντας μεγάλες τιμές για εξελικτικά απομακρυσμένους κλάδους αποσαφηνίζοντας σχέσεις που η κλασική μέθοδος θα αποτύγχανε (Lemoine et al., 2018). Με το συγκεκριμένο δεδομένο χρησιμοποιήθηκε το δέντρο με τις τιμές TBE προκειμένου να συμπεριλάβουμε πιο μακρινά ομόλογα στις διάφορες ομάδες παράγοντας τις λεγόμενες διευρημένες ομάδες (extended groups), όπου αυτό είναι δυνατό, αυξάνοντας με αυτό το τρόπο την ευαισθησία των μεμονωμένων HMM.

Στο φυλογενετικό δένδρο παρατηρούμε ότι σχεδόν όλες οι ακολουθίες μυκήτων σχηματίζουν ένα εξελικτικό group Figure 15. Στο κλάδο των μυκητιακών ακολουθιών υπάρχουν ακολουθίες από την *Adineta vaga* (bdelloid rotifer) που αποτελεί ένα πρωτοστόμιο. Θεωρήθηκε ότι η ύπαρξη του συγκεκριμένου οργανισμού στο συγκεκριμένο κλάδο οφείλεται σε γεγονός οριζόντιας μεταφοράς από τους μύκητες, γεγονός που υποστηρίζεται από τη βιβλιογραφία καθώς έχουν παρατηρηθεί εκτεταμένα περιστασιακά οριζόντια μεταφοράς γονιδίων στο συγκεκριμένο φύλο από μύκητες (Gladyshev et al., 2008).

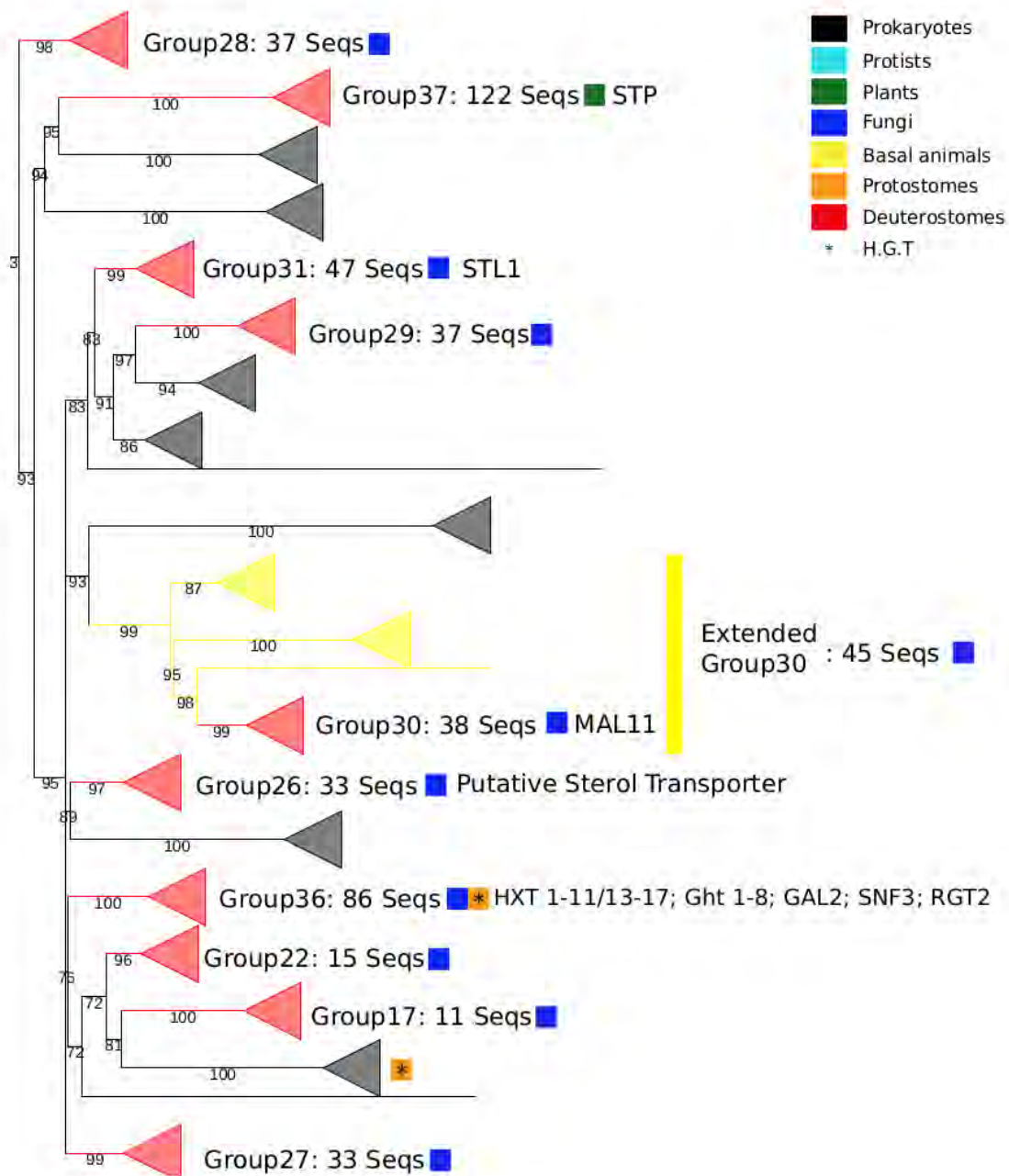


Figure 15: Ο κλάδος που περιέχει της ακολουθίες των μυκήτων. Στα φυλογενετικά δένδρα με κόκκινο και κίτρινο εμφανίζονται οι χαρακτηρισμένες εξελικτικές ομάδες με κίτρινο χρώμα υποδεικνύονται οι διευρυμένες ομάδες που ανιχνεύθηκαν μέσω των τιμών TBE. Οι ακολουθίες που δεν ανήκουν σε κάποια εξελικτική ομάδα έχουν γκρι χρώμα. Τα διάφορα βασίλεια απεικονίζονται ως χρωματισμένα τετράγωνα όπως φαίνεται και στο άνω δεξί τμήμα της εικόνα και ο αστερίσκος υποδηλώνει την ύπαρξη οριζόντιας μεταφοράς γονιδίων (*Adineta vaga*).

