

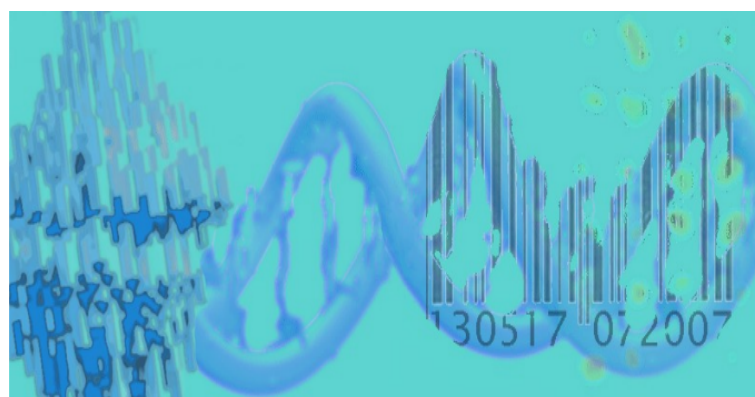
ΣΧΟΛΗ ΕΠΙΣΤΗΜΗΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ.
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Ανάπτυξη μαθηματικού υποβάθρου για τη βιοπληροφορική

Πτυχιακή εργασία

ΤΟΥ

Σταύρου Μαραγγοτίδη



Ονοματεπώνυμο επιβλέποντος καθηγητή: **ΠΑΝΑΓΙΩΤΗΣ ΒΛΑΜΟΣ**

Ορεστιάδα, 2023

ΠΝΕΥΜΑΤΙΚΑ ΔΙΚΑΙΩΜΑΤΑ

Copyright © [Σταύρος Μαραγγοτίδης, 2023]

Με επιφύλαξη παντός δικαιώματος

Απαγορεύεται η αντιγραφή μέρους ή του συνόλου της παρούσας διατριβής. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για εκπαιδευτικό ή ερευνητικό σκοπό, μη κερδοσκοπικού χαρακτήρα, με την προϋπόθεση να αναφέρεται η πηγή προέλευσης.

Copyright © [Stavros Marangotidis, 2023]

The approval of the Bachelor's Thesis by the Department of Primary Education, Department of Informatics of the Ionian University, does not necessarily indicate the acceptance of the views of the author.

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ

Δηλώνω υπεύθυνα και γνωρίζοντας τις κυρώσεις του Ν. 2121/1993 περί της Πνευματικής Ιδιοκτησίας, ότι η παρούσα Πτυχιακή εργασία αποτελεί ένα εξ ολοκλήρου αποτέλεσμα δικής μου ερευνητικής εργασίας, δεν αποτελεί προϊόν αντιγραφής αλλά ούτε και προέρχεται από ανάθεση σε τρίτους. Όλες οι πηγές, που χρησιμοποιήθηκαν (κάθε είδους, μορφής και προέλευσης) για τη συγγραφή της, συμπεριλαμβάνονται μέσα στη βιβλιογραφία αναλυτικά.

ΣΤΑΥΡΟΣ ΜΑΡΑΓΓΟΤΙΔΗΣ

Υπογραφή

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να εκφράσω τις ευχαριστίες μου σε όλους/ες όσοι/ες με βοήθησαν στο δύσκολο αυτό εγχείρημα σύνθεσης κι ολοκλήρωσης της Πτυχιακής μου εργασίας. Αρχικά θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή της παρούσας εργασίας μου, τον κύριο ΠΑΝΑΓΙΩΤΗ ΒΛΑΜΟ, του οποίου η (υπο-)στήριξη ήταν καθοριστική. Οι πολύτιμες συμβουλές του, η εμπιστοσύνη του και η μεταξύ μας επικοινωνία-συνεργασία συνέβαλαν καίρια στο να την ολοκληρώσω.

Τέλος θα ήθελα να ευχαριστήσω τους/τις φίλους/ες, την οικογένειά μου και όλους/ες όσοι/ες με στήριξαν ψυχολογικά και ηθικά και υποστηρικτικά.

ΜΑΡΑΓΓΟΤΙΔΗΣ ΣΤΑΥΡΟΣ

Ιούνιος, 2023.

Πίνακας περιεχομένων

<u>ΠΝΕΥΜΑΤΙΚΑ ΔΙΚΑΙΩΜΑΤΑ</u>	1
<u>ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ</u>	2
<u>ΕΥΧΑΡΙΣΤΙΕΣ</u>	3
<u>ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ</u>	6
<u>ΠΕΡΙΛΗΨΗ</u>	7
<u>ABSTRACT</u>	8
<u>ΛΕΞΕΙΣ-ΚΛΕΙΔΙΑ</u>	9
<u>KEYWORDS</u>	9
<u>ΚΕΦΑΛΑΙΟ 1 – ΕΙΣΑΓΩΓΗ</u>	10
<u>1.1 ΣΚΟΠΟΣ ΕΡΓΑΣΙΑΣ</u>	14
<u>1.2 ΔΙΑΡΘΡΩΣΗ ΕΡΓΑΣΙΑΣ</u>	15
<u>ΚΕΦΑΛΑΙΟ 2 -ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ</u>	16
<u>2.1 ΟΡΙΣΜΟΣ & ΕΡΜΗΝΕΙΑ ΤΗΣ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ</u>	16
<u>2.2 ΒΑΣΙΚΟΙ ΤΟΜΕΙΣ ΤΗΣ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ</u>	20
<u>2.3 ΑΝΑΓΚΑΙΟΤΗΤΑ ΤΗΣ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ</u>	21
<u>2.4 ΣΤΟΧΟΙ ΤΗΣ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ</u>	22
<u>2.5 Ο ΔΙΕΠΙΣΤΗΜΟΝΙΚΟΣ ΧΑΡΑΚΤΗΡΑΣ ΤΗΣ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ</u>	24
<u>2.6 ΠΕΔΙΑ ΕΡΕΥΝΑΣ ΤΗΣ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ</u>	25
<u>2.7 ΕΡΓΑΛΕΙΑ ΤΗΣ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ</u>	35
<u>2.7.1 Ο αλγόριθμος FASTA</u>	36
<u>2.7.2 Ο αλγόριθμος BLAST</u>	38
<u>2.7.3 Το Clustal</u>	42
<u>2.8 ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ</u>	43

<u>ΚΕΦΑΛΑΙΟ 3 – ΤΑ ΜΑΘΗΜΑΤΙΚΑ</u>	47
<u>3.1 ΓΡΑΜΜΙΚΗ ΑΛΓΕΒΡΑ</u>	47
<u>3.2 ΔΙΑΦΟΡΙΚΕΣ ΕΞΙΣΩΣΕΙΣ</u>	49
<u>3.3 ΣΤΑΤΙΣΤΙΚΗ</u>	49
<u>3.4 ΔΙΑΚΡΙΤΑ ΜΑΘΗΜΑΤΙΚΑ</u>	50
<u>ΚΕΦΑΛΑΙΟ 4 - Η ΣΥΜΒΟΛΗ-ΣΥΝΕΙΣΦΟΡΑ ΤΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΣΤΗ</u> <u>ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ</u>	51
<u>ΣΥΜΠΕΡΑΣΜΑΤΑ</u>	55
<u>ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ</u>	56

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικ.	1	–	Η	Βιοπληροφορική	σελ. 15
Εικ.	2	–	Το σύστημα	της διπλής έλικας	σελ. 18
Εικ.	3	–	Η συνεργασία	της Βιοπληροφορικής με άλλες επιστήμες	σελ. 20
Εικ. 4–	Το προφίλ ενός διεπιστημονικού ερευνητή, που ασχολείται με την ανάλυση βιολογικών δεδομένων με χρήση της πληροφορικής και της στατιστικής σελ. 21				
Εικ.	5	-	Η	τεχνολογία microarrays	σελ. 24
Εικ. 6 -	Σχηματική απεικόνιση των σταδίων ανάλυσης ενός πειράματος γονιδιακής έκφρασης από την αποκομιδή των πρωτογενών δεδομένων ως τη δημιουργία ομάδων γονιδίων με χαρακτηριστικά πρότυπα έκφρασης σελ. 34				
Εικ. 7 –	Διαγραμματική σχεδίαση (τα κύρια στάδια-βήματα) του αλγόριθμου Fasta σελ. 38				
Εικ.	8	–	Σχηματική αποτύπωση	του αλγόριθμου Blast	σελ. 40
Εικ.	9	–	Τα βασικά στάδια-βήματα	του Blast	σελ. 41
Εικ.	10	–	Εξατομικευμένη	Ιατρική	σελ. 45

