



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΙΑΣ

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΥΓΕΙΑΣ

ΤΜΗΜΑ ΒΙΟΧΗΜΕΙΑΣ ΚΑΙ ΒΙΟΤΕΧΝΟΛΟΓΙΑΣ

Βιοπληροφορική και Εξελικτική Ανάλυση Μεταφορέων Σακχάρων και η Σχέση τους με την Προσαρμοστικότητα των Ειδών

Γεωργία Καλογιάννη

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΥΠΕΥΘΥΝΟΣ

Αμούτζιας Γρηγόριος, Αναπληρωτής Καθηγητής Βιοπληροφορικής με έμφαση στη Μικροβιολογία, Τμήμα Βιοχημείας και Βιοτεχνολογίας, Σχολή Επιστημών Υγείας, Πανεπιστήμιο Θεσσαλίας

ΜΕΛΗ ΕΠΙΤΡΟΠΗΣ

Αμούτζιας Γρηγόριος, Αναπληρωτής Καθηγητής Βιοπληροφορικής με έμφαση στη Μικροβιολογία, Τμήμα Βιοχημείας και Βιοτεχνολογίας, Σχολή Επιστημών Υγείας, Πανεπιστήμιο Θεσσαλίας

Δημήτριος Μόσιαλος, Αναπληρωτής Καθηγητής Βιοτεχνολογίας Μικροβίων, Τμήμα Βιοχημείας και Βιοτεχνολογίας, Σχολή Επιστημών Υγείας, Πανεπιστήμιο Θεσσαλίας

Ευστάθιος Φριλίγγος, Καθηγητής Βιολογικής Χημείας, Τμήμα Ιατρικής, Σχολή Επιστημών Υγείας, Πανεπιστήμιο Ιωαννίνων

Λάρισα 2022



UNIVERSITY OF
THESSALY

FACULTY OF LIFE SCIENCES

DEPARTMENT OF BIOCHEMISTRY AND BIOTECHNOLOGY

Bioinformatics and Evolutionary Analysis of Sugar Transporters and Their Relationship with Species Adaptation

Georgia Kalogiannis

UNDERGRADUATE THESIS

SUPERVISOR

Grigoris Amoutzias, Associate Professor of Bioinformatics with emphasis in Microbiology at the Department of Biochemistry and Biotechnology, University of Thessaly

BOARD MEMBERS

Grigoris Amoutzias, Associate Professor of Bioinformatics with emphasis in Microbiology at the Department of Biochemistry and Biotechnology, University of Thessaly

Dimitris Mosialos, Associate Professor in Microbial Biotechnology at the Department Biochemistry and Biotechnology, University of Thessaly

Stathis Frilingos, Professor of Biological Chemistry at the Department of Medicine, University of Ioannina

Larissa 2022

Ευχαριστίες

Η παρούσα πτυχιακή εργασία αποτελεί το τέλος της φοιτητικής μου σταδιοδρομίας στο τμήμα Βιοχημείας και Βιοτεχνολογίας. Η διενέργεια της έρευνας πραγματοποιήθηκε στο εργαστήριο Βιοπληροφορικής, του τμήματος Βιοχημείας και Βιοτεχνολογίας (TBB), της Σχολής Επιστημών Υγείας του Πανεπιστημίου Θεσσαλίας (ΠΘ) υπό τον καθηγητή Γρηγόριο Αμούτζια, τον οποίο και ευχαριστώ από καρδιάς για τη φιλοξενία, τη βοήθεια, την καθοδήγηση και την ευκαιρία που μου έδωσε.

Από τη θέση αυτή, θα ήθελα ακόμη να ευχαριστήσω τα μέλη της τριμελούς επιτροπής μου για την πολύτιμη καθοδήγηση τους, αλλά και για την τιμή που μου κάνουν να παρουσιάσω την εργασία μου. Επιπλέον, θα ήθελα να δώσω τις θερμότερες των ευχών μου στην ολομέλεια των μελών Δ.Ε.Π. του τμήματος Βιοχημείας και Βιοτεχνολογίας που μου έδωσαν την ευκαιρία να μάθω και κατά συνέπεια να εξελιχθώ και σαν άνθρωπος αλλά και σαν φοιτήτρια.

Τελειώνοντας, θα ήθελα να ευχαριστήσω τον Μάριο Νικολαΐδη και την Διακογεωργίου Αλεξάνδρα, που χάρη στο δικό τους έργο δημιουργήθηκε αυτή η διπλωματική εργασία. Επιπλέον, θα ήθελα να χαιρετήσω την αρωγή που μου έδωσε ο κος Νικολαΐδης, που χωρίς αυτόν, την τεχνογνωσία και την εμπειρία που μου εμπότισε, η πτυχιακή αυτή σίγουρα δε θα ήταν η ίδια.

Κλείνοντας θα ήθελα να ευχαριστήσω για τα όσα είμαι σαν άνθρωπος την οικογένεια μου και ιδιαίτερα τον πατέρα μου, που χωρίς εκείνον η εμπειρίες μου θα ήταν πιο θολές, καθώς και να αφιερώσω σε αυτόν αλλά και στο αγαπημένο μου *Arabidopsis Thaliana* την παρούσα εργασία.

Περιεχόμενα

Ευχαριστίες.....	3
Περιεχόμενα	4
Ευρετήριο Εικόνων.....	5
Ευρετήριο Πινάκων	5
Περίληψη	6
Abstract	7
1. Εισαγωγή	8
1.1. Πλασματική Μembrάνη	8
1.2. Διαπερατότητα Διπλοστιβάδας Λιπιδίων	8
1.3. Διάχυση Μέσω της Κυτταρικής Μembrάνης	9
1.3.1. Παθητική Διάχυση.....	10
1.3.2. Διευκολυνόμενη Διάχυση	10
1.3.3. Πρωτεΐνες-Μεταφορείς	10
1.3.4. Ενεργός Μεταφορά.....	11
1.4. Μεταφορείς Σακχάρων και Οικογένειες	12
1.5. Δομή και Ακολουθία Μεταφορέων Σακχάρων	12
1.6. Αποτελέσματα Εξελικτικών Αναλύσεων Μεταφορέων Γλυκόζης.....	15
2. Υλικά και Μέθοδοι	17
2.1. Λήψη Δεδομένων	17
2.2. Εργαλεία, Λογισμικό, Αλγόριθμοι και Πλατφόρμες	18
2.2.1. Linux.....	18
2.2.2. Python3.....	19
2.2.3. Perl	19
2.2.4. MFS Search Tool	19
2.2.4.1. Hidden Markov Models	19
2.2.5. Εργαλείο Εύρεσης Γονιδιώματος Πυρήνα	19
2.2.6. Φυλογενετική Ανάλυση	20
2.2.7. Επεξεργασία Εικόνων και Βιβλιογραφίας.....	21
3. Αποτελέσματα - Συζήτηση.....	22
Συνομειύσεις.....	33
Βιβλιογραφικές Παραπομπές.....	34

Ευρετήριο Εικόνων

Εικόνα 1. Φωσφολιπιδική διπλοστοιβάδα.....	8
Εικόνα 2. Τρόποι διάχυσης.....	9
Εικόνα 3. Κρυσταλλική δομή της ανθρώπινης πρωτεΐνης-μεταφορέα γλυκόζης GLUT1	11
Εικόνα 4. Οι τυπικές δομές των μεταφορέων σακχάρων.....	13
Εικόνα 5. Glut domains	14
Εικόνα 6. Φυλογενετική ανάλυση GLUT θηλαστικών.....	16
Εικόνα 7. Τελικό φυλογενετικό δέντρο	28
Εικόνα 8. Πρώτο φυλογενετικό δέντρο για την βεβαίωση της ορθότητας των τοπολογιών	29
Εικόνα 9. Δεύτερο φυλογενετικό δέντρο για την βεβαίωση της ορθότητας των τοπολογιών	30
Εικόνα 10. Φυλογενωμικό δένδρο μεταξύ στελεχών του σακχαρομύκητα <i>S. cerevisiae</i>	31
Εικόνα 11. Μεγέθυνση φυλογενωμικού δέντρου με κατάδειξη των σημείων ενδιαφέροντος.....	32

Ευρετήριο Πινάκων

Πίνακας 1. Τα 132 Accessions που επιστράφηκαν κατά την αναζήτηση της Ensembl Fungi. ...	17
Πίνακας 2. Συγκεντρωτικός πίνακας πληροφοριών των 132 στελεχών Σακχαρομυκήτων που επιστράφηκαν από την αναζήτηση.....	22
Πίνακας 3. Συγκεντρωτικός πίνακας των σακχαρομυκήτων που πληρούσαν τα κριτήρια αξιοπιστίας που αναλύθηκαν.....	25

Περίληψη

Ο σκοπός της παρούσας πτυχιακής εργασίας είναι η μελέτη των μεταφορέων σακχάρων και η σχέση τους με την προσαρμοστικότητα των ειδών, και ειδικότερα των σακχαρομυκήτων. Τα σάκχαρα αποτελούν την κύρια πηγή ενέργειας όλων των οργανισμών και για αυτό οι πρωτεΐνες-μεταφορείς τους είναι από τους σημαντικότερους μεταφορείς στα κύτταρα. Χρησιμοποιώντας δεδομένα από δημόσια αποθετήρια, καθώς και μια πληθώρα Βιοπληροφορικών εργαλείων, λογισμικών και αλγορίθμων, καταλήξαμε στη δημιουργία δύο ενημερωμένων συγκεντρωτικών πινάκων που αποτέλεσαν τη βάση για την εξελικτική ανάλυση που ακολούθησε. Αρχικά, απεικονίστηκαν οι εξελικτικές σχέσεις μεταξύ ειδών σακχαρομυκήτων, καθώς και εξελικτικές σχέσεις μεταξύ διαφορετικών στελεχών του *Saccharomyces cerevisiae*. Στη συνέχεια εντοπίστηκαν οι μεταφορείς σακχάρων σε αυτά τα γονιδιώματα και έγινε φυλογενετική ανάλυσή τους, ώστε να διαπιστωθεί εάν οι μεταφορείς αυτοί είναι εξελικτικά σταθεροί ή υφίστανται εκτεταμένους γονιδιακούς διπλασιασμούς που πιθανά να σχετίζονται με την προσαρμογή τους σε περιβάλλοντα.

Abstract

The purpose of this thesis is the study of sugar transporters and their relationship with the adaptability of species, with a focus on *Saccharomyces*. Sugars are the main source of energy for all organisms and therefore their transporter proteins are among the most important proteins of cells. Using data from public repositories, as well as a plethora of Bioinformatics tools, software and algorithms, we ended up creating two annotated tables that formed the basis for the evolutionary analysis that followed. By analyzing phylogenetic trees, evolutionary relationships between different species of *Saccharomyces*, as well as evolutionary relationships between different strains of *Saccharomyces cerevisiae*, were investigated. Phylogenetic analysis of the sugar transporters was then performed to investigate whether these proteins are evolutionarily stable or undergo extensive gene duplications that may be related to species adaptations.

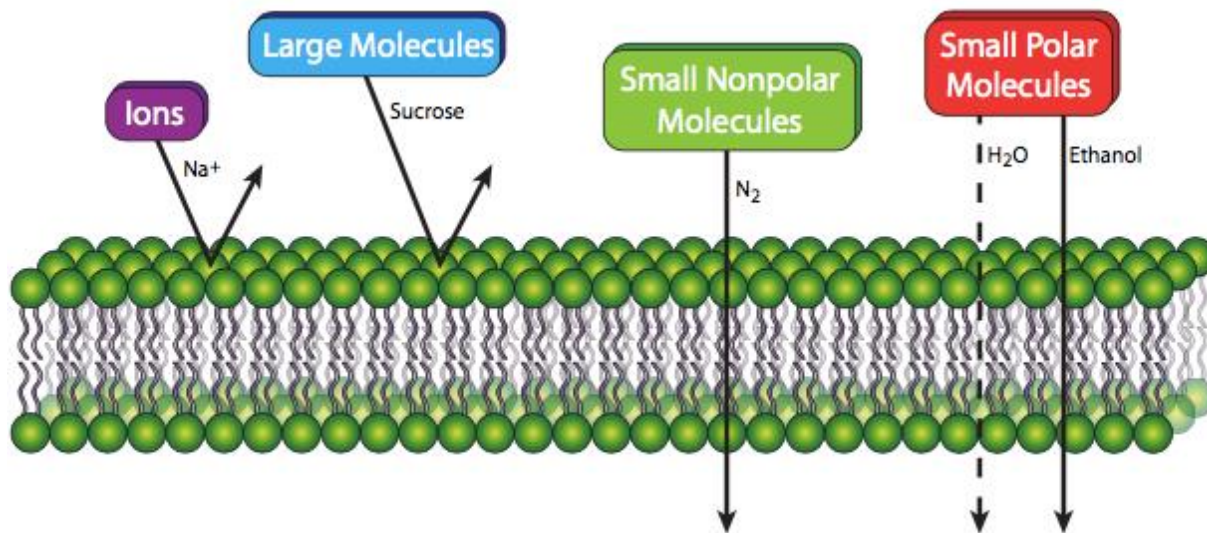
1. Εισαγωγή

1.1. Πλασματική Μεμβράνη

Όλα τα κύτταρα πρέπει να είναι σε θέση να ανταλλάσσουν υλικά με το εξωκυτταρικό περιβάλλον με έναν ελεγχόμενο τρόπο, προκειμένου να επιβιώνουν. Η πλασματική μεμβράνη (ή κυτταρική μεμβράνη) παρέχει προστασία σε ένα κύτταρο και επιτελεί ακόμη ποικίλες σημαντικές βιολογικές λειτουργίες. Συγκεκριμένα η πλασματική μεμβράνη ρυθμίζει τη μεταφορά θρεπτικών ουσιών εντός του κυττάρου και επίσης τη μεταφορά παραπροϊόντων εκτός από αυτό. Τα δομικά στοιχεία της πλασματικής μεμβράνης είναι τα φωσφολιπίδια, τα οποία σχηματίζουν μια διπλή στοιβάδα με πολική υδρόφιλη κεφαλή στραμμένη προς το υδατικό περιβάλλον μέσα και έξω από το κύτταρο και υδρόφοβες ουρές στο εσωτερικό (Εικόνα 1).

1.2. Διαπερατότητα Διπλοστιβάδας Λιπιδίων

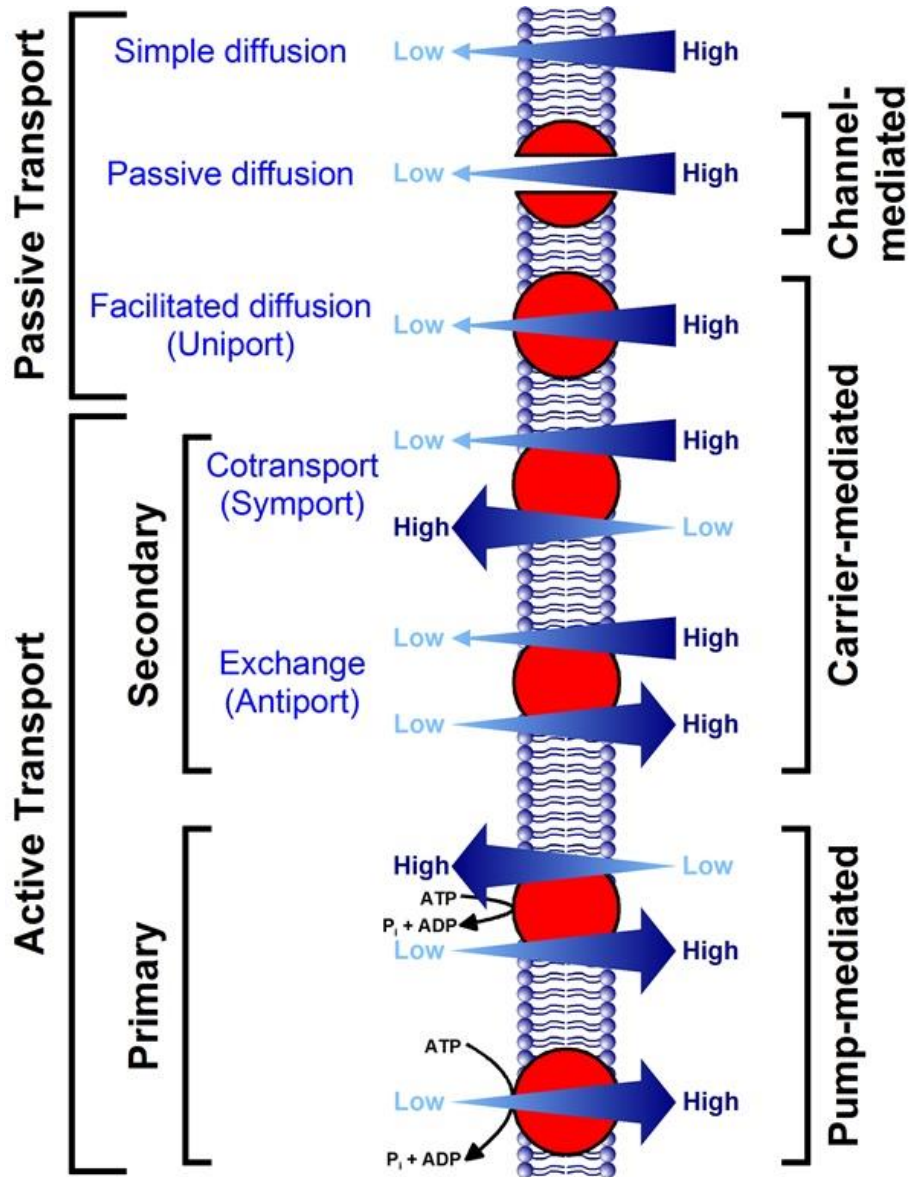
Οι βιολογικές μεμβράνες είναι κατ' ουσία ημιπερατές διπλοστιβάδες λιπιδίων· η ημιπερατότητα αναφέρεται στο γεγονός ότι λόγω της χημικής τους σύστασης και της δομής της φωσφολιπιδικής διπλοστιβάδας (υδρόφοβος πυρήνας), μόνο ορισμένα και λιποδιαλυτά μόρια και κάποια μικρά μόρια έχουν τη δυνατότητα να περάσουν ελεύθερα από τη λιπιδική διπλοστιβάδα. Μεγάλα πολικά μόρια και ιόντα δεν μπορούν να τη διαπεράσουν. Πιο συγκεκριμένα, το αν ένα μόριο μπορεί να περάσει μέσα από την μεμβράνη εξαρτάται από το μέγεθός του και την ηλεκτρική φύση του (το φορτίο του). Η μεμβράνη είναι εξαιρετικά διαπερατή από μη πολικά (λιποδιαλυτά) μόρια. Η διαπερατότητα της μεμβράνης σε πολικά (υδατοδιαλυτά) μόρια είναι πολύ χαμηλή - ιδιαίτερα χαμηλή σε μεγάλα πολικά μόρια. Επίσης είναι πολύ χαμηλή και σε φορτισμένα μοριακά είδη (ιόντα). Το πέρασμα των περισσότερων μορίων και ιόντων υποβοηθείται από την παρουσία ειδικών πρωτεϊνών μεταφοράς στη μεμβράνη.