Έχει προταθεί ότι οι GLUT των θηλαστικών χωρίζονται σε τρεις τάξεις (GLUT Class I, Class II και Class III) οι οποίες δεν συγκεντρώνονται σε ένα μονοφυλετικό κλάδο (Wilson-O'Brien et al., 2010). Στην ανάλυση μας βρήκαμε ότι οι τάξεις I και II Figure 16 εμφανίζονται σε δύο διαφορετικούς κλάδους αλλά με σαφή σχέση καθώς ο πατρικός κόμβος υποστηρίζεται με τιμή bootstrap 95%. Η ομάδα 34 που περιέχει τους GLUT της τάξης I εμφανίζεται συγγενική με την ομάδα 20 που περιέχει τον Glut1 από τη *Drosophilla melanogaster* και άλλων πρωτοστόμιων ζώων, όπως υποστηρίζεται και από την υπάρχουσα βιβλιογραφία (Wilson-O'Brien et al., 2010). Ο συγγενικός τους κλάδος που περιέχει τους class II GLUT εμφανίζει ομόλογα μόνο από δευτεροστόμια ζώα και πιο συγκεκριμένα στην εξελικτική γραμμή των κρυνιωτών.

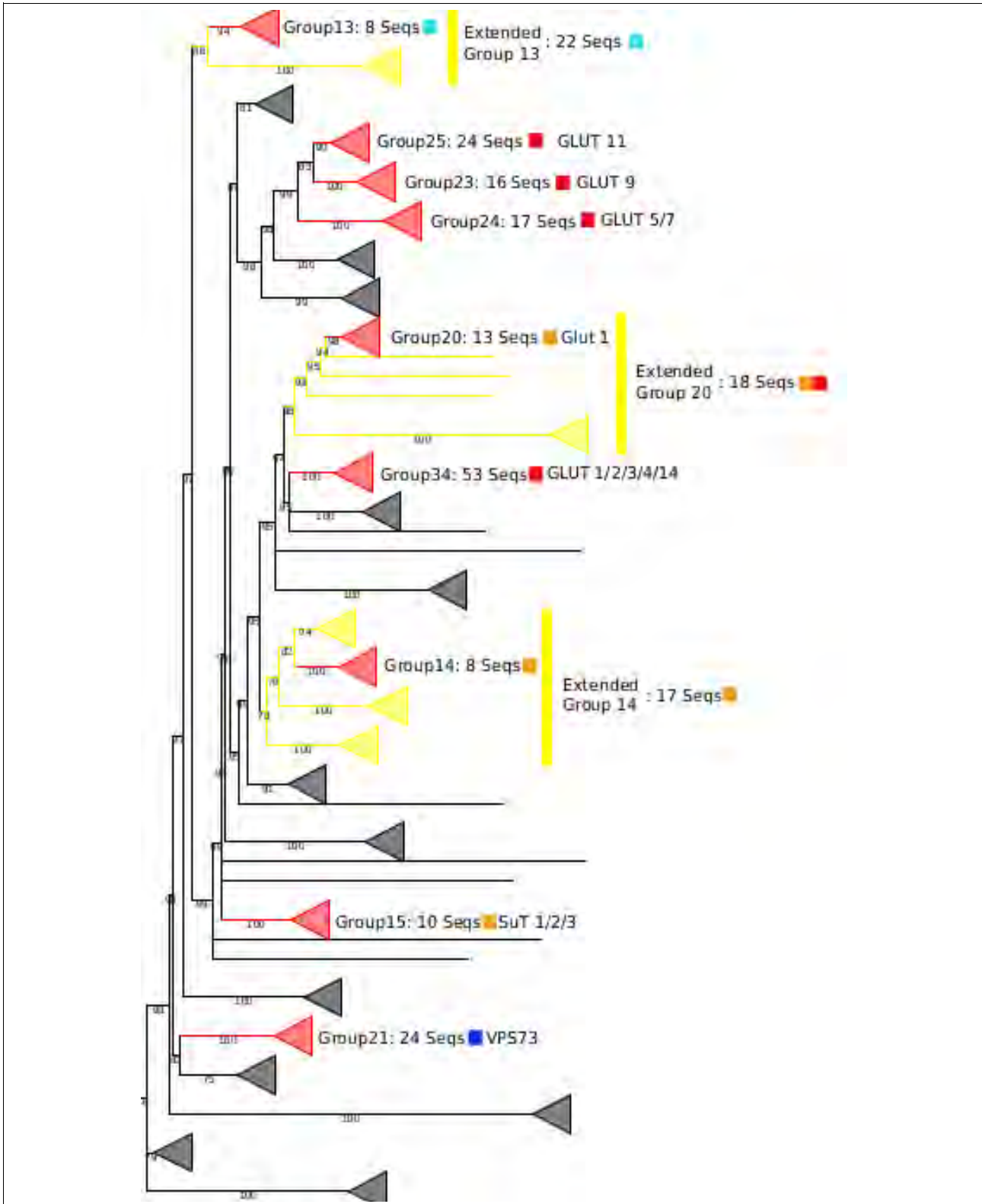


Figure 16: Η τοπολογία των GLUT της κλάσης 1 & 2.

Ο κλάδος που περιέχει τους GLUT 6 & 8 Figure 17 εμπεριέχει αναγνωρισμένες ομάδες στις οποίες κυριαρχούν πρωτοστόμιοι οργανισμοί και πιο συγκεκριμένα αρθρόποδα, ωστόσο στον κλάδο εμπεριέχονται και μακρινές ομόλογες ακολουθίες από κνιδάρια και πλακόζωα αν και δεν έχουν αναγνωριστεί να ανήκουν σε κάποια ομάδα σύμφωνα με τα κριτήρια μας. Οι (Wilson-O'Brien et al., 2010) στην εξελικτική ανάλυση που πραγματοποίησαν στους GLUT αναφέρουν ότι οι συγκεκριμένοι μεταφορείς έχουν κοντινά ομόλογα από διάφορα ασπόνδυλα, ωστόσο τονίζουν ότι οι συγκεκριμένοι μεταφορείς μπορεί να έχουν μακρινότερα ομόλογα αλλά λόγω έλλειψης δεδομένων ένα τέτοιο συμπέρασμα δεν είναι σίγουρο. Σε μια μελέτη που δημοσιεύθηκε φέτος οι ερευνητές συμπεριέλαβαν περισσότερους οργανισμούς, αποκτώντας μια γενικότερη εικόνα για τις φυλογενετικές σχέσεις των GLUT και των ομολόγων τους στα μετάζωα (Jia et al., 2019). Στη μελέτη του 2019 οι ερευνητές ανακάλυψαν ότι οι GLUT 6 & 8 του ανθρώπου εμφανίζουν μεγάλη ομοιότητα με τους μεταφορείς τρεχαλόζης (Tret 1&2) της *Drosophila melanogaster* και φυλογενετικά τους κατατάσσουν σε μια ομάδα ξεχωριστή από τους GLUT 10 & 12, που θεωρούνται τα πιο κοντινά ανθρώπινα ομόλογα (Deng and Yan, 2016). Η υπάρχουσα βιβλιογραφία φυλογενετικών αναλύσεων για τους GLUT ενισχύει τα αποτελέσματα μας αν και πρόκειται για μελέτες μικρότερης εμβέλειας (με λιγότερο αριθμό οργανισμών). Τελικά μπορούμε να συμπεράνουμε ότι οι συγκεκριμένοι μεταφορείς υπήρχαν στον πρόγονο των ζώων με αμφίπλευρη συμμετρία και των πλακόζωων περίπου 948 εκατομμύρια χρόνια πριν. Στη κλάση 3 βρίσκονται και οι GLUT 10 & 12, ωστόσο ο GLUT 10 δεν πέρασε το κριτήριο ύπαρξης 7/11 meme motifs (το πλήθος των meme motif του ήταν 6/12) και ο GLUT 12 τοποθετείται σε ένα διαφορετικό εξελικτικό γκρουπ (γκρουπ 8).

