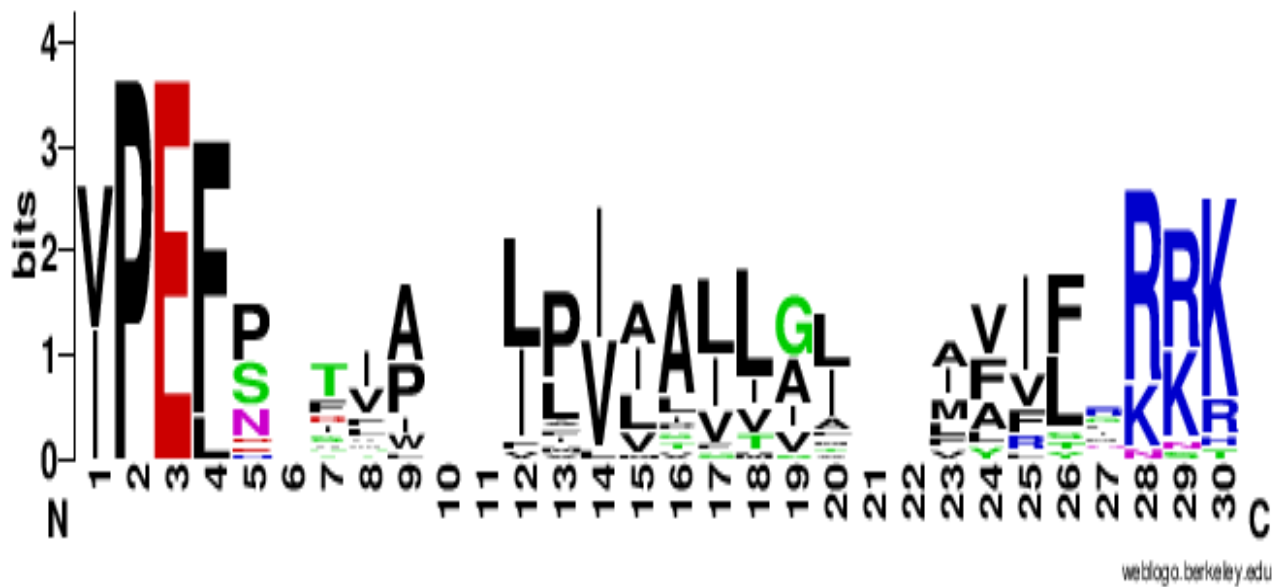




**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ**

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ  
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ ΒΙΟΪΑΤΡΙΚΗ**

**ΠΡΟΓΝΩΣΗ ΤΩΝ ΠΡΟΣΔΕΔΕΜΕΝΩΝ ΣΤΟ ΒΑΚΤΗΡΙΑΚΟ ΤΟΙΧΩΜΑ ΠΡΩΤΕΪΝΩΝ**



**ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ**

**ΜΠΑΓΚΟΣ ΠΑΝΤΕΛΗΣ, ΑΝΑΠΛΗΡΩΤΗΣ ΚΑΘΗΓΗΤΗΣ**

**ΕΛΕΝΗ Μ. ΝΥΔΡΙΩΤΗ  
(ΑΜ: 00423)**

**ΛΑΜΙΑ 2016**



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ**  
**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**  
**ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ**  
**ΒΙΟΪΑΤΡΙΚΗ**

**ΠΡΟΓΝΩΣΗ ΤΩΝ ΠΡΟΣΔΕΔΕΜΕΝΩΝ ΣΤΟ ΒΑΚΤΗΡΙΑΚΟ ΤΟΙΧΩΜΑ ΠΡΩΤΕΪΝΩΝ**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**  
**ΕΛΕΝΗ ΝΥΔΡΙΩΤΗ**

**ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ**  
**Δρ ΠΑΝΤΕΛΗΣ ΜΠΑΓΚΟΣ, ΑΝΑΠΛΗΡΩΤΗΣ ΚΑΘΗΓΗΤΗΣ**

**ΛΑΜΙΑ, 2016**

## **ΠΡΟΓΝΩΣΗ ΤΩΝ ΠΡΟΣΔΕΔΕΜΕΝΩΝ ΣΤΟ ΒΑΚΤΗΡΙΑΚΟ ΤΟΙΧΩΜΑ ΠΡΩΤΕΪΝΩΝ**

### **ΤΡΙΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:**

Δρ ΠΑΝΤΕΛΗΣ ΜΠΑΓΚΟΣ, ΑΝΑΠΛΗΡΩΤΗΣ ΚΑΘΗΓΗΤΗΣ

Δρ ΒΑΣΙΛΗΣ ΠΛΑΓΙΑΝΑΚΟΣ, ΑΝΑΠΛΗΡΩΤΗΣ ΚΑΘΗΓΗΤΗΣ

Δρ ΜΑΡΙΑ ΑΔΑΜ, ΕΠΙΚΟΥΡΟΣ ΚΑΘΗΓΗΤΡΙΑ

Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις <sup>(1)</sup>, που προβλέπονται από της διατάξεις της παρ. 6 του άρθρου 22 του Ν. 1599/1986, δηλώνω ότι:

1. Δεν παραθέτω κομμάτια βιβλίων ή άρθρων ή εργασιών άλλων αυτολεξεί **χωρίς να τα περικλείω σε εισαγωγικά** και χωρίς να αναφέρω το συγγραφέα, τη χρονολογία, τη σελίδα. Η αυτολεξεί παράθεση χωρίς εισαγωγικά χωρίς αναφορά στην πηγή, είναι λογοκλοπή. Πέραν της αυτολεξεί παράθεσης, λογοκλοπή θεωρείται και η παράφραση εδαφίων από έργα άλλων, συμπεριλαμβανομένων και έργων συμφοιτητών μου, καθώς και η παράθεση στοιχείων που άλλοι συνέλεξαν ή επεξεργάστηκαν, χωρίς αναφορά στην πηγή. Αναφέρω πάντοτε με πληρότητα την πηγή κάτω από τον πίνακα ή σχέδιο, όπως στα παραθέματα.
2. Δέχομαι ότι η αυτολεξεί **παράθεση χωρίς εισαγωγικά**, ακόμα κι αν συνοδεύεται από αναφορά στην πηγή σε κάποιο άλλο σημείο του κειμένου ή στο τέλος του, είναι αντιγραφή. Η αναφορά στην πηγή στο τέλος π.χ. μιας παραγράφου ή μιας σελίδας, δεν δικαιολογεί συρραφή εδαφίων έργου άλλου συγγραφέα, έστω και παραφρασμένων, και παρουσίασή τους ως δική μου εργασία.
3. Δέχομαι ότι υπάρχει επίσης περιορισμός στο μέγεθος και στη συχνότητα των παραθεμάτων που μπορώ να εντάξω στην εργασία μου εντός εισαγωγικών. Κάθε μεγάλο παράθεμα (π.χ. σε πίνακα ή πλαίσιο, κλπ), προϋποθέτει ειδικές ρυθμίσεις, και όταν δημοσιεύεται προϋποθέτει την άδεια του συγγραφέα ή του εκδότη. Το ίδιο και οι πίνακες και τα σχέδια
4. Δέχομαι όλες τις συνέπειες σε περίπτωση λογοκλοπής ή αντιγραφής.

Ημερομηνία: ...../...../20.....

Ο – Η Δηλ.

(Υπογραφή)

(1) «Όποιος εν γνώσει του δηλώνει ψευδή γεγονότα ή αρνείται ή αποκρύπτει τα αληθινά με έγγραφη υπεύθυνη δήλωση του άρθρου 8 παρ. 4 Ν. 1599/1986 τιμωρείται με φυλάκιση τουλάχιστον τριών μηνών. Εάν ο υπαίτιος αυτών των πράξεων σκόπευε να προσπορίσει στον εαυτόν του ή σε άλλον περιουσιακό όφελος βλάπτοντας τρίτον ή σκόπευε να βλάψει άλλον, τιμωρείται με κάθειρξη μέχρι 10 ετών.

***Στη μητέρα μου...***

## **Ευχαριστίες**

Μετά το τέλος της πτυχιακής μου εργασίας θα ήθελα να ευχαριστήσω θερμά όσους συνέβαλαν στη διεκπαιρέωσή της. Πρώτα από όλους, θα ήθελα να πω ένα μεγάλο ευχαριστώ στον επιβλέποντα καθηγητή μου, κ. Παντελή Μπάγκο για τις γνώσεις και την καθοδήγηση που μου προσέφερε. Οι ιδέες του και η παρότρυνσή του μου έδωσαν τη θέληση να ασχοληθώ μελλοντικά με την έρευνα.

Ευχαριστώ επίσης τον κ. Κώστα Τσιρίγο για την αμέριστη βοήθειά του καθόλη τη διάρκεια εκπόνησης της εργασίας μου. Μου παρείχε πολύτιμες συμβουλές και υλικό οποιαδήποτε στιγμή κι αν του ζητήθηκαν.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου και τους φίλους μου για την αμέριστη συμπαράσταση και στήριξη που έδειξαν σε αυτή την προσπάθειά μου. Ιδιαίτερα θέλω να πω ένα μεγάλο ευχαριστώ στη φίλη μου, τη Γιωργία.

## Περίληψη

Οι επιφανειακές πρωτεΐνες των Αρχαίων έχει διαπιστωθεί ότι παίζουν σπουδαίο ρόλο στην εκτέλεση των βιολογικών τους λειτουργιών και βοηθούν στην αλληλεπίδρασή τους με το περιβάλλον. Επίσης, τους παρέχουν σταθερότητα και συμβάλλουν στην επιβίωσή τους. Οι πρωτεΐνες αυτές στο καρβοξυτελικό τους άκρο φέρουν ένα πεπτιδίο στόχευσης, το οποίο περιέχει ένα από τα εξής χαρακτηριστικά μοτίβα: PGF, VPxxxxP, PIP, PEF και ακολουθείται από μία υδρόφοβη περιοχή και τέλος μία φορτισμένη ουρά. Ύστερα από μελέτες που έχουν διεξαχθεί έχει βρεθεί ότι οι πρωτεΐνες αυτές κόβονται στο χαρακτηριστικό μοτίβο από ειδικά ένζυμα τα οποία ονομάζονται Αρχαιοσορτάσες. Σκοπός της συγκεκριμένης εργασίας είναι να αναπτυχθεί ένα εργαλείο πρόγνωσης που θα είναι ικανό να εντοπίζει με ακρίβεια τις πρωτεΐνες που βρίσκονται στο τοίχωμα των Αρχαίων και περιέχουν ένα από τα παραπάνω χαρακτηριστικά μοτίβα. Συγκεντρώθηκαν συνολικά έξι πρωτεϊνικές οικογένειες στις οποίες εφαρμόστηκε το πιθανοθεωρητικό μοντέλο προφίλ Hidden Markov Model (pHMM) κι έγινε σύγκριση μεταξύ αυτών και όλων των γνωστών γονιδιωμάτων των Αρχαίων. Με βάση τα αποτελέσματα που πήραμε καταλήξαμε στο συμπέρασμα ότι το εργαλείο αυτό μπορεί να κάνει σωστό διαχωρισμό πρωτεϊνών σε πολύ μεγάλο βαθμό. Ωστόσο, επειδή βρέθηκαν πρωτεΐνες που δεν έδωσαν επαρκή στοιχεία, μελλοντικά, θα γίνει προσπάθεια κατασκευής του κλασικού Hidden Markov Model διότι είναι πιο ευαίσθητο. Η πρόγνωση των πρωτεϊνών στο κυτταρικό τοίχωμα αποτελεί πολύ σημαντικό στοιχείο στην έρευνα για τις λειτουργίες των Αρχαίων όχι μόνο επειδή μας παρέχει πολύτιμες πληροφορίες για τους βιολογικούς μηχανισμούς που χρησιμοποιούν αλλά και γιατί μπορεί να βρει εφαρμογή στη βιομηχανία και ειδικότερα στην παρασκευή φαρμάκων.

**Λέξεις - Κλειδιά:** Αρχαιοσορτάσες, προφίλ Hidden Markov Model (pHMM), γονιδίωμα

## **Abstract**

The surface proteins of the Archaea have been found to play an important role in the performance of their biological functions and help their interaction with the environment. Also, they provide them stability and contribute to their survival. These proteins, at their C-terminus, have a targeting peptide, which contains the PGF, VPxxxxP, PIP or PEF pattern, and is followed by a hydrophobic region and a charged tail in the end. After conducted studies, it has been found that these proteins were cleaved to the characteristic pattern of specific enzymes called Archaeosortases. The purpose of this project is to develop a prediction tool that is able to accurately detect the proteins found in the wall of the Archaea and contain one of the above characteristic patterns. Six protein families were gathered in which the probabilistic model profile Hidden Markov Model (pHMM) was applied and a comparison was made between them and all known genomes of the Archaea. Based on the obtained results, we concluded that this tool can do a proper separation in the proteins. However, because some proteins did not give enough information, a future effort will be made in order to construct the classical Hidden Markov Model which is more sensitive. The prediction of the proteins in the cell wall is a very important element in the investigation of the functions of the Archaea not only because it gives us valuable information about the biological mechanisms they use but also because it can find application in industry and particularly in the pharmaceutical manufacture.

**Key – Words:** Archaea, Archaeosortases, profile Hidden Markov Model (pHMM), genome



## Περιεχόμενα

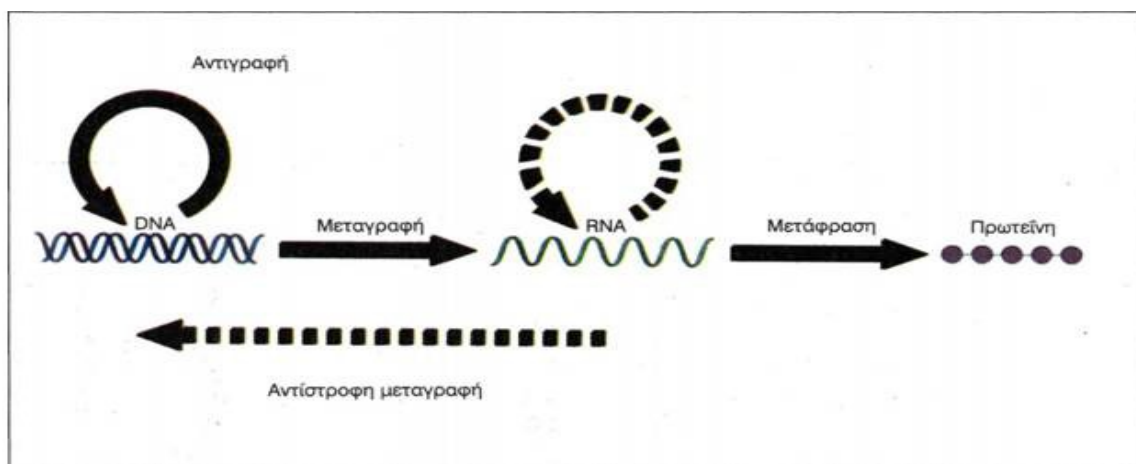
1.	Εισαγωγή.....σελ.	10
1.1	Πρωτεΐνες.....σελ.	11
1.2	Άλλες μορφές ζωής.....σελ.	11
1.3	Αρχαία.....σελ.	13
1.3.1	Μορφολογία Αρχαίων.....σελ.	13
1.3.2	Κυτταρικό τούχωμα και κυτταρική μεμβράνη.....σελ.	14
1.3.3	Ομάδες Αρχαίων.....σελ.	17
1.4	Σορτάσες/ Αρχαιοσορτάσες.....σελ.	22
1.4.1	Σορτάσες.....σελ.	22
1.4.2	Μηχανισμός προσκόλλησης των LPxTG πρωτεϊνών στο κυτταρικό τούχωμα.....σελ.	23
1.4.3	Αρχαιοσορτάσες.....σελ.	24
1.4.4	Μηχανισμός προσκόλλησης των Αρχαιοσορτασών στο κυτταρικό τούχωμα.....σελ.	26
2.	Μεθοδολογία.....σελ.	27
2.1	Συλλογή δεδομένων για την ανεύρεση των υποστρωμάτων των πρωτεϊνικών οικογενειών.....σελ.	27
2.2	Κατασκευή του προφίλ Hidden Markov Model.....σελ.	33
2.2.1	Μαρκοβιανά Μοντέλα (Hidden Markov Model).....σελ.	33
2.2.2	Profiles Hidden Markov Models.....σελ.	35
3.	Αποτελέσματα – Συμπεράσματα.....σελ.	40
4.	Βιβλιογραφικές Αναφορές.....σελ.	44
5.	Παράρτημα Α.....σελ.	47
6.	Παράρτημα Β.....σελ.	49
7.	Παράρτημα Γ.....σελ.	54
8.	Παράρτημα Δ.....σελ.	65

## Εισαγωγή

### 1.1 Πρωτεΐνες

Οι πρωτεΐνες θεωρούνται τα πιο διαδεδομένα και πολυδιάστατα, τόσο στη μορφή όσο και στη λειτουργία τους, μακρομόρια. Αποτελούνται από διαδοχικά αμινοξέα τα οποία ενώνονται μεταξύ τους με πεπτιδικούς δεσμούς σχηματίζοντας μια γραμμική αλυσίδα που ονομάζεται πολυπεπίδιο. Κάθε αμινοξύ έχει τη δική του φυσικοχημική συμπεριφορά, η οποία καθορίζεται από την πλευρική του ομάδα. Οι πρωτεΐνες συντίθενται στα ριβοσώματα και λόγω της ευελιξίας στην συμπεριφορά τους διαδραματίζουν ένα πολύ σημαντικό ρόλο στην οργάνωση και την λειτουργία των κυττάρων, εφόσον μπορούν να αποτελέσουν ένζυμα, συστατικά μεμβρανών, ορμόνες, ινίδια, μέσα επικοινωνίας των κυττάρων κλπ.

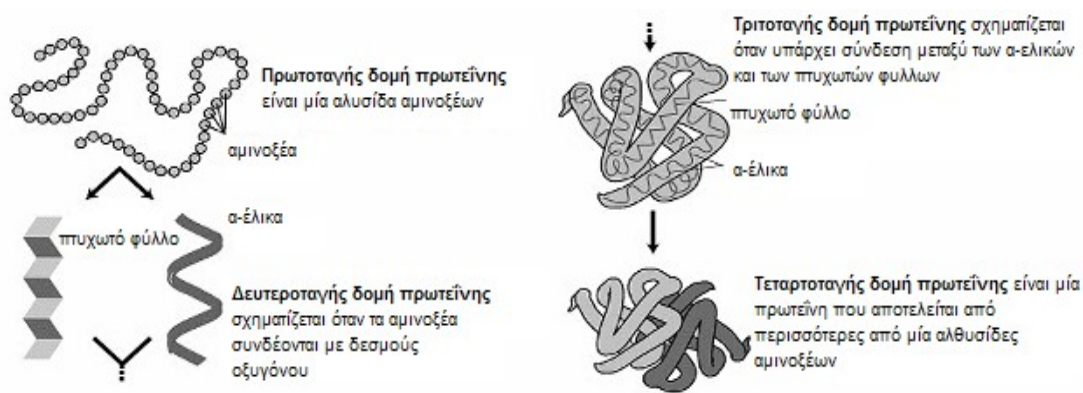
Σύμφωνα με το κεντρικό δόγμα της μοριακής βιολογίας [1, 2] οι πρωτεΐνες συντίθενται μονοσήμαντα, με βάση τη γενετική πληροφορία που εμπεριέχεται στο DNA και στη συνέχεια, μέσω του RNA, η πληροφορία αυτή μεταβιβάζεται στην πρωτεϊνοσυνθετική μηχανή των ριβοσωμάτων. Τέλος, με τη διαδικασία της μετάφρασης οδηγούμαστε στο τελικό προϊόν, την πρωτεΐνη.



**Εικόνα 1.1:** Το κεντρικό δόγμα της μοριακής βιολογίας. Τα συνεχή βέλη αναπαριστούν την κατεύθυνση με την οποία η γενετική πληροφορία, που είναι καταγραμμένη στο μόριο του DNA, «ρέει» προς τις πρωτεΐνες [21].

Ο τρόπος με τον οποίο μπορούν να συνδυαστούν τα 20 αμινοξέα στη φύση είναι άπειρος κι αυτή η ικανότητά τους τα καθιστά μοναδικά. Η κεντρική δομή κάθε αμινοξέος είναι η ίδια, δηλαδή αποτελείται από το άλφα άτομο άνθρακα το οποίο είναι ενωμένο με μία αμινομάδα, μία καρβοξυλομάδα, ένα υδρογόνο και μία πλευρική ομάδα. Η μόνη διαφορά που υπάρχει εντοπίζεται στην

πλευρική ομάδα, η οποία διαφέρει από αμινοξύ σε αμινοξύ και έτσι του προσδίδει τις χαρακτηριστικές του ιδιότητες. Αρχικά, οι πρωτεΐνες που βρίσκονται στα ριβοσώματα εμφανίζονται ως μία γραμμική αλληλουχία αμινοξέων (πρωτοταγής δομή) που στη συνέχεια σταθεροποιείται με ασθενικούς δεσμούς υδρογόνου (δευτεροταγής δομή) και τελικά η πολυπεπτιδική αλυσίδα υφίσταται ένα ακόμη πιο περίπλοκο δίκλωμα (πτύχωση), το οποίο καλείται τριτοταγής δομή. Στο στάδιο αυτό η πρωτεΐνη αποκτά την τελική της στερεοδιάταξη στο χώρο. Επιπλέον, υπάρχουν και περιπτώσεις κατά τις οποίες όμοιες ή διαφορετικές πολυπεπτιδικές αλυσίδες που έχουν αναδιπλωθεί μπορούν συχνά να συνενώνονται μεταξύ τους σχηματίζοντας μεγαλύτερα πρωτεϊνικά σύμπλοκα. Το τελικό σχήμα που αποκτά το πρωτεϊνικό σύμπλοκο στο χώρο αποτελεί την τεταρτοταγή δομή της πρωτεΐνης.



**Εικόνα 1.2:** Σχηματική αναπαράσταση των τεσσάρων μορφών που μπορεί να πάρει μια πολυπεπτιδική αλυσίδα, πρωτοταγής, δευτεροταγής, τριτοταγής και τεταρτοταγής δομή [30].

## 1.2 Άλλες μορφές ζωής

Όλοι οι οργανισμοί που σήμερα διαβιούν στη Γη, πιστεύεται ότι έχουν προέλθει από μια κοινή προγονική κυτταρική μορφή που εμφανίστηκε πριν από τουλάχιστον 3,8 δισεκατομμύρια χρόνια. Η παραδοχή αυτή στηρίζεται σε συγκριτικές μελέτες κυτταρικής και μοριακής βιολογίας, βιοχημείας και γενετικής, οι οποίες υποδεικνύουν ότι όλα τα διαφορετικά είδη οργανισμών διαθέτουν τα ίδια συστατικά (άτομα, μόρια, βιομόρια), τον ίδιο γενετικό κώδικα, την ίδια κυτταρική οργάνωση και τους ίδιους μηχανισμούς μεταβολισμού. Η εξελικτική πορεία από το πρωτόγονο αυτό «πρωτοκύτταρο» ή «πρωτοοργανισμό» προς τους μετέπειτα προκαρυωτικούς και ευκαρυωτικούς οργανισμούς, υπήρξε προϊόν συνδυασμού εξελικτικών μηχανισμών που σχετίζονταν με τη δημιουργία μεταλλαγών και την εγκαθίδρυσή τους σε επίπεδο πληθυσμού [3].

Οι έμβιοι οργανισμοί διακρίνονται σε δύο μεγάλες κατηγορίες, τους προκαρυωτικούς και τους ευκαρυωτικούς οργανισμούς, οι οποίες καθορίζονται από τη δομή των κυττάρων τους και τη

φυσιολογία τους. Με όρους ταξινομικής, κατατάσσονται σε τρεις επικράτειες: Αρχαία, Ευβακτήρια ή Βακτήρια και Ευκαρυωτικά ή Ευκάρυα. Τα προκαρυωτικά και τα ευκαρυωτικά κύτταρα έχουν παρόμοια χημική δομή, υπό την έννοια ότι και τα δύο περιέχουν νουκλεϊκά οξέα, πρωτεΐνες, λιπίδια και υδατάνθρακες. Χρησιμοποιούν τα ίδια είδη χημικών αντιδράσεων για να μεταβολίσουν την τροφή τους, να κατασκευάσουν τις πρωτεΐνες τους και να αποθηκεύσουν ενέργεια. Οι κύριες διαφορές τους εντοπίζονται στη δομή των τοιχωμάτων και των κυτταροπλασματικών μεμβρανών και στην απουσία οργανιδίων. Ωστόσο, η θεμελιώδης διαφορά τους βασίζεται στην ύπαρξη ή όχι πυρήνα. Πιο συγκεκριμένα, οι οργανισμοί των οποίων τα κύτταρα έχουν πυρήνα αποκαλούνται ευκαρυωτικοί, ενώ εκείνοι που αποτελούνται από κύτταρα χωρίς πυρήνα αποκαλούνται προκαρυωτικοί.

Τα βασικότερα χαρακτηριστικά των προκαρυωτών είναι τα εξής: Το DNA τους δεν εσωκλείεται μέσα σε μεμβράνη. Συνήθως είναι ένα μοναδικό κυκλικό χρωμόσωμα το οποίο συνδέεται με διάφορες πρωτεΐνες. Τα κυτταρικά τους τοιχώματα σχεδόν πάντα περιέχουν το σύνθετο πολυσακχαρίτη πεπτιδογλυκάνη και στερούνται οργανιδίων που περιβάλλονται από μεμβράνη. Επίσης, διαιρούνται συνήθως με διχοτόμηση. Κατά τη διάρκεια αυτής της διαδικασίας, το DNA αντιγράφεται και το μητρικό κύτταρο διαχωρίζεται σε δύο θυγατρικά κύτταρα [4].

**Πίνακας 1:** Σύγκριση προκαρυωτών και ευκαρυωτών.

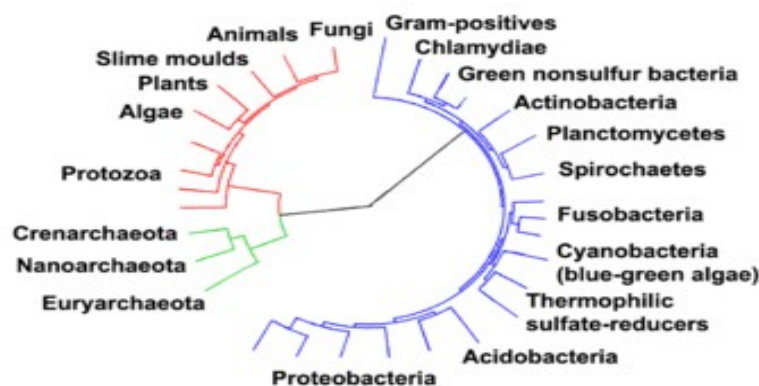
	<b>Προκαρυωτικοί</b>	<b>Ευκαρυωτικοί</b>
Ύπαρξη κυτταρικού πυρήνα	Όχι	Ναι
Κυτταρικά οργανίδια (μιτοχόνδρια-χλωροπλάστες)	Όχι	Ναι
Τρόπος αναπαραγωγής	Σχάση	Μίτωση ή μείωση
Μεταβολισμός	Αναερόβιος και αερόβιος	Αερόβιος
Κυτταρική οργάνωση	Μονοκύτταροι	Κυρίως πολυκύτταροι με διαφοροποίηση κυττάρων

Τα βασικότερα χαρακτηριστικά των ευκαρυωτών είναι τα εξής: Το DNA τους βρίσκεται στον πυρήνα του κυττάρου, ο οποίος χωρίζεται από το κυτταρόπλασμα με μια πυρηνική μεμβράνη. Το DNA βρίσκεται σε πολλά χρωμοσώματα και συνδέεται στενά με χρωμοσωμικές πρωτεΐνες που αποκαλούνται ιστόνες και μη ιστόνες. Επιπλέον, οι ευκαρυώτες έχουν αρκετά διαφορετικά οργανίδια που περιβάλλονται από μεμβράνη, συμπεριλαμβανομένων των μιτοχονδρίων, του ενδοπλασματικού δικτύου, του συμπλέγματος Golgi, των λυσοσωμάτων και συχνά των χλωροπλάστων. Η κυτταρική διαίρεσή τους περιλαμβάνει τη μίτωση, κατά την οποία τα χρωμοσώματα αντιγράφονται και όλα τα νέα χρωμοσώματα σχηματίζουν ένα καινούριο πυρήνα. Ακολουθεί η διαίρεση του κυτταροπλάσματος και των οργανιδίων σε δύο ίσα μέρη και προκύπτουν δύο πανομοιότυπα θυγατρικά κύτταρα [4].