Εικόνα 1. Φωσφολιπιδική διπλοστιβάδα. Μια φωσφολιπιδική διπλοστιβάδα είναι εγγενώς ημιπερατή. Εκτός από τα λιποδιαλυτά μόρια (στεροειδή, O₂, CO₂, κ.λπ.) και μερικά πολύ μικρά πολικά μόρια (νερό, ουρία, αιθανόλη, γλυκερόλη), η διέλευση ιόντων (Na⁺) και των περισσότερων πολικών μορίων (σουκρόζη) μέσω βιολογικών μεμβρανών απαιτεί την παρουσία ενσωματωμένων πρωτεϊνών μεταφοράς που λειτουργούν ως πρωτεΐνες μεταφοράς [1].

1.3. Διάχυση Μέσω της Κυτταρικής Μεμβράνης

Όπως αναφέρθηκε, πολλά μόρια έχουν την ικανότητα να διαπερνούν την κυτταρική μεμβράνη μέσω της διάχυσης. Ωστόσο, όσον αφορά την κυτταρική μεμβράνη αυτή η διαδικασία καταλύεται από εξειδικευμένες πρωτεΐνες μεταφορείς. Ένα μόριο μπορεί να διαπεράσει την κυτταρική μεμβράνη με τρεις (Εικόνα 2) διακριτούς τρόπους: α) Μέσω παθητικής διάχυσης, β) διευκολυνόμενης διάχυσης, γ) ενεργητικής μεταφοράς [2].



Εικόνα 2. Τρόποι διάχυσης. Μόνο ένας περιορισμένος αριθμός μορίων μπορεί να διασχίσει τις βιολογικές μεμβράνες χωρίς τη βοήθεια μεταφορικών πρωτεϊνών. Η πλειοψηφία των βιολογικών μορίων και ιόντων χρησιμοποιούν πρωτεΐνες μεταφοράς για να διασχίσουν τη μεμβράνη. Τα συστήματα μεταφοράς μπορεί να είναι παθητικά ή ενεργά [3].

1.3.1. Παθητική Διάχυση

Η παθητική διάχυση είναι μια μορφή παθητικής μεταφοράς ιόντων και χημικών ενώσεων μικρού μοριακού βάρους, μέσω της κυτταρικής μεμβράνης, χωρίς κατανάλωση ενέργειας. Η διαδικασία αυτή συμβαίνει αμφίπλευρα της κυτταρικής μεμβράνης, δηλαδή μεταξύ του κυτταροπλάσματος και του εξωκυττάρου υγρού. Τα σωματίδια κινούνται (συνήθως) από την περιοχή υψηλή συγκέντρωσης προς την περιοχή χαμηλής συγκέντρωσης. Η ταχύτητα μετακίνησης είναι ανάλογη της διαφοράς συγκεντρώσεων, του φορτίου και του μεγέθους τους. Η διάχυση σταματά όταν τα σωματίδια κατανεμηθούν ομοιόμορφα στις δύο περιοχές και επέλθει κατάσταση ισορροπίας.

Μεγαλύτερα μόρια αδυνατούν να περάσουν παθητικά από τη μεμβράνη και έτσι καθίσταται απαραίτητη η διαδικασία της διευκολυνόμενης διάχυσης [2].

1.3.2. Διευκολυνόμενη Διάχυση

Η διευκολυνόμενη διάχυση αποτελεί ένα είδος διαμεσολαβούμενης μεταφοράς ουσιών (όπως η γλυκόζη) χωρίς κατανάλωση ενέργειας, με τη βοήθεια δύο ειδικών δομικών πρωτεϊνών, που ονομάζονται πρωτεΐνες-μεταφορείς και πρωτεΐνες-διαυλοί.

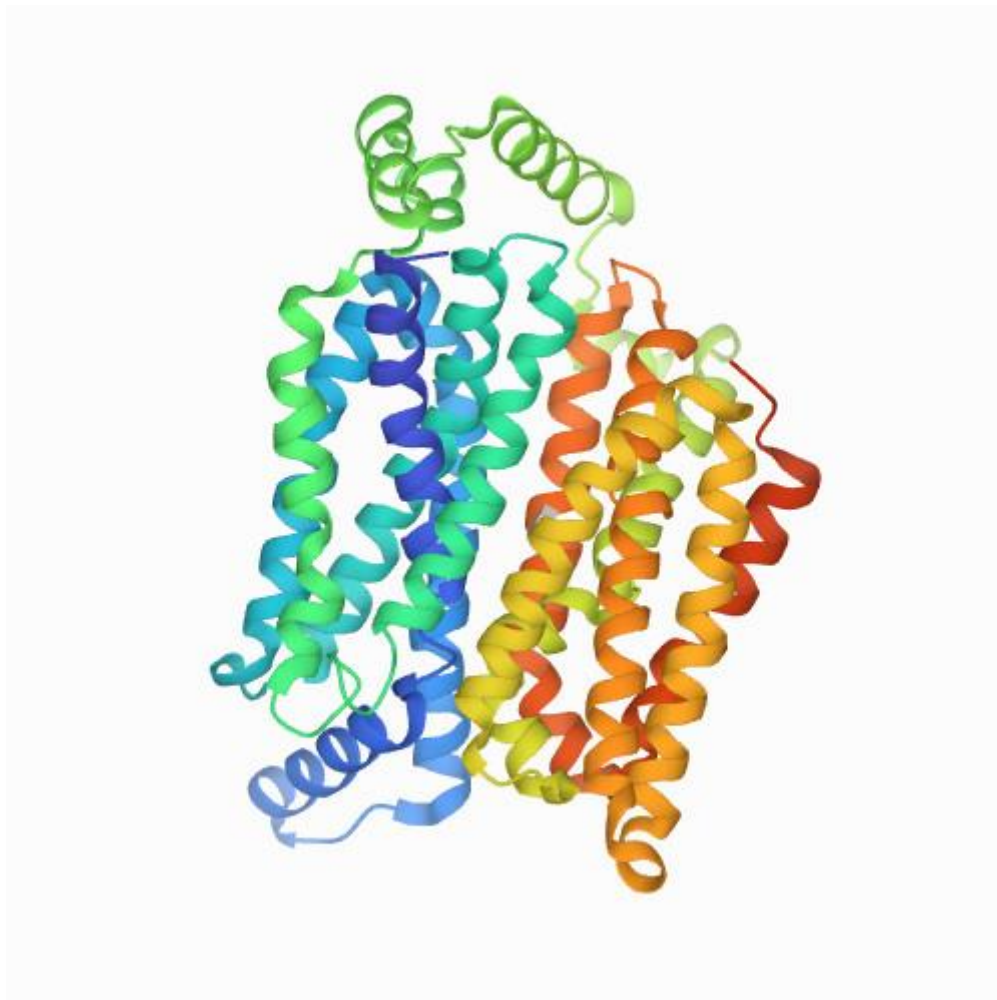
Στην πρώτη περίπτωση των πρωτεϊνών-μεταφορέων, η μεταφερόμενη ουσία συνδέεται αρχικά σε μία ειδική θέση του μεταφορέα, που εκτίθεται στη μία πλευρά της μεμβράνης και εξαιτίας της σύνδεσης αυτής, μεταβάλλεται η στερεοδιαμόρφωση ενός τμήματος του μεταφορέα, με αποτέλεσμα τη μετατόπιση της θέσης, πάνω στη οποία βρίσκεται συνδεδεμένη η ουσία, προς την άλλη πλευρά της μεμβράνης. Η διαδικασία ολοκληρώνεται με την αποδέσμευση της ουσίας από τον μεταφορέα.

Στη δεύτερη περίπτωση, δηλαδή των πρωτεϊνών-διαύλων, δεν αλλάζουν διαμόρφωση αλλά παραμένουν ανοιχτοί και υπάρχει ελεύθερη διάχυση μορίων αναλόγως του μεγέθους και του φορτίου. Στη διευκολυνόμενη διάχυση, όπως και στην απλή διάχυση, η ροή του μορίου γίνεται από την περιοχή υψηλής συγκέντρωσης προς την περιοχή χαμηλής συγκέντρωσης, μέχρι να εξισορροπηθούν οι συγκεντρώσεις του μορίου και στις δύο πλευρές της μεμβράνης. Όταν επέρχεται ισορροπία στη διευκολυνόμενη διάχυση, ο αριθμός των μορίων που δεσμεύεται στην εξωτερική πλευρά της μεμβράνης και απελευθερώνεται στην εσωτερική, είναι ίσος με τον αριθμό των μορίων που δεσμεύονται στην κυτταροπλασματική πλευρά και αποδεσμεύονται στο εξωκυττάριο περιβάλλον [2].

1.3.3. Πρωτεΐνες-Μεταφορείς

Οι πρωτεΐνες-μεταφορείς υπάρχουν σε όλους τους έμβιους οργανισμούς, οπότε χρήζουν ιδιαίτερου ενδιαφέροντος στην εξελικτική βιολογία. Οι πρωτεΐνες-μεταφορείς (Εικόνα 3) δεσμεύουν τη συγκεκριμένη διαλυμένη ουσία που πρόκειται να μεταφερθεί και υφίστανται μια σειρά διαμορφωτικών αλλαγών για τη μεταφορά της δεσμευμένης διαλυμένης ουσίας κατά μήκος της μεμβράνης. Όταν η στερεοδομή των πρωτεϊνών-μεταφορέων αναδιαμορφώνεται ουσιαστικά περνάει από τρία στάδια:

α) Πρώτα η ουσία εισέρχεται στο κέντρο δέσμευσης, όπου ο μεταφορέας αλλάζει διαμόρφωση → β) απελευθέρωση της ουσίας εντός του κυττάρου → γ) επιστροφή του μεταφορέα στην αρχική του κατάσταση [2,4].



Εικόνα 3. Κρυσταλλική δομή της ανθρώπινης πρωτεΐνης-μεταφορέα γλυκόζης GLUT1 [5]. Περίθλαση ακτινών-X σε ανάλυση 3.17 Å με συνολικό μοριακό βάρος 56.04 kDa (κωδικός PDB: 4PYP). Οι εικόνες 3&4 προέρχονται από την Protein Data Bank - PDB [6].

1.3.4. Ενεργός Μεταφορά

Όταν πρέπει να επιτευχθεί η μεταφορά ενός μορίου στην αντίθετη κατεύθυνση από αυτή που καθορίζονται από τις ηλεκτροχημικές διαβαθμίσεις μεταξύ των πλευρών της μεμβράνης, τότε αυτή η λειτουργία ονομάζεται ενεργός μεταφορά. Αντίθετα με την παθητική και τη διευκολυνόμενη διάχυση που δεν απαιτούνταν κατανάλωση ενέργειας, η μεταφορά αυτή είναι ενεργοβόρα διαδικασία [2,4].

Η ενεργός μεταφορά που ωθείται από την υδρόλυση του ATP πραγματοποιείται από αντλίες ιόντων και ένα τέτοιο παράδειγμα είναι η Na^+/K^+ -ATPase (Αντλία Νατρίου-Καλίου) [2].

Από την ενεργό μεταφορά μέσω υδρόλυσης ATP, η μεταφορά ενός μορίου αντίθετα στη διαβάθμιση συγκέντρωσης του είναι μια διαδικασία που χρειάζεται την κατανάλωση ενέργειας. Υπάρχουν όμως και άλλοι μεταφορείς που μεταφέρουν προς τη μη ευνοϊκώς ενεργειακή κατεύθυνση χρησιμοποιώντας συζευγμένη μεταφορά μεταξύ αυτών και ενός ακόμη μορίου το οποίο μεταφέρεται προς την ευνοϊκή κατεύθυνση. Υπάρχουν τρεις κατηγορίες μεταφορέων της μεμβράνης: α) οι συμμεταφορείς, οι οποίοι μεταφέρουν δύο μόρια προς την ίδια κατεύθυνση, β)

οι μονομεταφορείς, που μεταφέρουν μόνο ένα μόριο είτε εσωτερικά είτε εξωτερικά της μεμβράνης και γ) οι αντιμεταφορείς, οι οποίοι μεταφέρουν δύο μόρια προς αντίθετες κατευθύνσεις [2] (Εικόνα 2).

1.4. Μεταφορείς Σακχάρων και Οικογένειες

Η γλυκόζη (glucose - $C_6H_{12}O_6$) είναι η κύρια πηγή ενέργειας για τη ζωή στη γη. Επιτελεί επίσης τη λειτουργία ως σημαντικός πρόδρομος για τη σύνθεση βιομορίων και παίζει σημαντικό ρόλο στη σηματοδότηση των κυττάρων. Οι διατροφικοί υδατάνθρακες περιέχουν μεγάλη ποσότητα γλυκόζης με τη μορφή μονοσακχαρίτη ή ολιγο- και πολυσακχαριτών, όπως μαλτόζη, άμυλο και γλυκογόνο [7]. Ο ανθρώπινος εγκέφαλος, ο οποίος αντιπροσωπεύει γενικά περίπου το 2% της συνολικής μάζας σώματος ενός ενήλικα, καταναλώνει περίπου το 25% της προσφοράς γλυκόζης. Πολλαπλοί μεταφορείς γλυκόζης έχουν εξελιχθεί [8,9].

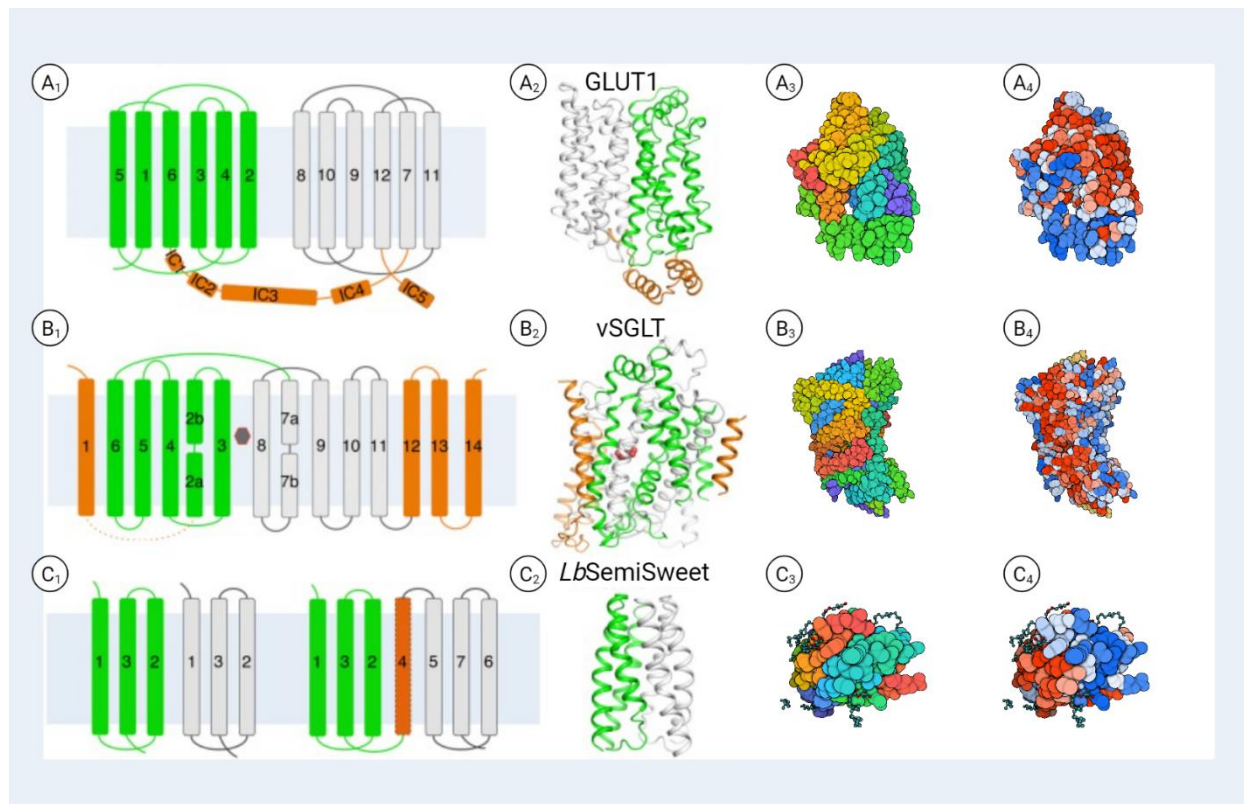
Έχουν ταυτοποιηθεί τρεις οικογένειες ευκαρυωτικών μεταφορέων σακχάρων, συμπεριλαμβανομένων των πιο αυστηρά μελετημένων GLUTs (SLC2) που συγκαταλέγονται στην οικογένεια Sugar Porter Family της υπεροικογένειας Major Facilitator Superfamily (MFS), των συμμεταφορέων νατρίου-γλυκόζης SGLTs (SLC5) και των SWEET (SLC50) [10]. Οι τρεις αυτές οικογένειες μεταφορέων γλυκόζης έχουν διακριτές λειτουργίες και μηχανισμούς. Οι GLUTs κατανέμονται παντού και καταλύουν τη διευκολυνόμενη διάχυση της γλυκόζης σύμφωνα με τη βαθμίδα συγκέντρωσής της. Οι SGLTs εκφράζονται κυρίως στο έντερο και τα νεφρά και αξιοποιούν την ενέργεια που απελευθερώνεται από την κατηφορική ροή του Na^+ για να οδηγήσει τη μετατόπιση της γλυκόζης έναντι της βαθμίδας συγκέντρωσής της στη μεμβράνη. Τέλος, οι SWEETs ευθύνονται κυρίως για την εκροή και την ενδοκυτταρική διακίνηση σακχάρων στα φυτά (δεν έχει αποδειχθεί ακόμη το ίδιο στα θηλαστικά) [8,11,12].