Εικ.	11	–	Ορθολογικός	σχεδιασμός	φαρμάκων
.....σελ. 45					

ΠΕΡΙΛΗΨΗ

Η Βιοπληροφορική είναι ένας επιστημονικός κλάδος εντός του οποίου έχουν συμπράξει οι επιστήμες της βιολογίας με την πληροφορική, τη στατιστική και τα μαθηματικά υπό ένα πλαίσιο καινοτόμο, το οποίο δίνει τη δυνατότητα εξίσου καινοτόμων μεθόδων προσέγγισης, ανάλυσης των διαφόρων βιολογικής φύσης προβλημάτων αλλά και αντίληψης του έμβιου κόσμου. Η Βιοπληροφορική όμως ως επί το πλείστον αξιοποιήθηκε περισσότερο στα της έρευνας της βιολογίας αλλά και των επιστημών υγείας με την αρωγή ιδιαίτερα των μαθηματικών, γεγονός που έδωσε τεράστια ώθηση ειδικά στην ιατρική και τη φαρμακολογία και στην προσπάθειά τους να καταπολεμήσουν διάφορες ασθένειες πιο στοχευμένα. Αυτός άλλωστε ήταν και ο κύριος σκοπός της παρούσας εργασίας, δηλαδή να μελετήσει βιβλιογραφικά την επιστήμη της Βιοπληροφορικής και τους στόχους της, με έμφαση στη συμβολή των μαθηματικών στη Βιοπληροφορική και τα εξαγόμενα οφέλη τους. Ο πρωταρχικός στόχος της Βιοπληροφορικής λοιπόν είναι να εφαρμοστούν οι κανόνες των επιστημών της τεχνολογίας και της πληροφορικής, προκειμένου να γίνει εφικτό να πραγματοποιηθούν τα μεγαλύτερα και τα πιο πολύπλοκα δεδομένα των επιστημών ζωής ως περισσότερα αντιληπτά κι εύχρηστα, κι έτσι να διαφανεί και η μέγιστη ισχύς όλων αυτών των δεδομένων. Μεταξύ των συμπερασμάτων, που κατέληξε η παρούσα Βιβλιογραφικής φύσης έρευνα ήταν πως πράγματι τα μαθηματικά

έχουν συμβάλει καταλυτικά στην επιστήμη της Βιοπληροφορικής και την προσπάθεια αυτής να βοηθήσει ειδικά στον χώρο της ιατρικής.

ABSTRACT

Bioinformatics is a scientific department, in which the sciences of biology have colluded with the computer science, statistics and mathematics under an innovative framework. This framework has the potential for equally innovative approaching methods, analysis of the various biological problems and perception of the living world. Bioinformatics was mostly used in biology research and in health science with the help of mathematics in particular, a fact that gave a huge boost especially in medicine and pharmacology and their efforts to deal with various diseases. The main objective of this research was to bibliographically study the science of Bioinformatics and its goals, emphasizing on the contribution of mathematics to Bioinformatics and their derived benefits. So, the primary objective of Bioinformatics is to apply the rules of the sciences of technology and computer science, in order to process the largest and most complicated data of life sciences as more conceivable, easy to use and in that way, the maximum power of all those data will be transparent. Among the conclusions reached by the bibliographic research was the fact that mathematics has indeed catalytically contributed to the science of Bioinformatics and its effort to help, especially in the field of medicine.

ΛΕΞΕΙΣ-ΚΛΕΙΔΙΑ: Βιοπληροφορική, Βιολογία, μαθηματικά,
ηλεκτρονικοί υπολογιστές, μεθοδολογία.

KEYWORDS: Bioinformatics, Biology, mathematics, computers, methodology.

ΚΕΦΑΛΑΙΟ 1 – ΕΙΣΑΓΩΓΗ

Η ραγδαία ανάπτυξη της επιστήμης της Πληροφορικής και των ΤΠΕ γενικότερα, που παρατηρείται στα τελευταία χρόνια στο χώρο της επιστημονικής κοινότητας, έδειξε πως οι δυνατότητες και οι επιλογές που παρέχει, θα ήταν πολύ χρήσιμο να αξιοποιηθούν αποτελεσματικά κι από άλλες επιστήμες όπως τη Βιολογία, τη Βιοτεχνολογία και την Ιατρική. Μάλιστα η μοριακή βιολογία με την ανάλυση και την ενδεδειγμένη χαρτογράφηση με ολοένα και πιο αυξημένα γονιδιώματα οδήγησε προς τη συγκέντρωση αναρίθμητων βιολογικών δεδομένων-ευρημάτων. Οι πρώτες όμως συζητήσεις κι ενέργειες σχετικά με την υλοποίηση χαρτογράφησης του ανθρώπινου γονιδιώματος συνάμα διεξήχθησαν κατά το έτος 1986, όπου κι έγινε και η πρώτη παρουσίαση του προσχεδίου κατά το έτος 2001 με δύο (2) προγράμματα, τα οποία και ήταν τα εξής, 1^ο το IHGSC (International Human Genome Sequencing Consortium) και 2^ο η Celera με τον Venter¹. Τα δύο αυτά προγράμματα κατά το έτος 2001 δηλαδή το HGP και η Celera ανακοινώνουν επίσημα το προσχέδιο του ανθρώπινου γονιδιώματος σε Nature και Science αντιστοίχως².

¹ <https://opencourses.auth.gr>

² Στο ίδιο.

Όμως πιο αναλυτικά στο σημείο αυτό, θα ήταν εξόχως χρήσιμο να παρατεθούν κάποια στοιχεία σχετικά με τα γονίδια. Αυτά λοιπόν τα γονίδια, τα οποία κι εποπτεύουν μια οριοθετημένη λειτουργία ή και χαρακτήρα, έχουν τεθεί ως ένα μικρό μέρος του DNA. Το καθένα εξ αυτών γονίδιο ενέχει και μιας δικής του θέσης επάνω σε ένα ορισμένο χρωμόσωμα. Η προαναφερθείσα «χαρτογράφηση» του γονιδιώματος έχει σχέση με την προσπάθεια, που έγινε, ώστε να γίνει εφικτό να «αποκωδικοποιηθεί» αλλά και να καταγραφεί η μοναδική θέση που ενείχε το εκάστοτε γονίδιο, καθώς και η αλληλουχία των βάσεων του DNA προκειμένου να καταστούν όλες αυτές οι πληροφορίες που εμπεριέχονται σε αυτά και γνωστές (Geoffrey & Robert, 2011). Η γονιδιωματική έχει τεθεί άλλωστε ως ένα τμήμα της γενετικής επιστήμης, που επεξεργάζεται το γονιδίωμα των οργανισμών, εντάσσοντας και την αλληλούχιση αλλά και την χαρτογράφηση του γονιδιώματος. Σε όλα τα τελευταία χρόνια λοιπόν έχουν γίνει πάρα πολλές προσπάθειες προκειμένου να χαρτογραφηθεί το ανθρώπινο DNA. Η χαρτογράφηση έκανε τα πρώτα της βήματα, όπως ακριβώς καταγράφηκαν παραπάνω, δηλαδή ξεκίνησε κατά το έτος 1986 όμως επίσημο χαρακτήρα έλαβαν οι κινήσεις κατά το έτος 1990 με την πρωτοβουλία από πλευράς της ακαδημαϊκής κοινότητας, και με την οργάνωση του Προγράμματος (Καταγραφής) του Ανθρώπινου Γονιδιώματος, γνωστό ως Human Genome Project (HGP), αλλά την ιδιωτική ενέργεια με την εταιρεία Celera Genomics. Κατά το έτος 2000 και με τη συμβολή του αλγορίθμου αλληλούχισης BLAT (BLAST-like alignment tool) έγινε πράξη ένα πρώτο γενικό σχέδιο αλληλούχισης του ανθρώπινου γονιδιώματος (draft genome). Ακολούθησε κατά το έτος 2001, αφενός η ομάδα του Human Genome Project που έκανε γνωστά τα αποτελέσματά της στο περιοδικό Science, αφετέρου η ομάδα της Celera Genomics που κινήθηκε παρομοίως ανακοινώνοντας τα αποτελέσματά της στο περιοδικό Nature. Το όλο Πρόγραμμα (Καταγραφής) του Ανθρώπινου Γονιδιώματος ολοκλήρωσε τις εργασίες του κατά το έτος 2004. Έως και στις μέρες μας πάντως έχει ολοκληρωθεί η αλληλούχιση του ανθρώπινου γονιδιώματος, ένα στοιχείο που (καθ-)οδήγησε στην αποκόμιση μιας πλειάδας