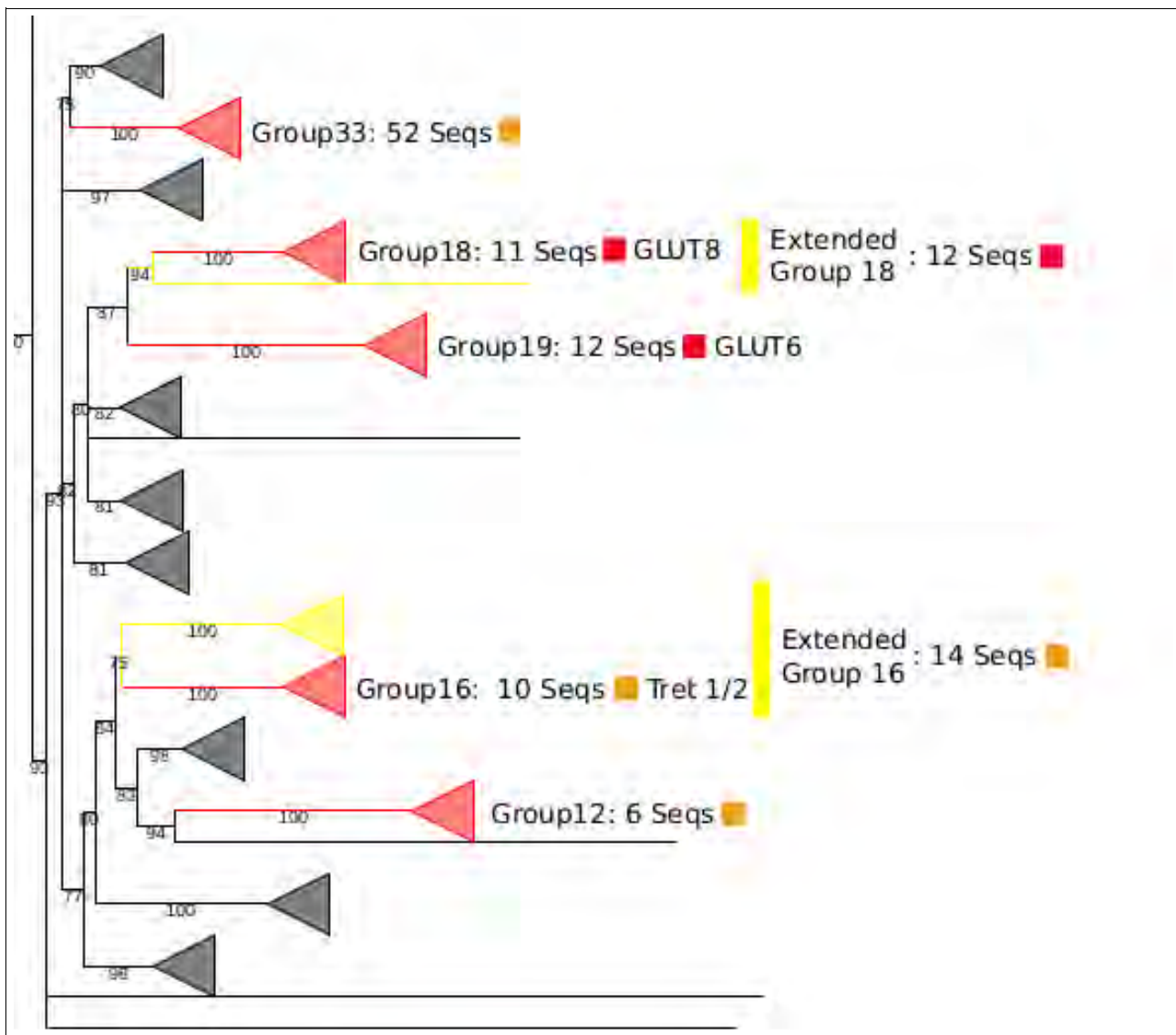


Figure 17: Η τοπολογία μερικών GLUT της τάξης 3.

Οι μεταφορές που μεταφέρουν ινοσιτόλη και μυο-ινοσιτόλη από όλους τους οργανισμούς συγκεντρώνονται σε μια μεγάλη ομάδα το Group 11, που αποτελεί και έναν μονοφυλετικό κλάδο με πολύ υψηλή αξιοπιστία (99% bootstrap). Η συγκεκριμένη ομάδα φανερώνει ότι οι μεταφορές ινοσιτόλης σε αντίθεση με τους GLUT και τα κοντινά τους ομόλογα, που υπάρχουν στους πολυκύτταρους οργανισμούς, εμφανίζονται τόσο σε μονοκύτταρους ευκαρυώτες (*Bigelowiella natans*; cryptophyta) όσο και σε πολυκύτταρους.