Οι GLUTs που συγκαταλέγονται στην οικογένεια Sugar Porter Family της υπεροικογένειας Major Facilitator Superfamily (MFS), ταξινομούνται σε τρεις κλάσεις [13]. Μεταξύ των διαφορετικών οικογενειών μεταφορέων, μόνο δύο απαντώνται παντού σε όλες τις ταξινομήσεις οργανισμών. Αυτές είναι η υπεροικογένεια ATP-Binding Cassette (ABC) και η Major Facilitator Superfamily (MFS). Οι μεταφορείς MFS είναι δευτερογενείς πολυπεπτιδικοί ενεργοί μεταφορείς ικανοί να μεταφέρουν μόνο μικρές διαλυμένες ουσίες σε απόκριση σε βαθμίδες χημειοσμωτικών ιόντων [14,15]. Οι MFS είναι η μεγαλύτερη και πιο ποικιλόμορφη υπεροικογένεια δευτερογενών ενεργών μεταφορέων που βρίσκονται παντού σε όλους τους ζωντανούς οργανισμούς. Τα μέλη των MFS πιστεύεται ότι είναι μία από τις παλαιότερες οικογένειες πρωτεϊνών στη Γη, καθώς υπάρχουν περισσότερα από 3 δισεκατομμύρια χρόνια πριν [16]. Η βάση δεδομένων ταξινόμησης μεταφορέων (<http://www.tcdb.org>) [17], η οποία περιλαμβάνει αναγνωρισμένες και υποθετικές πρωτεΐνες μεταφοράς μεμβράνης, ταξινομεί το MFS με βάση τη φυλογένεση και τη λειτουργία σε 16 διαφορετικές οικογένειες, με 89 υποοικογένειες και 1244 σχολιασμένες πρωτεΐνες [17].

1.5. Δομή και Ακολουθία Μεταφορέων Σακχάρων

Οι μεταφορείς σακχάρων ανάλογα με την κατηγορία τους έχουν μια χαρακτηριστική στερεοδιάταξη. Σε επίπεδο τριτοταγούς δομής η χαρακτηριστική στερεοδιάταξη τους καθορίζεται από την κατηγορία που ανήκουν και απαντάται σε κάθε μέλος της. Οι δομές των SGLT, SWEET

και MFS έχουν αρκετά διαφορετικά χαρακτηριστικά και κάποιες ομοιότητες (Εικόνα 4) [18]. Οι θέσεις στις οποίες προσδένονται είναι τοποθετημένες στο κέντρο της πλασματικής μεμβράνης, στη διεπιφάνεια μεταξύ των δύο λιπιδικών στιβάδων. Ακόμη, παρατηρούνται κατάλοιπα αρωματικών αμινοξικών ακολουθιών που επιτελούν στην επιλεκτική δέσμευση του υποστρώματος στο μεταφορέα, ή στη δημιουργία φυσικών φραγμών [10]. Οι GLUTs έχουν την τυπική δομή της MFS, οι SGLTs της APC, ενώ οι SWEETs στους ευκαρυωτικούς οργανισμούς έχουν 7-TMs ενώ τα βακτηριακά ομόλογα μόλις 3-TM (SemiSWEETs) [19].



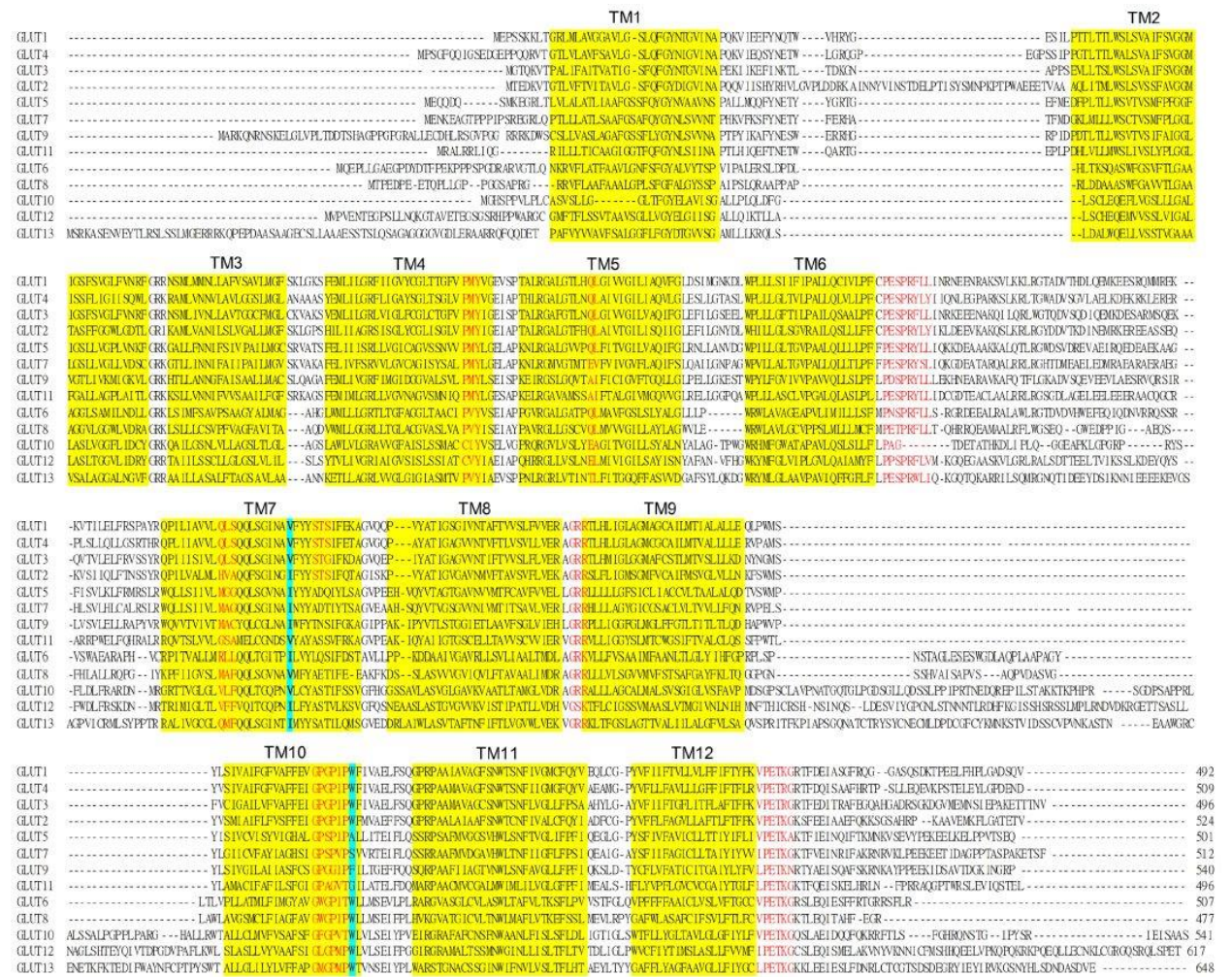
Εικόνα 4. Οι τυπικές δομές των μεταφορέων σακχάρων. A₁) Δομή των GLUTs. Διαθέτουν 12 διαμεμβρανικές έλικες, οι οποίες μοιράζονται σε δύο αυτοτελείς δομικές και λειτουργικές περιοχές, το N-τελικό (πράσινο) και το C-τελικό (λευκό) και μια ενδοκυτταρική έλικα που ονομάζεται ICH (πορτοκαλί). A₂) Η δομή του ανθρώπινου GLUT1. A₃) Απεικόνιση της ανθρώπινης GLUT1, χρωματισμένη βάσει της ακολουθίας της. A₄) Απεικόνιση της ανθρώπινης GLUT1 χρωματισμένη αναλόγως της υδροφοβικότητας της αμινοξικής της ακολουθίας (κωδικός PDB: 4PYP).

B₁) Δομή των SGLTs. Οι διαμεμβρανικές περιοχές TM2-11 συνιστούν «5+5» ανάστροφες επαναλήψεις (πράσινο και λευκό) της LeuT αναδίπλωσης του vSGLT. Το μαύρο εξαγώνιο υποδηλώνει το υπόστρωμα του μεταφορέα. B₂) Η κλειστή διαμόρφωση του vSGLT. B₃) Απεικόνιση της ανθρώπινης vSGLT, χρωματισμένη βάσει της ακολουθίας της. B₄) Απεικόνιση της ανθρώπινης vSGLT χρωματισμένη αναλόγως της υδροφοβικότητας της αμινοξικής της ακολουθίας (κωδικός PDB: 3DH4).

C₁) Δομή των SWEETs. 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. C₂) Η δομή του *LbSemiSWEET*. C₃) Απεικόνιση της ανθρώπινης *LbSemiSWEET*, χρωματισμένη βάσει της ακολουθίας της. C₄) Απεικόνιση της ανθρώπινης *LbSemiSWEET* χρωματισμένη αναλόγως της υδροφοβικότητας της αμινοξικής της ακολουθίας (κωδικός PDB: 4QNC). Η παρούσα εικόνα στηρίχτηκε στο έργο των Deng and Yan, 2016 και η περαιτέρω επεξεργασία της έγινε με τα εργαλεία που βρίσκονται στη σελίδα <https://BioRender.com> (19/01/2022).

Χαρακτηριστικές δομικές διαφορές στους GLUTs αποδόθηκαν με σαφήνεια στα αποτελέσματα της δουλειάς των Long W. και Cheeseman C.I. 2015. Ξεκινώντας από την αμινοξική τους ακολουθία, αποτελέσματα hydrophathy plot αποκάλυψαν -όπως είδαμε και προηγουμένως- ότι όλες έχουν 12-14 διαμεμβρανικές περιοχές με τα δύο τους άκρα να συνδέονται μέσω ενός βρόγχου στο TM-6 και TM-7 [20] (Hydrophathy plot analysis είναι μια αναπαράσταση του υδρόφοβου χαρακτήρα μιας αλληλουχίας αμινοξέων και μπορεί να είναι χρήσιμη στην πρόβλεψη διαμεμβρανικών περιοχών, πιθανών αντιγονικών θέσεων και περιοχών που είναι πιθανό να εκτίθενται στην επιφάνεια). Η πολλαπλή στοίχιση των GLUT (Εικόνα 5) αποκαλύπτει τη συντήρηση μεταξύ των διαφορετικών πρωτεϊνών και κλάσεων των GLUT [20].

Παρόλες τις διαφορές τους στις αμινοξικές τους ακολουθίες, οι GLUT πρωτεΐνες έχουν πολύ παρόμοια στερεοδιάταξη στο επίπεδο της τριτοταγούς τους δομής [20].



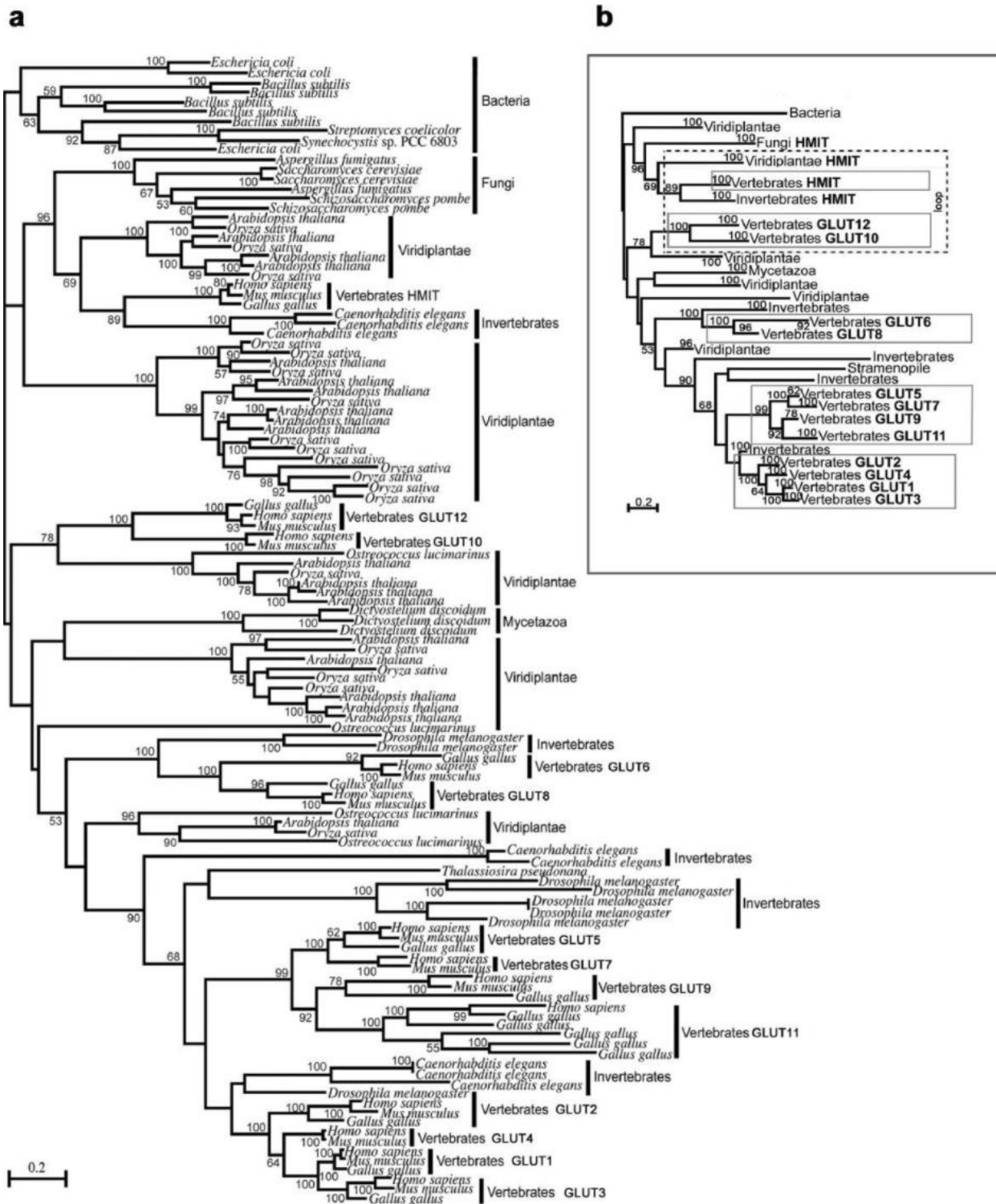
Εικόνα 5. Glut domains. Οι διαμεμβρανικές αυτοτελείς δομικές και λειτουργικές περιοχές (domain) επισημαίνονται με κίτρινο χρώμα με βάση τη κρυσταλλική δομή της GLUT1. Τα εξαιρετικά διατηρημένα υπολείμματα επισημαίνονται με κόκκινο χρώμα. Τα υπολείμματα που επισημαίνονται με μπλε στο TM-10 φαίνεται ότι είναι οι θέσεις αναγνώρισης/δέσμευσης της κυτοχλασίνης Β. Τα υπολείμματα που επισημαίνονται με μπλε χρώμα στο TM-7, υποδηλώνεται ότι είναι τα υδρόφοβα υπολείμματα που είναι υπεύθυνα για την επιλεκτικότητα του υποστρώματος. Η εικόνα παρατίθεται αυτούσια από τη δουλειά των Long W. και Cheeseman C.I. 2015. Ο πίνακας για την ευκολότερη ανάγνωση της παραπάνω εικόνας παρατίθεται στο Παράρτημα της παρούσας εργασίας.

1.6. Αποτελέσματα Εξελικτικών Αναλύσεων Μεταφορέων Γλυκόζης

Σαν παράδειγμα των εξελικτικών αναλύσεων στους μεταφορείς γλυκόζης θα δούμε τα αποτελέσματα της δουλειάς των Amy L. Wilson-O'Brien et al., "*Evolutionary ancestry and novel functions of the mammalian glucose transporter (GLUT) family*" [21]. Μέσα από μια συγκριτική γονιδιωματική ανάλυση παρουσιάζεται μια στατιστικά υποστηριζόμενη φυλογενετική ανάλυση όλων των ισομορφών των μεταφορέων γλυκόζης GLUT στα θηλαστικά.

Αποκαλύπτεται πως οι GLUT πρωτεΐνες χωρίζονται σε πέντε διακριτές κατηγορίες (αντι για των τριών που είδαμε παραπάνω) και αυτή η εξελικτική τους καταγωγή δίνει μια ολοκληρωμένη εικόνα για τη δομή, τη λειτουργία και τους μηχανισμούς εντός των διαφορετικών ομάδων [21] (**Εικόνα 6**).

Σκοπός της παρούσας εργασίας ήταν να γίνει μια φυλογενετική ανάλυση των μεταφορέων σακχάρων των MFS στους ζυμομύκητες και να διαπιστωθεί σε ποιο βαθμό γίνονται διπλασιασμοί αυτών των γονιδίων και εάν σχετίζονται με προσαρμογές των ζυμομυκήτων στα περιβάλλοντά τους.



Εικόνα 6. Φυλογενετική ανάλυση GLUT θηλαστικών. Η αυτούσια εικόνα από τα αποτελέσματα της έρευνας των Wilson-O'Brien et al. “*Evolutionary ancestry and novel functions of the mammalian glucose transporter (GLUT) family*”, και παρουσιάζεται η φυλογενετική ανάλυση των GLUT των θηλαστικών.

2. Υλικά και Μέθοδοι

2.1. Λήψη Δεδομένων

Για τη λήψη των δεδομένων που χρησιμοποιήθηκαν στην παρούσα εργασία περιηγηθήκαμε στη εξιδεικευμένη για μύκητες, βάση δεδομένων Ensembl Fungi (release 53 - Απρίλιος 2022) [22], <https://www.ensembl.org> και πιο συγκεκριμένα στην <https://fungi.ensembl.org/species.html>. Στο πεδίο αναζήτησης που μας προτρέπει να επιλέξουμε τα είδη που μας ενδιαφέρουν, πληκτρολογήσαμε τη λέξη “Saccharomyces” (σακχαρομύκητες), όπου η αναζήτηση μας απέδωσε 132 αποτελέσματα (26 Απριλίου 2022). Εκεί επιλέξαμε να κάνουμε εκφόρτωση όλων των δεδομένων που μας επέστρεψε η αναζήτηση μας (Download what you see). Οι μοναδικοί κωδικοί που μας επιστράφηκαν (Accessions) ήταν οι εξής:

Πίνακας 1. Τα 132 Accessions που επιστράφηκαν κατά την αναζήτηση της Ensembl Fungi.