### 1.3 Αρχαία

Στο μεγαλύτερο μέρος του 20ού αιώνα οι προκαρυώτες (κύτταρα χωρίς πυρήνα) θεωρήθηκαν ως μία ενιαία ομάδα οργανισμών και ταξινομήθηκαν με βάση τη μορφολογία, τη βιοχημεία και το μεταβολισμό τους. Το 1965, οι Linus Pauling και Emile Zuckerkandl πρότειναν για την καλύτερη μελέτη των προκαρυωτών αντί να εξετάζονται τα παραπάνω κριτήρια, να χρησιμοποιούνται οι αλληλουχίες των γονιδίων. Αυτή η προσέγγιση είναι γνωστή ως φυλογενετική κι αποτελεί την κύρια μέθοδο που εφαρμόζεται σήμερα. Αρχικά, θεωρείτο ότι τα Αρχαία ανήκουν στην ίδια ομάδα με τα βακτήρια όμως το 1977 οι Δρ. Carl Woese και Δρ. George E. Fox κατάφεραν να τα ταξινομήσουν σε ξεχωριστές ομάδες, σε φυλογενετικά δένδρα με βάση τις αλληλουχίες του ριβοσωμικού RNA (rRNA) των γονιδίων. Για να τονιστεί καλύτερα η διαφορά αυτή ο Δρ. Carl Woese πρότεινε ένα νέο σύστημα ταξινόμησης σε τρεις επικράτειες: την Επικράτεια των Ευκαρυωτών, την Επικράτεια των Βακτηρίων και την Επικράτεια των Αρχαίων [5].

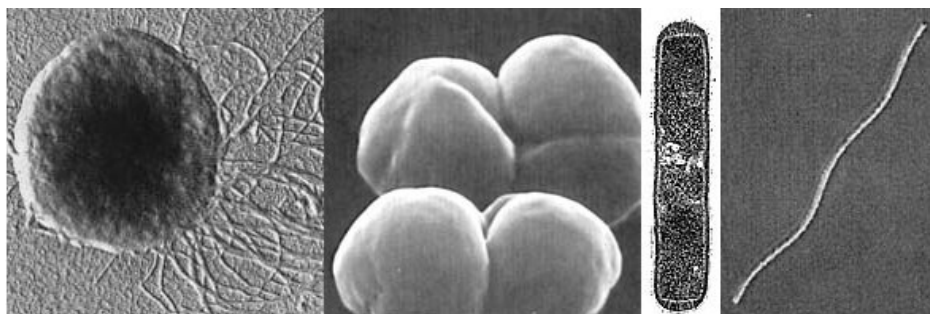
Τα περισσότερα Αρχαία ζουν σε ακραίες περιβαλλοντικές συνθήκες, στις οποίες καμία άλλη σύγχρονη μορφή ζωής δεν μπορεί να επιβιώσει αλλά οι οποίες πιθανόν να αντιπροσωπεύουν το κυρίαρχο περιβάλλον κατά την πρώτη περίοδο της εξέλιξης των προκαρυωτικών οργανισμών. Πρόκειται για κύτταρα που ζουν σε πολύ αλμυρό νερό, σε θερμές όξινες ηφαιστειακές πηγές, στον πυθμένα της θάλασσας, σε εγκαταστάσεις επεξεργασίας αποβλήτων και στο όξινο αναερόβιο περιβάλλον του στομάχου μιας αγελάδας, όπου αποδομούν την κυτταρίνη και παράγουν μεθάνιο. Πολλοί από τους παραπάνω βιότοπους θυμίζουν τις αντίξοες συνθήκες που πρέπει να επικρατούσαν στην αρχέγονη Γη, όταν πρωτοεμφανίστηκαν έμβια όντα, προτού η ατμόσφαιρα εμπλουτιστεί σε οξυγόνο [6, 7, 8].



**Εικόνα 1.3:** Φυλογενετικό δένδρο που δείχνει τη σχέση μεταξύ των Αρχαίων και των άλλων μορφών της ζωής. Οι ευκαρυώτες απεικονίζονται με κόκκινο χρώμα, τα αρχαία με πράσινο και τα βακτήρια με μπλε [10].

### 1.3.1 Μορφολογία Αρχαίων

Τα Αρχαία είναι μικροσκοπικοί οργανισμοί με διάμετρο που κυμαίνεται από 0.1 έως 15 μm. Εμφανίζουν μεγάλη ποικιλία ως προς το σχήμα ή τη μορφή τους. Μερικά είναι σφαιρικά, μία μορφή γνωστή ως κόκκος, μπορεί να είναι είτε απολύτως σφαιρικά είτε λοβωειδή. Μερικά έχουν σχήμα ράβδου και είναι γνωστά ως βάκιλοι. Επίσης, έχουν ανακαλυφθεί κάποια άλλα είδη με τριγωνικό ή τετραγωνικό σχήμα. Το σχήμα των Αρχαίων διατηρείται χάρη στην ύπαρξη του κυτταρικού τοιχώματος και του προκαρυωτικού κυτταροσκελετού. Ωστόσο, στην περίπτωση των *Thermoplasma* και *Ferroplasma* η έλλειψη κυτταρικού τοιχώματος δεν προσδίδει κάποιο συγκεκριμένο σχήμα. Επίσης, τα περισσότερα Αρχαία για να μετακινηθούν χρησιμοποιούν μαστίγια, τα οποία είναι προσκολλημένα στο κυτταρικό τοίχωμα [9].



**Εικόνα 1.4:** Από αριστερά, *Methanococcus janaschii*, κόκκος με πολυάριθμα μαστίγια συνδεδεμένα στη μία άκρη. Στο κέντρο-αριστερά, *Methanosarcina barkeri*, ένας λοβωειδής κόκκος χωρίς μαστίγια. Στο κέντρο δεξιά, *Methanothermus fervidus*, ένας μικρός βάκιλος χωρίς μαστίγια. Δεξιά, *Methanobacterium thermoautotrophicum*, επιμήκης βάκιλος [9].

### 1.3.2 Κυτταρικό τοίχωμα και κυτταρική μεμβράνη

Σε γενικές γραμμές, τα Αρχαία παρουσιάζουν πολλά κοινά στοιχεία με τα βακτήρια αλλά η χημική τους σύσταση τα καθιστά εντελώς διαφορετικά. Για παράδειγμα, και οι δύο αυτές ομάδες έχουν κυτταρική μεμβράνη, η οποία περιβάλλεται από κυτταρικό τοίχωμα, και μετακινούνται με τη βοήθεια μαστιγίων τα οποία είναι προσκολλημένα στο κυτταρικό τοίχωμα. Το κυτταρικό τοίχωμα των Αρχαίων σχηματίζεται από τις surface-layer πρωτεΐνες, οι οποίες συνθέτουν το S-layer. Το S-layer καλύπτει την εξωτερική επιφάνεια του κυττάρου παρέχοντάς του χημική και φυσική προστασία. Παράλληλα, αποτρέπει διάφορα μακρομόρια να έρθουν σε επαφή με την κυτταροπλασματική μεμβράνη. Ωστόσο, η χημική σύσταση του κυτταρικού τοιχώματος των βακτηρίων διαφέρει κατά πολύ από την αντίστοιχη των Αρχαίων. Πιο συγκεκριμένα, τα τοιχώματα των βακτηρίων αποτελούνται από πεπτιδογλυκάνη



ενώ των αρχαίων όχι. Κανένα από τα Αρχαία δεν έχει μουραμικό οξύ ή D-αμινο οξέα στα κυτταρικά του τοιχώματα, χημικά μόρια που είναι χαρακτηριστικά της πεπτιδογλυκάνης στα βακτήρια. Σε μερικά Αρχαία (*Methanobacterium*, *Methanothermus* και *Methanobrevibacter*) σχηματίζεται μια ψευδοπεπτιδογλυκάνη που αποτελείται από επαναλαμβανόμενα μόρια N-ακετυλογλουκοσαμίνης και N-ακετυλοταλοσαμινουρονικών οξέων και L-αμινο οξέων (όχι D-αμινο οξέων όπως στην πεπτιδογλυκάνη των ευβακτηρίων). Μερικά μεθανογόνα Αρχαία (π.χ. *Methanobolus*) και αρκετά ακραίως θερμοφιλά (π.χ. *Sulfolobus*, *Pyrodictium* και *Thermoproteus*) έχουν κυτταρικά τοιχώματα από γλυκοπρωτεΐνη, ενώ κάποια άλλα μεθανογόνα (π.χ. *Methanococcus* και *Methanomicrobium*) έχουν κυτταρικά τοιχώματα που αποτελούνται αποκλειστικά από πρωτεΐνη. Επίσης, τα μαστίγια των Αρχαίων λειτουργούν όπως και των βακτηρίων, δηλαδή τα μακριά στελέχη τους κινούνται περιστροφικά από τη βάση. Στα βακτήρια η απαιτούμενη ενέργεια για την κίνηση των μαστιγίων δίνεται από την κλίση της συγκέντρωσης των πρωτονίων κατά μήκος της μεμβράνης. Παρόλα αυτά, τα μαστίγια των Αρχαίων, διαφέρουν τόσο στην εξέλιξη όσο και στη σύνθεση. Τα δύο είδη μαστιγίων έχουν προέλθει από διαφορετικούς προγόνους. Τα βακτηριακά έχουν κοινό πρόγονο με το σύστημα τύπου έκκρισης III (type III secretion system), ενώ των Αρχαίων φαίνεται να εξελίχθηκαν από το type IV pili system. Σε αντίθεση με το βακτηριακό μαστίγιο, το οποίο είναι κοίλο και απαρτίζεται από υπομονάδες, που κινείται επάνω στον κεντρικό πόρο προς την άκρη του μαστιγίου, το μαστίγιο των Αρχαίων συντίθεται με την προσθήκη υπομονάδων στη βάση.

Η βασικότερη διαφορά ανάμεσα στα Αρχαία και στα βακτήρια, καθώς και σε όλα τα άλλα είδη κυττάρων, βρίσκεται στην κυτταρική μεμβράνη και εντοπίζεται σε τέσσερα διαφορετικά σημεία.

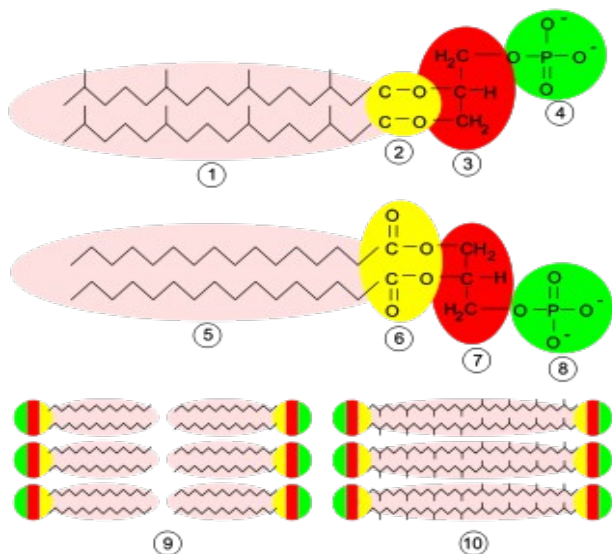
Πρώτον, η βασική ομάδα από την οποία κατασκευάζεται η κυτταρική μεμβράνη, είναι το φωσφολιπίδιο. Δηλαδή, ένα μόριο γλυκερόλης το οποίο έχει ένα φωσφορικό άλας που προστίθεται στο τέλος και δύο πλευρικές αλυσίδες που συνδέονται στο άλλο άκρο. Όταν συντίθεται η κυτταρική μεμβράνη, η γλυκερόλη και το φωσφορικό άκρο του μορίου βγαίνουν στην επιφάνεια και οι μακριές πλευρικές ομάδες πιέζονται στη μέση. Μ'αυτό τον τρόπο δημιουργείται ένα χημικό φράγμα κι επιτυγχάνεται η διατήρηση της χημικής ισορροπίας. Η γλυκερόλη που χρησιμοποιείται στα Αρχαία αποτελεί στερεοϊσομερές της αντίστοιχης γλυκερόλης η οποία χρησιμοποιείται για τη σύνθεση της μεμβράνης στα βακτήρια και τους ευκαρυώτες. Δύο μόρια που είναι το ένα καθρέφτης του άλλου, θεωρούνται στερεοϊσομερή. Αυτή η ιδιότητα, στην περίπτωση της γλυκερόλης των Αρχαίων και των άλλων οργανισμών, σημαίνει ότι δεν μπορούμε να μετατρέψουμε το ένα μόριο στο άλλο με απλή περιστροφή. Ενώ τα βακτήρια και τα ευκαρυωτικά κύτταρα έχουν D-γλυκερόλη (D-glycerol) στις μεμβράνες τους, τα Αρχαία έχουν L-γλυκερόλη (L-glycerol) στις δικές τους. Τα χημικά συστατικά του κυττάρου πρέπει να κατασκευαστούν από ένζυμα και αυτή η "χειρομορφία" (chirality) του κυττάρου

καθορίζεται από το σχήμα αυτών των ενζύμων. Έτσι, ένα κύτταρο που παίρνει μία μορφή αδυνατεί να δημιουργήσει την άλλη μορφή.

Δεύτερον, στα βακτήρια και τους ευκαρυώτες, οι πλευρικές αλυσίδες των φωσφολιπιδίων αποτελούνται από λιπαρά οξέα. Συνήθως σχηματίζονται από 16 μέχρι 18 άτομα άνθρακα. Αντιθέτως, τα Αρχαία για το σχηματισμό των φωσφολιπιδίων της μεμβράνης τους δεν χρησιμοποιούν λιπαρά οξέα. Αντ' αυτού, οι πλευρικές τους ομάδες αποτελούνται από 20 άτομα άνθρακα προερχόμενα από ισοπρένιο. Πιθανόν οι διακλαδισμένες αλυσίδες ισοπρενίου να εμποδίζουν τα Αρχαία να καταστραφούν σε πολύ υψηλές θερμοκρασίες.

Τρίτον, στις μεμβράνες των βακτηρίων και των ευκαρυωτών αναπτύσσονται εστερικοί δεσμοί, ενώ στην Αρχαίων αιθερικοί δεσμοί. Από χημικής απόψεως οι αιθερικοί δεσμοί είναι πιο ανθεκτικοί από τους αντίστοιχους εστερικούς, γεγονός που προσφέρει στα Αρχαία παραπάνω σταθερότητα. Αυτή η ιδιότητα μπορεί να τα βοηθήσει να επιβιώσουν σε ακραίες συνθήκες, όπως σε πολύ υψηλές θερμοκρασίες και σε όξινο ή αλκαλικό περιβάλλον.

Τέταρτον, σε ορισμένα Αρχαία η διπλοστοιβάδα των λιπιδίων έχει αντικαταθεί με μονοστοιβάδα. Στην πραγματικότητα, έχουν ενωθεί οι ουρές των δύο ανεξάρτητων μορίων φωσφολιπιδίου σε ένα μόριο με δύο πολικές κεφαλές. Αυτή η σύντηξη μπορεί να καταστήσει τις μεμβράνες τους πιο άκαμπτες και να τα βοηθήσει να επιβιώσουν σε πολύ σκληρές συνθήκες [9, 10].



**Εικόνα 1.5:** Απεικόνιση πρωτεϊνικών δομών. Επάνω: φωσφολιπίδιο από Αρχαία: 1. Αλυσίδα ισοπρενίου, 2. Αιθερικός δεσμός, 3. Τμήμα L-γλυκερόλης, 4. Φωσφορική ομάδα. Μέση: βακτηριακό ή ευκαρυωτικό φωσφολιπίδιο: 5. Αλυσίδες λιπαρών οξέων, 6. Εστερικοί δεσμοί, 7. τμήμα D-γλυκερόλης, 8. Φωσφορική ομάδα. Κάτω: 9. Λιπιδική διπλοστοιβάδα βακτηρίων ή ευκαρυωτών, 10. Λιπιδική μονοστοιβάδα ορισμένων Αρχαίων [10].



**Πίνακας 2:** Σύγκριση μεταξύ βακτηρίων και Αρχαίων.

Χαρακτηριστικά	Βακτήρια	Αρχαία
Συστατικά κυτταρικού τοιχώματος	Μουρεΐνη	Ψευδομουρεΐνη, πρωτεΐνες, πολυσακχαρίτες
Στρογγυλές και επίπεδες δομές	-	+
Ενδοσπόρια	+	-
Ειδικά συνένζυμα	-	+
Ευκαρυωτική RNA πολυμεράση	-	+
Ολοκληρωμένη φωτοσύνθεση	+	-
Μεθανογένεση	-	+
Μέγιστη θερμοκρασία ανάπτυξης	90°C	110 °C
Αναστέλλονται από αντιβιοτικά	+	-

### 1.3.3 Ομάδες Αρχαίων

Σύμφωνα με το Bergey's Manual τα Αρχαία διαιρούνται σε πέντε ομάδες: τα μεθανογόνα Αρχαία, τους αναγωγείς τωνθειικών, τα ακραιώς αλόφιλα, τα Αρχαία άνευ κυτταρικού τοιχώματος και τα ακραιώς θερμοοξειδόφιλα [6]. Η κ. Α. Τζώρτζη αναφέρει στην εργασία της αναφορικά με τις ομάδες των Αρχαίων [31]:

#### *Μεθανογόνα Αρχαία*

Τα μεθανογόνα οφείλουν την ονομασία τους στον μοναδικό τύπο ενεργειακού μεταβολισμού κατά τον οποίο το  $H_2$  χρησιμοποιείται για να ανάγει το  $CO_2$  σε μεθάνιο ( $CH_4$ ). Είναι αυστηρά αναερόβια και δηλητηριάζονται από το οξυγόνο. Ζουν σε βάλτους, όπου άλλα μικρόβια έχουν καταναλώσει όλο το οξυγόνο. Το μεθάνιο εξέρχεται με τη μορφή φυσαλίδων από τα έλη αυτά. Τα μεθανογόνα είναι επίσης σημαντικοί αποικοδομητές φυτικών υπολειμάτων. Τα ακραιώς θερμοφιλα μεθανογόνα, όπως το *Methanothermus fervidus*, είναι ουδετερόφιλα αυτότροφα που αναπτύσσονται σε  $H_2$  και  $CO_2$ . Έχουν απομονωθεί από θερμές πηγές όπου το pH ήταν περίπου 6, αν και το pH ήταν πολύ όξινο στην επιφάνεια. Πολλά μεθανογόνα μπορούν να ανάγουν το θείο σε  $H_2S$ . Γενικά, ο μεταβολισμόςθειούχων συστατικών σε διάφορες μορφές είναι συνήθης στα Αρχαία. Πιστεύεται ότι οι μεθανογόνοι προκαρυώτες ήταν οι οργανισμοί που επικρατούσαν στην αρχέγονη Γη, λόγω των μεγάλων ποσοτήτων διοξειδίου του άνθρακα που υπήρχαν τότε και του ελάχιστου οξυγόνου [6, 7, 11].

Τα κύρια γένη των μεθανογόνων που έχουν προσδιοριστεί είναι αυτά σε σχήμα ράβδου (*Methanobacterium*, *Methanobacillus*) και σε σχήμα σφαιράς (*Methanococcus*, *Methanosarcina*). Ένα από τα βασικότερα χαρακτηριστικά της μεθανογενοῦς ομάδας βακτηρίων είναι ο αρκετά περιορισμένος αριθμός υποστρωμάτων που χρησιμοποιούν ως πηγές ενέργειας για το σχηματισμό μεθανίου. Μέχρι σήμερα, τα υποστρώματα τα οποία είναι γνωστό ότι μπορούν να χρησιμοποιηθούν από τα μεθανογόνα

Αρχαία είναι: το μυρμηκικό οξύ, η μεθυλαμίνη, διοξειδίο του άνθρακα, υδρογόνο, φορμαλδεΐδη, οξικό οξύ, μεθανόλη, μεθυλαμίνες και μονοξειδίο του άνθρακα. Τα πιο σημαντικά βακτήρια της μεθανογενούς ομάδας είναι εκείνα που καταναλώνουν υδρογόνο και οξικό οξύ. Οι κυριότεροι δρόμοι σχηματισμού του μεθανίου είναι: [12, 13, 14, 15]

1. η μετατροπή του υδρογόνου και του διοξειδίου του άνθρακα σε μεθάνιο και νερό
2. η μετατροπή του οξικού οξέος σε μεθάνιο και διοξειδίο του άνθρακα.

Τα μεθανογόνα Αρχαία χρησιμοποιούνται κατεξοχήν στην κατεργασία των λυμάτων.

### *Αναγωγείς θειικών*

Μόνο ένα γένος (*Archaeoglobus*) έχει αναγνωριστεί ως αναγωγέας θειικών. Αυτός ο οργανισμός, ο οποίος απομονώθηκε από θαλάσσιες υδροθερμικές πηγές, είναι ξεχωριστός για την ικανότητά του να χρησιμοποιεί θείο σαν δέκτη ηλεκτρονίων. Κύτταρα από *Archaeoglobus* αναπτύσσονται σε πολύ υψηλές θερμοκρασίες, 92°C, με βέλτιστη θερμοκρασία τους 83°C. Το γένος αυτό παράγει συνένζυμο  $F_{420}$  και μεθανοπερίνη, δύο συνένζυμα που συναντώνται στα μεθανογόνα Αρχαία και σχετίζονται με την παραγωγή εθανίου. Φυλογενετικά, το *Archaeoglobus* βρίσκεται μεταξύ των μεθανογόνων και των ακραίως θερμοοξειδόφιλων Αρχαίων και μπορεί να είναι ένας ενδιάμεσος σύνδεσμος μεταξύ των δύο αυτών ομάδων [6].

### *Ακραίως αλόφιλα*

Οι οργανισμοί αυτοί επιβιώνουν σε περιβάλλοντα όπου οι συγκεντρώσεις του άλατος είναι πολύ υψηλές, περίπου επτά φορές μεγαλύτερες από τη συγκέντρωση άλατος στη θάλασσα. Τα αλόφιλα μπορούν να επιβιώνουν σε τέτοιες συνθήκες διότι τα ένζυμά τους έχουν προσαρμοστεί στο να απαιτούν υψηλές συγκεντρώσεις άλατος για να δράσουν. Τα ριβοσώματα και οι πρωτεΐνες του κυτταρικού τοιχώματος αυτών των Αρχαίων απαιτούν για την σταθεροποίησή τους υψηλές συγκεντρώσεις άλατος [6]. Τα Αρχαία αυτά δεν μπορούν να ζυμώσουν τους υδατάνθρακες και δεν περιέχουν χλωροφύλλη. Γι' αυτό το λόγο, αρχικώς, υποστηρίχθηκε ότι η ενέργειά τους προέρχεται από οξειδωτική φωσφορυλίωση. Η συναρπαστική ανακάλυψη ενός νέου συστήματος φωτοφωσφορυλίωσης προέκυψε από τη μελέτη της κυτταρικής μεμβράνης του *Halobacterium halibium*. Αποικίες αλοφίλων σχηματίζουν ένα ροζ αφρό που το χρώμα του οφείλεται στη χρωστική βακτηριοροδοψίνη. Η βακτηριοροδοψίνη υπάρχει στην πλασματική μεμβράνη, όπου απορροφά φως και χρησιμοποιεί την ενέργεια για να αντλεί ιόντα υδρογόνου έξω από το κύτταρο. Η κλίση ιόντων υδρογόνου οδηγεί τη σύνθεση του ATP. Αυτό είναι ο απλούστερος μηχανισμός φωτοφωσφορυλίωσης και τα αλόφιλα μελετώνται σε συστήματα-μοντέλα για τη μετατροπή της ηλιακής ενέργειας [4, 7].

Περαιτέρω μελέτες έδειξαν ότι το *H. halibium* μπορεί να αναπτυχθεί παρουσία είτε φωτός είτε οξυγόνου αλλά δεν μπορεί να αναπτυχθεί όταν τουλάχιστον το ένα εκ των δύο δεν είναι παρόν. Αυτό έδειξε ότι το *H. halibium* μπορεί και λαμβάνει ενέργεια με τη χρησιμοποίηση δύο συστημάτων, ένα που λειτουργεί παρουσία οξυγόνου (οξειδωτική φωσφορυλίωση) και ένα που λειτουργεί παρουσία φωτός (κάποιου είδους φωτοφωσφορυλίωσης). Το ποσοστό σύνθεσης ATP από το *H. halibium* βρέθηκε να είναι υψηλότερο όταν τα κύτταρα προσλαμβάνουν φως με μήκος κύματος μεταξύ 550 και 600 nm, το οποίο αντιστοιχεί ακριβώς στο φάσμα απορρόφησης της βακτηριοροδοψίνης. Οι ερευνητές υπέθεσαν ότι όπως η χλωροφύλλη ενεργεί ως αντλία πρωτονίων για να δημιουργήσει μία κλίση πρωτονίων στην κυτταροπλασματική μεμβράνη, έτσι και η βακτηριοροδοψίνη δημιουργεί μια κλίση πρωτονίων στην πορφυρή μεμβράνη (Η πορφυρή μεμβράνη εμφανίζεται ως ευδιάκριτα εξαγωνικά δικτυωτά πλέγματα μέσα στην κυτταροπλασματική μεμβράνη). Αυτή η κλίση πρωτονίων μπορεί να οδηγήσει στη σύνθεση ATP ή άλλων διαλυτών μεταφορέων.