από οφέλη για όλη την ανθρωπότητα και συνάμα έδωσε και την ευκαιρία στη διαμόρφωση κι ανάπτυξη διαφόρων άλλων επιστημονικών κλάδων (Κελεφιώτης, 2002 · Geoffrey & Robert, 2011). Στο σημείο αυτό όμως πρέπει να αναφερθεί το γεγονός πως το ανθρώπινο γονιδίωμα προστατεύεται και από τις νομοθεσίες όλων των χωρών, και μάλιστα σε διεθνές επίπεδο, βάσει των άρθρων 1 και 2 αντίστοιχα, της Οικουμενικής Διακήρυξης της UNESCO για την προστασία του ανθρώπινου γονιδιώματος αλλά και της ανθρώπινης αξιοπρέπειας (UNESCO Universal declaration on the human genome and human rights), το οποίο αποτελεί ουσιαστικά ένα κείμενο με μη νομικά δεσμευτικό χαρακτήρα (soft law), μιας και αξιοποιείται με τη «συνείδηση δικαίου» και υπό το πρίσμα πως ενέχει δέσμευσης, εν τέλει το ανθρώπινο γονιδίωμα έχει τεθεί ως ένα είδος κληρονομιάς για όλη την ανθρωπότητα. Κι αυτό φυσικά μεταφράζεται συμβολικά, ότι δηλαδή ο εκάστοτε άνθρωπος διατηρεί το αναφαίρετο δικαίωμά του να είναι σεβαστός κι ο ίδιος όπως επίσης και η αξιοπρέπειά του και η μοναδικότητά του ανεξάρτητα από τα όποια τυχόν γενετικά του γνωρίσματα. Σε επίπεδο ελληνικής νομοθεσίας προστατεύεται εν γένει ο σεβασμός και η προστασία της αξίας αλλά και της μοναδικότητας της κάθε ανθρώπινης προσωπικότητας (άρθρο 2 παρ. 1 του Ελληνικού Συντάγματος) (Φραγκουδάκη, 2007).

Σύμφωνα άλλωστε και με την έρευνα των Καναβάκη & Ξαϊδάρα (2001) η χαρτογράφηση του ανθρώπινου γονιδιώματος (βλ. Εικ. 1) όντας έως και στις μέρες μας το πιο μεγάλο τόλμημα σε επίπεδο της ιστορίας της Βιολογίας, πρέπει να τονιστεί πως αυτό μετέβαλε όλη την ιατρική πράξη. Ειδικότερα το ανθρώπινο γονιδίωμα συμπεριλαμβάνει πλήθος από γενετικές κι αποτελείται από περίπου τρία (3) δισεκατομμύρια βάσεις, οι οποίες και μοιράζονται στα 22 αυτό-σωματικά χρωμοσώματα αλλά και τα δύο (2) χρωμοσώματα φύλου. Έχει εκτιμηθεί μάλιστα πως ήδη υπάρχουν από 50.000 έως και 100.000 γονίδια, εκ των οποίων όλες όσες περιοχές είναι κωδικοποιημένες αυτές και καταλαμβάνουν το 2% με 3% του συνόλου του DNA ως του κύριου τμήματος της ζωής. Το ανθρώπινο γονιδίωμα αποτελεί ένα γνώρισμα του ανθρώπινου

είδους, δεν ενέχει μιας απόλυτης ομοιότητας, όμως σε μια πλειάδα γενετικών τοποθεσιών παρουσιάζονται από 2 έως και 3 αλληλόμορφα, οι λεγόμενοι αλληλομορφισμοί. Ιστορικά θα ήταν χρήσιμο να αναφερθούν ορισμένα στοιχεία, ότι δηλαδή η χαρτογράφηση του ανθρώπινου γονιδιώματος ξεκίνησε κατά το έτος 1990 με πρωτοβουλία της ακαδημαϊκής κοινότητας όμως στη συνέχεια σειρά πήρε και η ιδιωτική πρωτοβουλία κατά το έτος 1998. Οι δύο αυτές ομάδες λοιπόν παρουσίασαν τα αποτελέσματά τους σε έναν κοινό χρόνο (2000 και 2001) -όπως ήδη προαναφέρθηκε-, με τα οφέλη έως και στις μέρες μας να σχετίζονται με τη γενετικής φύσης διάγνωση αλλά και τις νέες – καινοτόμες πρακτικές θεραπείας (Καναβάκης & Ξαϊδάρα, 2001). Πιο συγκεκριμένα η γενετική διάγνωση αξιοποιείται πολύ στην πρόληψη προγεννητικής και προεμφυτευτικής διάγνωσης καθώς και την προκλινική διάγνωση μονογονιαδιακών νοσημάτων. Ένα πολύ σημαντικό στοιχείο, που πρέπει να αναφερθεί, είναι το γεγονός πως 1.000 περίπου γενετικά νοσήματα έχουν καταφέρει να διερευνηθούν σε μοριακό επίπεδο, στη βάση πως έγινε εφικτό να απομονωθεί το κύριο – υπεύθυνο γονίδιο για την εκδήλωση του νοσήματος καθώς και οι αντίστοιχες μεταλλάξεις. Σε ό, τι είχε να κάνει τα των πολυγονιαδικών νοσημάτων, ο εντοπισμός της γενετικής προδιάθεσης συνδυαστικά με την προσπάθεια αποφυγής ορισμένων περιβαλλοντικών επιρροών λέχθηκε πως θα οδηγούσε στον περιορισμό του κινδύνου εμφάνισης του νοσήματος ή ακόμη και στην αναστολή της κλιμάκωσης της όλης του κλινικής εικόνας. Σε ό, τι αφορά τώρα στην προσδοκώμενη θεραπεία των νοσημάτων υπάρχει διάχυτη η ελπίδα να καλυτερεύσει με τη συμβολή της οργάνωσης και της παραγωγής των κατάλληλων φαρμάκων, προκειμένου με τον τρόπο αυτόν να μπορέσει να ενισχυθεί – αυξηθεί η αποτελεσματικότητα και συνάμα να μειωθούν και οι όποιες ανεπιθύμητες παρενέργειες των φαρμάκων. Επιπρόσθετα, η επιδιωκόμενη αυτή γονιδιακή θεραπευτική αγωγή εκτιμάται πως θα καταφέρει να διασφαλίσει μια καθόλα και δη καταλυτική διαχείριση – καταπολέμηση των γενετικών νοσημάτων και τα νουκλεϊκά οξέα παράλληλα, υπό μια μέθοδο ανάλογης με εκείνης των μονοκλωνικών αντισωμάτων, κι έτσι

θα αξιοποιηθούν σε πολλά άλλα ακόμη νοσήματα (Καναβάκης & Ξαϊδάρη, 2001). Μια πολύ στοχευμένη βοήθεια στην ολοκλήρωση της χαρτογράφησης του ανθρώπινου γονιδιώματος δείχνει να προσφέρει η συγκριτική γονιδιωματική, εν ολίγοις επρόκειτο για την παράλληλης μορφής χαρτογράφηση του γονιδιώματος των πρότυπων όμως οργανισμών, η οποία με τη σειρά συμβάλλει μέσω των παρεχόμενων πληροφοριών της στην πολύπλευρη αναζήτηση γονιδίων που τίθενται ως υπεύθυνα για την πρόκληση κυρίως των πολυγονιδιακών νοσημάτων κι επίσης για το άμεσο «περιβάλλον» των γονιδίων. Τέλος σχετικά με τα των ηθικών, νομικών και κοινωνικών συνεπειών εκ της παραπάνω μελετούμενης χαρτογράφησης του ανθρώπινου γονιδιώματος, συνηθέστερα επεξεργάζονται όλα αυτά (όπως π.χ. το πρόγραμμα Legal κλπ.), έχοντας ως απώτερο σκοπό να γίνει εφικτό να θεσπιστεί ένα ανάλογο νομικό πλαίσιο και να υιοθετηθούν ακόμη συγκεκριμένοι κώδικες κοινωνικής συμπεριφοράς, οι οποίοι με τη σειρά τους θα μπορούν να διασφαλίζουν το απόρρητο με εχεμύθεια, τα ατομικά δεδομένα – στοιχεία αλλά και την ιδιοκτησία της κάθε γενετικής πληροφορίας με ένα απόλυτο σεβασμό προς όλα δικαιώματα των ανθρώπων (Καναβάκης & Ξαϊδάρη, 2001).