GCA_001298375.2	GCA_000976935.2	GCA_000977985.2	GCA_000976275.4
GCA_001413975.1	GCA_000976965.3	GCA_000978015.2	GCA_000976305.2
GCA_003046715.1	GCA_000976995.2	GCA_000978045.2	GCA_000976335.2
GCA_010356925.1	GCA_000977025.2	GCA_000978075.3	GCA_000976365.2
GCA_001634645.1	GCA_000977055.3	GCA_000978105.2	GCA_000181435.1
GCA_002804325.1	GCA_000977085.2	GCA_000978135.2	GCA_000976395.2
GCA_003046695.1	GCA_000977115.3	GCA_000978165.2	GCA_000976425.2
GCA_000146045.2	GCA_000977145.2	GCA_000978195.2	GCA_000976455.3
GCA_003046745.1	GCA_000977175.3	GCA_000978225.2	GCA_000976485.2
GCA_010356935.1	GCA_000977205.2	GCA_000978255.2	GCA_000976515.2
GCA_000269885.1	GCA_000977235.2	GCA_000978285.2	GCA_000976545.2
GCA_000218975.1	GCA_000977265.3	GCA_000978315.2	GCA_000976575.4
GCA_000190255.1	GCA_000977295.2	GCA_000978345.2	GCA_000976605.3
GCA_000326005.1	GCA_000977325.2	GCA_000975735.4	GCA_000662435.2
GCA_000260735.1	GCA_000977355.3	GCA_000975765.2	GCA_000976665.2
GCA_000325965.1	GCA_000977385.2	GCA_000975585.2	GCA_001983315.1
GCA_000568295.1	GCA_000977415.2	GCA_000975615.2	GCA_002335645.1
GCA_000568055.1	GCA_000977445.2	GCA_000975795.2	GCA_014132395.1
GCA_000568365.1	GCA_000977475.3	GCA_000975825.2	GCA_009014655.1
GCA_000149365.1	GCA_000977535.2	GCA_000975855.3	GCA_000326105.1
GCA_000190235.1	GCA_000977565.2	GCA_000975885.2	GCA_001298625.1
GCA_000190215.1	GCA_000977595.2	GCA_000975915.2	GCA_000167075.2
GCA_000975645.3	GCA_000977625.2	GCA_000975945.2	GCA_011022315.1
GCA_000975675.2	GCA_000977655.2	GCA_000975975.3	GCA_000004155.2
GCA_000975705.2	GCA_000977685.2	GCA_000976005.2	GCA_000149845.2
GCA_000976695.3	GCA_000977745.2	GCA_000976035.2	GCA_000150505.2
GCA_000976725.3	GCA_000977775.3	GCA_000976065.2	GCA_000002945.2
GCA_000976755.2	GCA_000977805.2	GCA_000976095.2	GCA_900162805.1
GCA_000976785.2	GCA_000977835.2	GCA_000976125.2	GCA_000530735.1
GCA_000976815.3	GCA_000977865.2	GCA_000976155.4	GCA_005406105.1
GCA_000976845.3	GCA_000977895.2	GCA_000976185.2	GCA_001984395.2
GCA_000976875.3	GCA_000977925.2	GCA_000976215.2	GCA_000026365.1
GCA_000976905.2	GCA_000977955.2	GCA_000976245.2	GCA_001972345.1

Για να κάνουμε αυτόματη και μαζική ανάκτηση των πληροφοριών που μας ενδιαφέρουν από τα 132 προσπελασμένα accessions χρησιμοποιήσαμε την εφαρμογή Batch Entrez [23] για να κάνουμε ανάκτηση περαιτέρω πληροφοριών καθώς και των Protein FASTA αρχείων που θα χρησιμοποιούσαμε στην πρωτεωμική ανάλυση μας παρακάτω (Στη βιοπληροφορική και στη βιοχημεία, η μορφή FASTA είναι μια μορφή κειμένου και βοηθάει στην αναπαράσταση νουκλεοτιδικών αλληλουχιών ή αμινοξέων ακολουθιών, στις οποίες τα νουκλεοτίδια ή τα αμινοξέα αναπαρίστανται χρησιμοποιώντας κωδικούς ενός γράμματος. Η μορφή FASTA επίσης επιτρέπει στα ονόματα των ακολουθιών σχόλια στην πρώτη γραμμή της). Δημιουργώντας ένα αρχείο txt το οποίο περιείχε μόνο τις 132 μοναδικές καταχωρήσεις (Πίνακας 1), η εφαρμογή Batch Entrez μας επέστρεψε 132 νέες καταχωρήσεις που αντιστοιχούν σε αυτές που πήραμε από την Ensembl Fungi (Η έξοδος της εφαρμογής μας έδωσε το εξής μήνυμα:

```
Received lines: 132
Rejected lines: 0
Removed duplicates: 0
Passed to Entrez: 132
Retrieve records for 132 UID\(s\)
```

Πατώντας στον σύνδεσμο οδηγούμαστε στην ιστοσελίδα του NCBI (The National Center for Biotechnology Information). Στο παρόν site επιλέγουμε να κατεβάσουμε τα Protein FASTA αρχεία που θα χρησιμοποιήσουμε.

2.2. Εργαλεία, Λογισμικό, Αλγόριθμοι και Πλατφόρμες

Για την περάτωση της παρούσας εργασίας χρησιμοποιήθηκε μια πληθώρα υπολογιστικών, βιοπληροφορικών και γραφιστικών εργαλείων. Σε αυτό το κεφάλαιο θα γίνει μια σύντομη παρουσίαση όλων των μεθόδων που χρησιμοποιήθηκαν.

2.2.1. Linux

Τα GNU Linux είναι ένα λειτουργικό σύστημα που αποτελείται από ελεύθερο λογισμικό. Η χρήση του είναι παρόμοια με αυτή του Unix, αλλά όλος ο πηγαίος κώδικας του έχει γραφτεί από την αρχή ως ελεύθερο λογισμικό υπό την ελεύθερη άδεια χρήσης GNU General Public License. Έχουν γραφτεί κυρίως στις γλώσσες προγραμματισμού C και Assembly. Χρησιμοποιούνται κατά κόρον στην επιστήμη της βιοπληροφορικής για τους εξής λόγους: Α) Πολλά εργαλεία έχουν γραφτεί μόνο για Linux, Β) πολλά εργαλεία είναι αργά για τα Windows, Γ) τα Linux έχουν build-in γλώσσες προγραμματισμού, Δ) είναι εύκολη η δημιουργία ροών εργασίας στα περιβάλλοντα Linux και τέλος, Ε) είναι δωρεάν διαθέσιμα. Επιπλέον τα Linux δίνουν τη δυνατότητα στον χρήστη να χρησιμοποιεί έναν διερμηνέα γραμμής εντολών (Command Line Interpreter ή Shell) ο οποίος ονομάζεται BASH. Μέσω του BASH ο χρήστης μπορεί να κάνει απλές λειτουργίες χωρίς να χρειάζεται κάποια γλώσσα προγραμματισμού όπως για παράδειγμα: να δημιουργήσει βρόγχους, να φτιάξει φακέλους και υποφακέλους που πολύ γρήγορα μπορεί να τους μετονομάζει και να τους τροποποιεί, να πλοηγηθεί στον υπολογιστή του και το σημαντικότερο, να μπορεί να καλεί κομμάτια από txt και να τα συνενώνει χωρίς να χρειάζεται εξειδικευμένες γνώσεις όπως π.χ. χειρισμός κειμένου (text manipulation) σε γλώσσα Perl.

2.2.2. Python3

Η Python (έκδοση 3.10.5) είναι μια ευρέως χρησιμοποιούμενη γενικής χρήσης, υψηλού επιπέδου γλώσσα προγραμματισμού στον τομέα της βιοπληροφορικής. Η σχεδιαστική της φιλοσοφία δίνει έμφαση στην αναγνωσιμότητα του κώδικα και η σύνταξή της επιτρέπει στους προγραμματιστές να εκφράζουν έννοιες σε λιγότερες γραμμές κώδικα από ό,τι θα ήταν δυνατό σε γλώσσες όπως η C++ ή η Java. Η Python υποστηρίζει πολλαπλά παραδείγματα προγραμματισμού, όπως αντικειμενοστραφή. Μαζί με την R Language είναι η γλώσσα που έχει τις περισσότερες βιβλιοθήκες βιοπληροφορικών εργαλείων.

2.2.3. Perl

Η Perl (έκδοση 5.36.0) είναι μια υψηλού επιπέδου, γενικής χρήσης γλώσσα προγραμματισμού που αναπτύχθηκε αρχικά για χειρισμό κειμένου. Δανείζεται πολλές δυνατότητες από την C και τα Shell. Βρίσκεται διαθέσιμη στη σελίδα <https://www.perl.org/>. Ενδείκνυται για τον χειρισμό δεδομένων και χρησιμοποιείται ιδιαίτερα στις βιοπληροφορικές αναλύσεις.

2.2.4. MFS Search Tool

Για την ανίχνευση και την ταξινόμηση των μεταφορέων σακχάρων χρησιμοποιήθηκε ένα εργαλείο γραμμένο στη γλώσσα προγραμματισμού Python. Το workflow ακολουθεί την εξής μορφή: Α) Ο αλγόριθμος για την αναγνώριση και ταξινόμηση των μεταφορέων εν αρχή πραγματοποιεί hmmscan έναντι κρυφών Μαρκοβιανών μοντέλων (Hidden Markov Models, HMMs), Β) η δοθείσα ακολουθία εν συνεχεία υποβάλλεται σε hmmscan έναντι μοντέλων που δημιουργούνται μέσω του εργαλείου. Παράλληλα στο ίδιο στάδιο πραγματοποιείται εξελικτική ταξινόμηση καθώς και έλεγχος πιστότητας της δοθείσας αμινοξικής ακολουθίας, Γ) τέλος, εάν περάσουν το “τεστ”, οι μεταφορείς σακχάρων ταξινομούνται σε putative και high confidence [11]. Επίσης, ο συγκεκριμένος αλγόριθμος που αναπτύχθηκε στο εργαστήριο του Αναπληρωτή Καθηγητή Γρηγόριου Αμούτζια, από τον κ. Μάριο Νικολαΐδη, με την αρωγή του κ. Παναγιώτη Βλασταρίδη υλοποιήθηκε και ως διαδικτυακό εργαλείο σε server. Το εργαλείο επιτρέπει στον χρήστη την ικανότητα να κατεβάσει αποτελέσματα στον δικό του υπολογιστή σε tabular μορφή [11].

2.2.4.1. Hidden Markov Models

Στην Υπολογιστική Βιολογία, ένα κρυφό μοντέλο Markov (Hidden Markov Model, HMM) είναι μια στατιστική προσέγγιση που χρησιμοποιείται συχνά για τη μοντελοποίηση βιολογικών ακολουθιών. Κατά την εφαρμογή της, μια ακολουθία μοντελοποιείται ως έξοδος μιας διακριτής στοχαστικής διαδικασίας, η οποία εξελίσσεται μέσω μιας σειράς καταστάσεων που είναι “κρυμμένες” από τον παρατηρητή. Κάθε τέτοια κρυφή κατάσταση εκπροσωπείται από ένα σύμβολο και αντιπροσωπεύει μια στοιχειώδη μονάδα των μοντελοποιημένων δεδομένων, π.χ. στην περίπτωση μιας αλληλουχίας πρωτεΐνης, αντιπροσωπεύει ένα αμινοξύ [24].

2.2.5. Εργαλείο Εύρεσης Γονιδιώματος Πυρήνα

Στην εργασία με τίτλο: “*Comparative Analysis of the Core Proteomes among the Pseudomonas Major Evolutionary Groups Reveals Species-Specific Adaptations for Pseudomonas aeruginosa and Pseudomonas chlororaphis*”, οι Nikolaidis et al. [25] για να προσδιορίσουν τον πυρήνα των πρωτεωμάτων ενός συνόλου μικροοργανισμών ανέπτυξαν 3 αλγορίθμους στη γλώσσα Python που χρησιμοποιούν την τοπική στοίχιση BLAST

χρησιμοποιώντας τα εργαλεία Biopython και blast+. Η τεχνική αυτή χρησιμοποιήθηκε για την εύρεση των ορθολόγων γονιδίων μεταξύ των διαφόρων πρωτεωμάτων και είχε αναπτυχθεί από το εργαστήριο μας και έχει ήδη εφαρμοστεί για την εξελικτική ανάλυση βακτηρίων του γένους *Pseudomonas*. Αφού ένας χρήστης ορίσει ένα πρωτέωμα αναφοράς πραγματοποιείται εύρεση ορθολόγων στα άλλα γονιδιώματα/πρωτεώματα με ανταποδοτικό Blast (Best Reciprocal Blast). Επίσης, εφάρμοσαν ένα ακόμη πρόσθετο κριτήριο για το καλύτερο και αυστηρότερο φιλτράρισμα των προβληματικών ορθολόγων. Στη συνέχεια δημιουργείται ένας πίνακας των καλύτερων αμοιβαίων BLAST hits, όπου οι σειρές αντιστοιχούν στις πρωτεΐνες εκείνες του πρωτεώματος αναφοράς που έχει οριστεί από τον χρήστη και οι στήλες αντιστοιχούν στα υπόλοιπα πρωτεώματα του συνόλου υπό ανάλυση. Στη συνέχεια δημιουργείται και δεύτερος πίνακας που περιλαμβάνει μόνο τις πρωτεΐνες που υπάρχουν σε όλα τα αναλυόμενα πρωτεώματα και αντιστοιχεί στο πραγματικό πρωτέωμα-πυρήνα της ομάδας. Τέλος, τα script πραγματοποιούν πολλαπλές στοιχίσεις με το λογισμικό MUSCLE για κάθε αναγνωρισμένη ομάδα ορθολόγων, συνενώνοντας τους σε μια υπερ-στοίχιση και γίνεται φιλτράρισμα των κακώς στοιχισμένων στηλών με το λογισμικό gblocks. Από την πολλαπλή στοιχίση έπειτα το πρόγραμμα επιχειρεί να δημιουργήσει ένα Neighbor Joining φυλογονιδιωματικό δέντρο με το μοντέλο Kimura με 500 bootstraps [25].

2.2.6. Φυλογενετική Ανάλυση

Το SeaView [26] στην πιο σύγχρονη έκδοση του (version 5.0.5) είναι ένα πρόγραμμα που μπορεί να λειτουργήσει σε πολλαπλές πλατφόρμες (MS Windows, Linux, MacOS X) και έχει σχεδιαστεί για να διευκολύνει την πολλαπλή στοίχιση και τη δημιουργία φυλογενετικών δέντρων από δεδομένα μοριακής αλληλουχίας μέσω της χρήσης μιας γραφικής διεπαφής χρήστη. Το SeaView συνδυάζει όλες τις λειτουργίες των προηγούμενων εκδόσεων του (όπως π.χ. SeaView και Phylo_win) και τις επεκτείνει προσθέτοντας πρόσβαση δικτύου σε βάσεις δεδομένων ακολουθιών, επιτρέπει την πολλαπλή στοίχιση, διευκολύνει στη δημιουργία δέντρων (Maximum Likelihood) μέγιστης πιθανοφάνειας [ή και φειδωλότητας (Maximum Parsimony)] με PhyML και τέλος δίνει πρόσβαση στις εξής εντολές: Α) προβολή, Β) εκτύπωση, και Γ) αντιγραφή ένριζων (ή μη), φυλογενετικών δέντρων. Σε σχέση με την ευρεία παρούσα προσφορά εργαλείων και αλγορίθμων για φυλογενετικές αναλύσεις, το SeaView είναι ιδιαίτερα χρήσιμο για διδασκαλία και για περιστασιακούς χρήστες. Το SeaView είναι δωρεάν διαθέσιμο στη διεύθυνση <http://pbil.univ-lyon1.fr/software/seaview>. Το SeaView ακόμη έχει δύο μεθόδους δημιουργίας δέντρων αποστάσεων, με Neighbor-Joining και BioNJ. Ακόμη έχει την δυνατότητα να αξιολογήσει την εμπιστοσύνη στους κλάδους του κάθε δέντρου με την μέθοδο bootstrap που αναλύεται ενδελεχώς παρακάτω. Τα παραγόμενα δένδρα δύναται να αποθηκευτούν για περαιτέρω ανάλυση ή και σχολιασμό στις εξής μορφές: Newick format, ή σε SVG, PDF ή PostScript αρχεία [12]. Μπορεί να χρησιμοποιήσει τους αλγορίθμους MUSCLE και Clustal Omega για να επιτελέσει πολλαπλή στοίχιση [11].

Στο έργο του ο Gascuel [27] προτείνει μια βελτιωμένη έκδοση του αλγορίθμου Neighbor-Joining (NJ) των Saitou και Nei [28]. Αυτός ο νέος αλγόριθμος έχει το όνομα BIONJ [27].

Η μορφή δέντρου Newick είναι ένας τρόπος αναπαράστασης θεωρητικών δέντρων χρησιμοποιώντας παρενθέσεις και κόμματα. Χρησιμοποιείται ευρέως στις βιολογικές επιστήμες και ιδιαίτερα για τη διαδικασία της φυλογένεσης. Εκτός από την μορφή Newick τα φυλογενετικά δένδρα μπορεί να αποδοθούν και σε άλλες μορφές, όπως π.χ. Nexus, PhyloXML και TGF.

Ο αλγόριθμος MUSCLE (MULTiple Sequence Comparison by Log- Expectation), είναι ένα πρόγραμμα το οποίο επιτρέπει στην δημιουργία πολλαπλών στοιχίσεων αλληλουχιών πρωτεϊνών. Το MUSCLE προσφέρει μια σειρά επιλογών που παρέχουν βελτιωμένη ταχύτητα και ακρίβεια στοίχισης σε σύγκριση με τα υπόλοιπα διαθέσιμα προγράμματα. Το MUSCLE είναι δωρεάν διαθέσιμο στη διεύθυνση <http://www.drive5.com/muscle> [29].

Το TreeDyn είναι ένα εργαλείο επεξεργασίας, οπτικοποίησης και σχολιασμού φυλογενετικών δέντρων που περιλαμβάνει εργαλεία για χειρισμό και σχολιασμό δέντρων και χρησιμοποιεί μεταπληροφορίες μέσω δυναμικών γραφικών τελεστών ή σεναρίων για να βοηθήσει στην ανάλυση και τους σχολιασμούς μεμονωμένων δέντρων ή συλλογών δέντρων [30]. Παρέχει μια πληθώρα εργαλείων για την περαιτέρω ανάλυση παραγόμενων δέντρων (όπως μεγέθυνση, σμίκρυνση, χρωματισμό, collapse κ.α.) και δέχεται αρχεία σε μορφή Newick ή και TGF (μια μορφή αρχείου του TreeDyn).