Τα ακραίως αλόφιλα Αρχαία μπορούν να χρησιμοποιηθούν στην κατασκευή ταχύτερων και εξυπνότερων Η/Υ. Στο κέντρο μοριακής ηλεκτρονικής του πανεπιστημίου των Συρακουσών, ο Robert Birge, καλλιέργησε *Halobacterium* για πέντε ημέρες σε δεξαμενές πέντε λίτρων και απομόνωσε την βακτηριοροδοψίνη από τα κύτταρα. Ο Birge ανέπτυξε ένα νέο μικροϋπολογιστή από ένα λεπτό στρώμα βακτηριοροδοψίνης. Οι συμβατικοί υπολογιστές αποθηκεύουν πληροφορίες σε λεπτούς σχηματισμούς από πυρίτιο. Οι υπολογιστές επεξεργάζονται τις πληροφορίες αυτές καθώς τα ηλεκτρόνια του ηλεκτρικού ρεύματος ανοιγοκλείνουν διακόπτες πυριτίου, δίνοντας σε κάθε διακόπτη την τιμή μηδέν ή ένα. Τα ηλεκτρόνια που περνούν μέσω ενός διακόπτη αντιπροσωπεύουν το 1 και ο διακόπτης που σταματά τη ροή ηλεκτρονίων αντιπροσωπεύει το 0. Εντούτοις, το πυρίτιο δεν μπορεί να αντιδράσει πολύ γρήγορα στις μεταβολές του ηλεκτρικού ρεύματος και έτσι δεν μπορεί να χρησιμοποιηθεί με ευχέρεια σε εφαρμογές τεχνητής νοημοσύνης ή ρομπότ. Αντίθετα, το τσιπ βακτηριοροδοψίνης είναι σε θέση να αποθηκεύσει περισσότερες πληροφορίες από ένα τσιπ πυριτίου και να επεξεργαστεί τις πληροφορίες γρηγορότερα, όπως και ο ανθρώπινος εγκέφαλος. Το τσιπ βακτηριοροδοψίνης λειτουργεί με φως, που διαθέτει μεγαλύτερη ταχύτητα από αυτήν της ροής ηλεκτρονίων. Το πράσινο φως αναγκάζει την πρωτεΐνη να διπλώσει. Μια διπλωμένη πρωτεΐνη αντιπροσωπεύει το 1, ενώ μια ξετυλιγμένη πρωτεΐνη το 0. Φως από λέιζερ χρησιμοποιείται για "να δει" τη διαμόρφωση της πρωτεΐνης [4].

Μέχρι στιγμής, το πρωτεϊνικό τσιπ πρέπει να αποθηκεύεται σε θερμοκρασία  $-4^{\circ}\text{C}$  ώστε να είναι ικανό να διατηρεί τη δομή του, αλλά ο Birge και οι συνάδελφοί του είναι αισιόδοξοι ότι θα επιλύσουν αυτό το πρόβλημα. Ένας πρωτεϊνικός επεξεργαστής για στρατιωτικά ραντάρ έχει ήδη κατασκευαστεί από Ρώσους επιστήμονες και οι ΗΠΑ χρησιμοποιούν τα πρωτεϊνικά τσιπ στα στρατιωτικά τους

αεροπλάνα. Εάν ένα τέτοιο αεροπλάνο συντριβεί, θα καταστραφεί μαζί του και το σύστημα ψύξης, καταστρέφοντας κατά συνέπεια το τσιπ και προστατεύοντας έτσι τις διαβαθμισμένες πληροφορίες από την υποκλοπή. Τελικά αυτά τα μικρότερα, ταχύτερα και υψηλότερης χωρητικότητας τσιπ θα βοηθήσουν στην ανάπτυξη υπολογιστών που θα εκτελούν πιο πολύπλοκες λειτουργίες παρόμοιες με την ανθρώπινη νοημοσύνη, όπως όραση για ανθρώπους με τύφλωση [4].

### *Άνευ κυτταρικού τοιχώματος Αρχαία*

Το Thermoplasma είναι ένας προκαρυωτικός οργανισμός που στερείται κυτταρικού τοιχώματος και αναπτύσσεται σε σωρούς ανθρακούχων απορριμάτων. Το περιβάλλον αυτό είναι πλούσιο σε FeS, άνθρακα και οργανικά παραπροϊόντα που θερμαίνονται πολύ και οξειδώνονται μέσω αυτοανάφλεξης των οργανικών προϊόντων. Το Thermoplasma, ένα θερμοοξειδόφιλο Αρχαίο που ευδοκίμει στους 55°C και σε pH 2. Μπορεί και ακμάζει σε αυτό το περιβάλλον χρησιμοποιώντας τα οργανικά προϊόντα που παίρνει από το κάρβουνο. Η κυτταρική μεμβράνη του Thermoplasma περιέχει έναν χημικώς μοναδικό λιποπολυσακχαρίτη που αποτελείται από μια τετρααιθερική λιπογλυκάνη η οποία περιέχει μονάδες γλυκόζης και μανόζης. Μαζί με τις γλυκοπρωτεΐνες της μεμβράνης, αυτό και άλλα μόρια της μεμβράνης καθιστούν το Thermoplasma ικανό, χωρίς κυτταρικό τοίχωμα, να αντέχει στο χαμηλό pH και την υψηλή θερμοκρασία του περιβάλλοντος στο οποίο ζει. Φυλογενετικά, το Thermoplasma προσεγγίζει περισσότερο τα μεθανογόνα και ακραίως αλόφιλα αρχαιοβακτήρια παρά τα ακραίως θερμοοξειδόφιλα που μεταβολίζουν θείο [6].

Οι ερευνητές στο Εθνικό Εργαστήριο του Όουκ Ριτζ στο Tennessee ταυτοποίησαν δύο Αρχαία που ζουν κοντά σε υδροθερμικές ωκεάνιες πηγές και που θα μπορούσαν να δώσουν λύση στο πρόβλημα των ανανεώσιμων πηγών ενέργειας. Το Thermoplasma acidophilus και ο Pyrococcus furiosus μπορούν να παράγουν καύσιμα, αέριο υδρογόνο και έναν εξωκυττάριο πολυσακχαρίτη από γλυκόζη. Ένα επιπλέον όφελος είναι ότι ο πολυσακχαρίτης παράγεται σε μεγάλες ποσότητες και μπορεί να βρει βιομηχανικές εφαρμογές, όπως ηχητικό σε τρόφιμα ή άλλα προϊόντα [4].

### *Ακραίως θερμοοξειδόφιλα*

Τα ακραίως θερμοοξειδόφιλα Αρχαία περιλαμβάνουν τρεις τάξεις (Thermococcales, Thermoproteales, Sulfolobales) και τουλάχιστον εννέα γένη. Πολλοί από τους μικροοργανισμούς αυτής της ομάδας είναι οξειδόφιλοι και μεταβολίζουν το θείο. Τα Thermococcus και Pyrococcus, μέλη της τάξης Thermococcales, είναι υποχρεωτικά αναερόβιοι οργανισμοί που αναπτύσσονται κοντά σε υδροθερμικές πηγές στο βάθος της θάλασσας. Και οι δύο είναι χημειοαυτότροφοι οργανισμοί που αναπτύσσονται σε πρωτεΐνες, άμυλο και άλλα οργανικά υλικά. Χρησιμοποιούν το στοιχειακό θείο (S<sub>0</sub>)

σαν δέκτη ηλεκτρονίων, ανάγοντας το σε H<sub>2</sub>S κατά τη διάρκεια της αναερόβιας αναπνοής. Το Thermoproteus έχει παρόμοιες μεταβολικές ιδιότητες με τα Thermococcus και Pyrococcus. Αναπτύσσεται, δηλαδή, σε θερμοκρασίες από 60°C έως 96°C και σε τιμές pH μεταξύ του 1.7 και 6.5. Συναντάται σε πλούσια σε θείο, θερμά, υδατικά περιβάλλοντα όπως οι θερμές πηγές. Το Sulfolobus, μέλος της τάξης Sulfolobales, είναι ένα θερμοοξειδόφιλο που αναπτύσσεται καλύτερα σε θερμοκρασίες από 75°C έως 85°C και τιμές pH μεταξύ 2 και 3. Είναι ένας υποχρεωτικά αερόβιος οργανισμός. Ικανός να οξειδώνει το S<sup>0</sup> ή το H<sub>2</sub>S σε H<sub>2</sub>SO<sub>4</sub> και να ενσωματώνει CO<sub>2</sub> σε οργανικές μορφές του άνθρακα. Συναντάται σε θερμές θειούχες πηγές και σε παρόμοια περιβάλλοντα [6].

Οι DNA πολυμεράσες (ένζυμα που συνθέτουν DNA) που απομονώνονται από αυτά τα δύο Αρχαία χρησιμοποιούνται στην αλυσιδωτή αντίδραση πολυμεράσης (PCR), μια τεχνική που πολλαπλασιάζει και δημιουργεί πολλά αντίγραφα DNA. Στην PCR, το μονόκλωνο DNA προκύπτει με τη θέρμανση του δίκλωνου DNA στους 98°C. Μετά την ψύξη του η πολυμεράση μπορεί να αντιγράψει κάθε μονόκλωνο σκέλος. Η πολυμεράση του DNA από το Thermococcus litoralis, η οποία καλείται VentR, και από τον Pyrococcus, η οποία καλείται Deep VentR, δεν αποδιατάσσονται στους 98°C. Αυτά τα ένζυμα μπορούν να χρησιμοποιηθούν στους αυτόματους θερμοκυκλοποιητές για να επαναλάβουν κύκλους θέρμανσης και ψύξης που επιτρέπουν τη δημιουργία πολλών αντιγράφων DNA. Η VentR και η Deep VentR προσθέτουν βάσεις στο DNA με ρυθμό περίπου 1000 βάσεων ανά λεπτό [4].

**Πίνακας 3:** Αντιπροσωπευτικά είδη Αρχαίων ανά ομάδα.

Ομάδα	Αντιπροσωπευτικά είδη
Μεθανογόνα	Methanobacterium Methanobrevibacter Metanococcus Methanolobus Methanomicrobium Methanosarcina Methanospirillum Methanothermus
Αναγωγείς θειικών	Archaeoglobus
Ακραιώς αλδοφιλα	Halobacterium Halococcus Haloferax Natronobacterium Natronococcus
Άνευ κυτταρικού τοιχώματος Αρχαία	Thermoplasma
Ακραιώς θερμοοξειδόφιλα	Desulfurolobus Pyrococcus Pyrodictium Sulfolobus Thermococcus Thermoproteus

## 1.4 Σορτάσες/ Αρχαιοσορτάσες

### 1.4.1 Σορτάσες

Οι σορτάσες είναι μία ομάδα προκαρυωτικών ενζύμων, τρανσπεπτιδάσες (transpeptidases), οι οποίες τροποποιούν τις επιφανειακές πρωτεΐνες αναγνωρίζοντας και κόβοντας ένα καρβοξυτελικό σήμα διαλογής. Για τα περισσότερα υποστρώματα των πρωτεϊνικών ενζύμων, το σήμα αναγνώρισης περιλαμβάνει το μοτίβο LPXTG (Leu-Pro-any-Thr-Gly), στη συνέχεια μία υδροφοβική διαμεμβρανική ακολουθία και τέλος ακολουθεί ένα σύμπλεγμα αλκαλικών καταλοίπων. Το "κόψιμο" πραγματοποιείται μεταξύ της θρεονίνης (Thr) και της γλυκίνης (Gly) με παροδική προσκόλληση του υπολείμματος της θρεονίνης στην ενεργό θέση του υπολείμματος της κυστεΐνης. Αυτό ακολουθείται από μεταφορά αμινοξέων, μέσω των τρανσπεπτιδασών, τα οποία συνδέουν ομοιοπολικά την πρωτεΐνη με τα συστατικά του κυτταρικού τοιχώματος. Οι σορτάσες απαντώνται κυρίως στα Gram-θετικά βακτήρια και περιστασιακά στα Gram-αρνητικά. Σύμφωνα με την πτυχιακή εργασία της κ. Δ. Φιμερέλη οι σορτάσες μπορούν να χωριστούν με βάση την πρωτοταγή δομή τους και τα υποστρώματα τα οποία κόβουν σε πέντε οικογένειες στο στέλεχος *Staphylococcus aureus*. Οι οικογένειες αυτές είναι οι εξής: [16 - 21]

**Σορτάση A (SrtA):** Οι σορτάσες A έχει βρεθεί ότι κόβουν υποστρώματα που περιέχουν κυρίως το μοτίβο LPxTG. Πολλά δεδομένα δείχνουν ότι οι σορτάσες της οικογένειας A είναι housekeeping genes, δηλαδή παίζουν βασικό ρόλο στη λειτουργία του κυττάρου και ότι πολλές πρωτεΐνες που προσδένονται στο κυτταρικό τοίχωμα είναι υποστρώματα της οικογένειας A. Ένα πλήθος βακτηριακών στελεχών που περιλαμβάνουν τη σορτάση A είναι τα γένη *Bacillus*, *Enterococcus*, *Lactobacillus*, *Lactococcus*, *Listeria*, *Staphylococcus* και *Streptococcus*.

**Σορτάση B (SrtB):** Οι σορτάσες της οικογένειας B έχει βρεθεί ότι κόβουν πρωτεΐνες που περιλαμβάνουν το μοτίβο NrxTN, το οποίο διαφέρει πολύ από το LPxTG. Τα κύρια γένη βακτηρίων που περιέχουν την συγκεκριμένη οικογένεια καθώς και τα υποστρώματα της είναι τα *Staphylococcus aureus*, *Bacillus* και *Listeria*. Οι σορτάσες B φαίνεται να παίζουν ρόλο στον μεταβολισμό του σιδήρου εφόσον περιέχουν το NEAT domain, το οποίο σχετίζεται με την μεταφορά σιδήρου [21].

**Οικογένεια 3:** Αποτελεί τη μεγαλύτερη οικογένεια και φαίνεται να παίζει ένα πιο συγκεκριμένο ρόλο, προσδένοντας λιγότερες πρωτεΐνες από ότι η σορτάση A. Η οικογένεια αυτή έχει βρεθεί ότι προσδένει πρωτεΐνες που έχουν το ίδιο μοτίβο με τις πρωτεΐνες που προσδένει η σορτάση A. Η βασική διαφορά των δύο σορτασών βρίσκεται στο έκτο αμινοξύ του μοτίβου των πρωτεϊνών που κόβουν. Πιο συγκεκριμένα, στη σορτάση A έχει βρεθεί ότι το μοτίβο είναι το LpxTG ενώ στην οικογένεια 3 είναι

κυρίως το LPxTGG. Άλλες δύο διαφορές που κάνουν τις σορτάσες αυτές να ξεχωρίζουν είναι οι εξής: 1) στην οικογένεια 3, η έκφραση των ενζύμων της οικογένειας και των υποστρωμάτων της έχει βρεθεί ότι πιθανόν ρυθμίζεται ταυτόχρονα καθώς τα γονίδια τους είναι διπλανά. 2) Η σορτάση A και η οικογένεια 3 έχει βρεθεί ότι «κάθονται» στην μεμβράνη διαφορετικά και ειδικότερα ότι η σορτάση A είναι τύπου II μεμβρανική πρωτεΐνη ενώ η σορτάση της οικογένειας 3 είναι τύπου I.

**Οικογένεια 4:** Η σορτάση της οικογένειας 4 κόβει πρωτεΐνες που έχουν το μοτίβο το LPxTA, δηλαδή έχει αντικατασταθεί η γλυκίνη στην πέμπτη θέση του μοτίβου με αλανίνη. Εντοπίζεται κυρίως σε βακτήρια του γένους *Bacillus*.

**Οικογένεια 5:** Η σορτάση της οικογένειας 5 έχει σαν υπόστρωμα πρωτεΐνες με το μοτίβο LaxTG, δηλαδή έχει αντικατασταθεί η προλίνη στη δεύτερη θέση του μοτίβου με αλανίνη. Πολλά high-G+C Gram-θετικά βακτήρια έχουν αντικαταστήσει την σορτάση A με την σορτάση της οικογένειας 5. Εντοπίζεται κυρίως σε βακτήρια των γενών *Corynebacterium* και *Streptomyces*.

CWS motif	1	2	3	4	5	6
SrtA	L	P	[K E R V D]	T	G	[E D S]
SrtB	N	P	[Q K]	[T S]	[N G S]	[D A]
Family 3	[L I]	P	[K S E L A D S]	T	G	G
Family 4	L	P	[N D E E]	T	A	[T S]
Family 5	L	A	[E A E E V]	T	G	[A S G V]

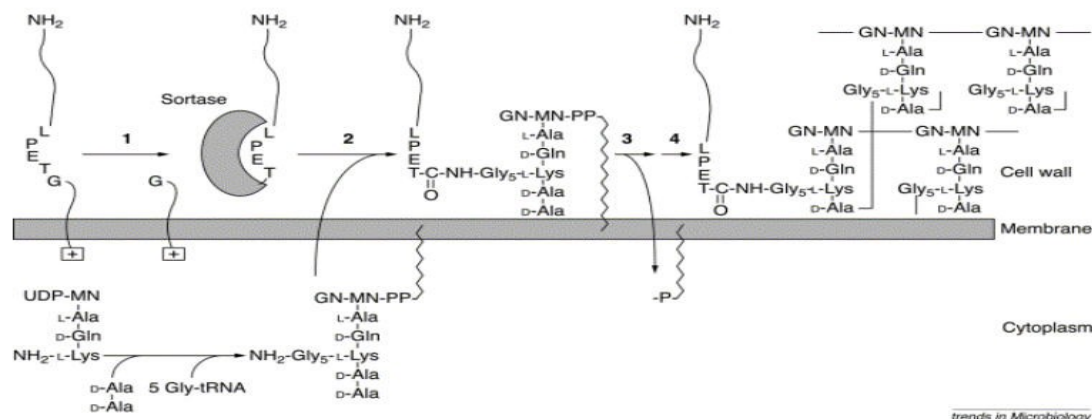
**Εικόνα 1.6:** Κατηγορίες σορτασών ανάλογα με την αμινοξική ακολουθία των πρωτεϊνών που κόβουν [21].

#### 1.4.2 Μηχανισμός προσκόλλησης των LPxTG πρωτεϊνών στο κυτταρικό τοίχωμα

Οι LPxTG και συναφείς πρωτεΐνες προσδένονται ομοιοπολικά στο κυτταρικό τοίχωμα με τη βοήθεια των σορτασών. Η διαδικασία της πρόσδεσης μπορεί να χωριστεί σε τέσσερα βήματα: Αρχικά, η πρωτεΐνη με το πεπτίδιο οδηγητή που διαθέτει, μπαίνει στο εκκριτικό μονοπάτι και στη συνέχεια το πεπτίδιο οδηγητής απομακρύνεται. Στο δεύτερο βήμα, η πρωτεΐνη συγκρατείται στο κυτταρικό τοίχωμα μέσω της υδρόφοβης περιοχής και της φορτισμένης ουράς που υπάρχουν μετά το μοτίβο



LPxTG. Στη συνέχεια, η σορτάση κόβει την πρωτεΐνη στο LPxTG μοτίβο μεταξύ θρεονίνης και γλυκίνης. Στο τέταρτο και τελευταίο βήμα σχηματίζεται αμιδικός δεσμός μεταξύ του καρβοξυτελικού άκρου της θρεονίνης και του αμινοτελικού άκρου της πενταγλυκίνης της πεπτιδογλυκάνης. Έτσι, η επιφανειακή πρωτεΐνη συνδέεται με την πεπτιδογλυκάνη, εισέρχεται στον φάκελο και μένει στην επιφάνεια [21, 22].



**Εικόνα 1.7:** Πρόσδεση πρωτεΐνης στο κυτταρικό τοίχωμα [21].

### 1.4.3 Αρχαιοσορτάσες

Οι Αρχαιοσορτάσες αποτελούν ένα κλασικό είδος διαμεμβρανικής πρωτεΐνης, δηλαδή περιέχουν εφτά αναδιπλώσεις κατά μήκος της μεμβράνης, με ενζυμική δράση. Μεταφέρουν στοχευμένα αμινοτελικά άκρα πρωτεϊνών σε φωσφολιπίδια με σκοπό το σχηματισμό του S-layer. Με βάση το χαρακτηριστικό μοτίβο στο οποίο κόβουν την πρωτεΐνη στόχο, χωρίζονται σε τέσσερις κατηγορίες: [23]

#### Αρχαιοσορτάση A/ PGF-CTERM:

Μία υπο-οικογένεια πρωτεϊνών, παρόμοια με τις εξωσορτάσες, εντοπίζεται σε είκοσι εννέα από τα πενήντα επτά γονιδιώματα του φύλου *Euryarchaeota* και πουθενά αλλού. Εξαιτίας αυτού του χαρακτηριστικού και της μακρινής ομολογίας με τις εξωσορτάσες, η υπο-οικογένεια αυτή θεωρείται ιδιαίτερη και καλείται Αρχαιοσορτάση A (ArtA). Τα αμινοξέα που απαρτίζουν το χαρακτηριστικό μοτίβο, για το σήμα της πρωτεϊνικής διαλογής, είναι τα εξής: προλίνη (Pro) – γλυκίνη (Gly) - φαινυλαλανίνη (Phe). Ως εκ τούτου το μοτίβο χαρακτηρίστηκε ως PGF-CTERM. Το PGF-CTERM μοντέλο εμφανίζεται ακριβώς στα ίδια είκοσι εννέα γονιδιώματα όπως και η Αρχαιοσορτάση A. Το γεγονός αυτό, το οποίο συμβαίνει στα *Halobacteria*, καθώς και οι αντίστοιχες σχέσεις ομολογίας μεταξύ των εξωσορτασών/ αρχαιοσορτασών και των μοτίβων PEP-CTERM/ PGF-CTERM υποδεικνύουν έντονα ότι το ζευγάρι ArtA και PGF-CTERM αντιπροσωπεύει ένα αρχαίο διατηρημένο σύστημα επεξεργασίας των πρωτεϊνών στο



γένος των Αρχαίων.

Οι πρωτεΐνες που περιλαμβάνουν το μοτίβο PGF-CTERM φαίνεται ότι έχουν ενζυμική λειτουργία, συμπεριφέρονται ως γλυκοπρωτεΐνες ή παίζουν δομικό ρόλο στο σχηματισμό του S-layer. Υπάρχουν γλυκοπρωτεΐνες από διάφορα είδη αλοβακτηρίων, οι οποίες γλυκοζυλιώνονται κοντά στο καρβοξυτελικό άκρο τους και σχηματίζουν το S-layer [23].

### **Αρχαιοσορτάση B/ VPxxxP-CTERM:**

Μέσα στην οικογένεια των Αρχαίων ανιχνεύτηκε ένα ακόμη σετ από απομακρυσμένα ομόλογα εξωσορτασών, τα οποία διαφέρουν από την Αρχαιοσορτάση A αλλά ταυτόχρονα είναι στενά συνδεδεμένα μαζί της. Η υπο-οικογένεια αυτή καλείται Αρχαιοσορτάση B (ArtB) κι εντοπίστηκε στους οργανισμούς *Methanohalophilus mahii* και *Methanohalobium evestigatum*. Αποτελεί ένα από τα δύο μεγαλύτερα μέρη της αρχαιοσορτάσης/ εξωσορτάσης. Στην εξωσορτάση αυτή, η Αρχαιοσορτάση A είναι επίσης παρούσα. Σε κάθε γονιδίωμα η πρωτεΐνη που κωδικοποιείται δίπλα στην ArtB εμφανίζει μία καρβοξυτελική περιοχή, η οποία είναι δομικά παρόμοια με το PEP-CTERM και το PGF-CTERM άκρο. Πιο συγκεκριμένα, πρόκειται για μία περιοχή που στο τέλος της αποτελείται από ένα διαμεμβρανικό κομμάτι και ένα σύμπλοκο από βασικά υπολείμματα. Ωστόσο, δεν έχει ταυτοποιηθεί κάποιο άλλο παρόμοιο σήμα (μοτίβο) πρωτεϊνικής διαλογής. Ύστερα από αναζήτηση που πραγματοποιήθηκε, προκειμένου να βρεθούν πρωτεΐνες οι οποίες περιέχουν παρόμοιες περιοχές με καρβοξυτελική αλληλουχία, εντοπίστηκαν πέντε ομόλογες πρωτεΐνες στον οργανισμό *Methanohalophilus mahii* κι εννιά στον *Methanohalobium evestigatum* με το χαρακτηριστικό μοτίβο VPxxxP-CTERM. Μέχρι στιγμής, δεν έχει βρεθεί κανένα άλλο παράδειγμα της υποτειθέμενης περιοχής VPxxxP-CTERM, για πρωτεϊνική διαλογή, σε οποιοδήποτε άλλο πλήρες μικροβιακό γονιδίωμα [23].

### **Αρχαιοσορτάση C/ PEF-CTERM:**

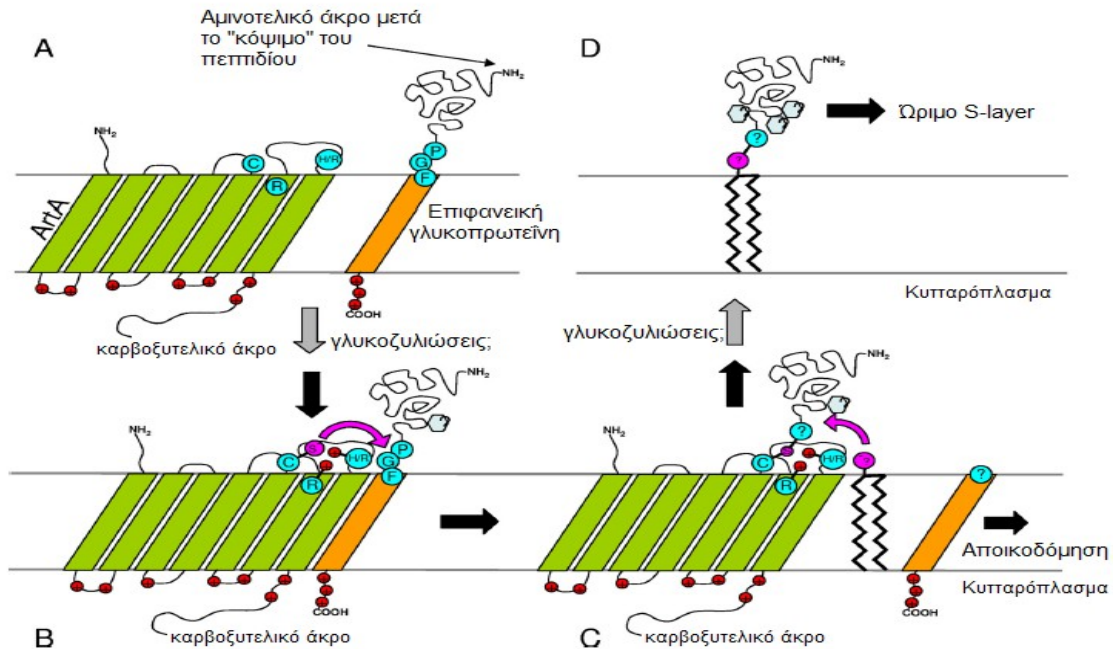
Η αναζήτηση για πρωτεΐνες με καρβοξυτελικά άκρα, τα οποία δεν μοιάζουν ιδιαίτερα με το PEP-CTERM άκρο, κατέδειξαν ένα νέο διαφορετικό, υποτειθέμενο σήμα πρωτεϊνικής διαλογής. Η περιοχή αυτή εντοπίζεται σε πέντε πρωτεΐνες στον οργανισμό *M. Mazei*, σε δέκα στον οργανισμό *Methanococcoides burtonii*, σε δώδεκα στον οργανισμό *Aciduliprofundum boonei* και σε τρεις στον *Ferroglobus placidus*. Το χαρακτηριστικό μοτίβο αποτελείται από βαλίνη (Val) ή ισολευκίνη (Ile) ακολουθούμενη από προλίνη (Pro), γλουταμινικό οξύ (Glu) και φαινυλαλανίνη (Phe) και σε δύο περιπτώσεις μπορεί να υπάρχει λευκίνη (Leu), με αποτέλεσμα η περιοχή αυτή να χαρακτηρίζεται ως PEF-CTERM. Η υπο-οικογένεια που εμφανίζει το παραπάνω χαρακτηριστικό καλείται Αρχαιοσορτάση C (ArtC) [23].