Θεωρήθηκε λοιπόν αναγκαίο να εντοπιστούν κάποιοι τρόποι για μια καθόλα αποτελεσματική και δη αποδοτική διαχείρισή των συλλεγέντων αναρίθμητων βιολογικών δεδομένων-ευρημάτων, και για αυτό και υιοθετήθηκε ο τομέας της Βιοπληροφορικής (Bioinformatics), ο οποίος και στάθηκε ως μια μείξη μεταξύ των επιστημών της Βιολογίας και των Ηλεκτρονικών Υπολογιστών (<https://www.youtube.com/watch?v=BrEs7zFVrMI>).

Πιο συγκεκριμένα, η επιστήμη της Βιοπληροφορικής είναι γνωστή τουλάχιστον εδώ και τρεις δεκαετίες, παρόλα αυτά ακόμη και στην εποχή μας την τωρινή την επανομαζόμενη ως εποχή της προόδου έχει διαπιστωθεί μια γενικότερη σύγχυση σε ό, τι αφορά στη λειτουργικότητά της. Μια μερίδα μάλιστα από βιολόγους δεν έχει σταματήσει να προσβλέπει τη Βιοπληροφορική ως ένα εργαλείο, ενώ μια άλλη μερίδα επιστημόνων της Πληροφορικής ταύτισε αυτόν τον κλάδο ως μία πορεία εξέλιξης νέων αλγορίθμων και συστημάτων. Η

Βιοπληροφορική στις μέρες μας λοιπόν έχει τεθεί ως ένας καθόλα διεπιστημονικός τομέας έρευνας σχετικά με τη διεπαφή ανάμεσα στην επιστήμη των υπολογιστών και της βιολογικής επιστήμης (Μπάγκος, 2015).

1.1 ΣΚΟΠΟΣ ΕΡΓΑΣΙΑΣ

Ο σκοπός της παρούσας πτυχιακής εργασίας είναι να μελετήσει βιβλιογραφικά την επιστήμη της Βιοπληροφορικής, με έμφαση στη συμβολή των μαθηματικών στη Βιοπληροφορική.

1.2 ΔΙΑΡΘΡΩΣΗ ΕΡΓΑΣΙΑΣ

Η παρούσα εργασία δομήθηκε ως εξής:

Αρχικά στο Κεφάλαιο 1 της Βιβλιογραφικής ανασκόπησης και στην Ενότητα 2.1 αποσαφηνίστηκε εννοιολογικά η έννοια της Βιοπληροφορικής, στην Ενότητα 2.2 καταγράφηκαν οι Βασικοί τομείς της Βιοπληροφορικής, στην Ενότητα 2.3 τονίστηκε η αναγκαιότητα της Βιοπληροφορικής, στην Ενότητα 2.4 προσδιορίστηκαν οι Στόχοι της Βιοπληροφορικής, στην Ενότητα 2.5 επισημάνθηκε ο Διεπιστημονικός χαρακτήρας της Βιοπληροφορικής, στην Ενότητα 2.6 αναπτύχθηκαν ενδελεχώς τα Πεδία έρευνας της Βιοπληροφορικής, στην Ενότητα 2.7 μελετήθηκαν τα Εργαλεία της Βιοπληροφορικής με τον αλγόριθμο Fasta να αναλύεται στην Υποενότητα 2.7.1, τον αλγόριθμο Blast να μελετάται στην Υποενότητα 2.7.2 και το Clustal στην Υποενότητα 2.7.3. Στην Ενότητα 2.8 περιγράφονται οι Εφαρμογές της Βιοπληροφορικής, ενώ στο

Κεφάλαιο 3 που έπεται μελετώνται τα Μαθηματικά και τα είδη του δηλαδή η Γραμμική Άλγεβρα στην Υποενότητα 3.1, οι Διαφορικές Εξισώσεις στην Υποενότητα 3.2, η Στατιστική στην Υποενότητα 3.3 και τα Διακριτά Μαθηματικά στην Υποενότητα 3.4. Εν συνεχεία στο Κεφάλαιο 4 γίνεται λόγος για τη συμβολή – συνεισφορά των Μαθηματικών στην επιστήμη της Βιοπληροφορικής. Τέλος καταγράφονται τα κυριότερα Συμπεράσματα και η παρούσα εργασία ολοκληρώνεται με την καταγραφή των Βιβλιογραφικών Αναφορών, που χρησιμοποιήθηκαν κατά τη σύνθεση της παρούσας Βιβλιογραφικής ανασκόπησης εργασίας μας.

ΚΕΦΑΛΑΙΟ 2 -ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

2.1 ΟΡΙΣΜΟΣ & ΕΡΜΗΝΕΙΑ ΤΗΣ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ

Ετυμολογικά ο όρος «Βιοπληροφορική» αποτελείται από το Βιο (μοριακή Βιολογία και την Πληροφορική (επιστήμη των υπολογιστών), με κατάληξη η έννοια της Βιοπληροφορικής να σημαίνει μία επιστήμη, η οποία τείνει να επιλύει διάφορα προβλήματα, τα οποία αναδύονται μέσα από τη Βιολογία αξιοποιώντας και τη μεθοδολογία από την επιστήμη των υπολογιστών (Αγγελίδης, 2015). Έτσι, η Βιοπληροφορική συνδέθηκε άμεσα και με άλλες επιστήμες, οι οποίες συνέδραμαν στο έργο της μεταξύ των οποίων ήταν η επιστήμη των υπολογιστών, τα μαθηματικά, η στατιστική, η Βιολογία, η ιατρική,

η χημεία και η φυσική. Επίσης πρέπει να επισημανθεί πως η Βιοπληροφορική συσχετίστηκε με την εξαγωγή των ακολουθιών του DNA εκ του γενετικού υλικού, την περιγραφή των ακολουθιών, την εμπέδωση της εποπτείας έκφρασης των γονιδίων και τη συσχέτιση μεταξύ της αλληλουχίας των αμινοξέων των πρωτεϊνών και της δόμησής τους. Το αντικείμενο στο οποίο επικεντρώθηκε ήταν η έμφασή της να αντιληφθεί τη λειτουργία των ζωντανών όντων, να συμβάλει στη σχεδίαση των φαρμάκων, στην ανίχνευση των γενετικών παραγόντων κινδύνου, στη γονιδιακή θεραπεία, στη γενετικής φύσης τροποποίηση και φυτών και ζώων καθώς και στη βελτίωση των εργαλείων του βιολογικού πολέμου (Βουλγαρίδης & Σαμαράς, 2012).

Πιο συγκεκριμένα, ο όρος «Βιοπληροφορική πρωτοχρησιμοποιήθηκε κατά το έτος 1977 από την ολλανδικής καταγωγής θεωρητική βιολόγο Ρ. Hogeweg, όταν αποφάσισε να επονομάσει το κύριο ερευνητικό της πεδίο ως *βιοπληροφορική* και μάλιστα δημιούργησε και μια ομάδα βιοπληροφορικής στο Πανεπιστήμιο της Ουτρέχτης (Attwood, Gisel, Eriksson & Bongcam-Rudloff, 2011). Στον ευρωπαϊκό χώρο όμως ο όρος αυτός της «Βιοπληροφορικής» αναγνωρίστηκε κατά το έτος περίπου του 1990 σε μια στιγμή οργάνωσης ενός συνεδρίου με τίτλο «Η Βιοπληροφορική στα 90s», το οποίο διεξήχθη στο Μάαστριχτ το έτος 1991. Μάλιστα στη διάρκεια όλης αυτής της περιόδου στις Ηνωμένες Πολιτείες της Αμερικής δημιουργήθηκε και το Εθνικό Κέντρο Βιοτεχνολογικής Πληροφόρησης (National Center for Biotechnology Information, NCBI) (Attwood, Gisel, Eriksson & Bongcam-Rudloff, 2011). Όμως παρόλες αυτές τις κινήσεις εξακολουθούσε να (εν-)υπάρχει ακόμη μια αίσθηση πως στην επιστημονική κοινότητα εξέλειπε μια ξεκάθαρη πρακτική ή καλύτερα μια σαφή «στρατηγική βιολογικής πληροφορικής», η οποία να μπορεί να συνδυάζει τη χρήση-εφαρμογή των ηλεκτρονικών υπολογιστών με το πεδίο της έρευνας στη βιολογία και πιο συγκεκριμένα στη μοριακή βιολογία (Attwood, Gisel, Eriksson & Bongcam-Rudloff, 2011).