Το Interactive Tree of Life [31] (το οποίο μπορεί να βρεθεί στην διεύθυνση <https://itol.embl.de>) είναι ένα διαδικτυακό εργαλείο για την εμφάνιση, το χειρισμό και τον σχολιασμό φυλογενετικών δέντρων. Είναι δωρεάν διαθέσιμο και ανοιχτό σε όλους και υποστηρίζει βελτιωμένες υπηρεσίες επί πληρωμή. Η έκδοση 5 έχει μια εντελώς νέα μηχανή προβολής δέντρων, μαζί με πολλές νέες δυνατότητες. Η μηχανή εμφάνισης και παρουσίασης ενός δέντρου υλοποιείται σε καθαρή JavaScript και χρησιμοποιεί το HTML5 Canvas για οπτικοποίηση. Στο iTOL 5, η μηχανή γραφικών ξαναγράφηκε πλήρως και βελτιστοποιήθηκε με υποστήριξη για οθόνες υψηλής ανάλυσης [31].

2.2.7. Επεξεργασία Εικόνων και Βιβλιογραφίας

Το διαδικτυακό εργαλείο BioRender, το οποίο έχει δωρεάν και επί πληρωμή εκδόσεις, επιτρέπει στον χρήστη την επεξεργασία εικόνων βιολογικού κυρίως ενδιαφέροντος. Έχει εκατοντάδες λειτουργίες και έτοιμα εικονίδια που μπορεί ο χρήστης να χρησιμοποιήσει για να οπτικοποιήσει τα αποτελέσματα της έρευνας του. Στην παρούσα διπλωματική εργασία χρησιμοποιήθηκε για την επεξεργασία εικόνων.

Για τη δημιουργία του πίνακα βιβλιογραφικών παραπομπών της παρούσας εργασίας χρησιμοποιήθηκε το πρόγραμμα EndNote, το οποίο παρέχει στον χρήστη την δυνατότητα να κατεβάσει τις παραπομπές που των ενδιαφέρουν, να δημιουργήσει μια βιβλιοθήκη και στη συνέχεια να παραθέτει τα σημεία τα οποία εν τέλη χρησιμοποιήθηκαν στο κείμενο του.

3. Αποτελέσματα - Συζήτηση

Στο κεφάλαιο Υλικά και Μέθοδοι πραγματοποιήθηκε σύντομη σύνοψη των εργαλείων και των αλγορίθμων που χρησιμοποιήθηκαν για την περάτωση της παρούσας διπλωματικής εργασίας. Στο κεφάλαιο αυτό θα γίνει ανάλυση της μεθόδου με την οποία εξήγαμε τα αποτελέσματα μας. Εν αρχή, αφού λάβαμε 132 μοναδικά accessions από τη βάση δεδομένων Ensembl Fungi (release 53 - Απρίλιος 2022), <https://www.ensembl.org> και πιο συγκεκριμένα στην <https://fungi.ensembl.org/species.html>, συμπληρώσαμε ένα αρχείο το οποίο περιείχε συγκεκριμένες κατηγορίες οι οποίες παρουσιάζονται παρακάτω. Τα αρχεία όπως ελήφθησαν από τη βάση δεδομένων βρίσκονταν σε δύο μορφές, ένα αρχείο annotation, καθώς και ένα αρχείο FASTA. Το τελικό αποτέλεσμα είναι ένας συγκεντρωτικός πίνακας ο οποίος περιέχει την πληροφορία του ονόματος των ειδών, το όνομα του στελέχους, τον ξεχωριστό αριθμό του καθενός, το βάθος ανάγνωσης της αλληλούχισης, το επίπεδο της κατασκευής του γονιδιώματος, καθώς και την τιμή N50 (Πίνακας 2). Ως contig εννοούμε τους επικαλυπτόμενους κλώνους ή τις συνεχόμενες αλληλουχίες οι οποίες συναρμολογούν σταδιακά μεγαλύτερες αλληλουχίες. Η ένωση των contigs δημιουργεί τα ικριώματα (scaffolds).

Πίνακας 2. Συγκεντρωτικός πίνακας πληροφοριών των 132 στελεχών Σακχαρομυκήτων που επιστράφηκαν από την αναζήτηση μας και έχουν εμπλουτιστεί με την εξής πληροφορία: Είδος, στέλεχος, ID, βάθος ανάγνωσης (Genome Coverage), επίπεδο κατασκευής γονιδιώματος (Assembly Level) και N50. Με ανοιχτό πράσινο σημειώνονται τα στελέχη τα οποία επιλέχθηκαν βάσει των παρακάτω παραμέτρων. Σε βαθύ πράσινο αντιστοιχεί το στέλεχος του *Saccharomyces cerevisiae* S288C, το οποίο αποτελεί το γονιδίωμα αναφοράς μας.

NO	Species	Strain	Taxid	Genome Coverage	Assembly Level	N50
1	<i>Schizosaccharomyces pombe</i> (fission yeast)	972h-	4896	-	Chromosome	4539804
2	<i>Schizosaccharomyces cryophilus</i> OY26 (ascomycetes)	OY26	653667	51x	Scaffold	1386795
3	<i>Zygosaccharomyces rouxii</i> (budding yeasts)	CBS 732	4956	-	Chromosome	1496342
4	<i>Saccharomyces cerevisiae</i> S288C (baker's yeast)	S288C	559292	-	Complete	924431
5	<i>Saccharomyces cerevisiae</i> RM11-1a (baker's yeast)	RM11-1a	285006	-	Scaffold	795018
6	<i>Schizosaccharomyces japonicus</i> yFS275 (ascomycetes)	yFS275	402676	47x	Scaffold	3866997
7	<i>Schizosaccharomyces octosporus</i> yFS286 (ascomycetes)	yFS286	483514	39x	Scaffold	3863862
8	<i>Saccharomyces kudriavzevii</i> IFO 1802 (budding yeasts)	IFO 1802	226230	3.4x	Scaffold	11253
9	<i>Saccharomyces cerevisiae</i> YJM789 (baker's yeast)	YJM789	307796	-	Scaffold	429709
10	<i>Saccharomyces cerevisiae</i> Vin13 (baker's yeast)	Vin13	764099	20x	Chromosome	700638
11	<i>Saccharomyces cerevisiae</i> VL3 (baker's yeast)	VL3	764100	20x	Chromosome	656188
12	<i>Saccharomyces cerevisiae</i> FostersB (baker's yeast)	FostersB	764102	20x	Chromosome	626897
13	<i>Saccharomyces cerevisiae</i> EC1118 (baker's yeast)	Lalvin EC1118	643680	-	Contig	657733
14	<i>Saccharomyces cerevisiae</i> Kyokai no. 7 (baker's yeast)	Kyokai no. 7	721032	9.1x	Scaffold	902266
15	<i>Saccharomyces cerevisiae</i> CEN.PK113-7D (baker's yeast)	CEN.PK113-7D	889517	18x	Chromosome	918791
16	<i>Saccharomyces cerevisiae</i> Lalvin QA23 (baker's yeast)	Lalvin QA23	764098	20x	Scaffold	182942
17	<i>Saccharomyces cerevisiae</i> FostersO (baker's yeast)	FostersO	764101	20x	Contig	195316
18	<i>Saccharomyces cerevisiae</i> x <i>Saccharomyces kudriavzevii</i> VIN7 (budding yeasts)	VIN7	1095631	20x	Contig	139679
19	<i>Zygosaccharomyces bailii</i> ISA1307 (budding yeasts)	SA1307	1355161	600x	Scaffold	322684
20	<i>Saccharomyces cerevisiae</i> P301 (baker's yeast)	P301	1182968	14x	Chromosome	548808
21	<i>Saccharomyces cerevisiae</i> P283 (baker's yeast)	P283	1177187	14x	Chromosome	509916
22	<i>Saccharomyces cerevisiae</i> R103 (baker's yeast)	R103	1182967	14x	Chromosome	375263
23	<i>Saccharomyces cerevisiae</i> YJM993 (baker's yeast)	YJM993	1294331	180x	Chromosome	913508
24	<i>Saccharomyces cerevisiae</i> YJM195 (baker's yeast)	YJM195	1294305	100x	Chromosome	916102
25	<i>Saccharomyces cerevisiae</i> YJM244 (baker's yeast)	YJM244	1294306	120x	Chromosome	906535
26	<i>Saccharomyces cerevisiae</i> YJM1078 (baker's yeast)	YJM1078	1296266	75x	Chromosome	907409

27	Saccharomyces cerevisiae YJM1083 (baker's yeast)	JM1083	1292971	500x	Chromosome	923843
28	Saccharomyces cerevisiae YJM1129 (baker's yeast)	YJM1129	1293430	180x	Chromosome	903327
29	Saccharomyces cerevisiae YJM189 (baker's yeast)	YJM189	1294303	400x	Chromosome	896353
30	Saccharomyces cerevisiae YJM193 (baker's yeast)	YJM193	1294304	60x	Chromosome	934902
31	Saccharomyces cerevisiae YJM248 (baker's yeast)	YJM248	1294307	90x	Chromosome	907630
32	Saccharomyces cerevisiae YJM270 (baker's yeast)	YJM270	1294308	110x	Chromosome	913025
33	Saccharomyces cerevisiae YJM271 (baker's yeast)	YJM271	1294309	90x	Chromosome	928255
34	Saccharomyces cerevisiae YJM320 (baker's yeast)	YJM320	947042	100x	Chromosome	905335
35	Saccharomyces cerevisiae YJM326 (baker's yeast)	YJM326	468558	400x	Chromosome	907133
36	Saccharomyces cerevisiae YJM428 (baker's yeast)	YJM428	947044	90x	Chromosome	910751
37	Saccharomyces cerevisiae YJM450 (baker's yeast)	YJM450	1294310	200x	Chromosome	914011
38	Saccharomyces cerevisiae YJM451 (baker's yeast)	YJM451	502869	75x	Chromosome	895934
39	Saccharomyces cerevisiae YJM453 (baker's yeast)	YJM453	1294311	110x	Chromosome	911881
40	Saccharomyces cerevisiae YJM456 (baker's yeast)	YJM456	1294312	190x	Chromosome	897427
41	Saccharomyces cerevisiae YJM470 (baker's yeast)	YJM470	1294313	180x	Chromosome	902232
42	Saccharomyces cerevisiae YJM541 (baker's yeast)	YJM541	1294314	100x	Chromosome	909353
43	Saccharomyces cerevisiae YJM554 (baker's yeast)	YJM554	1294315	-	Chromosome	918804
44	Saccharomyces cerevisiae YJM555 (baker's yeast)	YJM555	1294316	190x	Chromosome	909661
45	Saccharomyces cerevisiae YJM627 (baker's yeast)	YJM627	1294317	40x	Chromosome	905901
46	Saccharomyces cerevisiae YJM681 (baker's yeast)	YJM681	1294318	40x	Chromosome	906981
47	Saccharomyces cerevisiae YJM682 (baker's yeast)	YJM682	1294319	100x	Chromosome	896702
48	Saccharomyces cerevisiae YJM683 (baker's yeast)	YJM683	1294320	130x	Chromosome	900521
49	Saccharomyces cerevisiae YJM689 (baker's yeast)	YJM689	1294321	40x	Chromosome	895385
50	Saccharomyces cerevisiae YJM693 (baker's yeast)	YJM693	1294322	80x	Chromosome	907430
51	Saccharomyces cerevisiae YJM969 (baker's yeast)	YJM969	1294323	25x	Chromosome	924433
52	Saccharomyces cerevisiae YJM972 (baker's yeast)	YJM972	1294324	100x	Chromosome	902781
53	Saccharomyces cerevisiae YJM975 (baker's yeast)	YJM975	1294325	30x	Chromosome	904046
54	Saccharomyces cerevisiae YJM978 (baker's yeast)	YJM978	1294326	90x	Chromosome	902126
55	Saccharomyces cerevisiae YJM981 (baker's yeast)	YJM981	1294327	40x	Chromosome	916198
56	Saccharomyces cerevisiae YJM984 (baker's yeast)	YJM984	1294328	30x	Chromosome	910820
57	Saccharomyces cerevisiae YJM987 (baker's yeast)	YJM987	1294329	150x	Chromosome	912332
58	Saccharomyces cerevisiae YJM990 (baker's yeast)	YJM990	1294330	-	Chromosome	913759
59	Saccharomyces cerevisiae YJM996 (baker's yeast)	YJM996	1294332	180x	Chromosome	898411
60	Saccharomyces cerevisiae YJM1133 (baker's yeast)	YJM1133	1294333	60x	Chromosome	901770
61	Saccharomyces cerevisiae YJM1190 (baker's yeast)	YJM1190	1294334	30x	Chromosome	902060
62	Saccharomyces cerevisiae YJM1199 (baker's yeast)	YJM1199	1294335	60x	Chromosome	900262
63	Saccharomyces cerevisiae YJM1202 (baker's yeast)	YJM1202	1294336	650x	Chromosome	912629
64	Saccharomyces cerevisiae YJM1208 (baker's yeast)	YJM1208	1294337	220x	Chromosome	903091
65	Saccharomyces cerevisiae YJM1242 (baker's yeast)	YJM1242	1294338	-	Chromosome	894781
66	Saccharomyces cerevisiae YJM1244 (baker's yeast)	YJM1244	1294339	50x	Chromosome	911385
67	Saccharomyces cerevisiae YJM1248 (baker's yeast)	YJM1248	1294340	100x	Chromosome	924574
68	Saccharomyces cerevisiae YJM1250 (baker's yeast)	YJM1250	1294341	75x	Chromosome	1026598
69	Saccharomyces cerevisiae YJM1252 (baker's yeast)	YJM1252	1294342	500x	Chromosome	929116
70	Saccharomyces cerevisiae YJM1273 (baker's yeast)	YJM1273	1294343	400x	Chromosome	910208
71	Saccharomyces cerevisiae YJM1304 (baker's yeast)	YJM1304	1294344	750x	Chromosome	913052
72	Saccharomyces cerevisiae YJM1307 (baker's yeast)	YJM1307	1294345	-	Chromosome	907466
73	Saccharomyces cerevisiae YJM1311 (baker's yeast)	YJM1311	1294346	400x	Chromosome	903706
74	Saccharomyces cerevisiae YJM1326 (baker's yeast)	YJM1326	1294347	800x	Chromosome	913940
75	Saccharomyces cerevisiae YJM1332 (baker's yeast)	YJM1332	1294348	400x	Chromosome	923089
76	Saccharomyces cerevisiae YJM1336 (baker's yeast)	YJM1336	1294349	700x	Chromosome	901186
77	Saccharomyces cerevisiae YJM1338 (baker's yeast)	YJM1338	1294350	350x	Chromosome	895119
78	Saccharomyces cerevisiae YJM1341 (baker's yeast)	YJM1341	1294351	110x	Chromosome	912304
79	Saccharomyces cerevisiae YJM1342 (baker's yeast)	YJM1342	1294352	500x	Chromosome	912252
80	Saccharomyces cerevisiae YJM1355 (baker's yeast)	YJM1355	1294353	550x	Chromosome	912538

81	Saccharomyces cerevisiae YJM1356 (baker's yeast)	YJM1356	1294354	400x	Chromosome	910189
82	Saccharomyces cerevisiae YJM1381 (baker's yeast)	YJM1381	1294355	-	Chromosome	894919
83	Saccharomyces cerevisiae YJM1383 (baker's yeast)	YJM1383	1294356	650x	Chromosome	931094
84	Saccharomyces cerevisiae YJM1385 (baker's yeast)	YJM1385	1294357	350x	Chromosome	903329
85	Saccharomyces cerevisiae YJM1386 (baker's yeast)	YJM1386	1294358	120x	Chromosome	902760
86	Saccharomyces cerevisiae YJM1387 (baker's yeast)	YJM1387	1294359	200x	Chromosome	904346
87	Saccharomyces cerevisiae YJM1389 (baker's yeast)	YJM1389	1294361	50x	Chromosome	897207
88	Saccharomyces cerevisiae YJM1399 (baker's yeast)	YJM1399	1294362	110x	Chromosome	903451
89	Saccharomyces cerevisiae YJM1400 (baker's yeast)	YJM1400	1294363	700x	Chromosome	898273
90	Saccharomyces cerevisiae YJM1401 (baker's yeast)	YJM1401	1294364	500x	Chromosome	889931
91	Saccharomyces cerevisiae YJM1402 (baker's yeast)	YJM1402	1294365	200x	Chromosome	902709
92	Saccharomyces cerevisiae YJM1415 (baker's yeast)	YJM1415	1294366	600x	Chromosome	744674
93	Saccharomyces cerevisiae YJM1418 (baker's yeast)	YJM1418	1294368	600x	Chromosome	907734
94	Saccharomyces cerevisiae YJM1419 (baker's yeast)	YJM1419	1294369	80x	Chromosome	891878
95	Saccharomyces cerevisiae YJM1433 (baker's yeast)	YJM1433	1294370	75x	Chromosome	910803
96	Saccharomyces cerevisiae YJM1434 (baker's yeast)	YJM1434	1294371	40x	Chromosome	912768
97	Saccharomyces cerevisiae YJM1439 (baker's yeast)	YJM1439	1294372	50x	Chromosome	922967
98	Saccharomyces cerevisiae YJM1443 (baker's yeast)	YJM1443	1294373	50x	Chromosome	892329
99	Saccharomyces cerevisiae YJM1444 (baker's yeast)	YJM1444	1294374	50x	Chromosome	897256
100	Saccharomyces cerevisiae YJM1447 (baker's yeast)	YJM1447	1294375	80x	Chromosome	909714
101	Saccharomyces cerevisiae YJM1450 (baker's yeast)	YJM1450	1294376	120x	Chromosome	908914
102	Saccharomyces cerevisiae YJM1460 (baker's yeast)	YJM1460	1294377	400x	Chromosome	897146
103	Saccharomyces cerevisiae YJM1463 (baker's yeast)	YJM1463	1294378	200x	Chromosome	909835
104	Saccharomyces cerevisiae YJM1477 (baker's yeast)	YJM1477	1294379	500x	Chromosome	896592
105	Saccharomyces cerevisiae YJM1478 (baker's yeast)	YJM1478	1294380	400x	Chromosome	914917
106	Saccharomyces cerevisiae YJM1479 (baker's yeast)	YJM1479	1294381	200x	Chromosome	909920
107	Saccharomyces cerevisiae YJM1526 (baker's yeast)	YJM1526	1294382	400x	Chromosome	911172
108	Saccharomyces cerevisiae YJM1527 (baker's yeast)	YJM1527	1294383	100x	Chromosome	915190
109	Saccharomyces cerevisiae YJM1549 (baker's yeast)	YJM1549	1294384	300x	Chromosome	896171
110	Saccharomyces cerevisiae YJM1573 (baker's yeast)	YJM1573	1294385	180x	Chromosome	905398
111	Saccharomyces cerevisiae YJM1574 (baker's yeast)	YJM1574	1294386	400x	Chromosome	899661
112	Saccharomyces cerevisiae YJM1592 (baker's yeast)	YJM1592	1294387	50x	Chromosome	900534
113	Saccharomyces cerevisiae YJM1615 (baker's yeast)	YJM1615	1294388	150x	Chromosome	901735
114	Saccharomyces boulardii (nom. inval.) (budding yeasts)	biocodex	252598	100x	Chromosome	891313
115	Saccharomyces eubayanus (budding yeasts)	FM1318	1080349	681.6x	Chromosome	896107
116	Saccharomyces boulardii (nom. inval.) (budding yeasts)	unique28	252598	100x	Chromosome	909172
117	Saccharomyces cerevisiae (baker's yeast)	GLBRCY22-3	4932	1038x	Chromosome	908319
118	Zygosaccharomyces rouxii (budding yeasts)	NBRC 110957	4956	262x	Scaffold	1286020
119	Saccharomyces cerevisiae (baker's yeast)	131	4932	122x	Contig	111157
120	Zygosaccharomyces parabaillii (budding yeasts)	ATCC 60483	1365886	218x	Complete	1283838
121	Saccharomyces cerevisiae (baker's yeast)	Kagoshima	4932	324x	Contig	822716
122	Saccharomyces cerevisiae (baker's yeast)	I-328	4932	330x	Contig	62127
123	Saccharomyces cerevisiae (baker's yeast)	I-30	4932	323x	Contig	470360
124	Saccharomyces cerevisiae (baker's yeast)	I-566	4932	219x	Contig	58253
125	Saccharomyces cerevisiae (baker's yeast)	I-329	4932	294x	Contig	511336
126	Zygosaccharomyces mellis (budding yeasts)	Ca-7	42258	447x	Scaffold	304885
127	Saccharomyces cerevisiae (baker's yeast)	Pf-1	4932	162.1x	Scaffold	102829
128	Saccharomyces cerevisiae (baker's yeast)	INSC1005	4932	392x	Chromosome	929143
129	Saccharomyces cerevisiae (baker's yeast)	INSC1006	4932	191x	Chromosome	912391
130	Saccharomyces pastorianus (lager yeast)	CBS 1483	27292	72x	Chromosome	862309
131	Saccharomyces cerevisiae (baker's yeast)	P-684	4932	207.7x	Scaffold	883925
132	Zygosaccharomyces bailii (budding yeasts)	IST302	4954	154x	Scaffold	432084