## **Αρχαιοσορτάση D/ PIP-CTERM:**

Μία από τις υπο-οικογένειες των αρχαιοσορτασών, η οποία περιορίζεται σε πέντε είδη του οργανισμού *Methanococci*, ταυτοποιήθηκε και χαρακτηρίστηκε ως Αρχαιοσορτάση D. Το χαρακτηριστικό μοτίβο αναγνώρισης είναι το PIP-CTERM και απαντάται σε μία μόνο πρωτεΐνη για κάθε γονιδίωμα [23].

### **1.4.4 Μηχανισμός προσκόλλησης των Αρχαιοσορτασών στο κυτταρικό τοίχωμα**

Η Αρχαιοσορτάση A (ArtA) πλησιάζει την πρωτεΐνη – στόχο στο καρβοξυτελικό άκρο που φέρει το χαρακτηριστικό μοτίβο PGF. Πριν δράσει η ArtA, στα υποστρώματα έχει προηγηθεί Ο-γλυκοζυλίωση, η περιοχή είναι πλούσια σε θρεονίνη, εξασφαλίζοντας έτσι καλύτερη προσέγγιση της κυστεΐνης του “ενεργού κέντρου” με τον PGF – στόχο της διαμεμβρανικής γλυκοπρωτεΐνης. Η ακριβής θέση στην οποία κόβεται το PGF μοτίβο είναι άγνωστη. Μετά το κόψιμο όλο το αμινοτελικό άκρο του υποστρώματος, συμπεριλαμβανού και του κομμένου χαρακτηριστικού μοτίβου, προσδένεται παροδικά στην κυστεΐνη. Κατόπιν μεταφέρεται στη φωσφορική ομάδα ενός φωσφολιπιδίου (φωσφορικός διφαινυλεστεράς). Ταυτόχρονα το καρβοξυτελικό άκρο του υποστρώματος αποικοδομείται. Τέλος, το φωσφολιπίδιο, το οποίο περιέχει το αμινοτελικό άκρο του υποστρώματος, συμμετέχει στο σχηματισμό του ώριμου S-layer [23]. Όπως βλέπετε στην εικόνα 1.8 στο σχήμα A η ArtA (πράσινο) παρουσιάζεται ως πρωτεΐνη που εκτίνεται κατά μήκος της μεμβράνης διαπερνώντας τη, έχοντας εκθέσει το “ενεργό κέντρο” της κυστεΐνης προς την εξωτερική πλευρά της μεμβράνης. Η πρωτεΐνη – στόχος παρουσιάζεται με το πεπτίδιο – σηματοδότη να έχει ήδη αφαιρεθεί και με το καρβοξυτελικό της ακρο, που περιέχει το χαρακτηριστικό μοτίβο PGF, να διαπερνά την μεμβράνη (πορτοκαλί). Στο σχήμα B παρουσιάζεται η φυσική σύνδεση του PGF καρβοξυτελικού άκρου με τις έλικες της ArtA. Το “ενεργό κέντρο” του υποστρώματος είναι έτοιμο να κόψει τη γλυκοπρωτεΐνη. Σ' αυτό το σημείο μπορεί ήδη να έχουν συμβεί και άλλες μετα-μεταφραστικές διαδικασίες, όπως γλυκοζυλιώσεις. Η θέση στην οποία κόβεται το μοτίβο δεν είναι γνωστή. Το σχήμα C παρουσιάζει την πρωτεΐνη – στόχο μετά το κόψιμο, η οποία προσδένεται παροδικά στο υπόλειμμα της κυστεΐνης. Η απελευθέρωση της γλυκοπρωτεΐνης γίνεται με τη μεταφορά της επί του φωσφορικού διφαινυλεστερά. Ωστόσο, ένα άλλο ενδεχόμενο είναι να γίνεται μεταφορά νερού (υδρόλυση), με το τμήμα του λιπιδίου να προσδένεται αλλού. Στο μεταξύ, το κομμένο πεπτίδιο με το καρβοξυτελικό άκρο οδηγείται για αποικοδόμηση. Το σχήμα D παρουσιάζει τη γλυκοπρωτεΐνη, μετά την αλληλεπίδρασή της με την Αρχαιοσορτάση, η οποία είναι σε θέση να υποβληθεί σε περαιτέρω γλυκοζυλιώσεις μέχρι τελικά να μπορεί να σχηματίσει το ώριμο S-layer.



**Εικόνα 1.8:** Αναπαράσταση του τρόπου δράσης της Αρχαιοσορτάσης A στην κυτταρική επιφάνεια ενός αλοβακτηρίου [25].

## 2. Μεθοδολογία

### 2.1 Συλλογή δεδομένων για την ανεύρεση των υποστρωμάτων των πρωτεϊνικών οικογενειών

Για τη συλλογή των δεδομένων αναφορικά με τις οικογένειες των πρωτεϊνών που περιέχουν το χαρακτηριστικό μοτίβο ανατρέξαμε στη δουλειά του Δρ. D. Haft [23]. Αρχικά, συγκεντρώσαμε όλους τους κωδικούς των πρωτεϊνών που αναφέρονται στην εργασία του. Βρήκαμε συνολικά 55 κωδικούς από πρωτεΐνες, οι οποίες ανήκουν στη βάση δεδομένων του Ινστιτούτου J. Craig Venter (JCVI), την TIGRFAM.

Το Ινστιτούτο J. Craig Venter (JCVI) αποτελεί ένα μη κερδοσκοπικό ίδρυμα γονιδιωματικής, το οποίο ιδρύθηκε τον Οκτώβριο του 2006. Είναι αποτέλεσμα της ένωσης τεσσάρων οργανώσεων: The Institute for Genomic Research (TIGR) and The Center for the Advancement of Genomics (TCAG), The J. Craig Venter Science Foundation, The Joint Technology Center, and the Institute for Biological Energy Alternatives (IBEA). Η έρευνα του Ινστιτούτου περιλαμβάνει τομείς όπως η Γονιδιωματική ιατρική, η Μικροβιακή και Περιβαλλοντική Γονιδιωματική, η Γονιδιωματική των φυτών, οι Μολυσματικές ασθένειες, η Συνθετική Βιολογία και Βιοενέργεια καθώς και η Βιοπληροφορική.

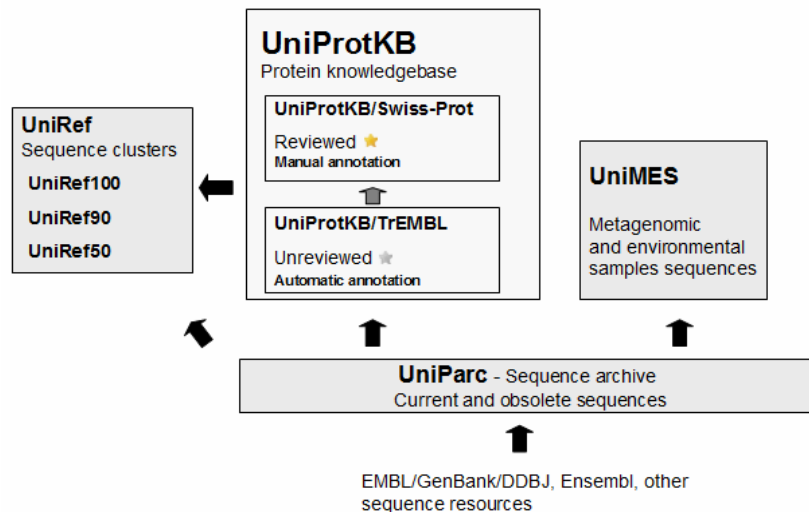
**Εικόνα 2.1:** Η αρχική σελίδα της επίσης ιστοσελίδας του Ινστιτούτου (<http://www.jcvi.org>) [24].

Στην ιστοσελίδα του JCVI περιλαμβάνεται η βάση δεδομένων TIGRFAM, η οποία είναι μέλος της InterPro. Η TIGRFAM, περιέχει πολλαπλές ευθυγραμμίσεις ακολουθιών, Hidden Markov Models (HMMs) για την ταξινόμηση της ακολουθίας των πρωτεϊνών και πληροφορίες σχετικά με τις πρωτεΐνες. Οι βιβλιοθήκες των HMM και λειπά σχετικά αρχεία με πληροφορίες είναι ελεύθερα διαθέσιμα για όλους τους χρήστες. Η συγκεκριμένη βάση δεδομένων περιλαμβάνει περίπου 4.488 οικογένειες πρωτεϊνών. Όλες οι σχετικές πληροφορίες διατίθενται στην ιστοσελίδα: <ftp://ftp.jcvi.org/pub/data/TIGRFAMs/>.

Από τα αποτελέσματα που πήραμε μόνο 34 είχαν σχέση με τις Αρχαιοσορτάσες και από αυτές 6 συνολικά αναφέρονταν σε πρωτεϊνικά υποστρώματα, τα οποία και αναζητούσαμε. Στη συνέχεια, συγκεντρώσαμε τους κωδικούς κάθε πρωτεΐνης (AC) και τα χρησιμοποιήσαμε ως είσοδο με σκοπό να τις αναζητήσουμε και να ανακτήσουμε τις αντίστοιχες εγγραφές στη UniProt καθώς και να κατεβάσουμε τα αντίστοιχα αρχεία [24].

Η UniProt (Universal Protein Resource) αποτελεί μια ολοκληρωμένη, ελεύθερα προσβάσιμη σε όλους, βάση δεδομένων σχετικά με την αλληλουχία των πρωτεϊνών. Παρέχει δυνατότητα περιήγησης από εγγραφή σε εγγραφή, με βάση οποιοδήποτε σχεδόν χαρακτηριστικό ή κριτήριο και περιέχει μεγάλη ποσότητα πληροφοριών σχετικά με τη βιολογική λειτουργία των πρωτεϊνών. Προήλθε από τη

συνεργασία των European Bioinformatics Institute (EBI), Swiss Institute of Bioinformatics (SIB) και Protein Information Resource (PIR). Επιπλέον, η UniProt συνδέεται και με τις παρακάτω βάσεις δεδομένων: UniProt Knowledgebase (UniProtKB), UniProt Reference Clusters (UniRef), και UniProt Archive (UniParc).



**Εικόνα 2.2:** Διάρθρωση της βάσης δεδομένων UniProt [21].

Η UniProt Knowledgebase (UniProtKB) αποτελεί το κέντρο των δραστηριοτήτων της UniProt και συνεχίζει την δουλειά της Swiss-Prot, TrEMBL, και PIR-PSD. Αποτελείται από δύο τμήματα, την UniProtKB/Swiss-Prot και την UniProtKB/TrEMBL. Η UniProtKB/Swiss-Prot περιέχει καλά σχολιασμένες ακολουθίες ενώ η UniProtKB/TrEMBL περιέχει ακολουθίες όπου έχει γίνει αυτοματοποιημένος σχολιασμός.

Για τις απαιτήσεις της εργασίας μας επιλέξαμε να πάρουμε ως αποτέλεσμα τα αρχεία σε μορφή FASTA, Mapping Table και UniProtKB. Για όσες οικογένειες δε βρέθηκαν αποτελέσματα μετά την αναζήτηση, κάναμε BLAST χρησιμοποιώντας και πάλι τους κωδικούς των πρωτεϊνών. Επιπλέον, μέσω της αναζήτησης με BLAST καταφέραμε να αποκτήσουμε ολόκληρη την αλληλουχία των αμινοξέων, τα οποία συνθέτουν την πρωτεΐνη, επειδή στα ευρήματα από τη βάση δεδομένων του JCVI περιλαμβάνονταν μόνο συγκεκριμένα τμήματα κι όχι η πρωτεΐνη στο σύνολό της.

Το BLAST (Basic Local Alignment Search Tool) αποτελεί ίσως τη σημαντικότερη δημοσίευση όλων των εποχών στον τομέα της πολλαπλής στοίχισης των αλληλουχιών. Βασίστηκε πάνω στις ανακαλύψεις για τη στατιστική κατανομή του σκορ (score) της τοπικής στοίχισης και συνέβαλε καθοριστικά στον τρόπο που διεξάγεται η αναζήτηση ομοιότητας σε βάσεις δεδομένων και η πολλαπλή στοίχιση. Αυτό συνέβη διότι ήταν πιο γρήγορος από κάθε άλλο αλγόριθμο επιτρέποντας έτσι ταχείες αναζητήσεις. Επιπλέον, έδινε για πρώτη φορά μια εκτίμηση για τη στατιστική σημαντικότητα των στοίχισεων. Οι αρχικές εκδόσεις του BLAST δεν επέτρεπαν την εισαγωγή κενών και έτσι ο αλγόριθμος



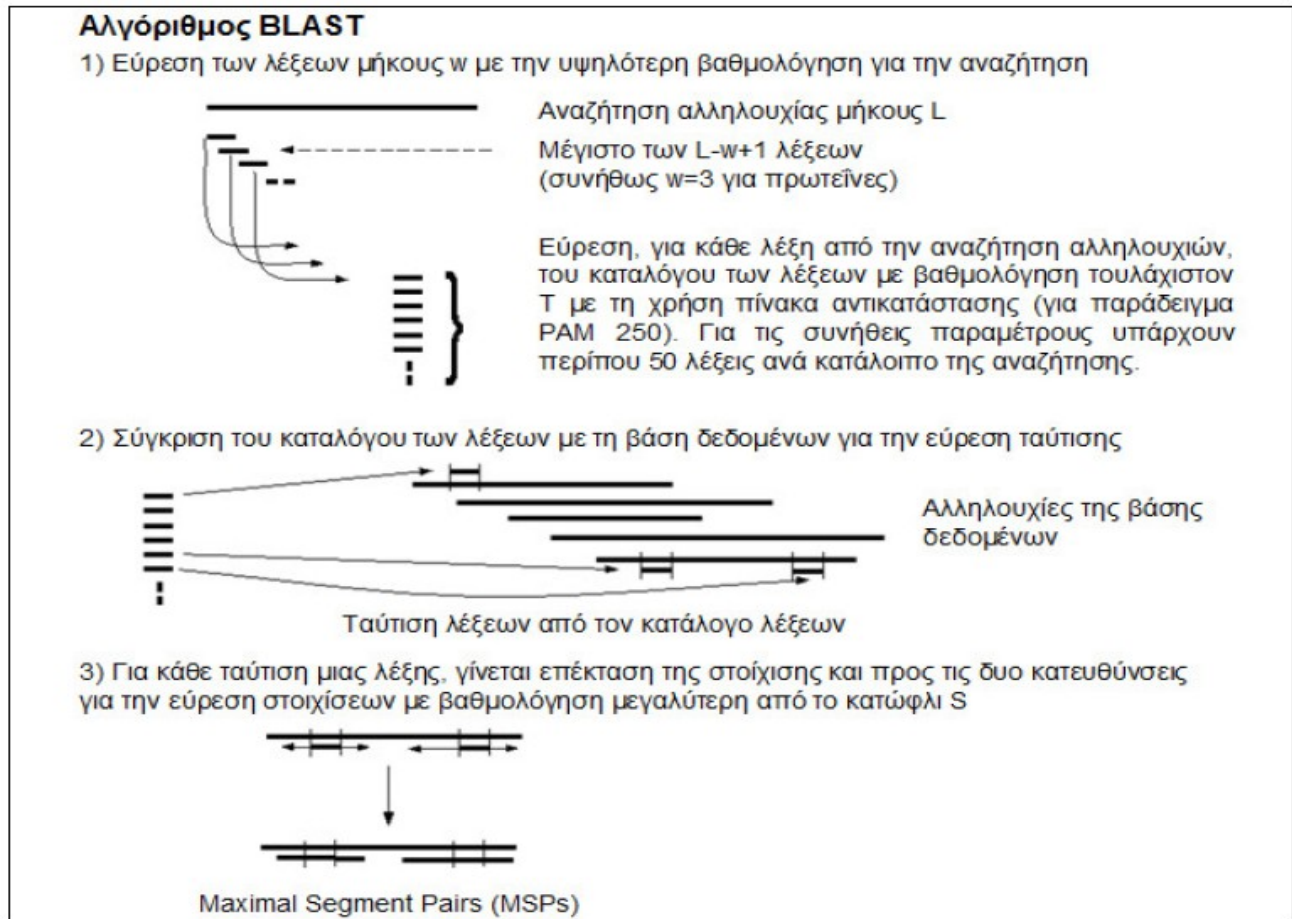
αποτελούσε ένα απλά εύχρηστο και γρήγορο εργαλείο το οποίο εντόπιζε όμοιες αλληλουχίες. Ωστόσο, από τη 2η έκδοση του προγράμματος και μετά προστέθηκε και η δυνατότητα εισαγωγής κενών με συνέπεια το BLAST να μπορεί να χρησιμοποιηθεί και σαν γενικό πρόγραμμα στοίχισης. Τέλος, οι νεότερες εκδόσεις του BLAST περιλαμβάνουν πολλές τροποποιήσεις που επιτρέπουν πιο ακριβείς υπολογισμούς με χρήση προφίλ και ειδικών ανά θέση πινάκων ομοιότητας (PSI-BLAST) [25, 26].

Γενικά η διαδικασία που ακολουθεί ο αλγόριθμος του BLAST είναι γρήγορη επειδή πολλοί παράμετροι είναι προϋπολογισμένες κι επιπλέον αποφεύγει να στοιχίσει αλληλουχίες της βάσης δεδομένων οι οποίες κρίνει ότι δεν έχουν σημαντική ομοιότητα. Η διαδικασία της σύγκρισης περιλαμβάνει τα εξής βήματα:

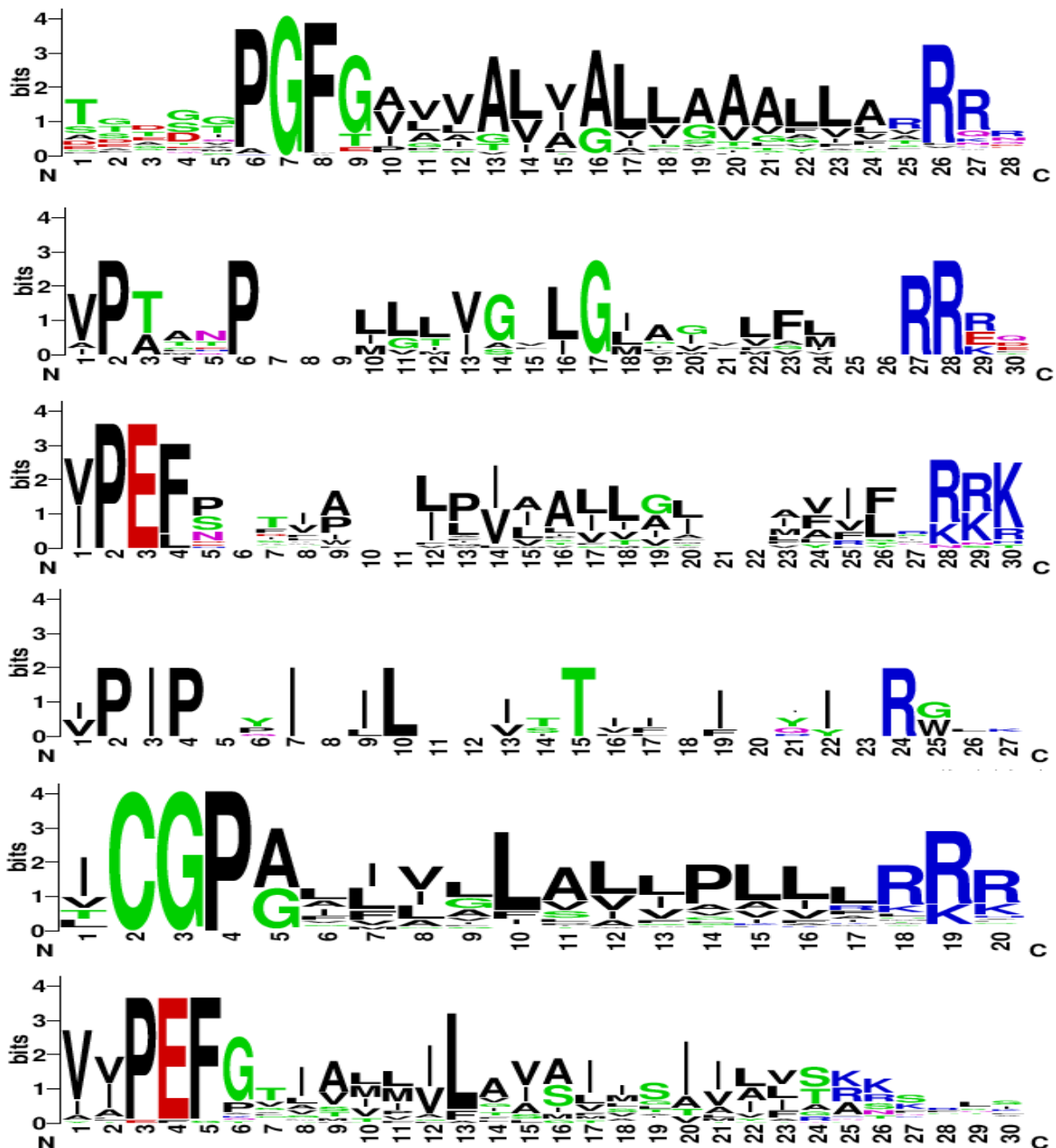
- Αρχικά, δημιουργείται ένας κατάλογος που περιλαμβάνει το σύνολο των λέξεων που θα ταίριαζαν με κάποια λέξη της άγνωστης αλληλουχίας και ξεπερνούν την τιμή κατωφλίου (προκαθορισμένη τιμή για πρωτεϊνικές αλληλουχίες  $T=13$ ).
- Στη συνέχεια, ο αλγόριθμος αναζητά αυτές τις λέξεις στις αλληλουχίες της βάσης δεδομένων και όταν εντοπίσει μια τέτοια λέξη, ξεκινάει μια διαδικασία επέκτασης του 'ευρήματος' προς τις δύο κατευθύνσεις, καθώς η βαθμολογία συνεχίζει και αυξάνει.
- Στο στάδιο αυτό οι περιοχές που εντοπίζονται με τη μέγιστη βαθμολογία είναι οι υποψήφιες περιοχές ομοιότητας (HSPs, high scoring pairs).
- Από όλα τα HSPs που εντοπίστηκαν καταγράφονται στα αποτελέσματα εκείνες οι περιοχές όπου η βαθμολογία ξεπερνά μια δεύτερη τιμή κατωφλίου  $S$ .
- Τελικά, αναφέρονται μόνο οι τοπικές ομοιότητες που εμφανίζουν υψηλή στατιστική σημαντικότητα [25].

Εφόσον συγκεντρώσαμε όλες τις απαραίτητες πληροφορίες κατασκευάσαμε τα λογότυπα που αντιστοιχούν σε κάθε πρωτεϊνική οικογένεια (sequence logos) με τη βοήθεια του εργαλείου weblogo. Πιο συγκεκριμένα, πρόκειται για μία διαδικτυακή εφαρμογή η οποία έχει σχεδιαστεί με σκοπό να κάνει τη δημιουργία των λογότυπων, τα οποία αντιστοιχούν σε συγκεκριμένες ακολουθίες, μία απλή και γρήγορη διαδικασία. Τα λογότυπα αυτά ουσιαστικά αποτελούν μία γραφική αναπαράσταση των πολλαπλών ακολουθιών των αμινοξέων ή των νουκλεϊκών οξέων. Κάθε λογότυπο αποτελείται από μία στοίβα συμβόλων. Μέσα στην ακολουθία υπάρχει ξεχωριστή στοίβα για κάθε σύμβολο. Το συνολικό ύψος της ακολουθίας υποδεικνύει το πόσο συντηρητική είναι η ακολουθία σε εκείνο το σημείο, ενώ το ύψος κάθε συμβόλου μέσα στη στοίβα δείχνει τη σχετική συχνότητα εμφάνισης κάθε αμινοξέος ή νουκλεϊκού οξέος στη συγκεκριμένη θέση. Εμείς, λοιπόν, συγκεντρώσαμε για κάθε πρωτεΐνη, η οποία είναι μέλος της εκάστοτε οικογένειας, το χαρακτηριστικό της μοτίβο (pattern) και τα δώσαμε ως είσοδο στο πρόγραμμα. Τότε αυτό μας επέστρεψε το λογότυπο στο οποίο αποτυπώνεται η ακολουθία

των αμινοξέων που απαρτίζουν το χαρακτηριστικό μοτίβο της κάθε οικογένειας. Έτσι, μπορούμε εύκολα να κατατάξουμε την κάθε οικογένεια σε ένα από τα διαφορετικά είδη των Αρχαιοσορτασών.



**Εικόνα 2.3:** Διαγραμματική απεικόνιση του αλγόριθμου BLAST [25].



**Εικόνα 2.4:** Λογότυπα για πρωτεϊνικά σήματα διαλογής ακολουθιών με καρβοξυτελικά άκρα. Τα υποθετικά σήματα διαλογής που παρουσιάζονται είναι τα εξής: σχήμα 1. PGF-CTERM (TIGR04126) ακολουθία που αντιστοιχεί στην Αρχαιοσορτάση A. Σχήμα 2. VPXXXP-CTERM (TIGR04143) ακολουθία που αντιστοιχεί στην Αρχαιοσορτάση B. Σχήμα 3. PEF-CTERM (TIGR03024) ακολουθία που αντιστοιχεί στην Αρχαιοσορτάση C. Σχήμα 4. PIP-CTERM (TIGR04173) ακολουθία που αντιστοιχεί στην Αρχαιοσορτάση D. Σχήμα 5. CGP-CTERM (TIGR04173) δεν είναι γνωστό σε ποιο είδος Αρχαιοσορτάσης ανήκει αλλά εντοπίζεται μόνο στον οργανισμό *Thermococcaceae*. Σχήμα 6. PEF-CTERM (TIGR04296) ακολουθία που αντιστοιχεί στην Αρχαιοσορτάση C [28].



## 2.2 Κατασκευή του προφίλ Hidden Markov Model

### 2.2.1 Μαρκοβιανά Μοντέλα (Hidden Markov Model)

Τα Hidden Markov Models είναι κατά βάση στοχαστικά μοντέλα, δηλαδή έχουν πιθανοθεωρητική ερμηνεία, τα οποία χρησιμοποιούνται για να περιγράψουν την εξέλιξη παρατηρήσιμων γεγονότων. Τα γεγονότα αυτά εξαρτώνται από ενδογενείς παράγοντες που δεν είναι άμεσα ορατοί. Αποτελούνται από ένα σύνολο κρυφών καταστάσεων, ένα σύνολο παρατηρούμενων συμβόλων και δυο σύνολα πιθανοτήτων, τις πιθανότητες μετάβασης και τις πιθανότητες εκπομπής ή εμφάνισης συμβόλων (emissions). Στα απλούστερα Μαρκοβιανά μοντέλα, όπως οι αλυσίδες Markov, οι καταστάσεις είναι άμεσα ορατές από τον παρατηρητή και ως εκ τούτου οι πιθανότητες μετάβασης των καταστάσεων αποτελούν τις μοναδικές παραμέτρους. Αντίθετα, σε ένα HMM οι καταστάσεις δεν είναι ορατές παρά μόνο οι έξοδοι, οι οποίες όμως εξαρτώνται από τις καταστάσεις [25].