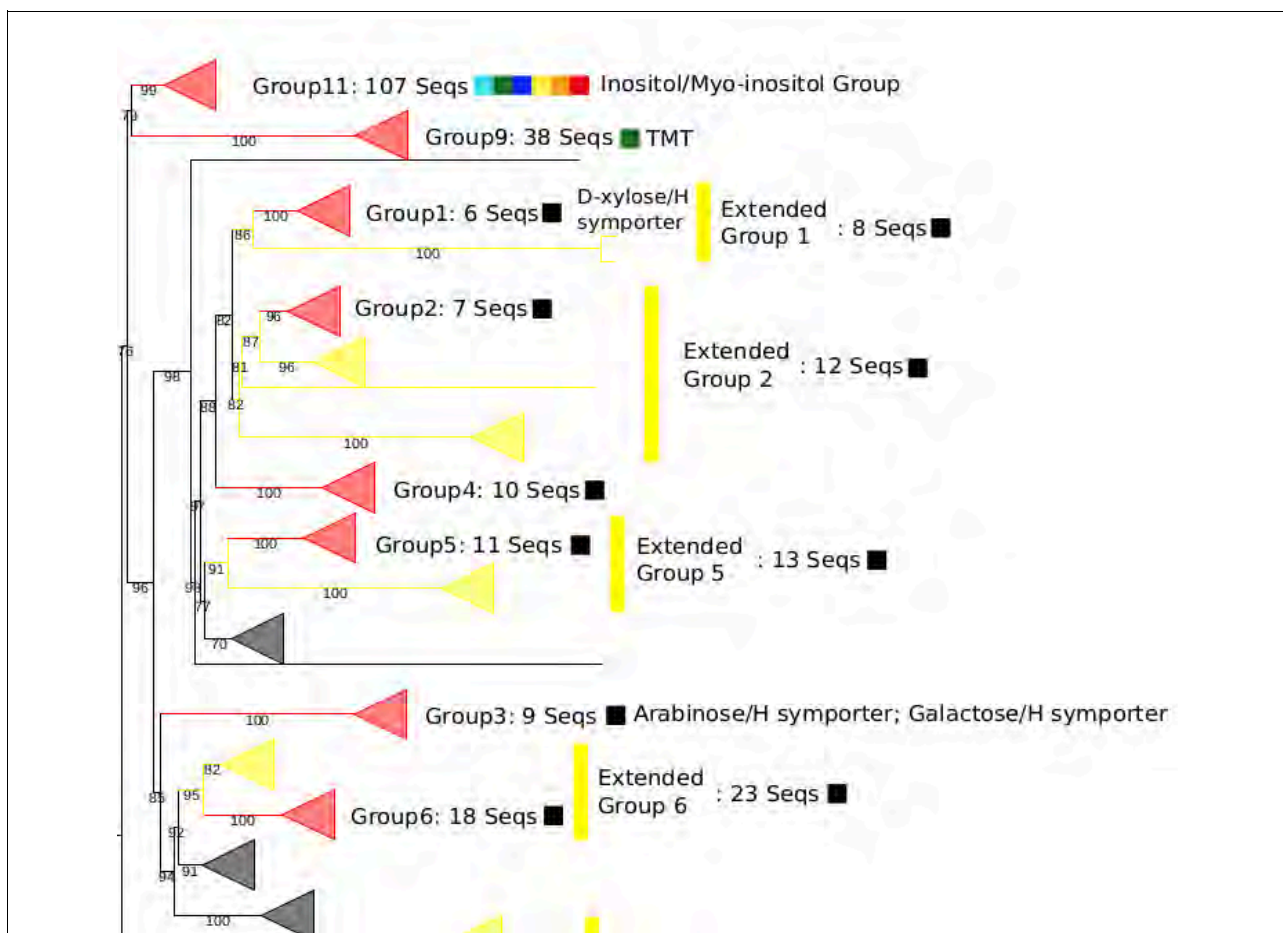


Figure 18: Ο μονοφυλετικός κλάδος των μεταφορέων ινοσιτόλης-μυοϊνοσιτόλης.

3.7 Εργαλείο ανίχνευσης και ταξινόμησης των μεταφορέων σακχάρων

Ένας από τους σκοπούς της παρούσας εργασίας ήταν η κατασκευή ενός εργαλείου για την εξελικτική κατηγοριοποίηση των μεταφορέων σακχάρων της υπεροικογένειας MFS. Με τη βοήθεια της φυλογενετικής ανάλυσης διακρίθηκαν οι προαναφερθείσες εξελικτικές ομάδες για τις οποίες και δημιουργήθηκαν Hidden Markov Models μεγάλης ευκρίνειας.

3.7.1 Σειρά διεργασιών (workflow) του εργαλείου

Ολόκληρη η σειρά διεργασιών που ακολουθήθηκε για την αναγνώριση μιας ακολουθίας επερώτησης ως μεταφορέας σακχάρων MFS καθώς και για τη κατηγοριοποίησή της σε μια από τις εξελικτικές ομάδες που βρήκαμε ενσωματώθηκε σε ένα εργαλείο γραμμένο στη γλώσσα προγραμματισμού Python3.6 Figure 19. Αρχικά η ακολουθία επερώτησης υποβάλλεται σε hhmScan έναντι των Hidden Markov Model της οικογένειας μεταφορέων σακχάρων 2.A.1.1. Στη συνέχεια η ακολουθία, αν ανιχνευθεί ως

μεταφορέας σακχάρων ομόλογος της 2.A.1.1, υποβάλλεται σε ένα δεύτερο hmmscan έναντι των Hidden Markov Model που έχουν δημιουργηθεί για τις 37 ταξινομικές ομάδες ακολουθούμενο από αναζήτηση των ευκαρυωτικών MEME motifs, μέσω του αλγορίθμου MAST, που έχουν δημιουργηθεί. Στο συγκεκριμένο στάδιο του αλγορίθμου πραγματοποιείται μια εξελεκτική ταξινόμηση αλλά και έλεγχος “πιστότητας” της ακολουθίας, καθώς εφαρμόζεται το κατώφλι 7/11 για τα μοτίβα όπως αναφέρθηκε και παραπάνω. Στη περίπτωση που η ακολουθία δεν περάσει τα αυστηρά κριτήρια πιστότητας μας λαμβάνει το χαρακτηρισμό putative.