Η εννοιολογική αποσαφήνιση μάλιστα, που αποδίδει το NCBI (National Center for Biotechnology Information) σχετικά με την βιοπληροφορική, είναι η

ακόλουθη: Η βιοπληροφορική αποτελεί έναν κλάδο μιας επιστήμης, εντός του οποίου οι δύο άλλες επιστήμες της βιολογίας και της πληροφορικής συγχωνεύονται σε μία μόνον επιστήμη, στην οποία και συνυπάρχουν και άλλες τρεις (3) πρόσθετες υπο-επιστήμες, που είναι οι εξής: 1. η ανάπτυξη νέων αλγορίθμων και στατιστικών μοντέλων σύμφωνα με τα οποία μοντέλα συνηθίζεται να μετρώνται οι συσχετίσεις ανάμεσα σε ένα πλήθος δεδομένων, η μελέτη-ανάλυση και η επεξήγηση πολυάριθμων μορφών δεδομένων, στα οποία με τη σειρά τους υπάγονται εκείνα των νουκλεοτιδικών και των αμινοξικών αλληλουχιών, των πρωτεϊνικών υπό-περιοχών καθώς και των πρωτεϊνικών δομών, 2. η ανάπτυξη κι εφαρμογή εκείνων των εργαλείων τα οποία και πραγματώνουν εν ολίγοις την πρόσβαση, και 3. η διαχείριση των διαφορετικών μορφών των πληροφοριών.

Έναν αρκετά συναφή εννοιολογικό ορισμό απέδωσαν και οι Luscombe, Greenbaum & Gerstein (2001), βάσει του οποίου ορισμού ως «Βιοπληροφορική» ετέθη η επιστήμη της βιολογίας ως μακρομόρια (όπως αυτά εννοούνται στον κλάδο της Χημείας), κι εξελικτικά η πρακτική εφαρμογή των υπολογιστικών τεχνολογιών (όπως αυτά προέρχονται από τους κλάδους των εφαρμοσμένων μαθηματικών, της πληροφορικής και της στατιστικής), προκειμένου να εμπεδωθούν-επεξηγηθούν και να δομηθούν σωστά όλες οι πληροφορίες, που έχουν σχέση με αυτά τα μόρια³. Ακόμη βάσει και του Fredj Tekaia (όπ. αναφ. στο Μπάγκος, 2015), η επιστήμη της βιοπληροφορικής τείνει ν'αποτελεί εκείνες τις μαθηματικές, στατιστικές αλλά και τις υπολογιστικές τεχνικές, οι οποίες αποσκοπούν στην επίλυση-εξάλειψη όλων των βιολογικής φύσης προβλημάτων υπό το πρίσμα χρήσης κι αξιοποίησης των νουκλεοτιδικών και αμινοξικών αλληλουχιών και όλων εκείνων των στοιχείων και πληροφοριών, που δείχνουν να σχετίζονται με αυτές.

Σύμφωνα λοιπόν με όλους τους ορισμούς, που αποδόθηκαν παραπάνω σχετικά με την επακριβή έννοια της επιστήμης της *Βιοπληροφορικής*, έχει γίνει

³ <https://www.ncbi.nlm.nih.gov/Class/MLACourse/Modules/MolBioReview/bioinformatics.html>

σαφές πως με την αρωγή άλλων διαφορετικών επιστημών όπως είναι τα μαθηματικά, η πληροφορική αλλά και η στατιστική έρευνα, η όλη ερευνητική διεργασία τείνει να αποκομίζει μια πολύ διεπιστημονικής φύσης οπτική. Κι όλα αυτά πραγματώνονται με την ανάλυση-μελέτη των ζωντανών οργανισμών να γίνεται πλέον και με τη συμβολή των υπολογιστικών τεχνικών εκτός από τις υπόλοιπες μεθόδους, οι οποίες έως και στις μέρες μας αξιοποιούνται στην έρευνα της βιολογίας, όπως δηλαδή είναι η παρατήρηση ή και η πειραματική διαδικασία στον πάγκο. Για αυτό άλλωστε κι εκτός της ορολογίας αυτής ως «Βιοπληροφορική» συνηθίζεται να γίνεται χρήση και του όρου «υπολογιστική βιολογία» (Μπάγκος, 2015). Και παρά το γεγονός πως έχει φανεί πλείστες φορές εντός της βιβλιογραφίας να εντοπίζονται ομοιότητες καθώς και διαφοροποιήσεις συνάμα ανάμεσα σε αυτούς τους δύο όρους, εντούτοις και οι δύο όροι έχουν γίνει αποδεκτοί όντες ένα στέρεοθεμέλιο της συνεργασίας μεταξύ αυτών των δύο επιστημών, που η βασική τους επιδίωξη είναι η αποθήκευση, η οργάνωση και η τακτική της ανάκτησης των βιολογικών δεδομένων-ευρημάτων (Bagos, Nikolaou, Liakopoulos & Tsirigos, 2010).

Ως «Βιοπληροφορική» (βλ. Εικ. 1) λοιπόν επονομάζεται ένας καθόλα επιστημονικός τομέας, εντός του οποίου συμπράττουν-συνεργάζονται οι επιστήμες της Βιολογίας, της Πληροφορικής, της Στατιστικής και των Μαθηματικών. Η σύμπραξη-συνεργασία αυτή αυτών ακριβώς των επιστημών μεταξύ τους έχει ως στόχο της να διερευνήσει αλλά και να εξερευνήσει διάφορες καινοτόμες μεθόδους σχετικά με την προσέγγιση κι ανάλυση διαφόρων βιολογικών προβλημάτων, καθώς και την αντίληψη των κύριων αρχών της Βιολογίας. Ουσιαστικά δηλαδή επρόκειτο για έναν γνωστικό τομέα με πολύ συγκεκριμένο όσο κι ένα ευρύ πεδίο εφαρμογών καθώς κι αλληλεπίδρασης με τη σύγχρονης μορφής δομική, μοριακή, πληθυσμιακή αλλά και περιβαλλοντική βιολογία. Η *Βιοπληροφορική* εν ολίγοις ενασχολείται με την αναπτυξιακή διαδικασία και την εφαρμογή συγκεκριμένων εργαλείων αναφορικά με τη διαχείριση των βιολογικών δεδομένων. Στην περίπτωση αυτήν η έννοια της διαχείρισης συμπεριλαμβάνει την αποθήκευση, την οργάνωση, τη

μελέτη-ανάλυση κι ακόμη και τη μέθοδο της οπτικοποίησης, δηλαδή της σκιαγράφησης της δόμησης των βιολογικών μακρομορίων – των δεδομένων αυτών. Θέτοντας μάλιστα και αυτά τα βιολογικά δεδομένα (το DNA, το RNA και τις πρωτεΐνες) ως μια ψηφιακή πληροφορία κι αξιοποιούμενης και της διαθέσιμης υπολογιστικής δύναμης φαίνεται να υιοθετεί τους αλγορίθμους, με στόχο την ανάλυση κι εξαγωγή επωφελών συμπερασμάτων αποδοτικού μάλιστα χαρακτήρα (Παπαϊωάννου, 2007).