Θεωρώντας μια πρωτεϊνική αλληλουχία  $x$  μήκους  $L$  καταλοίπων:

$$x = x_1, x_2, \dots, x_{L-1}, x_L$$

όπου με  $x_i$  συμβολίζουμε τις παρατηρήσεις που αποτελούνται από ένα εκ των 20 αμινοξέων, στο HMM οι παρατηρήσεις πλέον αποδεδμεύονται από τις καταστάσεις. Συνήθως η αλληλουχία των καταστάσεων συμβολίζεται ως μια συγκεκριμένη θέση  $i$  στην αλληλουχία, με  $\pi_i$ , και ονομάζεται «μονοπάτι» (path). Έτσι, οι δυο καταστάσεις  $k, l$  συνδέονται μέσω των πιθανοτήτων μετάβασης  $a_{kl}$ , σχηματίζοντας μια αλυσίδα Markov 1<sup>ης</sup> τάξης. Ο τυπικός ορισμός αυτών των πιθανοτήτων μεταβάσεως, είναι ο εξής:

$$a_{kl} = P(\langle \pi_i = l | \pi_{i-1} = k \rangle)$$

και συμβολίζει πολύ απλά την πιθανότητα, η κατάσταση  $k$  να δώσει μετάβαση (να προηγηθεί δηλαδή) προς την κατάσταση  $l$ . Αντίστοιχα με το απλό μοντέλο Markov, ορίζονται κι εδώ ειδικές καταστάσεις έναρξης και τερματισμού της αλληλουχίας, οι οποίες για συντομία ονομάζονται B (Begin)

$$a_{Bk} = P(\langle \pi_1 = k | B \rangle)$$

και E (End) αντίστοιχα

$$a_{kE} = P(\langle E | \pi_i = k \rangle)$$

Η σύνδεση των παρατηρηθέντων συμβόλων με τις καταστάσεις, γίνεται μέσω των πιθανοτήτων εμφάνισης συμβόλων:

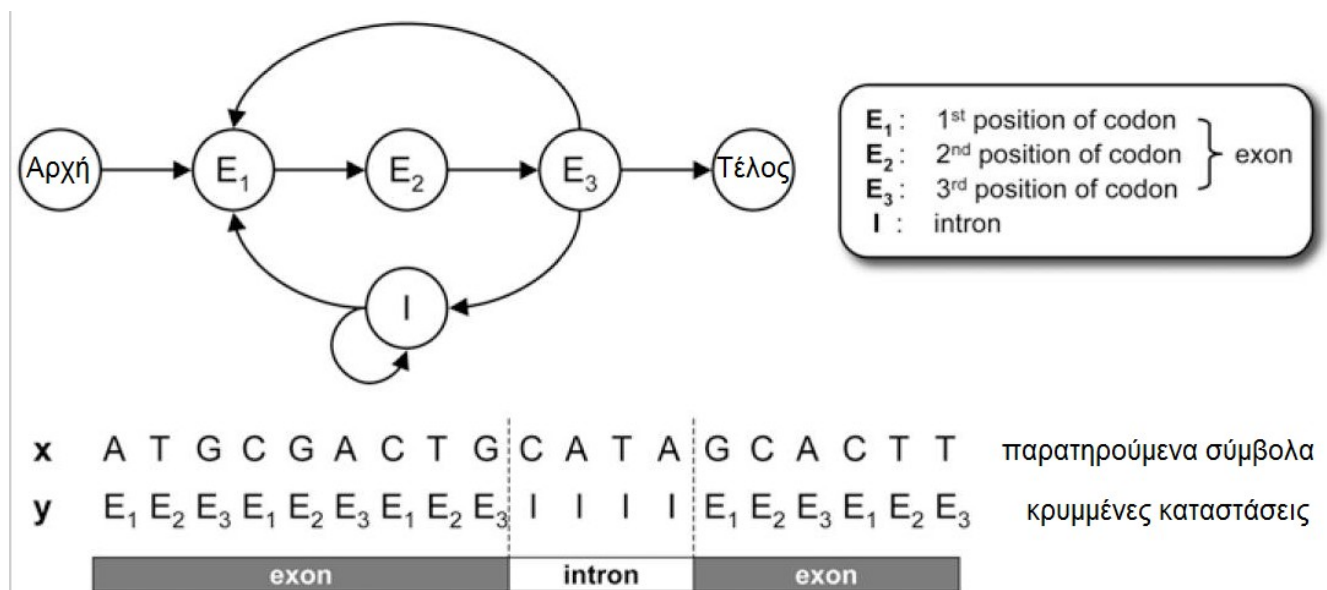
$$e_{kb} = P(\langle x_i = b | \pi_i = k \rangle)$$

οι οποίες, δηλώνουν την πιθανότητα εμφάνισης στη θέση  $i$  της αλληλουχίας, ενός συγκεκριμένου

συμβόλου  $b$ , δεδομένου ότι το σύστημα βρίσκεται στην κατάσταση  $k$ . Η από κοινού πιθανότητα μιας αλληλουχίας  $x$  και του μονοπατιού  $\pi$ , υπολογίζεται ως εξής: [25]

$$P(x, \pi) = P(x_L, x_{L-1}, \dots, x_1, \pi) = \alpha_{\beta\pi_1} \prod_{i=1}^L e_{\pi_i}(x_i) a_{\pi_i, \pi_{i+1}}$$

Τα Hidden Markov Models είναι πολύ γνωστά για την αποτελεσματικότητά τους στη μοντελοποίηση των συσχετίσεων μεταξύ γειτονικών συμβόλων κι έχουν χρησιμοποιηθεί εκτενώς σε διάφορους τομείς, κυρίως στην αναγνώριση ομιλίας και τις ψηφιακές τηλεπικοινωνίες, βρίσκουν διάφορες εφαρμογές και στη βιοπληροφορική. Ένα μεγάλο εύρος προβλημάτων σχετικά με την ανάλυση βιολογικών ακολουθιών έχει επωφεληθεί από αυτά. Για παράδειγμα: Τα HMMs και οι παραλλαγές τους έχουν χρησιμοποιηθεί για την πρόβλεψη γονιδίων, την ανά ζεύγη και πολλαπλή στοίχιση ακολουθιών, τη μοντελοποίηση των σφαλμάτων της ακολουθίας του DNA, την πρόβλεψη της δευτεροταγούς δομής των πρωτεϊνών, την πρόβλεψη του ncRNA, τη δομική ευθυγράμμιση του DNA, την επιτάχυνση αναδίπλωσης και ευθυγράμμισης του RNA, τον εντοπισμό των διαμεμβρανικών τμημάτων των πρωτεϊνών κλπ [29].

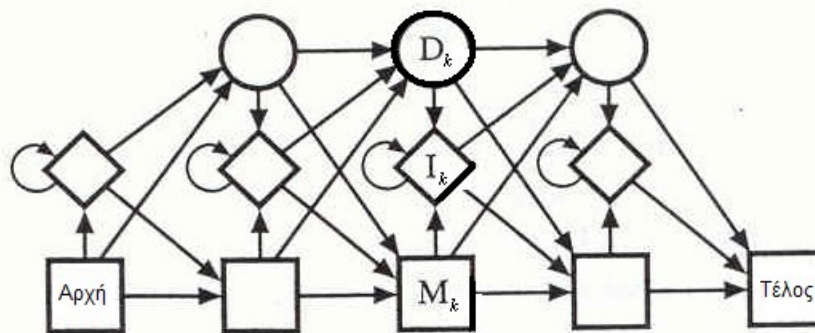


**Εικόνα 2.5:** Ένα απλό HMM για την μοντελοποίηση ευκαρυωτικών γονιδίων [29].

## 2.2.2 Profiles Hidden Markov Models:

Τα Profiles Hidden Markov Models (pHMM) αποτελούν μία ειδική κατηγορία του HMM, τα οποία εστιάζουν στην περιγραφή της πολλαπλής στοίχισης των αλληλουχιών. Σε κάθε κατάσταση περιγράφουν μία συγκεκριμένη θέση στην πολλαπλή στοίχιση. Αυτό έχει ως αποτέλεσμα η κατεύθυνση των μεταβάσεων να είναι μονόδρομη και το μοντέλο να χαρακτηρίζεται ως *left to right*. Αντίθετα, το κλασικό HMM είναι κυκλικό, δηλαδή το μοντέλο μπορεί να επισκεφθεί μία κατάσταση περισσότερες από μία φορές. Επιπλέον, μία πολύ σημαντική ιδιότητα του pHMM είναι ότι οι πιθανότητες εισαγωγής κενού και απαλοιφής μοντελοποιούνται με πιθανοθεωρητικό τρόπο λύνοντας έτσι το πρόβλημα επιλογής της ποινής για τα κενά.

Μία ακόμη σημαντική διαφορά του pHMM από το κλασικό μοντέλο αποτελεί η ύπαρξη σιωπηρών καταστάσεων. Πρόκειται για μία ειδική κατηγορία καταστάσεων στην οποία δεν εκπέμπεται κανένα σύμβολο (silent state) και χρησιμοποιούνται με σκοπό να περιγράψουν τη μετάβαση από ένα κατάλοιπο σε κάποιο άλλο, πιο απομακρυσμένο, χωρίς να χρειάζεται να γίνουν πολλαπλές μεταβάσεις. Μ'αυτό τον τρόπο μειώνεται και ο αριθμός των παραμέτρων του μοντέλου.



**Εικόνα 2.6:** Σχηματική αναπαράσταση ενός τυπικού profile Hidden Markov Model [25].

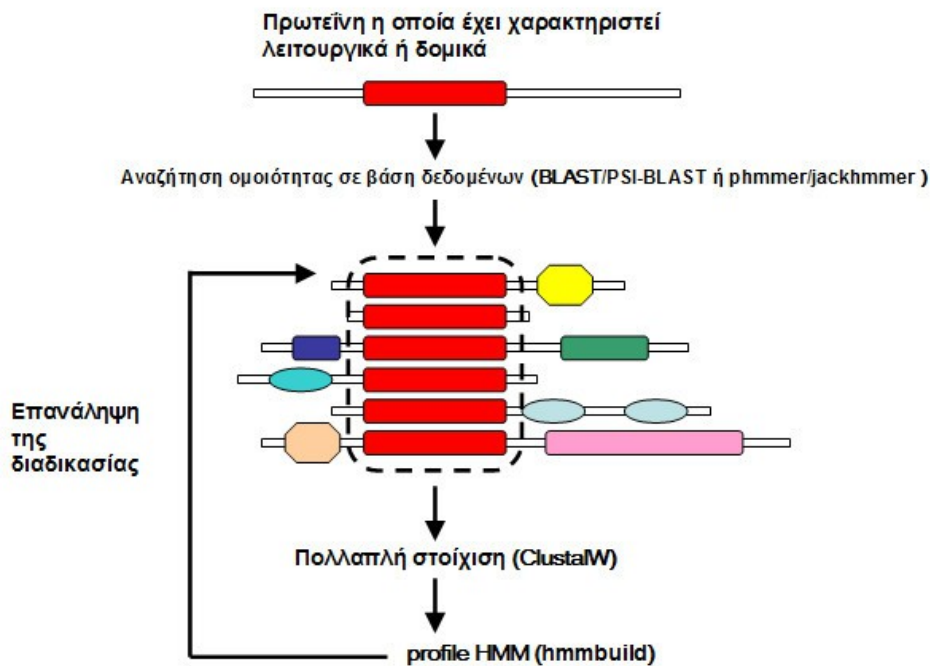
Οι καταστάσεις που παρατηρούνται σε ένα pHMM, πλην των καταστάσεων εκκίνησης και τερματισμού, είναι οι εξής:

- Καταστάσεις Ταύτισης (Match states)  $M_k$  τετράγωνα
- Καταστάσεις Εισαγωγής (Insertion states)  $I_k$  ρόμβοι
- Καταστάσεις Απαλοιφής (Deletion states)  $D_k$  κύκλοι

και συνδέονται με τις αντίστοιχες πιθανότητες μεταβάσεως, οι οποίες αναπαρίστανται με βέλη. Με παρόμοιο τρόπο ορίζονται οι πιθανότητες εμφάνισης συμβόλων, οι οποίες γεννούν τα σύμβολα σε κάθε κατάσταση. Έτσι υπάρχει και σ'αυτό το μοντέλο μια αλληλουχία καταστάσεων η οποία είναι

κρυφή και μία αλληλουχία συμβόλων που είναι φανερή και θεωρούμε ότι προκύπτει από την αλληλουχία των καταστάσεων. Επίσης, οι καταστάσεις ταύτισης αναφέρονται σε περιοχές που έχουν ομοιότητα, δηλαδή στοιχίζονται καλά, ενώ οι καταστάσεις εισαγωγής αντιστοιχούν σε περιοχές που δεν στοιχίζονται καλά κι έτσι έχουμε εισαγωγή χαρακτήρων. Οι περιοχές αυτές, οι οποίες δεν υπάρχουν στις υπόλοιπες αλληλουχίες, εμφανίζονται ως κενά τα οποία μοντελοποιούνται μέσω των σιωπηρών καταστάσεων απαλοιφής.

Η χρησιμότητα των ρHMM, ξεκινάει από τη δημιουργία πολλαπλών στοιχίσεων και φτάνει στη δημιουργία μοντέλων μέσω των οποίων μπορεί να γίνει ευαίσθητη αναζήτηση απομακρυσμένων ομολογιών. Με βάση το ρHMM και τη δημιουργία όλο και περισσότερων μοντέλων για την ταξινόμηση πρωτεϊνικών οικογενειών, έχουν κατασκευαστεί ειδικές βάσεις δεδομένων όπως η PFAM, με την οποία συντελείται ταξινόμηση μεγάλου μέρους της άγνωστης λειτουργίας των πρωτεϊνών που προσδιορίζονται καθημερινά, π.χ. με την αποκωδικοποίηση των γονιδιωμάτων και η TIGRFAM [25].



**Εικόνα 2.7:** Σχηματική αναπαράσταση της διαδικασίας χαρακτηρισμού μιας νέας πρωτεϊνικής οικογένειας [25].

## Το πακέτο λογισμικού HMMER

Το πιο δημοφιλές πακέτο λογισμικού για την κατασκευή Profile Hidden Markov Models είναι το HMMER. Ουσιαστικά πρόκειται για μία συλλογή προγραμμάτων, τα οποία διανέμονται ελεύθερα και παρέχουν δυνατότητα πρόσβασης στον πηγαίο κώδικα. Το πακέτο εφαρμόζεται σε μεγάλο εύρος βιολογικών δεδομένων.

Σύμφωνα με το βιβλίο "Βιοπληροφορική", η τρίτη έκδοση του HMMER περιλαμβάνει, μεταξύ άλλων, τα εξής προγράμματα:

- **hmmbuild:** Πρόγραμμα με το οποίο, ξεκινώντας από μια αρχική πολλαπλή στοίχιση, κατασκευάζεται ένα μοντέλο HMM το οποίο να την περιγράφει.
- **hmmalign:** Πρόγραμμα με το οποίο μια σειρά αλληλουχιών, οι οποίες προέρχονται από ένα HMM, στοιχίζονται σε μια πολλαπλή στοίχιση. Αυτό επιτυγχάνεται μέσω διαδοχικών στοιχίσεων των αλληλουχιών με το μοντέλο.
- **hmmsearch:** Πρόγραμμα το οποίο πραγματοποιεί αναζητήσεις ενός μοντέλου HMM έναντι μιας βάσης αλληλουχιών πρωτεϊνών.
- **phmmmer:** Πρόγραμμα το οποίο πραγματοποιεί αναζήτηση μια πρωτεϊνικής αλληλουχίας έναντι μιας βάσης δεδομένων πρωτεϊνών (ανάλογο με το BLASTP).
- **jackhmmmer:** Πρόγραμμα το οποίο πραγματοποιεί επαναληπτικές αναζητήσεις μια πρωτεϊνικής αλληλουχίας έναντι μιας βάσης δεδομένων πρωτεϊνών (ανάλογο με το PSIBLAST).
- **hmmsearch:** Πρόγραμμα με το οποίο πραγματοποιούνται αναζητήσεις μίας ή περισσότερων αλληλουχιών έναντι μιας βάσης δεδομένων από μοντέλα HMM. Σημαντικό είναι ότι αν έχουμε μια αλληλουχία και ένα HMM, τα δύο παραπάνω προγράμματα επιστρέφουν ακριβώς το ίδιο αποτέλεσμα. Αν διαφέρουν, είτε οι αλληλουχίες είτε τα μοντέλα, τότε δίνουν άλλο αποτέλεσμα λόγω του διαφορετικού τρόπου υπολογισμού της στατιστικής σημαντικότητας.
- **nhmmmer:** Πρόγραμμα που πραγματοποιεί αναζήτηση μιας αλληλουχίας DNA, μίας στοίχισης ή ενός pHMM, έναντι μιας βάσης αλληλουχιών DNA (ανάλογο με το BLASTN).
- **nhmmsearch:** Πρόγραμμα που πραγματοποιεί αναζήτηση μιας αλληλουχίας DNA έναντι μιας βάσης δεδομένων από DNA profile HMM.
- **hmmconvert:** Πρόγραμμα που μετατρέπει μοντέλα HMM από και προς τη μορφή του HMMER3.
- **hmmemit:** Πρόγραμμα με το οποίο 'εκπέμπεται' η καλύτερη, ανάλογα με τον ορισμό, αλληλουχία η οποία θα μπορούσε να παραχθεί από το μοντέλο.
- **hmmcompress:** Μετατρέπει μια βάση δεδομένων HMM σε δυαδικό κώδικα για το hmmsearch.
- **hmmstat:** Δείχνει συνοπτικά στατιστικά για μια βάση δεδομένων HMM.

Το παραπάνω μοντέλο διαφέρει από την τυπική έκδοση του Profile Hidden Markov Model επειδή δεν επιτρέπει να γίνονται μεταβάσεις από μία κατάσταση εισαγωγής κενού (I) σε κατάσταση απαλοιφής (D) και το αντίστροφο. Το πακέτο λογισμικού HMMER3 είναι διαθέσιμο στην ηλεκτρονική διεύθυνση: <http://hmmmer.org/download.html/> [25].

Για να κατασκευάσουμε το profile HMM, αρχικά προμηθευτήκαμε το πακέτο λογισμικού HMMER από την ιστοσελίδα <http://eddylab.org/>. Στη συνέχεια για κάθε μία από τις 6 οικογένειες πρωτεϊνών

τρέξαμε το πρόγραμμα hmmbuild ώστε να δημιουργήσουμε 6 profile HMM, τα οποία περιγράφουν τις πολλαπλές στοιχίσεις για κάθε οικογένεια. Τα αρχεία που πήραμε ως αποτέλεσμα τα ενώσαμε και φτιάξαμε μία βιβλιοθήκη που περιέχει όλες τις σχετικές πληροφορίες, όπως το όνομα της κάθε οικογένειας, το σύνολο των αμινοξέων που απαρτίζουν την αλληλουχία κλπ. Επίσης, από την ιστοσελίδα:

[ftp://ftp.uniprot.org/pub/databases/uniprot/current\\_release/knowledgebase/reference\\_proteomes/Archaea/](ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/reference_proteomes/Archaea/) ανασύραμε τον φάκελο "Archaea", ο οποίος περιέχει τα 131 πλήρως προσδιορισμένα πρωτεόματα των Αρχαίων, με σκοπό να αναζητήσουμε τις πολλαπλές αλληλουχίες τους έναντι της βάσης δεδομένων (βιβλιοθήκη) που αποτελείται από τα 6 μοντέλα HMM. Τα referenced proteomes είναι ένα υποσύνολο των πρωτεομάτων που έχουν επιλεγεί, είτε με το χέρι είτε αλγοριθμικά, με βάση κάποια κριτήρια ώστε να παρέχουν μία ευρεία κάλυψη του δέντρου της ζωής και μία αντιπροσωπευτική τομή της ταξονομικής ποικιλότητας που υπάρχει στη UniProt.

Για να πραγματοποιήσουμε αυτή τη διαδικασία χρησιμοποιήσαμε το πρόγραμμα hmmscan και ορίσαμε ένα πολύ αυστηρό e-value (e-value=0.001) ώστε να πάρουμε αποτελέσματα στατιστικώς σημαντικά. Τρέξαμε το πρόγραμμα χρησιμοποιώντας την εξής εντολή:

```
for i in Archaea/*.fasta; do # στο φάκελο Archaea για κάθε αρχείο με
                             κατάληξη .fasta
{
    ./hmm-3.1b2-linux-intel-ia32/binaries/hmmscan -E 0.001 LIBRARY.hmm
    $i > $i.RESULT_no_TC; # τρέξε το hmmscan με τιμή e-value 0.001 έναντι
                           του LIBRARY.hmm και βάλε τα αποτελέσματα στο
                           αρχείο με κατάληξη .RESULT_no_TC
}
done
```

Κατόπιν, με τη βοήθεια της γλώσσας προγραμματισμού Perl, δημιουργήσαμε ένα πρόγραμμα το οποίο μας δείχνει πόσες από τις πρωτεΐνες των 6 οικογενειών έκαναν "hit", εντοπίστηκαν δηλαδή, στα γονιδιώματα των Αρχαίων. Χρησιμοποιήσαμε την εντολή:

```
for i in Archaea/*RESULT_no_TC; do # για κάθε αρχείο RESULT_no_TC στο
                                     φάκελο Archaea
    perl count_res.pl $i; done > OLA_MAZI_RES_no_TC
# τρέξε το πρόγραμμα της perl με όνομα count_res.pl για κάθε i και βάλε τα
αποτελέσματα στο αρχείο με κατάληξη OLA_MAZI_RES_no_TC
```

Από τα αποτελέσματα που πήραμε παρατηρήσαμε ότι υπήρχαν πρωτεΐνες για τις οποίες γίνονταν πολλαπλά hit. Επομένως, το συγκεκριμένο μοντέλο δεν έκανε καλό διαχωρισμό στις πρωτεΐνες. Για να

λύσουμε αυτό το πρόβλημα αποφασίσαμε να ορίσουμε μία συγκεκριμένη τιμή κατωφλίου και να επαναλάβουμε την αναζήτηση. Η τιμή αυτή καλείται Trusted Cutoff (TC) κι αποτελεί τη χαμηλότερη τιμή των score μίας ακολουθίας και για τιμές υψηλότερες από αυτή δεν έχουμε false positive hits.

Για να μπορέσουμε να βρούμε την τιμή κατωφλίου για κάθε οικογένεια, τρέξαμε πάλι το πρόγραμμα hmmscan για κάθε ένα από τα αρχεία FASTA των 6 οικογενειών έναντι της βιβλιοθήκης μας, κρατώντας σταθερή την τιμή του e-value κι ακολουθήσαμε την εξής διαδικασία. Αρχικά, μετρήσαμε πόσες πρωτεΐνες έκαναν hit και δεν ανήκαν στην οικογένεια που εξετάζαμε. Καταγράψαμε τα score τους καθώς και τα score της εκάστοτε "σωστής" οικογένειας. Έπειτα για όσες οικογένειες βρέθηκαν "λανθασμένα" hits είχαμε δύο επιλογές. Στην πρώτη, αν όλες οι τιμές των score (λανθασμένες) ήταν μικρότερες από το ελάχιστο score της σωστής οικογένειας, τότε θα θεωρούσαμε ως Trusted Cutoff την τιμή του ελάχιστου score της σωστής οικογένειας. Αντίθετα, αν υπήρχε τιμή (λανθασμένη) μεγαλύτερη από την τιμή του ελάχιστου score της σωστής οικογένειας, τότε θα θεωρούσαμε ως Trusted Cutoff μία ελαφρώς μεγαλύτερη τιμή από το "λανθασμένο" score. Μ'αυτό τον τρόπο είμαστε σίγουροι ότι στις επόμενες αναζητήσεις θα έχουμε αποκλείσει το ενδεχόμενο να έχουμε hit από άλλες οικογένειες πέραν της εξεταζόμενης.

Έπειτα, ξαναφτιάξαμε τα αρχεία της βιβλιοθήκης προσθέτοντας τις αντίστοιχες τιμές κατωφλίου για κάθε οικογένεια. Ξανατρέξαμε το πρόγραμμα hmmscan για τα πρωτεόματα έναντι της βιβλιοθήκης κι αυτή τη φορά αντί της τιμής e-value χρησιμοποιήσαμε την τιμή κατωφλίου γράφοντας την εντολή:

```
for i in Archaea/*.fasta; do # στο φάκελο Archaea για κάθε αρχείο με
                             κατάληξη .fasta
{
    ./hmmmer-3.1b2-linux-intel-ia32/binaries/hmmscan --cut_tc LIBRARY.hmm
    $i > $i.RESULT_with_TC; # τρέξε το hmmscan με τιμή κατωφλίου έναντι
                             του LIBRARY.hmm και βάλε τα αποτελέσματα στο αρχείο
                             με κατάληξη .RESULT_with_TC
}
done
```

Αφού πήραμε τα αρχεία από τα 131 πρωτεόματα, τα χρησιμοποιήσαμε πάλι ως είσοδο στο πρόγραμμα της Perl με σκοπό να εντοπίσουμε αν υπήρχαν πάλι λανθασμένα hit ανάμεσα στις πρωτεΐνες. Χρησιμοποιήσαμε την εντολή:

```
for i in Archaea/*.RESULT_with_TC; do # για κάθε αρχείο RESULT_with_TC στο
                                       φάκελο Archaea
{
```

```
perl count_res.pl $i; done > OLA_MAZI_RES_with_TC # τρέξε το πρόγραμμα
της perl με όνομα count_res.pl για κάθε i και βάλε τα αποτελέσματα στο
αρχείο με κατάληξη OLA_MAZI_RES_with_TC
}
```

Στη συνέχεια, έπρεπε να ταυτοποιήσουμε τα πρωτεόματα οπότε από την ιστοσελίδα: [ftp://ftp.uniprot.org/pub/databases/uniprot/current\\_release/knowledgebase/reference\\_proteomes/README](ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/reference_proteomes/README) πήραμε τον κωδικό κάθε οργανισμού και τον αναζητήσαμε στο NCBI. Έτσι, βρήκαμε σε ποιο φύλο ανήκει ο κάθε οργανισμός και ποια είναι η ταξινόμική του. Ύστερα, καταγράψαμε σε ποιο φύλο ανήκει κάθε πρωτεΐνη από τις οικογένειες που εξετάζουμε.