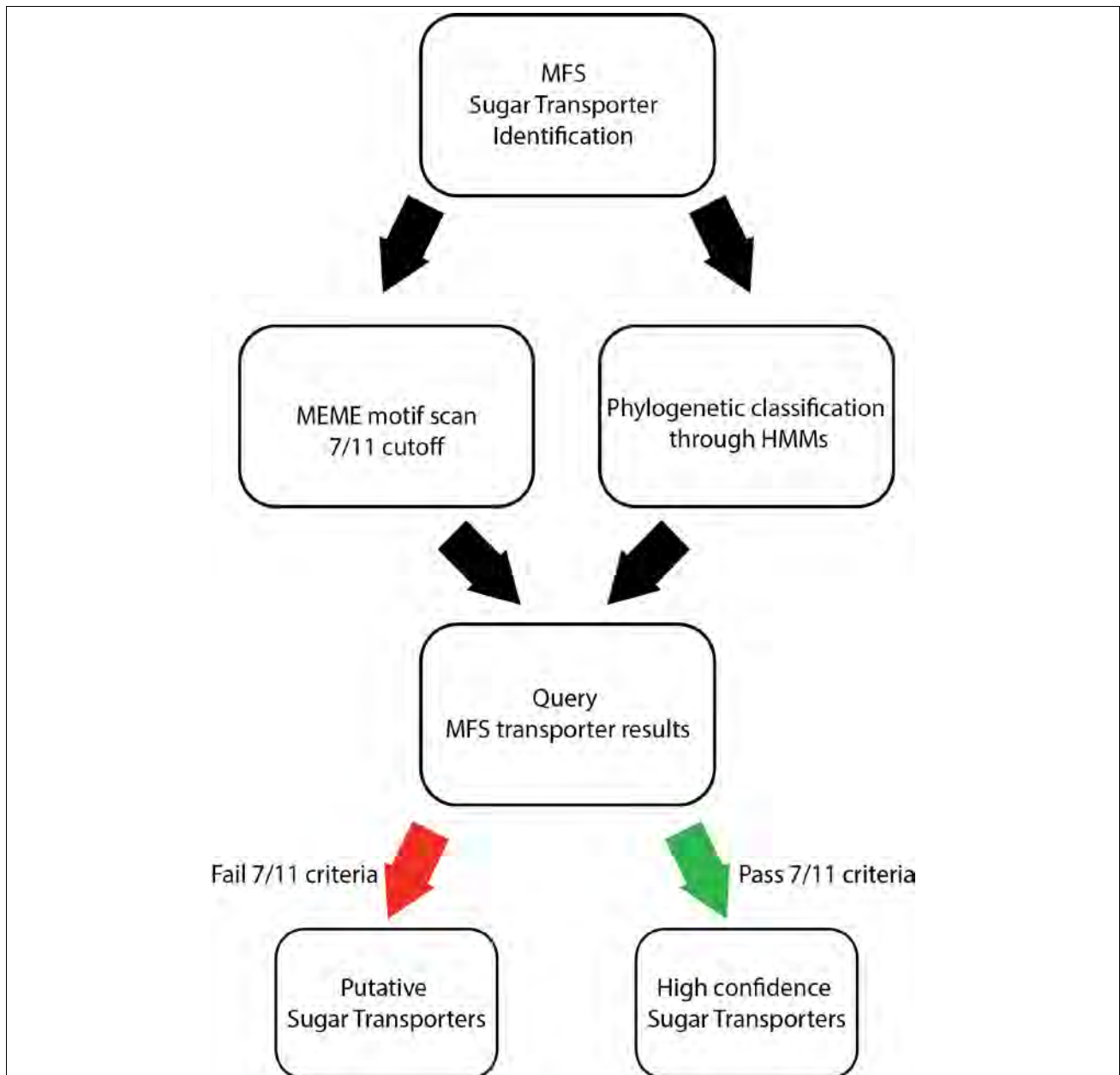


Figure 19: Η σειρά διεργασιών του εργαλείου ανίχνευσης και εξελεκτικής ταξινόμησης των μεταφορέων σακχάρων της MFS.

3.7.2 Ανάπτυξη ως διαδικτυακός server

Με τη βοήθεια του διδακτορικού φοιτητή Βλασταρίδη Παναγιώτη ο αλγόριθμος ανίχνευσης και εξελικτικής ταξινόμησης υλοποιήθηκε ως ένας διαδικτυακός server. Στο διαδικτυακό πλέον εργαλείο ο χρήστης έχει τη δυνατότητα να εισάγει κατευθείαν ακολουθίες προς ανάλυση σε fasta format ή να ανεβάσει ένα αρχείο fasta με πολλαπλές ακολουθίες. Τα αποτελέσματα παρουσιάζονται στο χρήστη όπως και στην εικόνα Figure 20. Για κάθε ακολουθία επερώτησης αναφέρεται η ταξινομική ομάδα στην οποία ανήκει σε συνδυασμό με το E-value, όπως στο παράδειγμα μας group34 και τα MEME motifs που ανιχνεύθηκαν. Πιο συγκεκριμένα τα motifs εμφανίζονται ως αριθμημένα χρωματιστά πλαίσια σε μια ειδική περιοχή, την οποία ο χρήστης μπορεί να μεγενθύνει σε επίπεδο αμινοξέος για να δει περισσότερες λεπτομέρειες, επίσης για το κάθε ένα motif παρέχονται πληροφορίες για την αλληλουχία του (logo), η στατιστική σημαντικότητά της εύρεσης του και η θέσεις αρχής και τέλους πάνω στην ακολουθία επερώτησης. Οι ακολουθίες που έχουν χαρακτηριστεί ως putative δίνονται στο χρήστη με το συγκεκριμένο χαρακτηρισμό. Ο χρήστης στη συνέχεια έχει την ικανότητα να κατεβάσει τα αποτελέσματα στον προσωπικό του υπολογιστή για περαιτέρω επεξεργασία σε tsv format.

Results

[Download results](#)

Protein	Subfamily/Evolutionary Cluster	E-value	Meme Motifs Found	Known Substrates
sp P11169 GTR3_HUMAN	group34	1.6e-259	15	-

Showing 1 - 1 of 1 items.

«» « 1 »»

Meme Motifs for Protein sp|P11169|GTR3_HUMAN

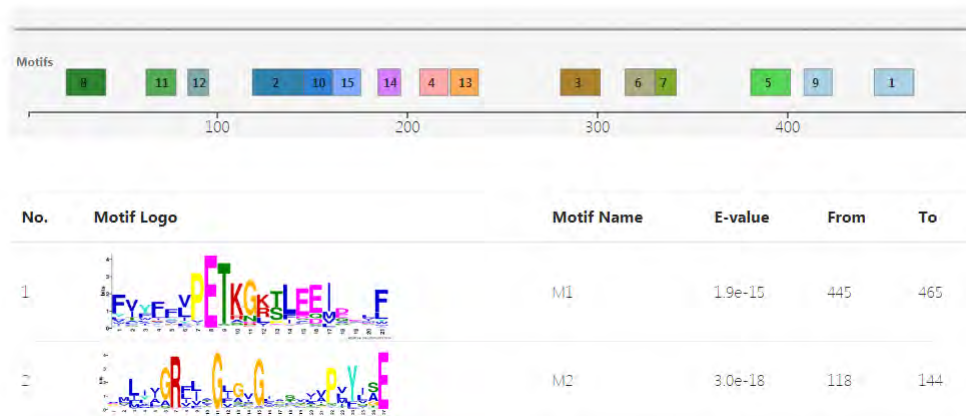


Figure 20: Χρήση του διαδικτυακού εργαλείου για την ανίχνευση MFS μεταφορέων σακχάρων. Στη συγκεκριμένη εικόνα φαίνονται τα αποτελέσματα για την αλληλουχία του GLUT3 (PDB entry:4zw9) (Deng et al., 2015). Με χρωματιστά πλαίσια εμφανίζονται τα MEME motifs που έχουν ανιχνευθεί και στη συνέχεια παρατίθενται πληροφορίες για αυτά.