Αφού ολοκληρώθηκε η κατασκευή του πρώτου πίνακα ο οποίος εμπεριείχε όλα τα είδη και στελέχη υπό διερεύνηση, πραγματοποιήσαμε περαιτέρω φιλτράρισμα στα δεδομένα μας, έτσι ώστε να καταλήξουμε σε ένα πιο υψηλής ποιότητας σύνολο δεδομένων το οποίο θα μας οδηγούσε σε ορθότερα συμπεράσματα. Συνεπώς, αφαιρέσαμε τα δεδομένα εκείνα των οποίων το επίπεδο κατασκευής γονιδιώματος δεν ήταν Chromosome ή Complete Genome (όπως π.χ., Contig και Scaffold). Ακόμη, οι μικροοργανισμοί οι οποίοι είχαν βάθος ανάγνωσης κατά την αλληλούχιση μικρότερο του 20x, αφαιρέθηκαν από το σύνολο δεδομένων μας.

Ο τρόπος με τον οποίο θέλαμε να εμπλουτίσουμε περαιτέρω τον πίνακα που προέκυψε από το διαχωρισμό των δεδομένων που αναλύονται στον Πίνακα 4, ήταν με τη χρήση του εργαλείου MFS Search Tool [11], το οποίο θα μας έδινε την πληροφορία του αριθμού των μεταφορέων σακχάρων για κάθε στέλεχος ενδιαφέροντος. Μετά από εισαγωγή των FASTA αρχείων του κάθε στελέχους, μας δόθηκε ο ακριβής αριθμός των μεταφορέων σακχάρων για τον κάθε ένα. Από αυτή τη διαδικασία προέκυψε ο τελικός συγκεντρωτικός πίνακας ο οποίος χρησιμοποιήθηκε για την περαιτέρω ανάλυση (Πίνακας 3).

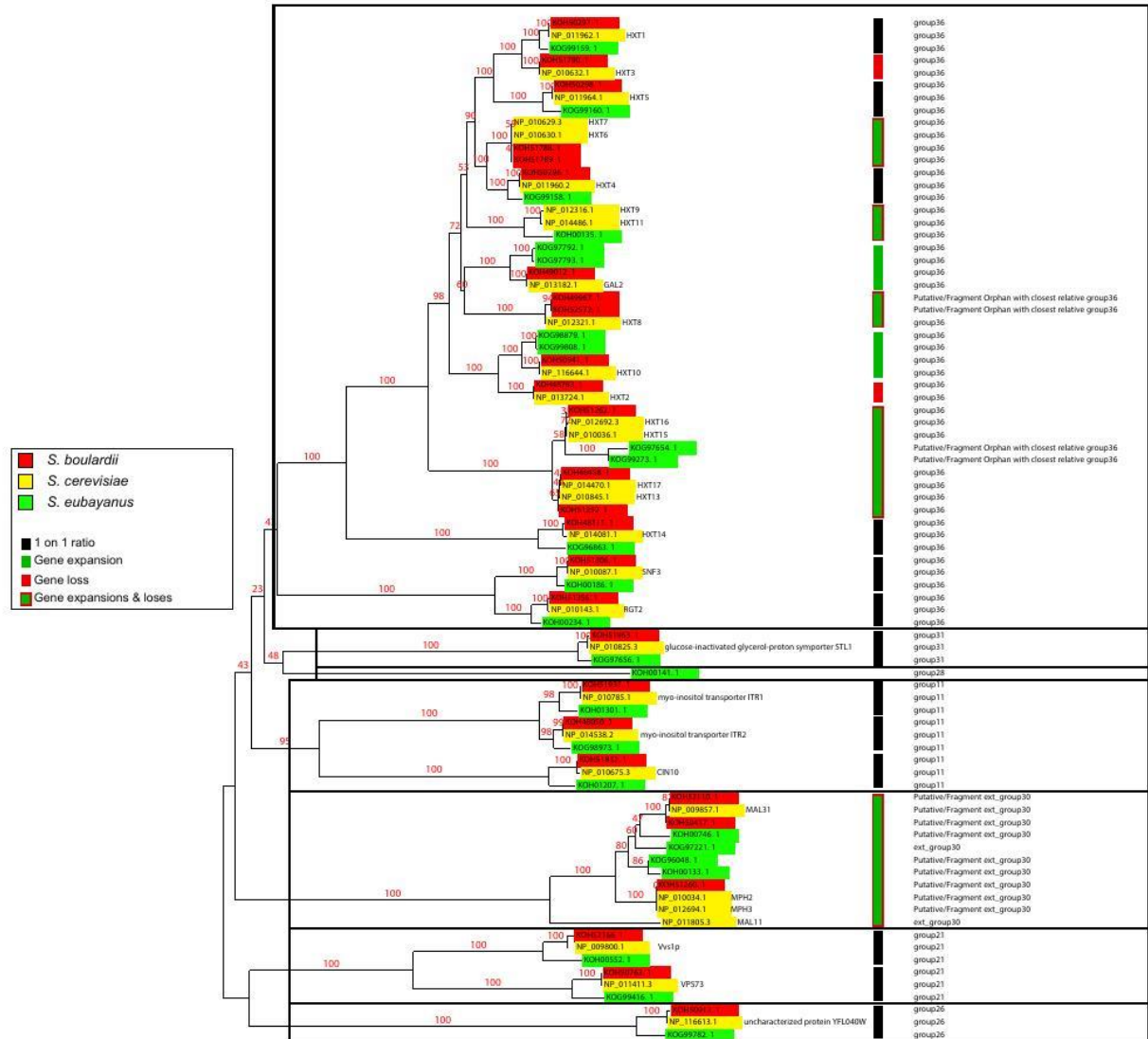
Πίνακας 3. Συγκεντρωτικός πίνακας των σακχαρομυκήτων που πληρούσαν τα κριτήρια αξιοπιστίας που αναλύθηκαν παραπάνω. Εν συγκρίσει με τον Πίνακα 4, προστέθηκε ένα ακόμη επίπεδο πληροφορίας όσον αφορά τον αριθμό των μεταφορέων γλυκόζης (STP Count) για κάθε στέλεχος ενδιαφέροντος. Με διακεκομμένο πλαίσιο περιγράφονται τα αποτελέσματα που προέκυψαν από το πρόγραμμα MFS Search Tool. Με ανοιχτό πράσινο περιγράφονται τα στελέχη τα οποία θα χρησιμοποιηθούν στα επόμενα επίπεδα της ανάλυσης μας (στοίχιση, φυλογένεση).

NO	Species	Strain	Taxid	STP Count	Genome Coverage	Assembly Level	N50
1	Saccharomyces cerevisiae YJM1399 (baker's yeast)	YJM1399	1294362	15	110x	Chromosome	903451
2	Saccharomyces cerevisiae YJM1252 (baker's yeast)	YJM1252	1294342	16	500x	Chromosome	929116
3	Saccharomyces cerevisiae YJM1342 (baker's yeast)	YJM1342	1294352	17	500x	Chromosome	912252
4	Saccharomyces cerevisiae YJM1443 (baker's yeast)	YJM1443	1294373	17	50x	Chromosome	892329
5	Saccharomyces cerevisiae YJM271 (baker's yeast)	YJM271	1294309	17	90x	Chromosome	928255
6	Saccharomyces cerevisiae YJM627 (baker's yeast)	YJM627	1294317	17	40x	Chromosome	905901
7	Saccharomyces cerevisiae YJM1083 (baker's yeast)	YJM1083	1292971	18	500x	Chromosome	923843
8	Saccharomyces cerevisiae YJM1332 (baker's yeast)	YJM1332	1294348	18	400x	Chromosome	923089
9	Saccharomyces cerevisiae YJM1338 (baker's yeast)	YJM1338	1294350	18	350x	Chromosome	895119
10	Saccharomyces cerevisiae YJM1355 (baker's yeast)	YJM1355	1294353	18	550x	Chromosome	912538
11	Saccharomyces cerevisiae YJM1389 (baker's yeast)	YJM1389	1294361	18	50x	Chromosome	897207
12	Saccharomyces cerevisiae YJM1439 (baker's yeast)	YJM1439	1294372	18	50x	Chromosome	922967
13	Saccharomyces cerevisiae YJM1447 (baker's yeast)	YJM1447	1294375	18	80x	Chromosome	909714
14	Saccharomyces cerevisiae YJM1479 (baker's yeast)	YJM1479	1294381	18	200x	Chromosome	909920
15	Saccharomyces cerevisiae YJM1527 (baker's yeast)	YJM1527	1294383	18	100x	Chromosome	915190
16	Saccharomyces cerevisiae YJM1592 (baker's yeast)	YJM1592	1294387	18	50x	Chromosome	900534
17	Saccharomyces cerevisiae YJM1248 (baker's yeast)	YJM1248	1294340	19	100x	Chromosome	924574
18	Saccharomyces cerevisiae YJM1400 (baker's yeast)	YJM1400	1294363	19	700x	Chromosome	898273
19	Saccharomyces cerevisiae YJM1444 (baker's yeast)	YJM1444	1294374	19	50x	Chromosome	897256
20	Saccharomyces cerevisiae YJM1463 (baker's yeast)	YJM1463	1294378	19	200x	Chromosome	909835
21	Saccharomyces cerevisiae YJM1477 (baker's yeast)	YJM1477	1294379	19	500x	Chromosome	896592
22	Saccharomyces cerevisiae YJM1478 (baker's yeast)	YJM1478	1294380	19	400x	Chromosome	914917
23	Saccharomyces cerevisiae YJM244 (baker's yeast)	YJM244	1294306	19	120x	Chromosome	906535
24	Saccharomyces cerevisiae YJM248 (baker's yeast)	YJM248	1294307	19	90x	Chromosome	907630
25	Saccharomyces cerevisiae YJM470 (baker's yeast)	YJM470	1294313	19	180x	Chromosome	902232
26	Saccharomyces cerevisiae YJM969 (baker's yeast)	YJM969	1294323	19	25x	Chromosome	924433
27	Saccharomyces cerevisiae YJM1078 (baker's yeast)	YJM1078	1296266	20	75x	Chromosome	907409
28	Saccharomyces cerevisiae YJM1129 (baker's yeast)	YJM1129	1293430	20	180x	Chromosome	903327
29	Saccharomyces cerevisiae YJM1208 (baker's yeast)	YJM1208	1294337	20	220x	Chromosome	903091

30	Saccharomyces cerevisiae YJM1341 (baker's yeast)	YJM1341	1294351	20	110x	Chromosome	912304
31	Saccharomyces cerevisiae YJM1434 (baker's yeast)	YJM1434	1294371	20	40x	Chromosome	912768
32	Saccharomyces cerevisiae YJM1615 (baker's yeast)	YJM1615	1294388	20	150x	Chromosome	901735
33	Saccharomyces cerevisiae YJM270 (baker's yeast)	YJM270	1294308	20	110x	Chromosome	913025
34	Saccharomyces cerevisiae YJM450 (baker's yeast)	YJM450	1294310	20	200x	Chromosome	914011
35	Saccharomyces cerevisiae YJM456 (baker's yeast)	YJM456	1294312	20	190x	Chromosome	897427
36	Saccharomyces cerevisiae YJM541 (baker's yeast)	YJM541	1294314	20	100x	Chromosome	909353
37	Saccharomyces cerevisiae YJM993 (baker's yeast)	YJM993	1294331	20	180x	Chromosome	913508
38	Saccharomyces cerevisiae YJM996 (baker's yeast)	YJM996	1294332	20	180x	Chromosome	898411
39	Saccharomyces cerevisiae YJM1202 (baker's yeast)	YJM1202	1294336	21	650x	Chromosome	912629
40	Saccharomyces cerevisiae YJM1387 (baker's yeast)	YJM1387	1294359	21	200x	Chromosome	904346
41	Saccharomyces cerevisiae YJM1415 (baker's yeast)	YJM1415	1294366	21	600x	Chromosome	744674
42	Saccharomyces cerevisiae YJM1418 (baker's yeast)	YJM1418	1294368	21	600x	Chromosome	907734
43	Saccharomyces cerevisiae YJM1433 (baker's yeast)	YJM1433	1294370	21	75x	Chromosome	910803
44	Saccharomyces cerevisiae YJM1526 (baker's yeast)	YJM1526	1294382	21	400x	Chromosome	911172
45	Saccharomyces cerevisiae YJM193 (baker's yeast)	YJM193	1294304	21	60x	Chromosome	934902
46	Saccharomyces cerevisiae YJM195 (baker's yeast)	YJM195	1294305	21	100x	Chromosome	916102
47	Saccharomyces cerevisiae YJM451 (baker's yeast)	YJM451	502869	21	75x	Chromosome	895934
48	Saccharomyces cerevisiae YJM975 (baker's yeast)	YJM975	1294325	21	30x	Chromosome	904046
49	Saccharomyces cerevisiae YJM1133 (baker's yeast)	YJM1133	1294333	22	60x	Chromosome	901770
50	Saccharomyces cerevisiae YJM1190 (baker's yeast)	YJM1190	1294334	22	30x	Chromosome	902060
51	Saccharomyces cerevisiae YJM1199 (baker's yeast)	YJM1199	1294335	22	60x	Chromosome	900262
52	Saccharomyces cerevisiae YJM1244 (baker's yeast)	YJM1244	1294339	22	50x	Chromosome	911385
53	Saccharomyces cerevisiae YJM1304 (baker's yeast)	YJM1304	1294344	22	750x	Chromosome	913052
54	Saccharomyces cerevisiae YJM1385 (baker's yeast)	YJM1385	1294357	22	350x	Chromosome	903329
55	Saccharomyces cerevisiae YJM1401 (baker's yeast)	YJM1401	1294364	22	500x	Chromosome	889931
56	Saccharomyces cerevisiae YJM1402 (baker's yeast)	YJM1402	1294365	22	200x	Chromosome	902709
57	Saccharomyces cerevisiae YJM1450 (baker's yeast)	YJM1450	1294376	22	120x	Chromosome	908914
58	Saccharomyces cerevisiae YJM1460 (baker's yeast)	YJM1460	1294377	22	400x	Chromosome	897146
59	Saccharomyces cerevisiae YJM1549 (baker's yeast)	YJM1549	1294384	22	300x	Chromosome	896171
60	Saccharomyces cerevisiae YJM320 (baker's yeast)	YJM320	947042	22	100x	Chromosome	905335
61	Saccharomyces cerevisiae YJM326 (baker's yeast)	YJM326	468558	22	400x	Chromosome	907133
62	Saccharomyces cerevisiae YJM428 (baker's yeast)	YJM428	947044	22	90x	Chromosome	910751
63	Saccharomyces cerevisiae YJM453 (baker's yeast)	YJM453	1294311	22	110x	Chromosome	911881
64	Saccharomyces cerevisiae YJM681 (baker's yeast)	YJM681	1294318	22	40x	Chromosome	906981
65	Saccharomyces cerevisiae YJM683 (baker's yeast)	YJM683	1294320	22	130x	Chromosome	900521
66	Saccharomyces cerevisiae YJM689 (baker's yeast)	YJM689	1294321	22	40x	Chromosome	895385
67	Saccharomyces cerevisiae YJM978 (baker's yeast)	YJM978	1294326	22	90x	Chromosome	902126
68	Saccharomyces cerevisiae YJM981 (baker's yeast)	YJM981	1294327	22	40x	Chromosome	916198
69	Saccharomyces cerevisiae YJM984 (baker's yeast)	YJM984	1294328	22	30x	Chromosome	910820
70	Saccharomyces cerevisiae (baker's yeast)	INSC1006	4932	23	191x	Chromosome	912391
71	Saccharomyces cerevisiae YJM1250 (baker's yeast)	YJM1250	1294341	23	75x	Chromosome	1026598
72	Saccharomyces cerevisiae YJM1311 (baker's yeast)	YJM1311	1294346	23	400x	Chromosome	903706
73	Saccharomyces cerevisiae YJM1383 (baker's yeast)	YJM1383	1294356	23	650x	Chromosome	931094
74	Saccharomyces cerevisiae YJM1419 (baker's yeast)	YJM1419	1294369	23	80x	Chromosome	891878
75	Saccharomyces cerevisiae YJM1573 (baker's yeast)	YJM1573	1294385	23	180x	Chromosome	905398
76	Saccharomyces cerevisiae YJM1574 (baker's yeast)	YJM1574	1294386	23	400x	Chromosome	899661
77	Saccharomyces cerevisiae YJM189 (baker's yeast)	YJM189	1294303	23	400x	Chromosome	896353
78	Saccharomyces cerevisiae YJM555 (baker's yeast)	YJM555	1294316	23	190x	Chromosome	909661
79	Saccharomyces cerevisiae YJM693 (baker's yeast)	YJM693	1294322	23	80x	Chromosome	907430
80	Saccharomyces cerevisiae YJM1273 (baker's yeast)	YJM1273	1294343	24	400x	Chromosome	910208
81	Saccharomyces cerevisiae YJM1326 (baker's yeast)	YJM1326	1294347	24	800x	Chromosome	913940
82	Saccharomyces cerevisiae YJM1336 (baker's yeast)	YJM1336	1294349	24	700x	Chromosome	901186
83	Saccharomyces cerevisiae YJM1356 (baker's yeast)	YJM1356	1294354	24	400x	Chromosome	910189