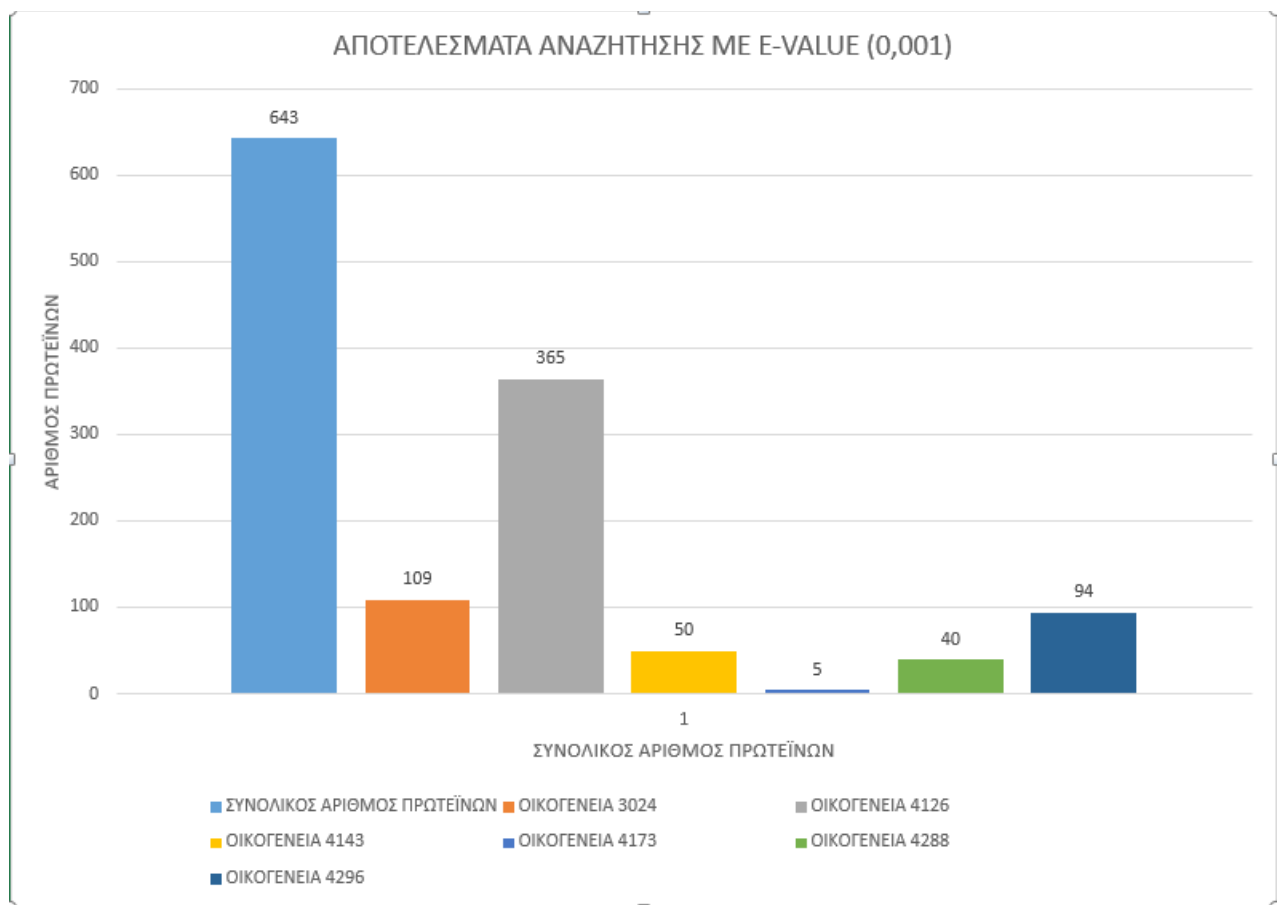
Τέλος, συλλέγοντας όλα τα παραπάνω δεδομένα φτιάξαμε τους τελικούς πίνακες οι οποίοι παρουσιάζουν συνοπτικά τα ευρήματα της αναζήτησής μας.

### **3. Αποτελέσματα – Συμπεράσματα**

Χρησιμοποιήσαμε τα 131 προσδιορισμένα πρωτεόματα των Αρχαίων, τα οποία ανασύραμε από τη UniProt, με σκοπό να κάνουμε διαχωρισμό των πρωτεϊνών για τις 6 οικογένειες που μας ενδιέφεραν. Ύστερα από καταγραφή και διερεύνηση των αποτελεσμάτων παρατηρήσαμε ότι από τις 309.096 πρωτεΐνες που περιέχονται στα πρωτεόματα, 643 πρωτεΐνες (0,21%) έκαναν μοναδικό hit, ενώ 20 πρωτεΐνες (0,006%) έκαναν πολλαπλά hits (Πίνακας 7 – Παράρτημα Γ). Επειδή για την εργασία μας μάς απασχολούσαν μόνο οι πρωτεΐνες που έκαναν μοναδικό hit αποφασίσαμε ότι το μοντέλο αυτό δεν έκανε καλό διαχωρισμό. Έτσι, αναζητήσαμε τις τιμές κατωφλίου για τις οποίες σε μία νέα αναζήτηση δε θα προέκυπτε σφάλμα, δηλαδή αποτελέσματα πρωτεϊνών με πολλαπλά hits.

Ύστερα από την πραγματοποίηση της δεύτερης αναζήτησης, με ορισμό της τιμής κατωφλίου, παρατηρήσαμε ότι από τις συνολικές 309.096 πρωτεΐνες των πρωτεομάτων μόνο 407 πρωτεΐνες (0,13%) έκαναν single hit, ενώ καμία δεν έδωσε πολλαπλό hit (Πίνακας 8 – Παράρτημα Δ). Εφόσον τα αποτελέσματα ήταν αποδεκτά, ταυτοποιήσαμε τους οργανισμούς των Αρχαίων και τους κατατάξαμε ταξονομικά. Από τους πίνακες 5,6 παρατηρούμε ότι και οι 6 πρωτεϊνικές οικογένειες εμφανίστηκαν στο φύλο Euryarchaeota και κάποιες άλλες στα φύλα Thaumarchaeota και Crenarchaeota με πολύ λιγότερες αντιστοιχίσεις. Δεν μπορούμε ακριβώς να εξηγήσουμε γιατί στα υπόλοιπα φύλα δεν είχαμε κανένα αποτέλεσμα. Πιθανόν σε αυτά τα φύλα να αντιστοιχίζονται υποστρώματα Αρχαιοσορτασών που φέρουν διαφορετικό χαρακτηριστικό μοτίβο από αυτά που εμείς εξετάσαμε. Επίσης, υπάρχει και το ενδεχόμενο να διαφέρει πολύ η υδρόφοβη περιοχή τους και το δικό μας εξειδικευμένο μοντέλο να μην κατορθώνει να τα εντοπίσει.

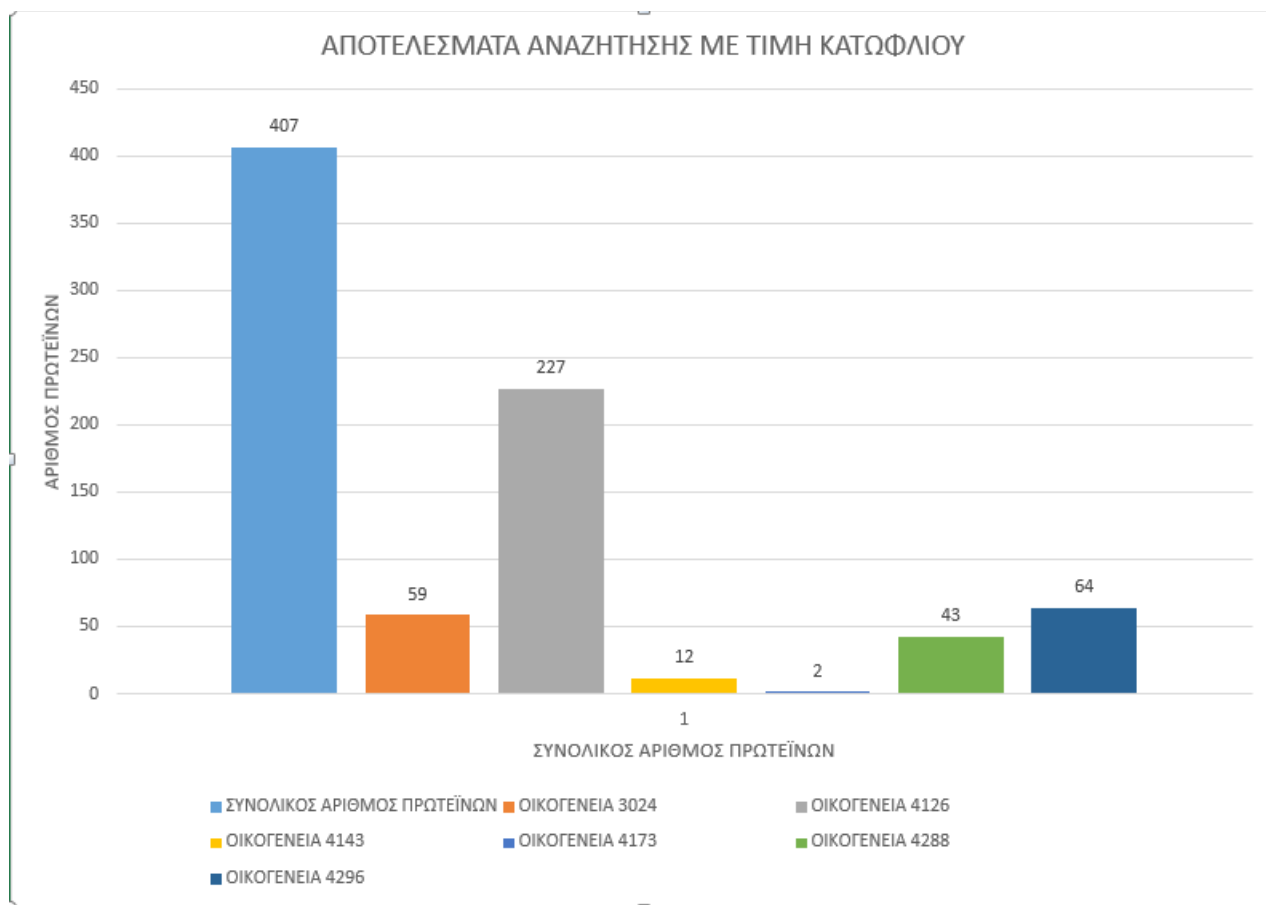




**Εικόνα 2.8:** Γραφική αναπαράσταση αποτελεσμάτων της αναζήτησης με e-value = 0.001

**Πίνακας 4:** Παρουσίαση των score κάθε οικογένειας και η επιλεγμένη τιμή κατωφλίου

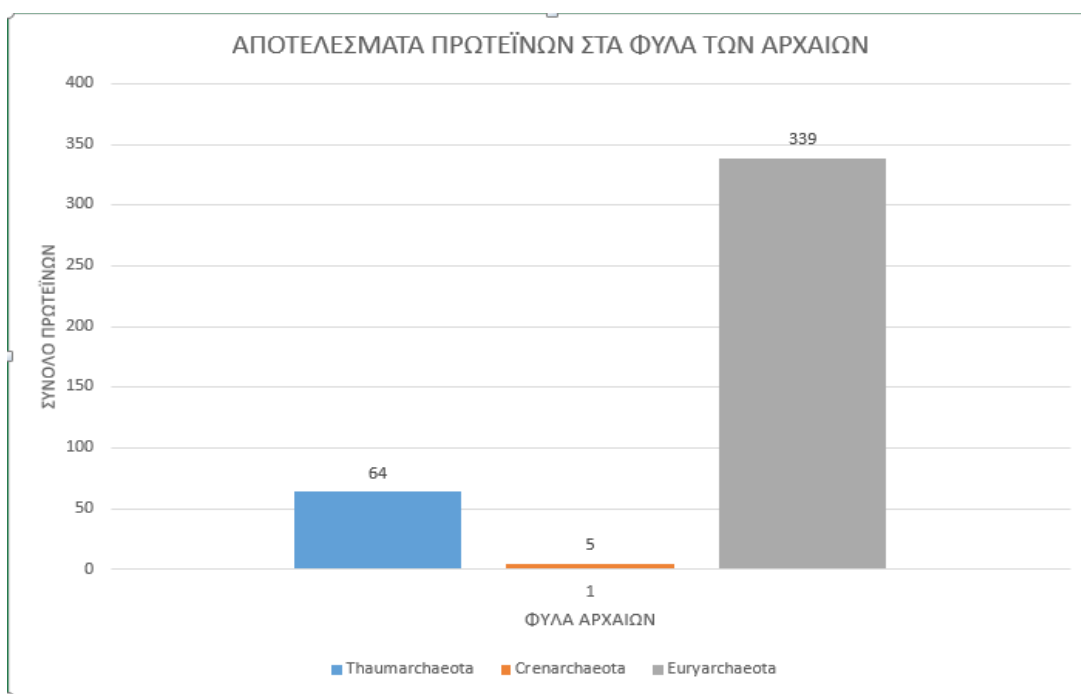
ΦΥΛΟ ΑΡΧΑΙΩΝ	ΟΙΚΟΓΕΝΕΙΑ 3024	ΟΙΚΟΓΕΝΕΙΑ 4126	ΟΙΚΟΓΕΝΕΙΑ 4143	ΟΙΚΟΓΕΝΕΙΑ 4173	ΟΙΚΟΓΕΝΕΙΑ 4288	ΟΙΚΟΓΕΝΕΙΑ 4296
Ευρυαρχαιοτά	56	227	12	2	40	2
Thaumarchaeota	1	0	0	0	1	61
Crenarchaeota	2	0	0	0	2	1
Nanoarchaeota	0	0	0	0	0	0
Korarchaeota	0	0	0	0	0	0
Lokiarchaeota	0	0	0	0	0	0
Βαθυρχειοτά	0	0	0	0	0	0



**Εικόνα 2.9:** Γραφική αναπαράσταση αποτελεσμάτων της αναζήτησης με τιμή κατωφλίου.

**Πίνακας 5:** Παρουσίαση των μελών κάθε οικογένειας ανά ταξινομική βαθμίδα των Αρχαίων

ΦΥΛΟ ΑΡΧΑΙΩΝ	ΟΙΚΟΓΕΝΕΙΑ 3024	ΟΙΚΟΓΕΝΕΙΑ 4126	ΟΙΚΟΓΕΝΕΙΑ 4143	ΟΙΚΟΓΕΝΕΙΑ 4173	ΟΙΚΟΓΕΝΕΙΑ 4288	ΟΙΚΟΓΕΝΕΙΑ 4296
Ευρυαρχαιοτά	56	227	12	2	40	2
Thaumarchaeota	1	0	0	0	1	61
Crenarchaeota	2	0	0	0	2	1
Nanoarchaeota	0	0	0	0	0	0
Korarchaeota	0	0	0	0	0	0
Lokiarchaeota	0	0	0	0	0	0
Βαθυρχειοτά	0	0	0	0	0	0



**Εικόνα 2.10:** Γραφική αναπαράσταση των αποτελεσμάτων των πρωτεϊνών που ανήκουν στα φύλα των Αρχαίων.

**Πίνακας 6:** Παρουσίαση των οικογενειών ανά ταξονομική βαθμίδα των Αρχαίων

ΦΥΛΟ	ΣΥΝΟΛΟ	ΣΤΗΝ ΟΙΚΟΓΕΝ. 3024	ΣΤΗΝ ΟΙΚΟΓΕΝ. 4126	ΣΤΗΝ ΟΙΚΟΓΕΝ. 4143	ΣΤΗΝ ΟΙΚΟΓΕΝ. 4173	ΣΤΗΝ ΟΙΚΟΓΕΝ. 4288	ΣΤΗΝ ΟΙΚΟΓΕΝ. 4296
Euryarchaeota	89	0.63	2.55	0.13	0.22	0.45	0.22
Thaumarchaeota	6	0.16	0	0	0	0.16	10.17
Crenarchaeota	28	0.07	0	0	0	0.71	0.36
Nanoarchaeota	1	0	0	0	0	0	0
Korarchaeota	1	0	0	0	0	0	0
Lokiarchaeota	1	0	0	0	0	0	0
Bathyarchaeota	2	0	0	0	0	0	0

Το προφίλ Hidden Markov Model που δημιουργήθηκε, κατάφερε να εντοπίσει συνολικά 407 πρωτεΐνες από τις 6 οικογένειες των Αρχαίων που φέρουν το χαρακτηριστικό σήμα διαλογής στο καρβοξυτελικό τους άκρο, σε 131 πρωτεόματα. Συγκρίνοντας τα ευρήματα της δεύτερης αναζήτησης με την πρώτη, προκύπτει ότι το μοντέλο μας μπορεί να κάνει διαχωρισμό των πρωτεϊνών σε πολύ μεγάλο βαθμό μιας και δε βρέθηκε κανένα αποτέλεσμα πρωτεΐνης με πολλαπλό hit. Κάτι τέτοιο ήταν αναμενόμενο μιας και η τιμή κατωφλίου μας εξασφάλισε ότι τα score από άλλες πρωτεΐνες, πλην της εξεταζόμενης, θα αποκλείονται. Επίσης, για τις οικογένειες TIGR04143 και TIGR04173 το ποσοστό των πρωτεϊνών που εντοπίστηκαν συνολικά ήταν πολύ μικρό (3%) και (0,5%) αντίστοιχα. Παρόλα αυτά θεωρείται ικανοποιητικό διότι οι πρωτεΐνες που απαρτίζουν τις παραπάνω οικογένειες είναι πολύ λίγες σε αριθμό. Τέλος, με τη βοήθεια του προφίλ HMM μπορέσαμε από τις απλές πολλαπλές στοιχίσεις που προμηθεύτηκαμε από την TIGRFAM να ταυτοποιήσουμε τις πρωτεΐνες στη UniProt, να τις αναζητήσουμε στα πρωτεόματα των Αρχαίων και κατόπιν να τις κατατάξουμε στις ταξονομικές βαθμίδες των Αρχαίων.

Παρόλο που το προφίλ Hidden Markov Model απεδείχθη πολύ χρήσιμο για το διαχωρισμό έχει ένα μειονέκτημα. Επειδή μοντελοποιεί κάθε στήλη της πολλαπλής στοιχίσης είναι πιθανό να έχει εξειδικευτεί ώστε να αναγνωρίζει συγκεκριμένες περιοχές μόνο και να παραβλέπει άλλες. Έτσι, δεν αποκλείεται να μην εντοπίζει μέλη από διάφορες οικογένειες πρωτεϊνών. Για να αντιμετωπίσουμε το πρόβλημα σκοπεύουμε να κατασκευάσουμε το κλασικό Hidden Markov Model. Είναι πιο γενικό και θα το εκπαιδεύσουμε ώστε να αναγνωρίζει με ακρίβεια τις περιοχές που περιέχουν το χαρακτηριστικό μοτίβο αλλά να μπορεί να διακρίνει και τις διαφορές στο κομμάτι της υδρόφοβης περιοχής. Συνεπώς θα είναι πιο ευαίσθητο και θα μπορέσει να μας δώσει περισσότερα αποτελέσματα. Για να επιτευχθεί αυτό, έχουμε ήδη κατασκευάσει πρόγραμμα το οποίο εντοπίζει τα αμινοξέα που βρίσκονται στο καρβοξυτελικό άκρο της ακολουθίας και σχηματίζουν το χαρακτηριστικό μοτίβο. Στη συνέχεια, τα αμινοξέα αφού εντοπιστούν, θα κατατάσσονται στις 3 περιοχές (χαρακτηριστικό μοτίβο, υδρόφοβη περιοχή και θετικά φορτισμένη) και θα υπολογίζεται το ποσοστό εμφάνισης κάθε αμινοξέος στην περιοχή αυτή. Τα στατιστικά αυτά θα χρησιμοποιηθούν ως δεδομένα για την εκπαίδευση του Hidden Markov Model. Η πρόβλεψη των πρωτεϊνών στο κυτταρικό τοίχωμα των Αρχαίων αποτελεί πολύ σημαντικό στοιχείο που θα ρίξει φως στους μηχανισμούς αγκυροβόλησης των Αρχαιοσορτασών, οι οποίοι τα βοηθούν να επιβιώσουν σε τόσο σκληρά περιβάλλοντα. Επίσης, μπορεί να παίξει καθοριστικό ρόλο στη βιομηχανία και ειδικότερα στον τομέα της παρασκευής φαρμάκων.

#### 4. Βιβλιογραφικές Αναφορές

1. F.H.Crick, *On protein synthesis*. Symp Soc Exp Biol, 1958. **12**: p. 138-63.
2. F.H.Crick, et al., *General nature of the genetic code for proteins*. Nature, 1961. **192**: p. 1227-32.
3. Λ.Χ.Μαργαρίτης, Β.Κ.Γαλανόπουλος, Κ.Ε.Κεραμάρης, Ε.Σ.Μαρίνος, Ι.Σ. Παπασιδέρη, Δ.Ι.Σραβοπόδης, Ι.Π.Τραγκάκος, «*Βιολογία Κυττάρου*», 4<sup>η</sup> Έκδοση, Ιατρικές εκδόσεις Λίτσας, Αθήνα, 2003.
4. G.J.Tortora, B.R.Funke, C.L.Case, «*Εισαγωγή στη Μικροβιολογία*», 1<sup>ος</sup> τόμος, Εκδόσεις Πασχαλίδης, Αθήνα, 2009.
5. Berkeley University. *Introduction to the Archaea, Life's extremists*. Retrieved from <http://www.ucmp.berkeley.edu/archaea/archaea.html>.
6. Daniel Lim, "*Microbiology*", 2<sup>nd</sup> edition, University of south Florida, MCB Companies, 2003.
7. Ε.Αλεξανδρή, «*Εισαγωγή στη Σύγχρονη Βιολογία*», ΕΚΠΑ, Αθήνα, 2000.
8. Alberts, Bray, Hopkin, Johnson, Lewis, Raff, Roberts, Wafer, «*Βασικές Αρχές Κυτταρικής Βιολογίας*», 1<sup>ος</sup> τόμος, 2<sup>η</sup> έκδοση, Ιατρικές εκδόσεις Π.Χ.Πασχαλίδης, Αθήνα, 2006.
9. Berkeley University. *Archaea: Morphology*. Retrieved from <http://www.ucmp.berkeley.edu/archaea/archaeamm.html>.
10. Wikipedia. *Archaea*. Retrieved from <https://en.wikipedia.org/wiki/Archaea>.
11. J.K.Kristjansson, "*Thermophilic bacteria*", CRC Press LLC, Florida, 1992.
12. Α.Τσελεπιδάκης, «*Αναερόβια χώνευση γλυκερίνης (απόβλητο της παραγωγής βιοντίζελ)*», Πτυχιακή εργασία, Ανώτατο τεχνολογικό ίδρυμα Κρήτης, Ηράκλειο, 2008.
13. Metcalf & Eddy, "*Wastewater Engineering*", 3<sup>rd</sup> edition, McGrawHill International Editions, New York, 1991.
14. W.W.Jr Eckenfelder, "*Industrial Water Pollution Control*", 2<sup>nd</sup> edition, McGrawHill International Editions, New York, 1989.
15. Α.Βλυσίδης, «*Αξιοποίηση βιομάζας-Αναερόβια χώνευση*», Ε.Μ.Π., Αθήνα , 2002.
16. D.Cabanes, et al., *Surface proteins and the pathogenic potential of Listeria monocytogenes*. Trends Microbiol, 2002. **10**(5): p. 238-45.
17. S.K.Mazmanian, H. Ton-That, and O. Schneewind, *Sortase-catalysed anchoring of surface proteins to the cell wall of Staphylococcus aureus*. Mol Microbiol, 2001. **40**(5): p. 1049-57.
18. J.Boekhorst, et al., *Genome-wide detection and analysis of cell wall-bound proteins with LPxTG-like sorting motifs*. J Bacteriol, 2005. **187**(14): p. 4928-34.

19. D.Comfort and R.T. Clubb, *A comparative genome analysis identifies distinct sorting pathways in gram-positive bacteria*. Infect Immun, 2004. **72**(5): p. 2710-22.
20. H.Ton-That, L.A. Marraffini, and O. Schneewind, *Protein sorting to the cell wall envelope of Gram-positive bacteria*. Biochim Biophys Acta, 2004. **1694**(1-3): p. 269-78.
21. Δ. Φιμερέλη, «Υπολογιστικές μελέτες των πρωτεϊνών του κυτταρικού τοιχώματος των θετικών κατά Gram βακτηρίων», Πτυχιακή εργασία, ΕΚΠΑ, Αθήνα, 2009.
22. W.W.Navarre and O. Schneewind, *Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope*. Microbiol Mol Biol Rev, 1999. **63**(1): p. 174-229.
23. Daniel H. Haft., et al., *Archaeosrtases and Exosortases Are Widely Distributed Systems Linking Membrane Transit with Posttranslational Modification*. Journal of Bacteriology, 2011. **194**: p. 36 – 48.
24. J. Craig Venter Institute. *TIGRFAMS TERMS*. Retrieved from <http://www.jcvi.org/cgi-bin/tigrfams/Terms.cgi>.
25. Π. Γ. Μπάγκος, «Βιοπληροφορική», 1<sup>η</sup> έκδοση , Ελληνικά Ακαδημαϊκά Ηλεκτρονικά Συγγράματα και Βοηθήματα, σ. 142-146, 276 – 277, 300 – 305. 2015.
26. Wikipedia. *BLAST*. Retrieved from <https://en.wikipedia.org/wiki/BLAST>.
27. Wikipedia. Hidden Markov Model. Retrieved from [https://en.wikipedia.org/wiki/Hidden\\_Markov\\_model](https://en.wikipedia.org/wiki/Hidden_Markov_model).
28. Retrieved from <http://weblogo.berkeley.edu/logo.cgi/>.
29. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2766791/>.
30. Retrieved from <https://eclass.uoa.gr/modules/document/file.php/BIOL199/>.
31. Τζώρτζη Αικατερίνη, «Αρχαιοβακτήρια, τι γνωρίζουμε και ποια είναι η σημασία τους στη ζωή», ΕΚΠΑ, Αθήνα, 2010.