Εικ. 1 – Η Βιοπληροφορική
Πηγή: Μπάγκος, 2015

2.2 ΒΑΣΙΚΟΙ ΤΟΜΕΙΣ ΤΗΣ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ

Σε ό, τι αφορά τώρα στους βασικούς τομείς της Βιοπληροφορικής, αυτοί είναι οι εξής (Αγγελίδης, 2015 · Κοτσιδά, 2008 · Shendure, Balasubramanian, Church, Gilbert, Rogers, Schloss & Waterston, 2017):

1. Βάσεις δεδομένων (Databases) με οργάνωση, αποθήκευση και αναζήτηση δεδομένων,

2. Αναλυτική διεργασία των ακολουθιών DNA, RNA και πρωτεϊνών (Sequence analysis) με στοίχιση ακολουθιών (με συγκριτική αποτίμηση μεταξύ των αντίστοιχων περιοχών, ή μεταξύ δύο ή και πιο πολλών ακολουθιών) και φυλογενετική ανάλυση με εξελικτικές σχέσεις ανάμεσα σε ομοειδή αντικείμενα (όπως είναι τα γονίδια, οι πρωτεΐνες και οι οργανισμοί),
3. Γονιδιακή ρύθμιση/αποτύπωση (Gene expression) με μελέτη των δεδομένων από μικροσυστοιχίες,
4. Δόμηση RNA/πρωτεϊνών (structural biology) με πρόβλεψη μιας μορφής δευτεροταγούς και τριτοταγούς δόμησης κι ανάλυση των πρωτεϊνικών επιφανειών που τείνουν να αναπτύσσουν μια αλληλεπίδραση μεταξύ τους,
5. Εξαγωγή δεδομένων εκ της βιβλιογραφίας (text mining),
6. Βιολογικής φύσης δίκτυα/οδοί, Βιολογικά συστήματα, και
7. Οντολογίες (Ontologies) με αξιοποίηση ενός ελεγχόμενου λεξιλογίου σε ό, τι αφορά στην περιγραφή των χαρακτηριστικών και των λειτουργικών μεθόδων των ομοειδών αντικειμένων όπως π.χ. είναι οι πρωτεΐνες⁴.

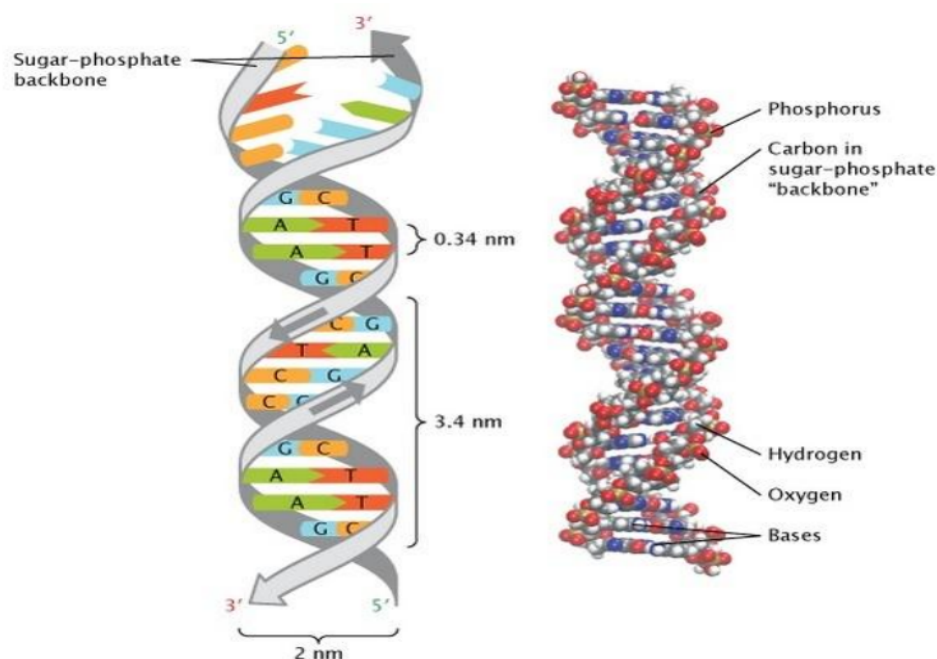
2.3 ΑΝΑΓΚΑΙΟΤΗΤΑ ΤΗΣ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ

Με σημείο αναφοράς λοιπόν τα προαναφερθέντα σε μια συνοπτική θεώρησή τους, η επιστήμη της Βιοπληροφορικής υιοθετεί μεθόδους της πληροφορικής (που προέρχονται από επιστημονικής φύσης πεδία όπως είναι τα εφαρμοσμένα μαθηματικά, η πληροφορική, η στατιστική), προκειμένου να αντιληφθεί και να (διο-)οργανώσει όλο το φάσμα των πληροφοριών που συσχετίζονται με τη μοριακή βιολογία ευρέως. Πιο συγκεκριμένα, η Βιοπληροφορική αξιοποιεί

⁴ <https://eclass.uowm.gr>

εύστοχα και δη λειτουργικά ένα πλήθος υπολογιστικών μεθόδων, προκειμένου να συλλέξει, να αποθηκεύσει σωστά κι εν τέλει να αναλύσει και να επεξεργαστεί τα διαθέσιμα βιολογικά δεδομένα (Κοσσιδά, 2008 · Baxevanis & Ouellette, 2004).

Η βαθμιαία πλην όμως σταθερά αυξανόμενη ανάπτυξη της μοριακής βιολογίας μετά την ανακάλυψη του συστήματος της διπλής έλικας του DNA (βλ. Εικ. 2) από τους Watson & Crick (1990)⁵, και παράλληλα με το ξεκίνημα της προσπάθειας να αποκωδικοποιηθεί το ανθρώπινο γονιδίωμα (1990), που ολοκληρώθηκε αρκετά χρόνια αργότερα (2003), κατέληξε στις μέρες μας να εμφανίζει την αναγκαιότητα συνένωσης της επιστήμης της Βιολογίας και με άλλες επιστήμες. Μάλιστα, στη βάση πως όλα σχεδόν τα αναδυθέντα νέα δεδομένα προϋπέθεταν τη χρήση των ηλεκτρονικών υπολογιστών, κι επιπρόσθετα, επειδή η αλματώδης εξέλιξη της επιστήμης της Πληροφορικής ειδικά τα τελευταία χρόνια έχει πάρει τεράστιες διαστάσεις με εξίσου τεράστιες λειτουργικές χρήσεις όπως άλλωστε και της επιστήμης της Βιολογίας, έτσι λοιπόν τέθηκε η επιτακτική χρησιμότητα της συνεργασίας μεταξύ των δύο αυτών επιστημών (Watson & Crick, 1990).



Εικ. 2 – Το σύστημα της διπλής έλικας

⁵ http://jupiter.chem.uoa.gr/chem/talks/talks_2017/double_helix.pdf

2.4 ΣΤΟΧΟΙ ΤΗΣ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ

Μία εκ των μεγαλύτερων προκλήσεων, που έχει να αντιμετωπίσει στις μέρες μας ο κλάδος της βιολογίας, είναι η στο μέγιστο αποδοτική εκμετάλλευση της γνώσης που (εν-)υπάρχει εντός ενός πλήθους από βιολογικά δεδομένα, τα οποία αναδύονται μέσω της επεξεργασίας των γονιδιωμάτων των διαφόρων οργανισμών. Αυτή λοιπόν η αναγκαιότητα για μια καθόλα πολύπλευρη αντίληψη της γνώσης τείνει να προσδιορίζει μερικώς και τους κύριους εκ των πολλών στόχων της βιοπληροφορικής, που είναι οι εξής (Βουλγαρίδης & Σαμαράς, 2012 · Lesk, 2021):