4 ΣΥΖΗΤΗΣΗ

Οι μεταφορείς σακχάρων της υπεροικογένειας MFS έχουν προσελκύσει το ενδιαφέρον πολλών ερευνητικών ομάδων και έχει πραγματοποιηθεί εκτεταμένη έρευνα για το ρόλο τους, για το μηχανισμό δράσης τους και για το πρότυπο έκφρασής τους. Στο γονιδίωμα της ντομάτας έχουν ανιχνευθεί 52 πιθανοί (putative) μεταφορείς σακχάρων, με τους 35 να έχουν δεδομένα EST (EST evidence) και οι 32 εξ'αυτών να ανήκουν σε μια εκ των επτά γνωστών υποοικογενειών (PMT, VGT, ESL, STP, pGLcT, INT, TMT). Στην ίδια εργασία πραγματοποιήθηκε ανάλυση γονιδιακής έκφρασης για ορισμένες πρωτεΐνες από κάθε υποοικογένεια. Οι STP εμφανίζουν διαφορεικό πρότυπο έκφρασης κατά την ανάπτυξη του φυτού με ορισμένα γονίδια να εκφράζονται στα νεαρά φρούτα και φύλλα ενώ άλλα να παίζουν σημαντικό ρόλο στη συσσώρευση σακχάρων στο ώριμο φρούτο. Ορισμένοι εκπρόσωποι των υποοικογενειών PMT, VGT, TMT & pGLcT εκφράζονταν σε όλα τα στάδια δημιουργίας του καρπού. Οι ESL φάνηκε να εκφράζονται κυρίως στους βλαστητικούς ιστούς (Reuscher et al., 2014). Στο σταφύλι βρέθηκαν 59 μεταφορείς σακχάρων MFS με ορισμένους να έχουν ιστοειδική έκφραση και πιο συγκεκριμένα ορισμένοι εκφράζονται σχεδόν σε όλους τους βλαστητικούς ιστούς, λίγοι εκφράζονται εξειδικευμένα στη ρίζα, στα ώριμα φύλλα και άλλοι τρεις συμβάλλουν στην ωρίμανση του καρπού (Afoufa-Bastien et al., 2010). Στο αχλάδι έχουν ανιχνευθεί 53 μεταφορείς, εκ των οποίων περίπου οι 38 εμφανίζουν μόνιμα υψηλό επίπεδο έκφρασης καθ'όλη τη διάρκεια ανάπτυξης του φρούτου (Li et al., 2015). Το φυτό cassava (*Manihot esculenta*), που καλλιεργείται για την παραλαβή των θρεπτικών ριζών του, εκφράζει 17 μεταφορείς που ανήκουν στην υποοικογένεια STP με 15 να εκφράζονται κατά την ανάπτυξη της ρίζας. Στην εξελικτική γραμμή των αγγειόσπερμων συνολικά παρατηρείται ένας μεγάλος αριθμός γονιδίων μεταφορέων που οφείλεται σε γονιδιακούς διπλασιασμούς.

Πρόσφατη μελέτη του *Phaedoncochleariae* (Mustard Leaf Beetle) (Stock et al., 2013) έχει αναγνωρίσει 68 μεταφορείς σακχάρων, χαρακτηρισμένους ως SLC2 (Glut) ή ως Tret (trehalose transporter). Μέσω πειραμάτων RNAi και αλληλούχισης RNA-seq έχει βρεθεί ότι ορισμένοι μεταφορείς σακχάρων υπερεκφράζονται στους αμυντικούς αδένες του εντόμου και η σίγηση ορισμένων οδηγεί στην έκφραση άλλων ώστε να αντισταθμιστεί η απώλεια της δραστηριότητας. Στο μεταξοσκώληκα (*Bombyx mori*) έχουν βρεθεί 62 γονίδια μεταφορείς σακχάρων με τους 43 να παρουσιάζουν ιστοειδική έκφραση, μεταξύ άλλων στον αδένα που παράγει μετάξι, στα χυμοκύτταρα, στο μεσέντερο και στα μαλπιγκιανά σωληνάκια. Αξίζει να αναφερθεί ότι ο μεγάλος αριθμός των ορθόλογων γονιδίων στην εξελικτική γραμμή των εντόμων οφείλεται σε εκτενείς γονιδιακούς διπλασιασμούς (gene duplication & tandem gene duplication), που είναι γνωστό ότι έχουν συμβεί στα εξάποδα όπως αναφέρεται και στην εργασία (Li et al., 2018).

Τα παρασιτικά πρώτιστα Apicomplexa χρησιμοποιούν τα σάκχαρα του ξενιστή για την κάλυψη των ενεργειακών αναγκών τους, η μεταφορά των οποίων γίνεται μέσω μεταφορέων σακχάρων που ανήκουν στην υπεροικογένεια MFS. Στο άρθρο (Mahmud and Kissinger, 2017) μέσω ανάλυσης 20 αντιπροσωπευτικών οργανισμών από 5 μεγάλες ταξινομικές ομάδες των συγκεκριμένων πρώτιστων βρέθηκε πως ο αριθμός των μεταφορέων σακχάρων στους συγκεκριμένους οργανισμούς ποικίλει από 1 μέχρι 8.

Επίσης αναγνωρίστηκε η ύπαρξη 6 μεγάλων ταξινομικών ομάδων, αριθμός που συμπίπτει με τα δικά μας δεδομένα για τα πρώτιστα.