## Παράρτημα Α

### Πρόγραμμα για τον εντοπισμό των οικογενειών που έκαναν hit μετά την σύγκριση των ρHMM και των πρωτεομάτων των Αρχαίων.

```
$count_family3024=0;          # orismos twn hitcounters gia kathe
                                oikogeneia

$count_family4126=0;
$count_family4143=0;
$count_family4173=0;
$count_family4288=0;
$count_family4296=0;

$total_proteins=0;
$proteins_with_single_hit=0;
$proteins_with_multiple_hit=0;

$infile = $ARGV[0];
open OUT, ">$infile.HITS_ONLY"; # to apotelesma tha apothikeuetai sto
                                arxeio me kataliksi infile.HITS_ONLY
$/="//\n";          # orizo to input_record_reparator gia kathe nea seira
open IN, $infile;   # anoigo to arxeio
while(<IN>)         # gia oso uparxon grammes sto arxeio kano tin parakatw
                    epanalipsi
{
    if($_ =~ /^Query:\s+(\.*)\s+\[L=\d+\]/m) # an gia kathe grammi tou
arxeioy poy ksekinaei apo Query: kai ikanopoiei thn sinthiki tote:
    {
        $prot_id=$1;      # sth metavliti prot_id anathetw o,ti periexetai
                            sto (.*
        $total_proteins++; # o metritis auksanei kata 1
    }

    if($_ !~/No hits detected that satisfy reporting thresholds/)
# Paravlepsi toy "No hits detected that satisfy reporting thresholds"
    {
        if($_ =~ /Scores for complete sequence \(score includes all
domains\)\: (.*)Domain annotation for each model/ms) # Krataei sti
metavliti all_hits opia pliroforia einai mesa sto (.* ) sxetika me ta hit
tw n proteinwn
        {
            $all_hits=$1;
            @split_hits=split("\n", $all_hits); #xwrizetw to all_hits
                                                kai to vazetw se mia lista

            print OUT $prot_id."\n"; # typonei sto arxeio ton kodiko t
is proteinis
            print OUT $all_hits."\n//\n"; # typwnei th lista me ta
hits

            $best_hit = $split_hits[4]; # h metavliti best_hit pairnei
ws timi o,ti vrisketai stin tetarth thesi ths listas "split_hits"
```



```

        if($best_hit=~/.*(family\d+).*/) # anakta to onoma
                                            tis oikogeneias
        {$family_name=$1;}
        $total_no_lines=$#split_hits;      # h metavliti
total_no_lines isoutai me ton ariumo twñ stoixeiwn pou vriskontai sti lista
        if($total_no_lines==4) # an o arithmos isoutai me 4
            tote simaninei oti exw monadiko hit
        { $proteins_with_single_hit++; }# metraei poses einai
            oi proteines pou ekanan monadiko
            hit
        elsif($total_no_lines>=5) # an o arithos einai
megaluteros h isos me 5 tote simainei oti stin oikogeneia exw pollapla hits
        { $proteins_with_multiple_hit++; } # metraei poses
einai oi proteines pou ekanan pollaplo hit

        if($family_name eq 'family3024')    {$count_family3024++;}
# afksanei ton metriti gia kathe oikogeneia
        elsif($family_name eq 'family4126') {$count_family4126++;}
        elsif($family_name eq 'family4143') {$count_family4143++;}
        elsif($family_name eq 'family4173') {$count_family4173++;}
        elsif($family_name eq 'family4288') {$count_family4288++;}
        elsif($family_name eq 'family4296') {$count_family4296++;}
    }
}
close IN;      # kleinei to arxeio pou diavazei
close OUT;    # kleinei to arxeio sto opoio apothikeuei
$/="\n";      # epanaferei to iput_record_separator se nea grami

$total_apo_hits = $proteins_with_single_hit+$proteins_with_multiple_hit;
#athroizei ta apotelesmata sxetika me ta hits (monadiko+pollaplo)
$total_apo_families = $count_family3024+$count_family4126+
$count_family4143+$count_family4173+$count_family4288+$count_family4296;
# athroizei ta apotelesmata apo tis oikogneies
if($total_apo_families != $total_apo_hits) {print "PROBLEM in $infile\n";}
# elegxei an oi metavlitες total_apo_hits kai otal_apo_families einai
diaforetikes. An einai diaforetikes tote emfanizei provlima

print $infile."\t".$total_proteins."\t".$proteins_with_single_hit."\t".
$proteins_with_multiple_hit."\t".      # emfanizei sto termatiko to onoma
tou arxeioy kai ta athroismata
    $count_family3024."\t".$count_family4126."\t".
$count_family4143."\t".
    $count_family4173."\t".$count_family4288."\t".
$count_family4296."\n";

```

## Παράρτημα Β

### Αποτελέσματα αναζήτησης των 6 οικογένειων πρωτεϊνών με τα pHMM.

ΚΩΔΙΚΟΣ ΠΡΩΤΕΪΝΗΣ	ΜΕΛΟΣ ΟΙΚΟΓΕΝΕΙΑΣ	TIGR03024	TIGR04126	TIGR04143	TIGR04173	TIGR04288	TIGR04296
Q8PS01	TIGR03024	41.3	-	-	-	-	6
Q12TD4	TIGR03024	39.5	-	-	-	-	-
Q12ZP5	TIGR03024	36.2	-	-	-	-	-
Q8PV56	TIGR03024	38.9	-	-	-	-	-
Q12WI2	TIGR03024	34	-	-	-	-	-3.9
B5ICW1	TIGR03024	33.7	-	-	-	-	-
B5IFE3	TIGR03024	32.1	-	-	-	-	-
Q12UJ4	TIGR03024	33.6	-	-	-	-	-
Q8PU64	TIGR03024	36.1	-	-	-	-	-
B5IFT4	TIGR03024	30.6	-	-	-	-	-
Q12YZ7	TIGR03024	33	-	-	-	-	-
B5IBJ1	TIGR03024	35	-	-	-	-	15.4
B5ID12	TIGR03024	22.9	-	-	-	-	-
B5IH14	TIGR03024	24	-	-	-	-	-
B5IER4	TIGR03024	26.4	-	-	-	-	-
B5ICX2	TIGR03024	21	-	-	-	-	-
B5IEI3	TIGR03024	20.2	-	-	-	-	-
B5IB07	TIGR03024	22	-	-	-	-	-
B5IEI2	TIGR03024	26.2	-	-	-	-	-
B5IBC0	TIGR03024	22.7	-	-	-	-	-
B5IBR5	TIGR03024	37	-	-	-	-	-
Q5UY99	TIGR04126	-	35.3	-	-	-	-
Q5V7F4	TIGR04126	-	42.8	-	-	-	-
C7NV88	TIGR04126	-	40.5	-	-	-	-
Q3INA1	TIGR04126	-	41.7	-	-	-	-
L9UPH7	TIGR04126	-	45.9	-	-	-	-
Q18KV5	TIGR04126	-	45.5	-	-	-	-
D4GTN4	TIGR04126	-	39.1	-	-	-	-
E4NU72	TIGR04126	-	39.4	-	-	-	-
P25062	TIGR04126	-	44.4	-	-	-	-
E4NR07	TIGR04126	-	40.5	-	-	-	-
C7P4P2	TIGR04126	-	41.4	-	-	-	-
B9LSV0	TIGR04126	-	47.7	-	-	-	-
D4GV93	TIGR04126	-	38.5	-	-	-	-
E4NQI6	TIGR04126	-	38	-	-	-	-
Q5V2C2	TIGR04126	-	36	-	-	-	-
Q9HPZ0	TIGR04126	-	26.1	-	-	-	-
B9LV01	TIGR04126	-	37.9	-	-	-	-
D8J931	TIGR04126	-	44.6	-	-	-	-
B0R8E4	TIGR04126	-	38.4	-	-	-	-
D2RYX7	TIGR04126	-	38.9	-	-	-	-

D3SUE8	TIGR04126	-	42.1	-	-	-	-
D2S3B3	TIGR04126	-	33.4	-	-	-	-
D2S2G3	TIGR04126	-	33.4	-	-	-	-
B9LS09	TIGR04126	-	38	-	-	-	-
B9LP11	TIGR04126	-	35.3	-	-	-	-
C7NMW0	TIGR04126	-	30.9	-	-	-	-
C7NVB4	TIGR04126	-	39	-	-	-	-
Q5UXY2	TIGR04126	-	41.8	-	-	-	-
Q5V3A2	TIGR04126	-	38.3	-	-	-	-
E4NQ90	TIGR04126	-	45.9	-	-	-	-
C7NXX0	TIGR04126	-	40.4	-	-	-	-
C7NPV2	TIGR04126	-	40.6	-	-	-	-
D4GTW5	TIGR04126	-	41	-	-	-	-
C7NUZ3	TIGR04126	-	24.8	-	-	-	-
E4NQC5	TIGR04126	-	37.3	-	-	-	-
D2RZP0	TIGR04126	-	28.9	-	-	-	-
D2RZN9	TIGR04126	-	23.5	-	-	-	-
D3SYM9	TIGR04126	-	36.9	-	-	-	-
D4GW17	TIGR04126	-	25	-	-	-	-
B9LPR1	TIGR04126	-	32.4	-	-	-	-
D8J2R9	TIGR04126	-	41.9	-	-	-	-
C7NZZ1	TIGR04126	-	28.8	-	-	-	-
B9LT73	TIGR04126	-	35	-	-	-	-
Q3IQJ1	TIGR04126	-	34.1	-	-	-	-
Q46GI5	TIGR04126	-	34.6	-	-	-	-
Q5UXE2	TIGR04126	-	30.2	-	-	-	-
C7NVU1	TIGR04126	-	31.1	-	-	-	-
E4NQH2	TIGR04126	-	29.9	-	-	-	-
D2RSD4	TIGR04126	-	31.2	-	-	-	-
Q3IQ12	TIGR04126	-	39.8	-	-	-	-
Q5V3R9	TIGR04126	-	31	-	-	-	-
B0R629	TIGR04126	-	26.9	-	-	-	-
C7NUZ4	TIGR04126	-	33.1	-	-	-	-
A7I5M9	TIGR04126	-	34.7	-	-	-	-
Q2FTS3	TIGR04126	-	34.8	-	-	-	-
A3CSY3	TIGR04126	-	30.6	-	-	-	-
C7NRY2	TIGR04126	-	36.8	-	-	-	-
E4NNB6	TIGR04126	-	30.9	-	-	-	-
B9LVB5	TIGR04126	-	38.7	-	-	-	-
D4GPK7	TIGR04126	-	30.2	-	-	-	-
L9UVE1	TIGR04126	-	31.8	-	-	-	-
D2RPQ4	TIGR04126	-	22.9	-	-	-	-
D3T2A7	TIGR04126	-	38	-	-	-	-
D3T211	TIGR04126	-	33.4	-	-	-	-
B9LPC5	TIGR04126	-	26.4	-	-	-	-
Q3IS59	TIGR04126	-	19.4	-	-	-	-

Q18E85	TIGR04126	-	25	-	-	-	-
O28785	TIGR04126	-	26.5	-	-	-	-
Q46BP2	TIGR04126	-	35.6	-	-	-	-
Q8PVI7	TIGR04126	-	36.5	-	-	-	-
D5E903	TIGR04126	-	36.9	-	-	-	-
D2RFI4	TIGR04126	-	38.1	-	-	-	-
A0B6Z3	TIGR04126	-	33.7	-	-	-	-
D1YZ52	TIGR04126	-	30.2	-	-	-	-
A3CT98	TIGR04126	-	31.4	-	-	-	-
Q8TQY5	TIGR04126	-	32.4	-	-	-	-
D5E785	TIGR04143	-	-	40.4	-	-	-
D7E9Z0	TIGR04143	-	-	39.1	-	-	-
D5E8S8	TIGR04143	-	-	42.5	-	-	-
D7EAX6	TIGR04143	-	-	46.4	-	-	-
D7EAY3	TIGR04143	-	-	44.9	-	-	-
D5E924	TIGR04143	-	-	41.9	-	-	-
D5E786	TIGR04143	-	-	43.7	-	-	-
D5EBB0	TIGR04143	-	-	45	-	-	-
D7EC12	TIGR04143	-	-	45.8	-	-	-
Q58863	TIGR04173	-	-	-	49.8	-	-
C9RIG1	TIGR04173	-	-	-	54.7	-	-
D5VRK5	TIGR04173	-	-	-	44.8	-	-
F6BDF1	TIGR04173	-	-	-	42.9	-	-
C7P902	TIGR04173	-	-	-	42.7	-	-
F8AJW8	TIGR04173	-	-	-	45.2	-	-
Q9V294	TIGR04288	-	-	-	-	16.9	-
Q9V0S2	TIGR04288	-	-	-	-	21.9	-
Q9V057	TIGR04288	-	-	-	-	25.2	-
Q9UZW3	TIGR04288	-	-	-	-	18.6	-
Q9UYG6	TIGR04288	-	-	-	-	25.3	-
O58071	TIGR04288	-	-	-	-	16.5	-
O58238	TIGR04288	-	-	-	-	28.6	-
O58537	TIGR04288	-	-	-	-	19.8	-
O58925	TIGR04288	-	-	-	-	25.4	-
O59112	TIGR04288	-	-	-	-	18.4	-
Q8U287	TIGR04288	-	-	-	-	24.9	-
Q8U0Z2	TIGR04288	-	-	-	-	24.8	-
Q8TZQ2	TIGR04288	-	7.3	-	-	27.4	-
B6YU14	TIGR04288	-	-	-	-	19.9	-
B6YX80	TIGR04288	-	-	-	-	27.7	-
B6YXX4	TIGR04288	-	-	-	-	19.6	-
B6YUB5	TIGR04288	-	-	-	-	22.7	-
B6YUG3	TIGR04288	-	-	-	-	20.1	-
B7R4L2	TIGR04288	-	-	-	-	18.7	-
B7R2C2	TIGR04288	-	-	-	-	26.2	-
B7R2L9	TIGR04288	-	-	-	-	26.5	-

B7R264	TIGR04288	-	-	-	-	26.4	-
B7R1S6	TIGR04288	-	-	-	-	26.6	-
B7R0H5	TIGR04288	-	-	-	-	31.2	-
B7R0F1	TIGR04288	-	-	-	-	18.1	-
B7R0I3	TIGR04288	-	-	-	-	27.5	-
C5A2S6	TIGR04288	-	-	-	-	24.1	-
C5A3S5	TIGR04288	-	-	-	-	25.7	-
C5A4E3	TIGR04288	-	-	-	-	28	-
C5A598	TIGR04288	-	-	-	-	21.9	-
C5A5B4	TIGR04288	-	-	-	-	25.1	-
C5A5C5	TIGR04288	-	-	-	-	21.4	-
C5A5Z8	TIGR04288	-	-	-	-	14.7	-
C5A6C3	TIGR04288	-	-	-	-	19.7	-
C5A6E0	TIGR04288	-	-	-	-	27.1	-
C5A6Z4	TIGR04288	-	-	-	-	26.7	-
C5A1H9	TIGR04288	-	-	-	-	19.3	-
C5A1J1	TIGR04288	-	-	-	-	31.2	-
C6A0Z0	TIGR04288	-	-	-	-	26.8	-
C6A197	TIGR04288	-	-	-	-	26.6	-
C6A4X3	TIGR04288	-	-	-	-	19.8	-
C6A5A8	TIGR04288	-	-	-	-	28.8	-
F0LN19	TIGR04288	-	-	-	-	19.3	-
F0LHU7	TIGR04288	-	-	-	-	23	-
F0LI96	TIGR04288	-	-	-	-	23.7	-
F0LJB0	TIGR04288	-	-	-	-	27.5	-
F4HMA8	TIGR04288	-	-	-	-	25.8	-
F4HMJ6	TIGR04288	-	-	-	-	21	-
F4HML0	TIGR04288	-	-	-	-	20.7	-
F4HLL4	TIGR04288	-	-	-	-	23.1	-
F8AHS7	TIGR04288	-	-	-	-	26	-
F8AIJ8	TIGR04288	-	-	-	-	26.7	-
F8AEY2	TIGR04288	-	-	-	-	24.7	-
F8AJG0	TIGR04288	-	-	-	-	28.1	-
G0HJI7	TIGR04288	-	-	-	-	20.3	-
G0HNI9	TIGR04288	-	-	-	-	26.9	-
G0HNI0	TIGR04288	-	-	-	-	24.6	-
G0HP26	TIGR04288	-	-	-	-	23.5	-
G0HL20	TIGR04288	-	-	-	-	21.8	-
G0HP41	TIGR04288	-	-	-	-	28.3	-
G0HPN6	TIGR04288	-	-	-	-	28.9	-
B7R256	TIGR04288	-	-	-	-	24.7	-
B7R3E2	TIGR04288	-	-	-	-	21.0	-
B7R3J9	TIGR04288	-	-	-	-	19.1	-
Q5JEC5	TIGR04288	-	-	-	-	19.5	-
Q5JG97	TIGR04288	-	-	-	-	28.7	-

Q5JHC9	TIGR04288	-	-	-	-	29.4	-
A0RXJ2	TIGR04296	-	-	-	-	-	26.8
A0RXX4	TIGR04296	-	-	-	-	-	24
A0RYJ2	TIGR04296	-	-	-	-	-	25.3
A0RW65	TIGR04296	-	-	-	-	-	39.2
A9A3S0	TIGR04296	-	-	-	-	-	33.9
A9A4Q1	TIGR04296	-	-	-	-	-	37.4
F3KKI4	TIGR04296	-	-	-	-	-	29.3
A9A1P7	TIGR04296	-	-	-	-	-	34.3
F3KKN4	TIGR04296	-	-	-	-	-	35.4
F9CWV2	TIGR04296	-	-	-	-	-	30.9
A9A5U5	TIGR04296	-	-	-	-	-	40.8
F3KKG2	TIGR04296	-	-	-	-	-	44.4
A9A2P5	TIGR04296	-	-	-	-	-	30.4
F3KLM7	TIGR04296	-	-	-	-	-	34.4
F9CZD2	TIGR04296	-	-	-	-	-	37
A9A2Q3	TIGR04296	14.1	-	-	-	-	41.7
A9A2A3	TIGR04296	-	-	-	-	-	40.2
F3KIK8	TIGR04296	-	-	-	-	-	36.7
F9CW87	TIGR04296	-	-	-	-	-	36.1
F9CW39	TIGR04296	-	-	-	-	-	46.5
A0RUQ7	TIGR04296	-	-	-	-	-	43
A0RYI9	TIGR04296	-	-	-	-	-	45.7
F3KKT2	TIGR04296	-	-	-	-	-	47.9
F3KMU2	TIGR04296	-	-	-	-	-	41.2
A0RXX2	TIGR04296	11.4	-	-	-	-	46
A9A1J0	TIGR04296	-	-	-	-	-	48.7
F9CYF5	TIGR04296	-	-	-	-	-	44.5
F3KIH8	TIGR04296	-	-	-	-	-	42.8
A9A454	TIGR04296	-	-	-	-	-	44.8
F3KKT4	TIGR04296	-	-	-	-	-	45.4
F9CWD5	TIGR04296	-	-	-	-	-	37.8
F9CYK2	TIGR04296	-	-	-	-	-	38.7

## Παράρτημα Γ

**Πίνακας 7:** ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΝΑΖΗΤΗΣΗΣ ΧΩΡΙΣ ΤΙΜΗ ΚΑΤΩΦΛΙΟΥ (ΜΕ E-VALUE).

ΟΡΓΑΝΙΣΜΟΣ	ΦΥΛΟ	ΟΛΙΚΟΣ ΑΡΙΘΜΟΣ ΠΡΩΤΕΪΝΩΝ	# ΠΡΩΤΕΪΝΩΝ ΜΕ 1 HIT	# ΠΡΩΤΕΪΝΩΝ ΜΕ > 1 HIT	#ΣΤΗΝ ΟΙΚΟΓΕΝΕΙΑ 3024	#ΣΤΗΝ ΟΙΚΟΓΕΝΕΙΑ 4126	#ΣΤΗΝ ΟΙΚΟΓΕΝΕΙΑ 4143	#ΣΤΗΝ ΟΙΚΟΓΕΝΕΙΑ 4173	#ΣΤΗΝ ΟΙΚΟΓΕΝΕΙΑ 4288	#ΣΤΗΝ ΟΙΚΟΓΕΝΕΙΑ 4296
Candidatus Bathyarchaeota archaeon BA2	Bathyarchaeota	1760	0	0	0	0	0	0	0	0
Candidatus Bathyarchaeota archaeon BA1	Bathyarchaeota	2401	0	0	0	0	0	0	0	0
miscellaneous Crenarchaeota group-15 archaeon DG-45	Crenarchaeota	1294	0	0	0	0	0	0	0	0
miscellaneous Crenarchaeota group-1 archaeon SG8-32-1	Crenarchaeota	1097	9	1	8	0	0	0	0	2
miscellaneous Crenarchaeota group-6 archaeon AD8-1	Crenarchaeota	1496	2	0	2	0	0	0	0	0
miscellaneous Crenarchaeota group archaeon SMTZ1-55	Crenarchaeota	1119	0	0	0	0	0	0	0	0
miscellaneous Crenarchaeota group archaeon SMTZ-80	Crenarchaeota	1357	1	0	1	0	0	0	0	0
Metallosphaera sedula (strain ATCC 51363 / DSM 5348)	Crenarchaeota	2246	1	0	1	0	0	0	0	0
Staphylothermus marinus (strain ATCC 43588 / DSM 3639 / JCM 9404 / F1)	Crenarchaeota	1566	0	0	0	0	0	0	0	0
Ignicoccus hospitalis (strain KIN4/I / DSM 18386 / JCM 14125)	Crenarchaeota	1427	1	0	0	1	0	0	0	0
Acidilobus saccharovorans (strain DSM 16705 / JCM 18335 / VKM B-2471 / 345-15)	Crenarchaeota	1495	1	0	0	1	0	0	0	0
Thermofilum pendens (strain Hrk 5)	Crenarchaeota	1870	0	0	0	0	0	0	0	0
Sulfolobus tokodaii (strain DSM 16993 / JCM 10545 / NBRC 100140 / 7)	Crenarchaeota	2762	0	0	0	0	0	0	0	0



Sulfolobus acidocaldarius (strain ATCC 33909 / DSM 639 / JCM 8929 / NBRC 15157 / NCIMB 11770)	Crenarchaeota	2187	4	0	0	2	1	1	0	0
--	---------------	------	---	---	---	---	---	---	---	---

Pyrolobus fumarii (strain DSM 11204 / 1A)	Crenarchaeota	1955	0	0	0	0	0	0	0	0
Caldivirga maquilensis (strain ATCC 700844 / DSM 13496 / JCM 10307 / IC-167)	Crenarchaeota	1959	0	0	0	0	0	0	0	0
Ignisphaera aggregans (strain DSM 17230 / JCM 13409 / AQ1.S1)	Crenarchaeota	1920	3	0	2	0	0	0	1	0
Sulfolobus solfataricus (strain ATCC 35092 / DSM 1617 / JCM 11322 / P2)	Crenarchaeota	2924	1	0	0	0	1	0	0	0
Thermosphaera aggregans (strain DSM 11486 / M11TL)	Crenarchaeota	1383	1	0	1	0	0	0	0	0
Pyrobaculum aerophilum (strain ATCC 51768 / IM2 / DSM 7523 / JCM 9630 / NBRC 100827)	Crenarchaeota	2588	0	0	0	0	0	0	0	0
Aeropyrum pernix (strain ATCC 700893 / DSM 11879 / JCM 9820 / NBRC 100138 / K1)	Crenarchaeota	1679	0	0	0	0	0	0	0	0
Hyperthermus butylicus (strain DSM 5456 / JCM 9403 / PLM1-5)	Crenarchaeota	1591	3	0	0	1	1	0	0	1
Thermoproteus tenax (strain ATCC 35583 / NBRC 100435 / JCM 9277 / Kra 1)	Crenarchaeota	1950	0	0	0	0	0	0	0	0
Thermogladius cellulolyticus (strain 1633)	Crenarchaeota	1409	0	0	0	0	0	0	0	0
Vulcanisaeta distributa (strain DSM 14429 / JCM 11212 / NBRC 100878 / IC-017)	Crenarchaeota	2485	0	0	0	0	0	0	0	0
Desulfurococcus kamchatkensis (strain 1221n / DSM 18924)	Crenarchaeota	1467	0	0	0	0	0	0	0	0
Fervidicoccus fontis (strain DSM 19380 / VKM B-2539 / Kam940)	Crenarchaeota	1377	0	0	0	0	0	0	0	0
Acidianus hospitalis (strain W1)	Crenarchaeota	2315	1	0	0	1	0	0	0	0
Caldisphaera lagunensis (strain DSM 15908 / JCM 11604 / IC-154)	Crenarchaeota	1470	2	0	0	1	0	0	0	1

Thermoproteus sp. AZ2	Crenarchaeota	1697	2	0	2	0	0	0	0	0
Methanocorpusculum labreanum (strain ATCC 43576 / DSM 4855 / Z)	Euryarchaeota	1728	5	0	0	3	2	0	0	0
Halalkalicoccus jeotgali (strain DSM 18796 / CECT 7217 / JCM 14584 / KCTC 4019 / B3)	Euryarchaeota	3756	2	0	0	2	0	0	0	0
Methanohalobium evestigatum (strain DSM 3721 / OCM 161 / Z-7303)	Euryarchaeota	2236	15	0	1	5	9	0	0	0
Picrophilus torridus (strain ATCC 700027 / DSM 9790 / JCM 10055 / NBRC 100828)	Euryarchaeota	1532	0	0	0	0	0	0	0	0
Methanlobus psychrophilus R15	Euryarchaeota	3111	16	3	13	6	0	0	0	0
Thermococcus kodakarensis (strain ATCC BAA-918 / JCM 12380 / KOD1) (Pyrococcus kodakaraensis (strain KOD1))	Euryarchaeota	2290	13	0	0	1	0	0	12	0
Halobacterium salinarum (strain ATCC 700922 / JCM 11081 / NRC-1) (Halobacterium halobium)	Euryarchaeota	2415	8	0	0	8	0	0	0	0
Methanococcus maripaludis (strain S2 / LL)	Euryarchaeota	1695	0	0	0	0	0	0	0	0
Methanosarcina mazei (strain ATCC BAA-159 / DSM 3647 / Goe1 / Go1 / JCM 11833 / OCM 88) (Methanosarcina frisia)	Euryarchaeota	3254	10	2	5	5	1	0	0	1
Methanocella arvoryzae (strain DSM 22066 / NBRC 105507 / MRE50)	Euryarchaeota	3064	7	0	0	7	0	0	0	0
Methanosaeta thermophila (strain DSM 6194 / JCM 14653 / NBRC 101360 / PT) (Methanotherix thermophila)	Euryarchaeota	1661	3	0	1	1	0	0	0	1
Halorubrum lacusprofundi (strain ATCC 49239 / DSM 5036 / JCM 8891 /	Euryarchaeota	3465	10	0	0	10	0	0	0	0

ACAM 34)										
Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) (Methanococcus jannaschii)	Euryarchaeota	1787	1	0	0	0	0	1	0	0

Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) (Methanococcus jannaschii)	Euryarchaeota	1787	1	0	0	0	0	1	0	0
Pyrococcus abyssi (strain GE5 / Orsay)	Euryarchaeota	1780	8	0	0	0	0	0	8	0
Pyrococcus furiosus (strain ATCC 43587 / DSM 3638 / JCM 8422 / Vc1)	Euryarchaeota	2044	8	1	1	0	0	0	8	0
Thermoplasma acidophilum (strain ATCC 25905 / DSM 1728 / JCM 9062 / NBRC 15155 / AMRC-C165)	Euryarchaeota	1481	1	0	1	0	0	0	0	0
Methanohalophilus mahii (strain ATCC 35705 / DSM 5219 / SLP)	Euryarchaeota	1972	15	0	1	9	5	0	0	0
Methanococcus aeolicus (strain Nankai-3 / ATCC BAA-1280)	Euryarchaeota	1480	0	0	0	0	0	0	0	0
Haloarcula marismortui (strain ATCC 43049 / DSM 3752 / JCM 8966 / VKM B-1809) (Halobacterium marismortui)	Euryarchaeota	4101	11	0	0	10	1	0	0	0
Aciduliprofundum boonei (strain DSM 19572 / T469)	Euryarchaeota	1532	13	1	13	0	1	0	0	0
Halomicrobium mukohataei (strain ATCC 700874 / DSM 12286 / JCM 9738 / NCIMB 13541) (Haloarcula mukohataei)	Euryarchaeota	3321	9	0	1	7	0	0	0	1
Methanopyrus kandleri (strain AV19 / DSM 6324 / JCM 9639 / NBRC 100938)	Euryarchaeota	1683	0	0	0	0	0	0	0	0
Natrialba magadii (strain ATCC 43099 / DSM 3394 / NCIMB 2190 / MS3) (Natronobacterium magadii)	Euryarchaeota	4167	8	0	1	7	0	0	0	0
Methanocella paludicola (strain DSM 17711 / JCM 13418 / NBRC 101707 / SANAE)	Euryarchaeota	2987	8	0	0	8	0	0	0	0