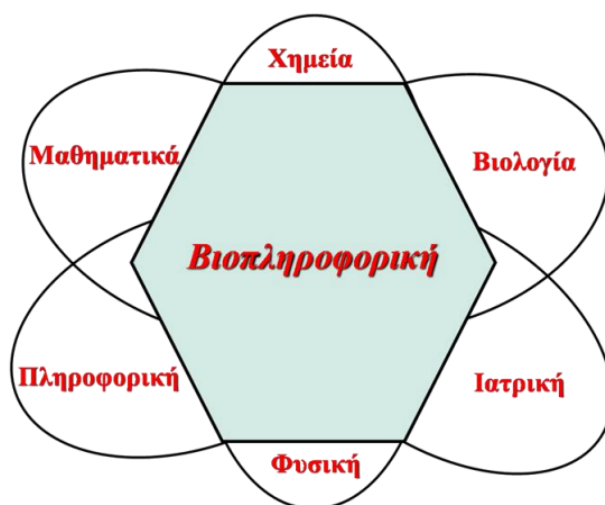
- Η κατάλληλη και στοχευμένη οργάνωση όλων των δεδομένων, προκειμένου η διαχείρισή τους να καταστεί και η πλέον ευκολότερη για όλους τους ερευνητές,
- Η εξέλιξη των εργαλείων με στόχο την επεξεργασία των συλλεγέντων δεδομένων και την ερμηνεία των εξαχθέντων αποτελεσμάτων,
- Η ανάπτυξη των αποδοτικών αλγορίθμων, ώστε να καταμετρηθεί η ομοιότητα ανάμεσα στις ακολουθίες, καθώς επίσης και να εκτιμηθούν οι σχέσεις ανάμεσα στα πολύ μεγάλα σύνολα των δεδομένων,
- Η διεύρυνση των πιλοτικών δεδομένων με τη συμβολή των προβλέψεων, όπως π.χ. με την πρόβλεψη της δόμησης των πρωτεϊνών,
- Τα εργαλεία για τη συσχέτιση των δεδομένων εκείνων, που επιδιώκεται να αξιοποιηθούν για την εμπέδωση των φυσικών φαινομένων, των προς εξέλιξη σχέσεων ανάμεσα στους οργανισμούς αλλά και τη γενετική προέλευση των ασθενειών.

Από τα παραπάνω γίνεται αντιληπτό πως το σύνολο των στόχων της βιοπληροφορικής -ως ένας κλάδος της βιολογίας που αναδύθηκε από την αναγκαιότητα χρήσης κι αξιοποίησης των τεχνικών της πληροφορικής στην επιστήμη της βιολογίας- είναι άρρηκτα συνδεδεμένο με τον σπουδαιότερο εκ των στόχων της βιολογίας, ο οποίος εν προκειμένω είναι η στο μέγιστο πλήρης εμπέδωση της δόμησης και της λειτουργικότητας των οργανισμών. Ο κύριος σκοπός της βιοπληροφορικής λοιπόν τίθεται να είναι η υλοποίηση των κανόνων των επιστημών της τεχνολογίας και της πληροφορικής, προκειμένου να καταφέρει να πράξει τα μεγαλύτερα, ποικίλα αλλά και τα πιο πολύπλοκα δεδομένα των επιστημών ζωής ως περισσότερα αντιληπτά και δη εύχρηστα, ώστε τοιουτοτρόπως να αναδυθεί και η μέγιστη ισχύς όλων αυτών των δεδομένων (National Institutes of Health, όπ. αναφ. στο Watson et al., 2007). Μάλιστα, όπως επισημαίνει και το NCBI (όπ. αναφ. στο Δελβενιώτη, 2017) «ο ύψιστος στόχος της βιοπληροφορικής είναι να κάνει δυνατή την ανακάλυψη νέων γνώσεων στην επιστήμη της βιολογίας, καθώς και να δημιουργήσει μία ολιστική αντίληψη η οποία θα ενοποιεί αυτές τις γνώσεις». Η επίτευξη εν προκειμένω όλων των προαναφερόμενων στόχων της βιοπληροφορικής συντείνει στην εξέλιξη κι άλλων κλάδων όπως για παράδειγμα είναι η κτηνοτροφία, η γεωργία καθώς και ο κλάδος της βιομηχανίας (Machluf & Yarden, 2013).

2.5 Ο ΔΙΕΠΙΣΤΗΜΟΝΙΚΟΣ ΧΑΡΑΚΤΗΡΑΣ ΤΗΣ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ

Μέσα από τη σύνθεση της παρούσας εργασίας σχετικά με το θέμα της επιστήμης της *Βιοπληροφορικής* έγινε έως τώρα σαφές, πως επρόκειτο για έναν άκρως διεπιστημονικής φύσης κλάδο, ο οποίος αξιοποιεί άλλες επιστήμες όπως είναι τα μαθηματικά, η πληροφορική και η στατιστική, προκειμένου να (επι-)λύσει ζητήματα του έμβιου κόσμου (βλ. Εικ. 3). Από αυτό διαφαίνεται και η ανάγκη να καλλιεργηθεί η συνεργατικότητα και η επικοινωνία μεταξύ όλων των

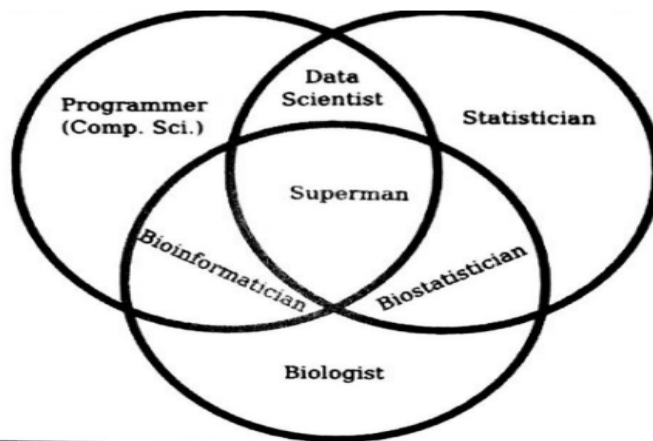
ειδικών των επιστημών αυτών όπως επίσης και από όλους τους κλάδους, ειδικά σε περιπτώσεις μάλιστα που κρίνονται ως εξόχως δύσκολες λόγω τεράστιων προβλημάτων (Δελβενιώτη, 2017 · Μπάγκος, 2015).



Εικ. 3 – Η συνεργασία της Βιοπληροφορικής με άλλες επιστήμες
Πηγή:

https://oeclass.aua.gr/openeclass/modules/document/file.php/OCDB100/GBT_3040_01_2h.pdf

Εν ολίγοις, η Βιοπληροφορική ή αλλιώς η επονομαζόμενη ως και η Υπολογιστική Βιολογία αποτελεί έναν διεπιστημονικό κλάδο, που προέρχεται από την επιστήμη της μοριακής βιολογίας και πιο συγκεκριμένα από την ανάλυση των βιολογικών αλληλουχιών και των δομών (βλ. Εικ. 4). Ο ορισμός, που αποδόθηκε, κρίνεται ως και ο πλέον συμβατός με τον ορισμό που δόθηκε από το NCBI, πως δηλαδή το διεπιστημονικό πεδίο έρευνας εστιάζει στη συγκέντρωση κι οργάνωση των δεδομένων, στην εξέλιξη και αξιοποίηση των υπολογιστικών εργαλείων με στόχο την επεξεργασία κι επεξήγηση των βιολογικών δεδομένων, και τέλος τη θέση της βιολογίας ως μιας επιστήμης της πληροφορίας και όχι μόνον ως μια απόρροια οριοθετημένων βιοφυσικών και βιοχημικών διεργασιών (<http://www.ncbi.nlm.nih.gov/Class/MLACourse/Modules/MolBioReview/bioinformatics.html>).



Εικ. 4– Το προφίλ ενός διεπιστημονικού ερευνητή, που ασχολείται με την ανάλυση βιολογικών δεδομένων με χρήση της πληροφορικής και της στατιστικής
Πηγή: Μπάγκος, 2015

2.6 ΠΕΔΙΑ ΕΡΕΥΝΑΣ ΤΗΣ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ

Μεταξύ των ερευνητικών πεδίων της Βιοπληροφορικής είναι τα εξής: (Βουλγαρίδης & Σαμαράς, 2012 · Κοσσιδά, 2002 · Lesk, 2021 · Shendure, Balasubramanian, Church, Gilbert, Rogers, Schloss & Waterston, 2017):

- *Ανάλυση Αλληλουχιών (sequence analysis)*: στα τέλη του 1970 υιοθετήθηκε η ιδέα μελέτης-επεξεργασίας όλων των γενετικών υλικών των οργανισμών. Μάλιστα ο επονομαζόμενος ως Βακτηριοφάγος Phi-X174 (ή ΦΧ174) είχε τεθεί ως ο πρώτος οργανισμός του οποίου το γονιδίωμα πέτυχε να χαρτογραφηθεί. Από τότε και έως και στις μέρες μας αυτό ακριβώς το γονιδίωμα έχει χαρτογραφηθεί αλλά κι αποθηκευθεί σε βάσεις δεδομένων, τα οποία δεδομένα επεξεργάζονται με τη συμβολή ενός κατάλληλου λογισμικού που δίνει τη δυνατότητα παροχής πολλών πληροφοριών, που είναι χρησιμότες για τον επακριβή προσδιορισμό των γονιδίων τα οποία κωδικοποιούν πρωτεΐνες, γονίδια RNA, δομικής φύσης μοτίβα καθώς και ρυθμιστικές περιοχές. Ειδικότερα, η ανάλυση ως προς την ακολουθία μπορεί να