84	Saccharomyces cerevisiae YJM682 (baker's yeast)	YJM682	1294319	24	100x	Chromosome	896702
85	Saccharomyces cerevisiae YJM972 (baker's yeast)	YJM972	1294324	24	100x	Chromosome	902781
86	Saccharomyces eubayanus (budding yeasts)	FM1318	1080349	25	681.6x	Chromosome	896107
87	Saccharomyces cerevisiae YJM1386 (baker's yeast)	YJM1386	1294358	25	120x	Chromosome	902760
88	Saccharomyces cerevisiae YJM987 (baker's yeast)	YJM987	1294329	25	150x	Chromosome	912332
89	Saccharomyces boulardii (nom. inval.)	unique28	252598	26	100x	Chromosome	909172
90	Saccharomyces boulardii (nom. inval.)	biocodex	252598	27	100x	Chromosome	891313
91	Saccharomyces cerevisiae (baker's yeast)	INSC1005	4932	29	392x	Chromosome	929143
92	Saccharomyces cerevisiae (baker's yeast)	GLBRCY22	4932	30	1038x	Chromosome	908319
93	Saccharomyces cerevisiae S288C (baker's yeast)	S288C	559292	30	REF (SGD)	Complete	924431

Με βάση τον παραπάνω πίνακα κατασκευάστηκαν τα παρακάτω φυλογενετικά δένδρα με χρήση των προγραμμάτων TreeDyN (**Εικόνα 7**) και iTOL (**Εικόνα 8, Εικόνα 9, Εικόνα 10**). Ως φυλογενετικό δένδρο στην, παρούσα εργασία, ορίζεται ένα δενδρόγραμμα στο οποίο απεικονίζονται οι εξελικτικές σχέσεις μεταξύ στελεχών, όπου στο μήκος του κάθε κλάδου αναδεικνύονται οι εξελικτικές διαφορές/αποστάσεις μεταξύ πρωτεϊνών όλων των συγκρινόμενων στελεχών/ειδών.



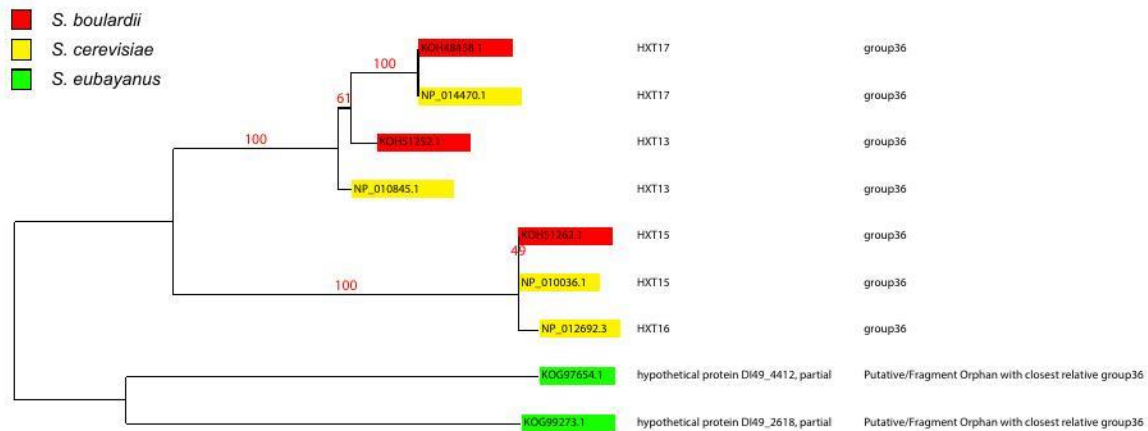
Εικόνα 7. Τελικό φυλογενετικό δέντρο. Στο αριστερό πλαίσιο με κόκκινο χρώμα αποδίδονται οι πρωτεΐνες του στελέχους *S. boulardii*, με κίτρινο χρώμα οι πρωτεΐνες του στελέχους *S. cerevisiae* και με πράσινο χρώμα οι πρωτεΐνες του στελέχους *S. eubayanus*.

Για τη δημιουργία αυτού του δένδρου (Εικόνα 7) χρησιμοποιήσαμε όλα τα FASTA αρχεία των συγκεκριμένων 3 μικροοργανισμών (*S. boulardii*, *S. cerevisiae*, *S. eubayanus*) από τους μεταφορείς γλυκόζης στη διάθεσή μας έτσι ώστε να οπτικοποιήσουμε τα σημεία στα οποία έχουν γίνει διπλασιασμοί, γονιδιακές απώλειες, ή ο συνδυασμός και των δύο καταστάσεων. Ο σκοπός του παραπάνω πίνακα είναι ο χαρακτηρισμός των μεταφορέων γλυκόζης ως προς το εάν αυτές οι πρωτεΐνες είναι μεταξύ τους ορθόλογες, παράλογες κ.ο.κ.

Στο φυλογενετικό δένδρο (Εικόνα 7) απεικονίζονται και οπτικοποιούνται αναλυτικά οι γενετικές αποστάσεις ανάμεσα στα ομόλογα μεταφορέων σακχάρων στους 3 μικροοργανισμούς ενδιαφέροντος (*S. boulardii*, *S. cerevisiae*, *S. eubayanus*). Έχει δημιουργηθεί με μέθοδο Poisson με 1000 bootstrap στο TreeDyn. Οι εξελικτικές τροποποιήσεις είναι οι εξής: Α) 1 on 1 ratio με μαύρο χρώμα, το οποίο σημαίνει ότι ανάμεσα στις πρωτεΐνες του συγκεκριμένου κλάδου δεν έχει συμβεί κανένα γεγονός που να υποδηλώνει διπλασιασμό, Β) γονιδιακή επέκταση με

πράσινο χρώμα, Γ) γονιδιακή απώλεια με κόκκινο χρώμα, και Δ) ο συνδυασμός γονιδιακής επέκτασης και απώλειας ταυτόχρονα με πράσινο χρώμα σε κόκκινο πλαίσιο.

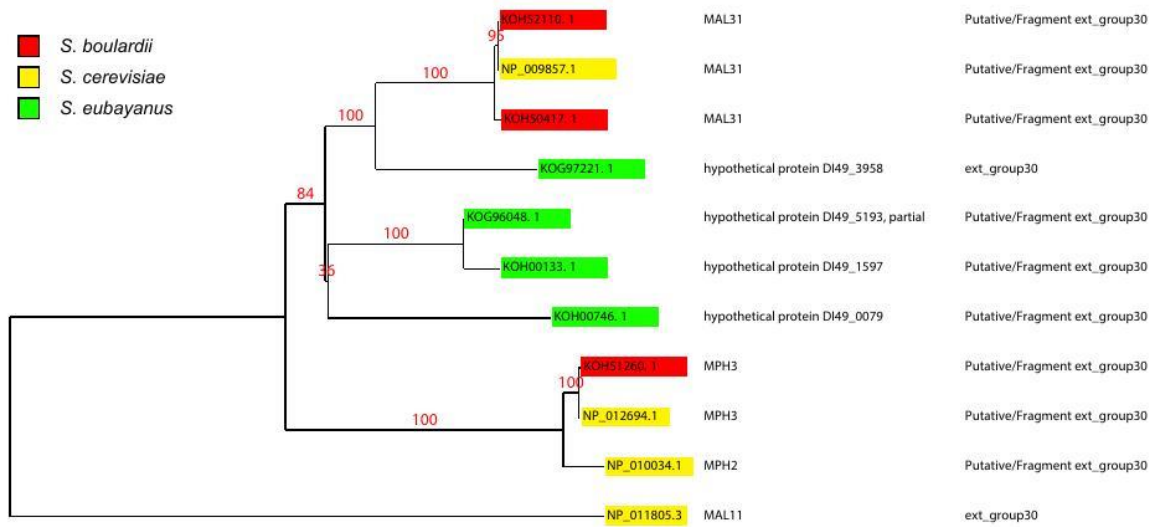
Από το συγκεκριμένο φυλογενετικό δένδρο επιλέξαμε τους κλάδους οι οποίοι εμφανίζουν εκτεταμένους διπλασιασμούς και απώλειες (**Εικόνα 8**, **Εικόνα 9**) για τους οποίους δημιουργήθηκαν επιπλέον φυλογενετικά δέντρα, ώστε να αποδίδεται καλύτερα η οπτικοποίηση των αποτελεσμάτων μας και να αποδοθεί καλύτερα η σχέση μεταξύ των γονιδίων μεταφορέων σακχάρων ανάμεσα στα είδη. Τα προσεχή φυλογενετικά δέντρα δημιουργήθηκαν στο πρόγραμμα iTOL με 1000 bootstrap.



Εικόνα 8. Πρώτο φυλογενετικό δέντρο για την βεβαίωση της ορθότητας των τοπολογιών. Στο αριστερό πλαίσιο με κόκκινο χρώμα αποδίδονται οι πρωτεΐνες του είδους *S. boulardii*, με κίτρινο χρώμα οι πρωτεΐνες του είδους *S. cerevisiae* και με πράσινο χρώμα οι πρωτεΐνες του είδους *S. eubayanus*.

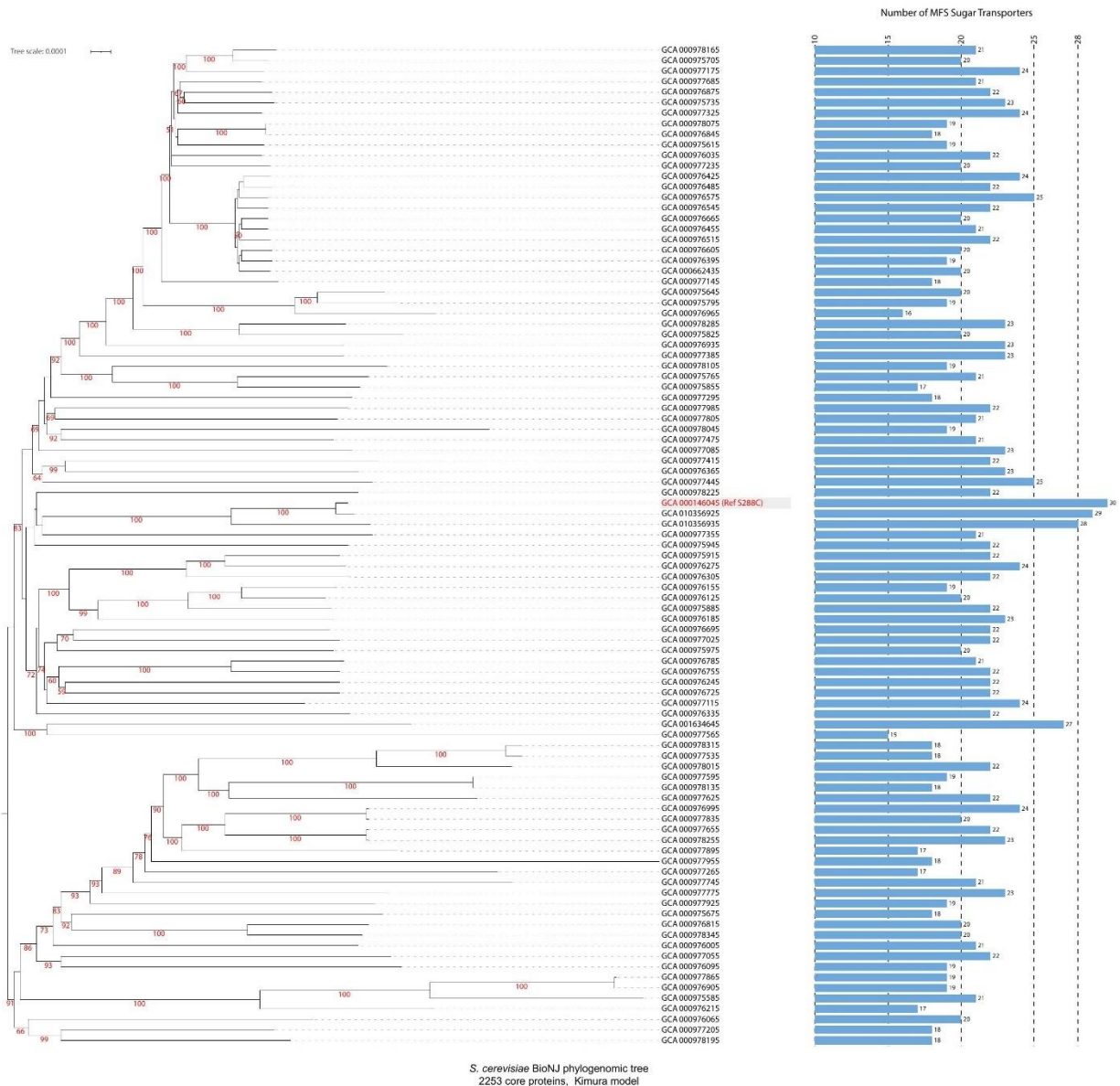
Στο παραπάνω φυλογενετικό δέντρο (**Εικόνα 8**) παρατηρούνται εξελικτικές τροποποιήσεις για τα γονίδια της οικογένειας *HXT* και χαρακτηρίζεται η σχέση μεταξύ αυτών. Το γονίδιο *HXT17* είναι ένας μεταφορέας εξόζης ο οποίος υποστηρίζει την ανάπτυξη και την πρόσληψη μαννιτόλης (mannitol) και σορβιτόλης (sorbitol) και καταστέλλεται από υψηλά επίπεδα γλυκόζης. Το *HXT17* έχει ένα παράλογο, το *HXT13*, που προέκυψε από έναν τμηματικό διπλασιασμό [32-35], όπως είδαμε και στο παραπάνω δέντρο (**Εικόνα 8**).

Το γονίδιο *HXT15* είναι ένας υποτιθέμενος διαμεμβρανικός μεταφορέας πολυόλης ο οποίος υποστηρίζει την ανάπτυξη και την πρόσληψη μαννιτόλης, σορβιτόλης και ξυλιτόλης. Η έκφραση του προκαλείται από χαμηλά επίπεδα γλυκόζης και καταστέλλεται από υψηλά επίπεδα γλυκόζης [32,35]. Το γονίδιο *HXT16* είναι και αυτός ένας υποτιθέμενος διαμεμβρανικός μεταφορέας πολυόλης, ο οποίος υποστηρίζει την ανάπτυξη και την πρόσληψη της σορβιτόλης και έχει ομοιότητα με μεταφορείς εξόζης. Η έκφραση του καταστέλλεται από τυχόν υψηλά επίπεδα γλυκόζης [32,35].



Εικόνα 9. Δεύτερο φυλογενετικό δέντρο για την βεβαίωση της ορθότητας των τοπολογιών. Στο αριστερό πλαίσιο με κόκκινο χρώμα αποδίδονται οι πρωτεΐνες του *S. boulardii*, με κίτρινο χρώμα οι πρωτεΐνες του *S. cerevisiae* και με πράσινο χρώμα οι πρωτεΐνες του *S. eubayanus*.

Με την ίδια λογική εργαστήκαμε και για γονίδια της οικογένειας *MAL*. Το γονίδιο *MAL31* μεταφράζεται σε μια περμεάση μαλτόζης και είναι μέλος της υπεροικογένειας των 12 διαμεμβρανικών μεταφορέων σακχάρων. Είναι επίσης πλήρως λειτουργικό στο στέλεχος αναφοράς S288C [36].



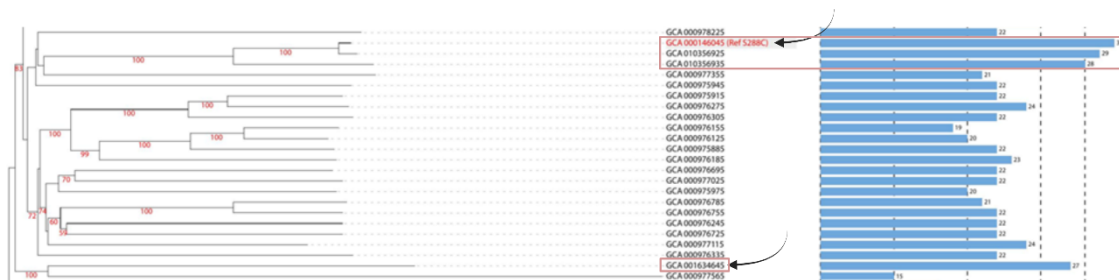
S. cerevisiae BioNJ phylogenomic tree
2253 core proteins, Kimura model

Εικόνα 10. Φυλογενωμικό δένδρο μεταξύ διαφορετικών στελεχών του σακχαρομύκητα *S. cerevisiae*. Το δένδρο έχει δημιουργηθεί βάση του μοντέλου Kimura με τον βελτισμένο αλγόριθμο BioNJ. Στα δεξιά της εικόνας υπάρχει ο αριθμός των MFS Sugar Transporters ανά στέλεχος *S. cerevisiae*. Όπως διαφαίνεται, υπάρχει μεγάλη μεταβολή στον αριθμό μεταφορέων μεταξύ των διαφόρων στελεχών του είδους.

Στο παραπάνω φυλογενωμικό δέντρο για τα διάφορα στελέχη του *S. cerevisiae* (**Εικόνα 10**) μπορούμε να δούμε ότι υπάρχει μεγάλη μεταβολή στον αριθμό μεταφορέων μεταξύ των διαφόρων στελεχών του είδους. Στο φυσικό ενδιαίτημα του ο *S. cerevisiae* (δηλαδή στην εξωτερική επιφάνεια καρπών) έχει εξελικτικό πλεονέκτημα έναντι άλλων ειδών, διότι εγκαθίσταται ταχύτατα στις φυτικές επιφάνειες και λόγω της δράσης του παρέχει προστασία στο χώρο εγκατάστασης του ενώ απορροφάει τα θρεπτικά συστατικά του περιβάλλοντος που τον

φιλοξενεί. Τεχνητό ενδιαίτημα του *cerevisiae* είναι και τα βαρέλια στα οποία χρησιμοποιείται για τη αλκοολική ζύμωση (μετατροπή υδατανθράκων σε αλκοόλη και CO₂) [37].