Η πιο πρόσφατη εξελικτική ανάλυση των GLUT στα μετάζωα (Jia et al., 2019) προτείνει την ύπαρξη τριών μεγάλων τάξεων. Οι GLUT 1-5, 7, 9, 11 & 14 συγκροτούν την πρώτη τάξη, οι GLUT 6 και 8 εμπεριέχονται στη δεύτερη τάξη και οι GLUT 10, 12 & 13 (HMIT) αποτελούν την τρίτη τάξη. Στη συγκεκριμένη ανάλυση οι πολλαπλές ακολουθίες των μεταζώων συγκρίθηκαν κυρίως μέσω δικτύων ομοιότητας αλληλουχιών (SSN; Sequence Similarity Networks) που αποτελούν ένα διαφορετικό μέσο ομαδοποίησης ακολουθιών σε σύγκριση με την υποστήριξη bootstrap που χρησιμοποιείται κατά κόρον και είχε χρησιμοποιηθεί σε παλαιότερη φυλογενετική ανάλυση των αντίστοιχων μεταφορέων (Wilson-O'Brien et al., 2010). Ωστόσο η τοπολογία των ακολουθιών εντός των ομάδων έρχεται σε συμφωνία τόσο με τα αποτελέσματα των (Wilson-O'Brien et al., 2010) όσο και τα αποτελέσματα της παρούσας εργασίας, όπου τα ορθόλογα των GLUT 1-4 σχηματίζουν μια υποομάδα με την πιο κοντινή της να είναι η ομάδα των GLUT 5, 7, 9, 11. Επίσης οι GLUT 6 & 8 αποτελούν ένα ξεχωριστό group με τα πιο κοντινά τους ομόλογα να είναι οι Tret 1 & 2 των εντόμων και οι GLUT 10 & 12 δημιουργούν μια ξεχωριστή ομάδα από τους υπόλοιπους GLUT, με πιο κοντινά ομόλογα τις ακολουθίες της υποομάδας HMIT.

Οι ακολουθίες που έχουν αναλυθεί από τους (Li et al., 2015; Liu et al., 2018; Mahmud and Kissinger, 2017; Reuscher et al., 2014; Stock et al., 2013) δεν έχουν ενσωματωθεί στην ανάλυση της εν λόγω εργασίας. Ωστόσο, η άντληση και η ταξινόμησή τους, μέσω του εργαλείου που αναπτύχθηκε, στις διάφορες εξελικτικές ομάδες που αναγνωρίστηκαν πρόκειται να χρησιμοποιηθεί για τον εμπλουτισμό των ομάδων αυτών με βιολογικά/λειτουργικά δεδομένα. Στο σύνολο της η εργασία έρχεται σε συμφωνία με τα δεδομένα από την υπάρχουσα βιβλιογραφία (Jia et al., 2019; Wilson-O'Brien et al., 2010) και ως εκ τούτου μπορεί να συμπληρώσει και να διευρύνει τα ήδη υπάρχοντα δεδομένα στο πεδίο της μελέτης των μεταφορέων σακχάρων. Τέλος, η ανάπτυξη του βιοπληροφικού εργαλείου της παρούσας μελέτης δύναται να διευκολύνει, να απλουστεύσει και να επιταχύνει τη διαδικασία εύρεσης μεταφορέων σακχάρων της υπεροικογένειας MFS και την επακόλουθη εξελικτική κατηγοριοποίηση αυτών, σημεία που αποτελούν πρώτα βήματα για την περαιτέρω μελέτη τους στους διάφορους οργανισμούς.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- About Perl - www.perl.org [WWW Document], n.d. URL <https://www.perl.org/about.html> (accessed 6.20.19).
- Afoufa-Bastien, D., Medici, A., Jeauffre, J., Coutos-Thévenot, P., Lemoine, R., Atanassova, R., Laloi, M., 2010. The Vitisvinifera sugar transporter gene family: phylogenetic overview and macroarray expression profiling. *BMC Plant Biol.* 10, 245. <https://doi.org/10.1186/1471-2229-10-245>
- Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S., 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. <https://doi.org/10.1093/nar/gkp335>
- Canonical, n.d. The leading operating system for PCs, IoT devices, servers and the cloud [WWW Document]. Ubuntu. URL <https://ubuntu.com/> (accessed 6.20.19).
- Chen, L.-Q., Cheung, L.S., Feng, L., Tanner, W., Frommer, W.B., 2015. Transport of Sugars. *Annu. Rev. Biochem.* 84, 865–894. <https://doi.org/10.1146/annurev-biochem-060614-033904>
- Chien, S., 2013. Figure 8, [Permeability of the cell membrane...]. [WWW Document]. URL <https://www.ncbi.nlm.nih.gov/books/NBK6156/figure/A76034/> (accessed 6.23.19).
- Cooper, Geoffrey M., and Robert E. Hausman, 2011. TO KYTTAPO Μία Μοριακή Προσέγγιση.
- Deng, D., Sun, P., Yan, C., Ke, M., Jiang, X., Xiong, L., Ren, W., Hirata, K., Yamamoto, M., Fan, S., Yan, N., 2015. Molecular basis of ligand recognition and transport by glucose transporters. *Nature* 526, 391–396. <https://doi.org/10.1038/nature14655>
- Deng, D., Yan, N., 2016. GLUT, SGLT, and SWEET: Structural and mechanistic investigations of the glucose transporters: Structural Investigations of the Glucose Transporters. *Protein Sci.* 25, 546–558. <https://doi.org/10.1002/pro.2858>
- Doidy, J., Grace, E., Kühn, C., Simon-Plas, F., Casieri, L., Wipf, D., 2012. Sugar transporters in plants and in their interactions with fungi. *Trends Plant Sci.* 17, 413–422. <https://doi.org/10.1016/j.tplants.2012.03.009>