Archaeoglobus profundus (strain DSM 5631 / JCM 9629 / NBRC 100127 / Av18)	Euryarchaeota	1813	4	0	1	3	0	0	0	0
Haloterrigena turkmenica (strain ATCC 51198 / DSM 5511 / NCIMB 13204 / VKM B-1734) (Halococcus turkmenicus)	Euryarchaeota	5070	15	0	0	15	0	0	0	0
Methanosphaera stadtmanae (strain ATCC 43021 / DSM 3091 / JCM 11832 / MCB-3)	Euryarchaeota	1531	0	0	0	0	0	0	0	0
Methanospirillum hungatei JF-1 (strain ATCC 27890 / DSM 864 / NBRC 100397 / JF-1)	Euryarchaeota	3070	3	0	0	1	2	0	0	0
Haloquadratum walsbyi (strain DSM 16790 / HBSQ001)	Euryarchaeota	2550	4	0	0	3	1	0	0	0
Methanococcoides burtonii (strain DSM 6242 / NBRC 107633 / OCM 468 / ACE-M)	Euryarchaeota	2200	14	1	10	5	0	0	0	0
Methanobrevibacter smithii (strain PS / ATCC 35061 / DSM 861)	Euryarchaeota	1778	1	0	0	0	1	0	0	0
Methanocaldococcus infernus (strain DSM 11812 / JCM 15783 / ME)	Euryarchaeota	1433	1	0	0	0	0	1	0	0
Methanoculleus marisnigri (strain ATCC 35101 / DSM 1498 / JR1)	Euryarchaeota	2458	3	0	0	3	0	0	0	0
Archaeoglobus fulgidus (strain ATCC 49558 / VC-16 / DSM 4304 / JCM 9628 / NBRC 100126)	Euryarchaeota	2396	11	0	0	10	0	0	1	0
Methanothermus fervidus (strain ATCC 43054 / DSM 2088 / JCM 10308 / V24 S)	Euryarchaeota	1277	0	0	0	0	0	0	0	0
Methanoregula boonei (strain 6A8)	Euryarchaeota	2431	4	0	0	2	2	0	0	0
Methanosphaerula palustris (strain ATCC BAA-1556 / DSM 19958 / E1-9c)	Euryarchaeota	2632	3	0	0	2	1	0	0	0
Methanosarcina	Euryarchaeota	4296	10	0	4	5	0	0	0	1

acetivorans (strain ATCC 35395 / DSM 2834 / JCM 12185 / C2A)										
Ferroglobus placidus (strain DSM 10642 / AEDIII2DO)	Euryarchaeota	2451	5	0	0	5	0	0	0	0
Natronomonas pharaonis (strain ATCC 35678 / DSM 2160) (Halobacterium pharaonis)	Euryarchaeota	2763	9	0	0	8	0	0	1	0
Halobacterium sp. DL1	Euryarchaeota	3243	7	0	0	7	0	0	0	0
Methanofollis liminatans DSM 4140	Euryarchaeota	2408	3	1	0	4	0	0	0	0
Methanothermobacter thermoautotrophicus (strain ATCC 29096 / DSM 1053 / JCM 10044 / NBRC 100330 / Delta H) (Methanobacterium thermoautotrophicum)	Euryarchaeota	1868	0	0	0	0	0	0	0	0
Methanoplanus limicola DSM 2279	Euryarchaeota	2907	7	0	1	4	1	0	0	1
Methanosaeta harundinacea (strain 6Ac)	Euryarchaeota	2341	4	0	0	2	1	0	0	1
Natrinema sp. (strain J7-2)	Euryarchaeota	4260	8	0	0	7	1	0	0	0
Methanolacinia petrolearia (strain DSM 11571 / OCM 486 / SEBR 4847) (Methanoplanus petrolearius)	Euryarchaeota	2767	10	0	2	7	0	0	0	1
Methanosalsum zhilinae (strain DSM 4017 / NBRC 107636 / OCM 62 / WeN5) (Methanohalophilus zhilinae)	Euryarchaeota	1956	16	0	0	12	4	0	0	0
Halogeometricum borinquense (strain ATCC 700274 / DSM 11551 / JCM 10706 / PR3)	Euryarchaeota	3865	21	0	0	20	1	0	0	0
Halopiger xanaduensis (strain DSM 18323 / JCM 14033 / SH-6)	Euryarchaeota	4191	11	0	0	10	0	0	0	1
Thermococcus barophilus (strain DSM 11836 / MP)	Euryarchaeota	2253	8	0	0	0	0	0	8	0
Methanobacterium lacus	Euryarchaeota	2477	1	0	1	0	0	0	0	0



(strain AL-21)										
Methanococcus voltae (strain ATCC BAA-1334 / A3)	Euryarchaeota	1653	1	0	0	0	0	1	0	0
Methanosaeta concilii (strain ATCC 5969 / DSM 3671 / JCM 10134 / NBRC 103675 / OCM 69 / GP-6) (Methanotherix concilii)	Euryarchaeota	2735	3	0	1	2	0	0	0	0
Archaeoglobus veneficus (strain DSM 11195 / SNP6)	Euryarchaeota	2049	5	0	0	5	0	0	0	0
Methanosarcina barkeri (strain Fusaro / DSM 804)	Euryarchaeota	3585	5	0	1	4	0	0	0	0
Haloferax volcanii (strain ATCC 29605 / DSM 3757 / JCM 8879 / NBRC 14742 / NCIMB 2012 / VKM B-1768 / DS2) (Halobacterium volcanii)	Euryarchaeota	3983	10	0	0	8	1	0	0	1
Methanobrevibacter ruminantium (strain ATCC 35063 / DSM 1093 / JCM 13430 / OCM 146 / M1) (Methanobacterium ruminantium)	Euryarchaeota	2204	1	0	0	0	0	0	0	1
Methanobacterium paludis (strain DSM 25820 / JCM 18151 / SWAN1)	Euryarchaeota	2377	1	0	1	0	0	0	0	0
Natronobacterium gregoryi (strain ATCC 43098 / CCM 3738 / NCIMB 2189 / SP2)	Euryarchaeota	3593	14	0	0	13	0	0	0	1
Methanoregula formicica (strain DSM 22288 / NBRC 105244 / SMSPP)	Euryarchaeota	2789	5	0	2	2	1	0	0	0
Natrinema pellirubrum (strain DSM 15624 / JCM 10476 / NCIMB 786)	Euryarchaeota	4094	6	0	0	5	1	0	0	0
Halovivax ruber (strain DSM 18193 / JCM 13892 / XH-70)	Euryarchaeota	3076	10	0	0	9	1	0	0	0
Methanomethylovorans hollandica (strain DSM 15978 / NBRC 107637 / DMS1)	Euryarchaeota	2531	3	0	0	2	0	0	0	1
Natronococcus occultus	Euryarchaeota	4115	6	0	0	6	0	0	0	0

SP4										
Natronomonas moolapensis (strain DSM 18674 / JCM 14361 / 8.8.11)	Euryarchaeota	2721	5	0	0	5	0	0	0	0
Thermoplasmatales archaeon BRNA1	Euryarchaeota	1525	1	0	0	0	1	0	0	0
Candidatus Methanomethylophilus alvus Mx1201	Euryarchaeota	1641	0	0	0	0	0	0	0	0
Archaeoglobus sulfaticallidus PM70-1	Euryarchaeota	2203	4	0	0	4	0	0	0	0
Methanobrevibacter sp. AbM4	Euryarchaeota	1661	1	0	0	1	0	0	0	0
Candidatus Methanomassiliicoccus intestinalis Isoire-Mx1	Euryarchaeota	1819	0	0	0	0	0	0	0	0
Salinarchaeum sp. Harcht-Bsk1	Euryarchaeota	2993	9	0	0	8	1	0	0	0
Halorhabdus tiamatea SARL4B	Euryarchaeota	2957	8	0	0	7	1	0	0	0
Methanobacterium sp. MB1	Euryarchaeota	2011	0	0	0	0	0	0	0	0
Thermoplasmatales archaeon I-plasma	Euryarchaeota	1685	1	0	0	0	0	0	0	1
Ferroplasma sp. Type II	Euryarchaeota	2636	0	0	0	0	0	0	0	0
Thermoplasmatales archaeon E-plasma	Euryarchaeota	1655	2	0	0	1	0	1	0	0
Thermoplasmatales archaeon Gpl	Euryarchaeota	1904	0	0	0	0	0	0	0	0
Thermoplasmatales archaeon A-plasma	Euryarchaeota	2267	0	0	0	0	0	0	0	0
Halostagnicola larsenii XH-48	Euryarchaeota	4058	14	0	1	12	1	0	0	0
Candidatus Methanoplasma termitum	Euryarchaeota	1393	0	0	0	0	0	0	0	0
Methanosarcina sp. MTP4	Euryarchaeota	3208	15	2	4	12	1	0	0	0
Geoglobus ahangari	Euryarchaeota	1961	8	0	2	6	0	0	0	0
Acidiplasma cupricumulans	Euryarchaeota	1680	0	0	0	0	0	0	0	0
Candidatus Methanoperedens sp. BLZ1	Euryarchaeota	4509	9	0	2	7	0	0	0	0
Halolamina pelagica	Euryarchaeota	3459	6	0	0	6	0	0	0	0
Methanolina sp. SDB	Euryarchaeota	2019	1	0	0	1	0	0	0	0
Methanosaeta sp. SDB	Euryarchaeota	1841	4	0	0	3	0	0	0	1
Methanomassiliicoccales archaeon RumEn M1	Euryarchaeota	1869	1	0	0	0	1	0	0	0
Methanoculleus sp. SDB	Euryarchaeota	2027	17	2	14	2	0	0	0	3
Korarchaeum	Korarchaeota	1599	0	0	0	0	0	0	0	0

cryptofilum (strain OPF8)										
Lokiarchaeum sp. GC14_75	Lokiarchaeota	5361	2	0	0	0	1	0	0	1
Nanoarchaeum equitans (strain Kin4-M)	Nanoarchaeota	535	1	0	1	0	0	0	0	0
Cenarchaeum symbiosum (strain A)	Thaumarchaeota	2016	8	1	0	1	0	0	0	8
Nitrosopumilus maritimus (strain SCM1)	Thaumarchaeota	1788	14	1	2	0	0	0	0	13
Candidatus Nitrosoarchaeum limnia SFB1	Thaumarchaeota	2031	14	1	0	0	0	0	0	15
<b>ΣYNOAO</b>	--	<b>309096</b>	<b>643</b>	<b>20</b>	<b>109</b>	<b>365</b>	<b>50</b>	<b>5</b>	<b>40</b>	<b>94</b>

## Παράρτημα Δ

**Πίνακας 8: ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΝΑΖΗΤΗΣΗΣ ΜΕ ΤΙΜΗ ΚΑΤΩΦΛΙΟΥ.**

ΟΡΓΑΝΙΣΜΟΣ	ΦΥΛΟ	ΟΛΙΚΟΣ ΑΡΙΘΜΟΣ ΠΡΩΤΕΪΝΩΝ	# ΠΡΩΤΕΪΝΩΝ ΜΕ 1 HIT	# ΠΡΩΤΕΪΝΩΝ ΜΕ > 1 HIT	#ΣΤΗΝ ΟΙΚΟΓΕΝΕΙΑ 3024	#ΣΤΗΝ ΟΙΚΟΓΕΝΕΙΑ 4126	#ΣΤΗΝ ΟΙΚΟΓΕΝΕΙΑ 4143	#ΣΤΗΝ ΟΙΚΟΓΕΝΕΙΑ 4173	#ΣΤΗΝ ΟΙΚΟΓΕΝΕΙΑ 4288	#ΣΤΗΝ ΟΙΚΟΓΕΝΕΙΑ 4296
Candidatus Bathyarchaeota archaeon BA2	Bathyarchaeota	1760	0	0	0	0	0	0	0	0
Candidatus Bathyarchaeota archaeon BA1	Bathyarchaeota	2401	0	0	0	0	0	0	0	0
miscellaneous Crenarchaeota group-15 archaeon DG-45	Crenarchaeota	1294	0	0	0	0	0	0	0	0
miscellaneous Crenarchaeota group-1 archaeon SG8-32-1	Crenarchaeota	1097	3	0	2	0	0	0	0	1
miscellaneous Crenarchaeota group-6 archaeon AD8-1	Crenarchaeota	1496	0	0	0	0	0	0	0	0
miscellaneous Crenarchaeota group archaeon SMTZ1-55	Crenarchaeota	1119	0	0	0	0	0	0	0	0
miscellaneous Crenarchaeota group archaeon SMTZ-80	Crenarchaeota	1357	0	0	0	0	0	0	0	0
Metallosphaera sedula (strain ATCC 51363 / DSM 5348)	Crenarchaeota	2246	0	0	0	0	0	0	0	0
Staphylothermus marinus (strain ATCC 43588 / DSM 3639 / JCM 9404 / F1)	Crenarchaeota	1566	0	0	0	0	0	0	0	0
Ignicoccus hospitalis (strain KIN4/I / DSM 18386 / JCM 14125)	Crenarchaeota	1427	0	0	0	0	0	0	0	0
Acidilobus saccharovorans (strain DSM 16705 / JCM 18335 / VKM B-2471 / 345-15)	Crenarchaeota	1495	0	0	0	0	0	0	0	0
Thermofilum pendens (strain Hrk 5)	Crenarchaeota	1870	0	0	0	0	0	0	0	0
Sulfolobus tokodaii (strain DSM 16993 / JCM 10545 / NBRC	Crenarchaeota	2762	0	0	0	0	0	0	0	0

100140 / 7)										
-------------	--	--	--	--	--	--	--	--	--	--

Sulfolobus acidocaldarius (strain ATCC 33909 / DSM 639 / JCM 8929 / NBRC 15157 / NCIMB 11770)	Crenarchaeota	2187	0	0	0	0	0	0	0	0
--	---------------	------	---	---	---	---	---	---	---	---







<i>Thermoproteus</i> sp. AZ2	Crenarchaeota	1697	0	0	0	0	0	0	0	0
<i>Methanocorpusculum labreanum</i> (strain ATCC 43576 / DSM 4855 / Z)	Euryarchaeota	1728	1	0	0	1	0	0	0	0
<i>Halalkalicoccus jeotgali</i> (strain DSM 18796 / CECT 7217 / JCM 14584 / KCTC 4019 / B3)	Euryarchaeota	3756	2	0	0	2	0	0	0	0
<i>Methanohalobium evestigatum</i> (strain DSM 3721 / OCM 161 / Z-7303)	Euryarchaeota	2236	9	0	0	2	7	0	0	0
<i>Picrophilus torridus</i> (strain ATCC 700027 / DSM 9790 / JCM 10055 / NBRC 100828)	Euryarchaeota	1532	0	0	0	0	0	0	0	0
<i>Methanobolus psychrophilus</i> R15	Euryarchaeota	3111	17	0	13	4	0	0	0	0
<i>Thermococcus kodakarensis</i> (strain ATCC BAA-918 / JCM 12380 / KOD1) ( <i>Pyrococcus kodakaraensis</i> (strain KOD1))	Euryarchaeota	2290	12	0	0	0	0	0	12	0
<i>Halobacterium salinarum</i> (strain ATCC 700922 / JCM 11081 / NRC-1) ( <i>Halobacterium halobium</i> )	Euryarchaeota	2415	4	0	0	4	0	0	0	0
<i>Methanococcus maripaludis</i> (strain S2 / LL)	Euryarchaeota	1695	0	0	0	0	0	0	0	0
<i>Methanosarcina mazei</i> (strain ATCC BAA-159 / DSM 3647 / Goe1 / Go1 / JCM 11833 / OCM 88) ( <i>Methanosarcina frisia</i> )	Euryarchaeota	3254	5	0	4	1	0	0	0	0
<i>Methanocella arvoryzae</i> (strain DSM 22066 / NBRC 105507 / MRE50)	Euryarchaeota	3064	4	0	0	4	0	0	0	0
<i>Methanosaeta thermophila</i> (strain DSM 6194 / JCM 14653 / NBRC 101360 / PT) ( <i>Methanotherix thermophila</i> )	Euryarchaeota	1661	1	0	0	1	0	0	0	0
<i>Halorubrum lacusprofundi</i> (strain ATCC 49239 / DSM 5036 / JCM 8891 /	Euryarchaeota	3465	9	0	0	9	0	0	0	0

ACAM 34)										
Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) (Methanococcus jannaschii)	Euryarchaeota	1787	1	0	0	0	0	1	0	0
Pyrococcus abyssi (strain GE5 / Orsay)	Euryarchaeota	1780	8	0	0	0	0	0	8	0
Pyrococcus furiosus (strain ATCC 43587 / DSM 3638 / JCM 8422 / Vc1)	Euryarchaeota	2044	8	0	0	0	0	0	8	0
Thermoplasma acidophilum (strain ATCC 25905 / DSM 1728 / JCM 9062 / NBRC 15155 / AMRC-C165)	Euryarchaeota	1481	0	0	0	0	0	0	0	0
Methanohalophilus mahii (strain ATCC 35705 / DSM 5219 / SLP)	Euryarchaeota	1972	8	0	0	3	5	0	0	0
Methanococcus aeolicus (strain Nankai-3 / ATCC BAA-1280)	Euryarchaeota	1480	0	0	0	0	0	0	0	0
Haloarcula marismortui (strain ATCC 43049 / DSM 3752 / JCM 8966 / VKM B-1809) (Halobacterium marismortui)	Euryarchaeota	4101	9	0	0	9	0	0	0	0
Aciduliprofundum boonei (strain DSM 19572 / T469)	Euryarchaeota	1532	12	0	12	0	0	0	0	0
Halomicrobium mukohataei (strain ATCC 700874 / DSM 12286 / JCM 9738 / NCIMB 13541) (Haloarcula mukohataei)	Euryarchaeota	3321	6	0	0	6	0	0	0	0
Methanopyrus kandleri (strain AV19 / DSM 6324 / JCM 9639 / NBRC 100938)	Euryarchaeota	1683	0	0	0	0	0	0	0	0
Natrialba magadii (strain ATCC 43099 / DSM 3394 / NCIMB 2190 / MS3) (Natronobacterium magadii)	Euryarchaeota	4167	6	0	0	6	0	0	0	0
Methanocella paludicola	Euryarchaeota	2987	4	0	0	4	0	0	0	0

(strain DSM 17711 / JCM 13418 / NBRC 101707 / SANAE)										
<i>Archaeoglobus profundus</i> (strain DSM 5631 / JCM 9629 / NBRC 100127 / Av18)	Euryarchaeota	1813	2	0	0	2	0	0	0	0
<i>Haloterrigena turkmenica</i> (strain ATCC 51198 / DSM 5511 / NCIMB 13204 / VKM B-1734) ( <i>Halococcus turkmenicus</i> )	Euryarchaeota	5070	12	0	0	12	0	0	0	0
<i>Methanosphaera stadtmanae</i> (strain ATCC 43021 / DSM 3091 / JCM 11832 / MCB-3)	Euryarchaeota	1531	0	0	0	0	0	0	0	0
<i>Methanospirillum hungatei</i> JF-1 (strain ATCC 27890 / DSM 864 / NBRC 100397 / JF-1)	Euryarchaeota	3070	1	0	0	1	0	0	0	0
<i>Haloquadratum walsbyi</i> (strain DSM 16790 / HBSQ001)	Euryarchaeota	2550	2	0	0	2	0	0	0	0
<i>Methanococcoides burtonii</i> (strain DSM 6242 / NBRC 107633 / OCM 468 / ACE-M)	Euryarchaeota	2200	13	0	10	3	0	0	0	0
<i>Methanobrevibacter smithii</i> (strain PS / ATCC 35061 / DSM 861)	Euryarchaeota	1778	0	0	0	0	0	0	0	0
<i>Methanocaldococcus infernus</i> (strain DSM 11812 / JCM 15783 / ME)	Euryarchaeota	1433	1	0	0	0	0	1	0	0
<i>Methanoculleus marisnigri</i> (strain ATCC 35101 / DSM 1498 / JR1)	Euryarchaeota	2458	2	0	0	2	0	0	0	0
<i>Archaeoglobus fulgidus</i> (strain ATCC 49558 / VC-16 / DSM 4304 / JCM 9628 / NBRC 100126)	Euryarchaeota	2396	6	0	0	5	0	0	1	0
<i>Methanothermus fervidus</i> (strain ATCC 43054 / DSM 2088 / JCM 10308 / V24 S)	Euryarchaeota	1277	0	0	0	0	0	0	0	0
<i>Methanoregula boonei</i> (strain 6A8)	Euryarchaeota	2431	1	0	0	1	0	0	0	0
<i>Methanosphaerula palustris</i> (strain ATCC	Euryarchaeota	2632	1	0	0	1	0	0	0	0

BAA-1556 / DSM 19958 / E1-9c)										
Methanosarcina acetivorans (strain ATCC 35395 / DSM 2834 / JCM 12185 / C2A)	Euryarchaeota	4296	5	0	2	3	0	0	0	0
Ferroglobus placidus (strain DSM 10642 / AEDII12DO)	Euryarchaeota	2451	2	0	0	2	0	0	0	0
Natronomonas pharaonis (strain ATCC 35678 / DSM 2160) (Halobacterium pharaonis)	Euryarchaeota	2763	7	0	0	6	0	0	1	0
Halobacterium sp. DL1	Euryarchaeota	3243	3	0	0	3	0	0	0	0
Methanofollis liminatans DSM 4140	Euryarchaeota	2408	1	0	0	1	0	0	0	0
Methanothermobacter thermautotrophicus (strain ATCC 29096 / DSM 1053 / JCM 10044 / NBRC 100330 / Delta H) (Methanobacterium thermoautotrophicum)	Euryarchaeota	1868	0	0	0	0	0	0	0	0
Methanoplanus limicola DSM 2279	Euryarchaeota	2907	2	0	1	1	0	0	0	0
Methanosaeta harundinacea (strain 6Ac)	Euryarchaeota	2341	1	0	0	1	0	0	0	0
Natrinema sp. (strain J7-2)	Euryarchaeota	4260	6	0	0	6	0	0	0	0
Methanolacinia petrolearia (strain DSM 11571 / OCM 486 / SEBR 4847) (Methanoplanus petrolearius)	Euryarchaeota	2767	4	0	0	4	0	0	0	0
Methanosalsum zhilinae (strain DSM 4017 / NBRC 107636 / OCM 62 / WeN5) (Methanohalophilus zhilinae)	Euryarchaeota	1956	5	0	0	5	0	0	0	0
Halogeometricum borinquense (strain ATCC 700274 / DSM 11551 / JCM 10706 / PR3)	Euryarchaeota	3865	18	0	0	18	0	0	0	0
Halopiger xanaduensis (strain DSM 18323 / JCM 14033 / SH-6)	Euryarchaeota	4191	6	0	0	6	0	0	0	0
Thermococcus	Euryarchaeota	2253	9	0	0	0	0	0	9	0

barophilus (strain DSM 11836 / MP)										
Methanobacterium lacus (strain AL-21)	Euryarchaeota	2477	0	0	0	0	0	0	0	0
Methanococcus voltae (strain ATCC BAA-1334 / A3)	Euryarchaeota	1653	0	0	0	0	0	0	0	0
Methanosaeta concilii (strain ATCC 5969 / DSM 3671 / JCM 10134 / NBRC 103675 / OCM 69 / GP-6) (Methanotherix concilii)	Euryarchaeota	2735	1	0	0	1	0	0	0	0
Archaeoglobus veneficus (strain DSM 11195 / SNP6)	Euryarchaeota	2049	4	0	0	4	0	0	0	0
Methanosarcina barkeri (strain Fusaro / DSM 804)	Euryarchaeota	3585	3	0	1	2	0	0	0	0
Haloferax volcanii (strain ATCC 29605 / DSM 3757 / JCM 8879 / NBRC 14742 / NCIMB 2012 / VKM B-1768 / DS2) (Halobacterium volcanii)	Euryarchaeota	3983	8	0	0	8	0	0	0	0
Methanobrevibacter ruminantium (strain ATCC 35063 / DSM 1093 / JCM 13430 / OCM 146 / M1) (Methanobacterium ruminantium)	Euryarchaeota	2204	0	0	0	0	0	0	0	0
Methanobacterium paludis (strain DSM 25820 / JCM 18151 / SWAN1)	Euryarchaeota	2377	1	0	0	0	0	0	1	0
Natronobacterium gregoryi (strain ATCC 43098 / CCM 3738 / NCIMB 2189 / SP2)	Euryarchaeota	3593	9	0	0	9	0	0	0	0
Methanoregula formicica (strain DSM 22288 / NBRC 105244 / SMSP)	Euryarchaeota	2789	2	0	0	2	0	0	0	0
Natrinema pellirubrum (strain DSM 15624 / JCM 10476 / NCIMB 786)	Euryarchaeota	4094	5	0	0	5	0	0	0	0
Halovivax ruber (strain DSM 18193 / JCM 13892 / XH-70)	Euryarchaeota	3076	6	0	0	6	0	0	0	0
Methanomethylovorans	Euryarchaeota	2531	2	0	0	2	0	0	0	0

hollandica (strain DSM 15978 / NBRC 107637 / DMS1)										
Natronococcus occultus SP4	Euryarchaeota	4115	4	0	0	4	0	0	0	0
Natronomonas moolapensis (strain DSM 18674 / JCM 14361 / 8.8.11)	Euryarchaeota	2721	3	0	0	3	0	0	0	0
Thermoplasmatales archaeon BRNA1	Euryarchaeota	1525	0	0	0	0	0	0	0	0
Candidatus Methanomethylophilus alvus Mx1201	Euryarchaeota	1641	0	0	0	0	0	0	0	0
Archaeoglobus sulfatocalidus PM70-1	Euryarchaeota	2203	2	0	0	2	0	0	0	0
Methanobrevibacter sp. AbM4	Euryarchaeota	1661	0	0	0	0	0	0	0	0
Candidatus Methanomassiliicoccus intestinalis Issoire-Mx1	Euryarchaeota	1819	0	0	0	0	0	0	0	0
Salinarchaeum sp. Harcht-Bsk1	Euryarchaeota	2993	7	0	0	7	0	0	0	0
Halorhabdus tiamatea SARL4B	Euryarchaeota	2957	5	0	0	5	0	0	0	0
Methanobacterium sp. MB1	Euryarchaeota	2011	0	0	0	0	0	0	0	0
Thermoplasmatales archaeon I-plasma	Euryarchaeota	1685	0	0	0	0	0	0	0	0
Ferroplasma sp. Type II	Euryarchaeota	2636	0	0	0	0	0	0	0	0
Thermoplasmatales archaeon E-plasma	Euryarchaeota	1655	0	0	0	0	0	0	0	0
Thermoplasmatales archaeon Gpl	Euryarchaeota	1904	0	0	0	0	0	0	0	0
Thermoplasmatales archaeon A-plasma	Euryarchaeota	2267	0	0	0	0	0	0	0	0
Halostagnicola larsenii XH-48	Euryarchaeota	4058	7	0	0	7	0	0	0	0
Candidatus Methanoplasma termitum	Euryarchaeota	1393	0	0	0	0	0	0	0	0
Methanosarcina sp. MTP4	Euryarchaeota	3208	8	0	4	4	0	0	0	0
Geoglobus ahangari	Euryarchaeota	1961	6	0	1	5	0	0	0	0
Acidiplasma cupricumulans	Euryarchaeota	1680	0	0	0	0	0	0	0	0
Candidatus Methanoperedens sp. BLZ1	Euryarchaeota	4509	4	0	2	2	0	0	0	0
Halolamina pelagica	Euryarchaeota	3459	5	0	0	5	0	0	0	0
Methanolinea sp. SDB	Euryarchaeota	2019	1	0	0	1	0	0	0	0
Methanosaeta sp. SDB	Euryarchaeota	1841	1	0	0	1	0	0	0	0

Methanomassiliicoccales archaeon RumEn M1	Euryarchaeota	1869	0	0	0	0	0	0	0	0
Methanoculleus sp. SDB	Euryarchaeota	2027	9	0	6	1	0	0	0	2
Korarchaeum cryptofilum (strain OPF8)	Korarchaeota	1599	0	0	0	0	0	0	0	0
Lokiarchaeum sp. GC14 75	Lokiarchaeota	5361	0	0	0	0	0	0	0	0
Nanoarchaeum equitans (strain Kin4-M)	Nanoarchaeota	535	0	0	0	0	0	0	0	0
Cenarchaeum symbiosum (strain A)	Thaumarchaeota	2016	8	0	0	0	0	0	0	8
Nitrosopumilus maritimus (strain SCM1)	Thaumarchaeota	1788	13	0	0	0	0	0	0	13
Candidatus Nitrosoarchaeum limnia SFB1	Thaumarchaeota	2031	14	0	0	0	0	0	0	14
<b>ΣΥΝΟΛΟ</b>	--	<b>309096</b>	<b>407</b>	<b>0</b>	<b>59</b>	<b>227</b>	<b>12</b>	<b>2</b>	<b>43</b>	<b>64</b>