Τρία στελέχη (GCA000146045-REF288C, GCA010356925, GCA010356935) έχουν αρκετά περισσότερους μεταφορείς γλυκόζης σε σχέση με τα υπόλοιπα στελέχη. Ο κλάδος αυτός μπορεί να εμφανίζει Lineage Specific Expansion (LSE). Λόγω όγκου δεδομένων δεν μπορούν να εξαχθούν ασφαλή συμπεράσματα. Παρόλα αυτά, ένα αξιοσημείωτο εύρημα ήταν το εξής: το στέλεχος GCA001634645, έχει και αυτό μεγαλύτερο αριθμό μεταφορέων γλυκόζης και βρίσκεται μακριά από τη συστάδα των τριών στελεχών GCA000146045-REF288C, GCA010356925 και GCA010356935 που έχουν περισσότερους από τον μέσο όρο (~21 Glucose Transporters/*cerevisiae* strain).



Εικόνα 11. Μεγέθυνση του φυλογενωμικού δέντρου (Εικόνα 10) με κατάδειξη των σημείων ενδιαφέροντος.

Μετά από περαιτέρω ανάλυση του GCA001634645 (ή αλλιώς BioProject Accession Number: PRJNA279877) στην δημοσίευση με τίτλο “*Genome Sequence and Analysis of a Stress-Tolerant, Wild-Derived Strain of Saccharomyces cerevisiae Used in Biofuels Research*” που αναρτήθηκε η ακολουθία του συγκεκριμένου strain, οι McIlwain et al., περιγράφουν τη σχεδόν πλήρη αλληλουχία του γονιδιώματος του GLBRCY22-3 (Y22-3), ενός στελέχους *S. cerevisiae* που προέρχεται από το ανθεκτικό στο στρες wild-type στέλεχος NRRL YB-210 και στη συνέχεια τροποποιήθηκε για να μεταβολίζει ξυλόζες (Xylose - C₅H₁₀O₅) [38]. Οι ξυλόζες είναι είδος πεντόζων και χρησιμοποιούνται συχνά σαν γλυκαντικές ουσίες στο φαγητό και η χρήση ξυλόζης βρίσκεται σε μεγάλη άνοδο τα τελευταία χρόνια με την βιομηχανική ανάπτυξη της χρήση βιομάζας. Κατά συνέπεια, η κρίση της ενέργειας και των τροφίμων μπορεί να λυθεί, εάν η ξυλόζη χρησιμοποιηθεί ως πηγή άνθρακα στη ζύμωση [39]. Ακόμη, επί παραδείγματι, στην παραγωγή βιοαιθανόλης, ο πιο ευρέως χρησιμοποιούμενος οργανισμός που παράγει αιθανόλη είναι ο ζυμομύκητας (*S. cerevisiae*). Η μαγιά δεν έχει την ικανότητα να μεταβολίζει γρήγορα την ξυλόζη. Για να λύσουν αυτό το πρόβλημα, πολλοί ερευνητές προσπαθούν να μεταμορφώσουν την υπάρχουσα διαδικασία μικροβιακού μεταβολισμού προκειμένου να αποκτήσουν βέλτιστα στέλεχη με γενετική μηχανική [40].

Ενώ όπως αναφέραμε παραπάνω, δεν μπορούμε με ασφάλεια να εξαγάγουμε σαφές συμπέρασμα ότι ο κλάδος των τριών στελεχών (GCA000146045-REF288C, GCA010356925, GCA010356935) είναι Lineage Specific Expansion, αλλά μπορούμε με αρκετή ασφάλεια να αποκλείσουμε ως απόκλιση τον GCA001634645, διότι είναι ένα γενετικά τροποποιημένο στέλεχος το οποίο χρησιμοποιείται στην βιομηχανία. Σαν μελλοντικός στόχος, για την εξαγωγή ασφαλέστερων συμπερασμάτων, θα έπρεπε να γίνει φυλογένεση σε ένα είδος συγγενές του *cerevisiae* (όπως π.χ. ο *boulardii*) ώστε να πραγματοποιηθεί σύγκριση των μεταφορέων γλυκόζης μεταξύ των δύο διαφορετικών ειδών [ή στην συγκεκριμένη περίπτωση των *boulardii* και *cerevisiae*, παραλλαγών (variants)]. Το *S. boulardii* και το *S. cerevisiae* είναι δύο στενά

συνδεδεμένα στελέχη που χρησιμοποιούνται είτε ως προβιοτικά είτε για την παρασκευή φαγητού και κρασιού. Τα δύο στελέχη παρόλο που είναι σχεδόν πανομοιότυπα σε μοριακό επίπεδο, το *S. boulardii* παρουσιάζει μεγαλύτερη φυσιολογική αντίσταση στη θερμότητα και τους όξινους στρεσογόνους παράγοντες [41]. Ακόμη, για την εξαγωγή ακόμη καλύτερων αποτελεσμάτων θα άξιζε στο μέλλον να γίνει φυλογενετική ανάλυση μεταξύ των στελεχών του φυλογενωμικού δέντρου για να δούμε πόσες από τις ενδείξεις που έχουμε είναι γονιδιακοί διπλασιασμοί.

Αξίζει να σημειωθεί επίσης ότι τα αποτελέσματα μας σχετικά με το γονίδιο *MAL11* αποτελούν αντικείμενο πρόσφατων ανακαλύψεων και νέων προβληματισμών. Στην πρόσφατη δημοσίευση τους, οι Sophie Claire de Valk και Robert Mans αποκρυσταλλώνουν μια σημαντική πρόκληση στην έρευνα των πρωτεϊνών μεταφορέων σε σχέση με την προσαρμοστικότητα των ειδών: Το αν μπορούμε να κατανοήσουμε πώς τα μεμονωμένα αμινοξικά υπολείμματα μπορούν να συμβάλλουν στη δομή και τη βιολογική λειτουργία των μεταφορέων (και κατά συνέπεια και των μεταφορέων γλυκόζης).

Οι υποκαταστάσεις (substitutions) αμινοξέων που οδηγούν σε επιλεκτικό πλεονέκτημα σε προσαρμοστικά πειράματα καθοδηγούμενης εξέλιξης μπορούν να παρέχουν πολύτιμες υποδείξεις για το ρόλο τους στις πρωτεϊνες μεταφοράς. Ειδικότερα σε αυτή τη μελέτη, εφάρμοσαν μια στρατηγική εξελικτικής μηχανικής για να αλλάξουν την ειδικότητα του υποστρώματος του συζευγμένου με πρωτόνιο μεταφορέα δισακχαρίτη Mal11 στο *S. cerevisiae*, ο οποίος έχει συγγένεια με σακχαρόζη, μαλτόζη και γλυκόζη. Η εισαγωγή του *MAL11* σε ένα στέλεχος χωρίς όλους τους άλλους μεταφορείς σακχάρων και δισακχαριδικές υδρολάσες αποκατέστησε την ανάπτυξη στη γλυκόζη, αλλά κατέστησε το στέλεχος εξαιρετικά ευαίσθητο στην παρουσία σακχαρόζης ή μαλτόζης [42].

Συνομειύσεις

BLAST	Basic Local Alignment Search Tool
GLUTs	Glucose Transporters
HGNC	HUGO Gene Nomenclature Committee
HMM	Hidden Markov Model
LSE	Lineage Specific Expansion
MFS	Major Facilitator Superfamily
MUSCLE	MUltiple Sequence Comparison by Log- Expectation
Na ⁺ /K ⁺ -ATPase	Αντλία Νατρίου-Καλίου
NCBI	The National Center for Biotechnology Information
NJ	Neighbor-Join
PDB	Protein Data Bank
SGLTs	Sodium-Driven Glucose Symporters
TCDB	The Transporter Classification Database

Βιβλιογραφικές Παραπομπές

1. E. V. Wong. Membrane Permeability. Available online: <https://bio.libretexts.org/@go/page/16107> (accessed on 20/01/2022).
2. Cooper, G.M.; Hausman, R.E. *The Cell: A Molecular Approach*; ASM Press: 2007.
3. <https://www.physiologyweb.com/>. Summary of Membrane Transport Processes. Available online: https://www.physiologyweb.com/lecture_notes/membrane_transport/membrane_transport_processes_summary.html (accessed on 20/01/2022).
4. Chaffey, N. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. Molecular biology of the cell. 4th edn. *Ann Bot* **2003**, *91*, 401-401, doi:10.1093/aob/mcg023.
5. Deng, D.; Xu, C.; Sun, P.; Wu, J.; Yan, C.; Hu, M.; Yan, N. Crystal structure of the human glucose transporter GLUT1. *Nature* **2014**, *510*, 121-125, doi:10.1038/nature13306.
6. Burley, S.K.; Bhikadiya, C.; Bi, C.; Bittrich, S.; Chen, L.; Crichlow, G.V.; Christie, C.H.; Dalenberg, K.; Di Costanzo, L.; Duarte, J.M.; et al. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Research* **2021**, *49*, D437-D451, doi:10.1093/nar/gkaa1038.
7. Hudson, C.S. Emil Fischer's discovery of the configuration of glucose. A semicentennial retrospect. *Journal of Chemical Education* **1941**, *18*, 353, doi:10.1021/ed018p353.
8. Deng, D.; Yan, N. GLUT, SGLT, and SWEET: Structural and mechanistic investigations of the glucose transporters. *Protein Sci* **2016**, *25*, 546-558, doi:10.1002/pro.2858.
9. Tutton, A.E. Synthesis of Glucose. *Nature* **1887**, *37*, 7-8, doi:10.1038/037007b0.
10. Chen, L.-Q.; Cheung, L.S.; Feng, L.; Tanner, W.; Frommer, W.B. Transport of Sugars. *Annual Review of Biochemistry* **2015**, *84*, 865-894, doi:10.1146/annurev-biochem-060614-033904.
11. Νικολαΐδης, Μ. Ανάπτυξη βιοπληροφορικού εργαλείου για τον εντοπισμό και την εξελικτική κατηγοριοποίηση των ευκαρυωτικών μεταφορέων σακχάρων της υπεροικογένειας MFS. Πανεπιστήμιο Θεσσαλίας, Λαμία, 2019.
12. Διακογεωργίου, Α. Βιοπληροφορική και εξελικτική ανάλυση των μεταφορέων σακχάρων της major facilitator superfamily. Πανεπιστήμιο Θεσσαλίας, Λαμία, 2018.
13. Thorens, B.; Mueckler, M. Glucose transporters in the 21st Century. *American Journal of Physiology-Endocrinology and Metabolism* **2009**, *298*, E141-E145, doi:10.1152/ajpendo.00712.2009.
14. Wang, S.C.; Davejan, P.; Hendargo, K.J.; Javadi-Razaz, I.; Chou, A.; Yee, D.C.; Ghazi, F.; Lam, K.J.K.; Conn, A.M.; Madrigal, A.; et al. Expansion of the Major Facilitator Superfamily (MFS) to include novel transporters as well as transmembrane-acting enzymes. *Biochim Biophys Acta Biomembr* **2020**, *1862*, 183277, doi:10.1016/j.bbamem.2020.183277.
15. Walmsley, A.R.; Barrett, M.P.; Bringaud, F.; Gould, G.W. Sugar transporters from bacteria, parasites and mammals: structure-activity relationships. *Trends in biochemical sciences* **1998**, *23*, 476-481, doi:10.1016/s0968-0004(98)01326-7.
16. Pao, S.S.; Paulsen, I.T.; Saier, M.H., Jr. Major facilitator superfamily. *Microbiol Mol Biol Rev* **1998**, *62*, 1-34, doi:10.1128/MMBR.62.1.1-34.1998.
17. Saier, M.H.; Reddy, V.S.; Moreno-Hagelsieb, G.; Hendargo, K.J.; Zhang, Y.; Iddamsetty, V.; Lam, K.J.K.; Tian, N.; Russum, S.; Wang, J.; et al. The Transporter Classification Database (TCDB): 2021 update. *Nucleic Acids Res* **2021**, *49*, D461-d467, doi:10.1093/nar/gkaa1004.
18. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Research* **2000**, *28*, 235-242, doi:10.1093/nar/28.1.235.

19. Shi, Y. Common Folds and Transport Mechanisms of Secondary Active Transporters. *Annual Review of Biophysics* **2013**, *42*, 51-72, doi:10.1146/annurev-biophys-083012-130429.
20. Cheeseman, C.; Long, W. Structure of, and functional insight into the GLUT family of membrane transporters. *Cell Health and Cytoskeleton* **2015**, *7*, 167, doi:10.2147/CHC.S60484.
21. Wilson-O'Brien, A.L.; Patron, N.; Rogers, S. Evolutionary ancestry and novel functions of the mammalian glucose transporter (GLUT) family. *BMC Evol Biol* **2010**, *10*, 152-152, doi:10.1186/1471-2148-10-152.
22. Cunningham, F.; Allen, J.E.; Allen, J.; Alvarez-Jarreta, J.; Amode, M R.; Armean, Irina M.; Austine-Orimoloye, O.; Azov, Andrey G.; Barnes, I.; Bennett, R.; et al. Ensembl 2022. *Nucleic Acids Research* **2022**, *50*, D988-D995, doi:10.1093/nar/gkab1049.
23. Batch Entrez. In *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2006; pp. 131-131.
24. Franzese, M.; Iuliano, A. Hidden Markov Models. In *Encyclopedia of Bioinformatics and Computational Biology*, Ranganathan, S., Gribskov, M., Nakai, K., Schönbach, C., Eds.; Academic Press: Oxford, 2019; pp. 753-762.
25. Nikolaidis, M.; Mossialos, D.; Oliver, S.G.; Amoutzias, G.D. Comparative Analysis of the Core Proteomes among the Pseudomonas Major Evolutionary Groups Reveals Species-Specific Adaptations for Pseudomonas aeruginosa and Pseudomonas chlororaphis. *Diversity* **2020**, *12*, 289.
26. Gouy, M.; Guindon, S.; Gascuel, O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **2010**, *27*, 221-224, doi:10.1093/molbev/msp259.
27. Gascuel, O. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol* **1997**, *14*, 685-695, doi:10.1093/oxfordjournals.molbev.a025808.
28. Saitou, N.; Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **1987**, *4*, 406-425, doi:10.1093/oxfordjournals.molbev.a040454.
29. Edgar, R.C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **2004**, *5*, 113, doi:10.1186/1471-2105-5-113.
30. Chevenet, F.; Brun, C.; Bañuls, A.L.; Jacq, B.; Christen, R. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* **2006**, *7*, 439, doi:10.1186/1471-2105-7-439.
31. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* **2021**, *49*, W293-W296, doi:10.1093/nar/gkab301.
32. Ozcan, S.; Johnston, M. Function and regulation of yeast hexose transporters. *Microbiol Mol Biol Rev* **1999**, *63*, 554-569, doi:10.1128/mmbr.63.3.554-569.1999.
33. Greatrix, B.W.; van Vuuren, H.J. Expression of the HXT13, HXT15 and HXT17 genes in Saccharomyces cerevisiae and stabilization of the HXT1 gene transcript by sugar-induced osmotic stress. *Curr Genet* **2006**, *49*, 205-217, doi:10.1007/s00294-005-0046-x.
34. Katju, V.; Farslow, J.C.; Bergthorsson, U. Variation in gene duplicates with low synonymous divergence in Saccharomyces cerevisiae relative to Caenorhabditis elegans. *Genome Biol* **2009**, *10*, R75, doi:10.1186/gb-2009-10-7-r75.
35. Jordan, P.; Choe, J.Y.; Boles, E.; Oreb, M. Hxt13, Hxt15, Hxt16 and Hxt17 from Saccharomyces cerevisiae represent a novel type of polyol transporters. *Sci Rep* **2016**, *6*, 23502, doi:10.1038/srep23502.
36. Chow, T.H.; Sollitti, P.; Marmur, J. Structure of the multigene family of MAL loci in Saccharomyces. *Mol Gen Genet* **1989**, *217*, 60-69, doi:10.1007/bf00330943.
37. Η Ομάδα γεωπόνων της Farmacon - Farmacon Team. Οι ζυμωμύκητες στην σύγχρονη Γεωργία. Available online: <https://blog.farmacon.gr/katigories/texniki->

[arthrografia/fytoprostasia/item/2703-oi-zymomykites-stin-sygxroni-georgia](#) (accessed on 01/07/2022).

38. McIlwain, S.J.; Peris, D.; Sardi, M.; Moskvina, O.V.; Zhan, F.; Myers, K.S.; Riley, N.M.; Buzzell, A.; Parreiras, L.S.; Ong, I.M.; et al. Genome Sequence and Analysis of a Stress-Tolerant, Wild-Derived Strain of *Saccharomyces cerevisiae* Used in Biofuels Research. *G3 (Bethesda)* **2016**, *6*, 1757-1766, doi:10.1534/g3.116.029389.
39. Chen, H.; Wang, L. Chapter 6 - Sugar Strategies for Biomass Biochemical Conversion. In *Technologies for Biochemical Conversion of Biomass*, Chen, H., Wang, L., Eds.; Academic Press: Oxford, 2017; pp. 137-164.
40. Hamacher, T.; Becker, J.; Gárdonyi, M.; Hahn-Hägerdal, B.; Boles, E. Characterization of the xylose-transporting properties of yeast hexose transporters and their influence on xylose utilization. *Microbiology (Reading)* **2002**, *148*, 2783-2788, doi:10.1099/00221287-148-9-2783.
41. Fietto, J.L.; Araújo, R.S.; Valadão, F.N.; Fietto, L.G.; Brandão, R.L.; Neves, M.J.; Gomes, F.C.; Nicoli, J.R.; Castro, I.M. Molecular and physiological comparisons between *Saccharomyces cerevisiae* and *Saccharomyces boulardii*. *Can J Microbiol* **2004**, *50*, 615-621, doi:10.1139/w04-050.
42. de Valk, S.C.; Mans, R. Novel Evolutionary Engineering Approach to Alter Substrate Specificity of Disaccharide Transporter Mal11 in *Saccharomyces cerevisiae*. *J Fungi (Basel)* **2022**, *8*, doi:10.3390/jof8040358.