- Eddy, S.R., 1998. Profile hidden Markov models. *Bioinformatics* 14, 755–763. <https://doi.org/10.1093/bioinformatics/14.9.755>
- Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>
- Eom, J.-S., Chen, L.-Q., Sosso, D., Julius, B.T., Lin, I., Qu, X.-Q., Braun, D.M., Frommer, W.B., 2015. SWEETs, transporters for intracellular and intercellular sugar translocation. *Curr. Opin. Plant Biol.* 25, 53–62. <https://doi.org/10.1016/j.pbi.2015.04.005>
- Feng, L., Frommer, W.B., 2015. Structure and function of SemiSWEET and SWEET sugar transporters. *Trends Biochem. Sci.* 40, 480–486. <https://doi.org/10.1016/j.tibs.2015.05.005>
- Gladyshev, E.A., Meselson, M., Arkhipova, I.R., 2008. Massive Horizontal Gene Transfer in Bdelloid Rotifers. *Science* 320, 1210–1213. <https://doi.org/10.1126/science.1156407>
- Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol. Biol. Evol.* 27, 221–224. <https://doi.org/10.1093/molbev/msp259>
- HMMER [WWW Document], n.d. URL <http://hmmer.org/> (accessed 6.20.19).
- Hoglund, P.J., Nordstrom, K.J.V., Schioth, H.B., Fredriksson, R., 2011. The Solute Carrier Families Have a Remarkably Long Evolutionary History with the Majority of the Human Families Present before Divergence of Bilaterian Species. *Mol. Biol. Evol.* 28, 1531–1541. <https://doi.org/10.1093/molbev/msq350>
- Hopfner, K.-P., 2016. Invited review: Architectures and mechanisms of ATP binding cassette proteins: Architectures and Mechanisms of ABC Proteins. *Biopolymers* 105, 492–504. <https://doi.org/10.1002/bip.22843>
- Huerta-Cepas, J., Serra, F., Bork, P., 2016. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* 33, 1635–1638. <https://doi.org/10.1093/molbev/msw046>
- Jia, B., Yuan, D.P., Lan, W.J., Xuan, Y.H., Jeon, C.O., 2019. New insight into the classification and evolution of glucose transporters in the Metazoa. *FASEB J.* 33, 7519–7528. <https://doi.org/10.1096/fj.201802617R>
- Johnson, D.A., Hill, J.P., Thomas, M.A., 2006. The monosaccharide transporter gene family in land plants is ancient and shows differential

- subfamily expression and expansion across lineages. *BMC Evol. Biol.* 20.
- Kulbacka, J., Choromańska, A., Rossowska, J., Weżgowiec, J., Saczko, J., Rols, M.-P., 2017. Cell Membrane Transport Mechanisms: Ion Channels and Electrical Properties of Cell Membranes, in: Kulbacka, J., Satkauskas, S. (Eds.), *Transport Across Natural and Modified Biological Membranes and Its Implications in Physiology and Therapy*. Springer International Publishing, Cham, pp. 39–58. https://doi.org/10.1007/978-3-319-56895-9_3
 - Lemoine, F., DomelevoEntfellner, J.-B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T., Gascuel, O., 2018. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* 556, 452–456. <https://doi.org/10.1038/s41586-018-0043-0>
 - Letunic, I., Bork, P., 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkz239>
 - Li, J., Zheng, D., Li, L., Qiao, X., Wei, S., Bai, B., Zhang, S., Wu, J., 2015. Genome-Wide Function, Evolutionary Characterization and Expression Analysis of Sugar Transporter Family Genes in Pear (*Pyrusbretschneideri*Rehd). *Plant Cell Physiol.* 56, 1721–1737. <https://doi.org/10.1093/pcp/pcv090>
 - Li, Z., Tiley, G.P., Galuska, S.R., Reardon, C.R., Kidder, T.I., Rundell, R.J., Barker, M.S., 2018. Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proc. Natl. Acad. Sci.* 115, 4713–4718. <https://doi.org/10.1073/pnas.1710791115>
 - Liu, Q., Dang, H., Chen, Z., Wu, J., Chen, Y., Chen, S., Luo, L., 2018. Genome-Wide Identification, Expression, and Functional Analysis of the Sugar Transporter Gene Family in Cassava (*Manihotesculenta*). *Int. J. Mol. Sci.* 19, 987. <https://doi.org/10.3390/ijms19040987>
 - Mahmud, O., Kissinger, J.C., 2017. Evolution of the Apicomplexan Sugar Transporter Gene Family Repertoire. *Int. J. Genomics* 2017, 1–9. <https://doi.org/10.1155/2017/1707231>
 - Pao, S.S., Paulsen, I.T., Saier, M.H., 1998. Major Facilitator Superfamily. *MICROBIOL MOL BIOL REV* 62, 34.
 - Perland, E., Fredriksson, R., 2017. Classification Systems of Secondary Active Transporters. *Trends Pharmacol. Sci.* 38, 305–315. <https://doi.org/10.1016/j.tips.2016.11.008>
 - Reddy, V.S., Shlykov, M.A., Castillo, R., Sun, E.I., Saier, M.H., 2012. The major facilitator superfamily (MFS) revisited: The major facilitator

- family revisited. *FEBS J.* 279, 2022–2035. <https://doi.org/10.1111/j.1742-4658.2012.08588.x>
- Reuscher, S., Akiyama, M., Yasuda, T., Makino, H., Aoki, K., Shibata, D., Shiratake, K., 2014. The Sugar Transporter Inventory of Tomato: Genome-Wide Identification and Expression Analysis. *Plant Cell Physiol.* 55, 1123–1141. <https://doi.org/10.1093/pcp/pcu052>
 - Stock, M., Gretscher, R.R., Groth, M., Eiserloh, S., Boland, W., Burse, A., 2013. Putative Sugar Transporters of the Mustard Leaf Beetle *Phaedoncochleariae*: Their Phylogeny and Role for Nutrient Supply in Larval Defensive Glands. *PLoS ONE* 8, e84461. <https://doi.org/10.1371/journal.pone.0084461>
 - Welcome to Python.org [WWW Document], n.d. . Python.org. URL <https://www.python.org/> (accessed 6.20.19).
 - Wilson-O'Brien, A.L., Patron, N., Rogers, S., 2010. Evolutionary ancestry and novel functions of the mammalian glucose transporter (GLUT) family. *BMC Evol. Biol.* 10, 152. <https://doi.org/10.1186/1471-2148-10-152>
 - Wright, E.M., Ghezzi, C., Loo, D.D.F., 2017. Novel and Unexpected Functions of SGLTs. *Physiology* 32, 435–443. <https://doi.org/10.1152/physiol.00021.2017>
 - Yan, N., 2017. A Glimpse of Membrane Transport through Structures—Advances in the Structural Biology of the GLUT Glucose Transporters. *J. Mol. Biol.*, John Kendrew's 100th Anniversary Special Edition 429, 2710–2725. <https://doi.org/10.1016/j.jmb.2017.07.009>
 - Yan, N., 2015. Structural Biology of the Major Facilitator Superfamily Transporters. *Annu. Rev. Biophys.* 44, 257–283. <https://doi.org/10.1146/annurev-biophys-060414-033901>
 - ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΕΞΕΛΙΚΤΙΚΗ ΑΝΑΛΥΣΗ ΤΩΝ ΜΕΤΑΦΟΡΕΩΝ ΣΑΚΧΑΡΩΝ ΤΗΣ MAJOR FACILITATOR SUPERFAMILY, 2018. . Διακογεωργίου Αλεξάνδρα.