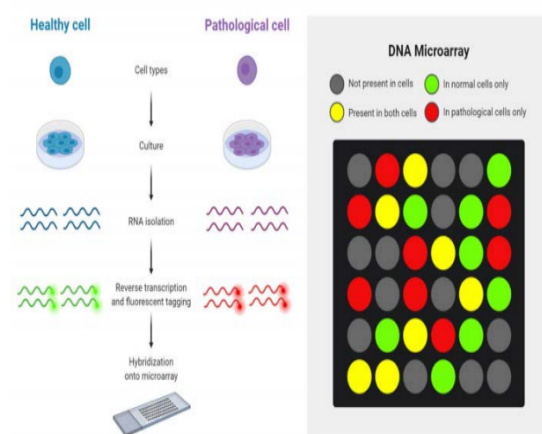
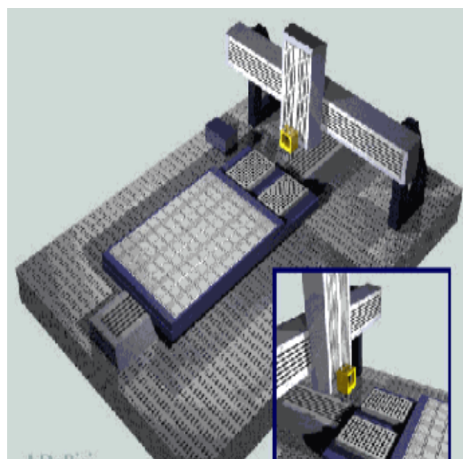
αποδώσει στοιχεία και πληροφορίες σχετικά με: 1. την εξελικτική πορεία κι όλες τις διεργασίες που πιθανόν (καθ-)οδήγησαν στην εδραίωση των ακολουθιών των βιομακρομορίων που εντοπίζονται στους σύγχρονους οργανισμούς, 2. Το ιστορικό εξέλιξης των βιολογικών μακρομορίων καθώς και τον καθορισμό των τυχόν «προγονικών» μορίων τους, 3. όλες τις περιοχές των ακολουθιών που έχουν τεθεί ως οι σπουδαιότερες για την τελική στερεοδιάταξη αλλά και τις φυσικοχημικής μορφής ιδιότητες των βιομακρομορίων, στις οποίες στηρίζονται και όλες οι βιολογικές λειτουργίες τους. Επιπρόσθετα, μια άλλη εφαρμογή της παραπάνω ανάλυσης αλληλουχιών είναι η ανίχνευση των γονιδίων εκείνων που βρίσκονται εντός των αλληλουχιών DNA (Genome Annotation). Το πρώτο λογισμικό ανίχνευσης των γονιδίων πρέπει να σημειωθεί πως διαμορφώθηκε κατά το έτος 1995 από τον Dr. Owen White κι εμπεριείχε ως κύριες λειτουργίες τους εκτός της ανίχνευσης γονιδίων και την ανίχνευση του t-RNA (μεταφορικό RNA), αλλά και την αντιστοίχιση πρόβλεψης της πλειονότητας των λειτουργιών σε καθένα εκ των εντοπιζόμενων γονιδίων. Φτάνοντας στη σημερινή εποχή μας και παρά το γεγονός πως η τεχνολογία έχει εξελιχθεί αλματωδώς, εντούτοις η πλειονότητα των συστημάτων μελέτης κι επεξεργασίας των αλληλουχιών δείχνουν πως λειτουργούν με συναφή τεχνική όπως εκείνου του Dr. White.

- *Συγκριτική γονιδιωματική (comparative genomics)*: αυτό το ερευνητικό πεδίο της συγκριτικής γονιδιωματικής εστιάζει με τη συγκριτική αποτίμηση των αλληλουχιών διαφορετικών οργανισμών και την προσπάθεια εξεύρεσης των μεταξύ τους αναδυόμενων σχέσεων. Αυτή η ερευνητική τεχνική μελετά τα κοινά σημεία αλλά και τις διαφοροποιήσεις ανάμεσα στις πρωτεΐνες, το RNA καθώς και τις ρυθμιστικές ακολουθίες (*) των υπό μελέτη οργανισμών. Όλες αυτές οι σχέσεις διαθέτουν τη δυνατότητα να παρέχουν χρήσιμες πληροφορίες και στοιχεία σχετικά και με την προαναφερόμενη εξελικτική συσχέτιση αλλά και με τη λειτουργία

αυτών των οργανισμών. Σε ό,τι αφορά στο ζήτημα της πολυπλοκότητας, που διαπιστώνεται να διαθέτει η γενετική εξέλιξη καθώς και τα οφέλη που εξάγονται εκ της εξεύρεσης των διαφορών ανάμεσα στα γονιδιώματα τίθεται ως μια πρόκληση για το σύνολο των ερευνητών των πεδίων και της βιοπληροφορικής αλλά και των μαθηματικών. Και για αυτό μάλιστα κι εξελίχθηκε μια ευρεία γκάμα από μοντέλα μαθηματικά καθώς και αλγόριθμοι, που αξιοποιούνται στις μέρες μας στα των ερευνών του κλάδου της συγκριτικής γονιδιωματικής με βασική επιδίωξη τη στο μέγιστο ασφαλή κι αποδοτική ανακάλυψη της γνώσης.

- *Ανάλυση γονιδιακής έκφρασης (Analysis of gene expression)*: αυτό το ερευνητικό πεδίο της γονιδιακής έκφρασης έχει σχέση με τη διεργασία εκείνη, η οποία δημιουργεί τη «μεταβολή» των κωδικοποιημένων πληροφοριών του γονιδίου στο λειτουργικό του αυτό προϊόν (ή πρωτεΐνη ή RNA). Όμως υπάρχουν πλείστες τεχνικές (κατα-)μέτρησης της γονιδιακής έκφρασης, εκ των οποίων οι γνωστότερες είναι οι 1. μικροσυστοιχίες (microarrays, βλ. Εικ. 5), 2. η τεχνική SAGE, 3. η τεχνική MPSS καθώς και 4. η τεχνική RNA-Seq. Πιο συγκεκριμένα, 1. οι μικροσυστοιχίες αξιοποιούν αλληλουχίες που έχουν αναδυθεί εκ της μελέτης-επεξεργασίας ορισμένων γονιδιωμάτων, τα οποία επιδιώκουν να εξακριβώσουν εκείνα τα γονίδια που σκιαγραφούνται σε έναν ορισμένο τύπο κυττάρου ενός οργανισμού, σε ένα ορισμένο χρόνο και υπό ορισμένες συγκυρίες. Στηρίζεται στη μέθοδο της υβριδοποίησης. 2. η τεχνική SAGE (serial analysis of gene expression) αξιοποιείται εν προκειμένω, προκειμένου να παραγάγει την εποπτική εικόνα ενός συνόλου εκ της γονιδιακής έκφρασης των κυττάρων. 3. η τεχνική MPSS (massively parallel signature sequencing) συνδέεται με συναφείς αρχές της προηγούμενης τεχνικής SAGE, όμως η εν λόγω τεχνική MPSS αξιοποιείται συνηθέστερα για την αναγνώριση της παρουσίας αλλά και την (κατά-)μέτρηση της ποσότητας του mRNA. 4. η τεχνική RNA-Seq (ή αλλιώς και ως WTSS) επικεντρώνεται στην αξιοποίηση των τεχνολογιών εκείνων,

οι οποίες κι επεξεργάζονται επισταμένως το cDNA (*), καταφέροντας την ανάκτηση πολλών πληροφοριών και στοιχείων σχετικά με το περιεχόμενο ενός δείγματος RNA (Morin et al., 2008, όπ. αναφ. στο Βουλγαρίδης & Σαμαράς, 2012). Μάλιστα οι τρεις (3) αυτές τελευταίες πρακτικές συνδέονται με την αλληλούχιση τμημάτων του RNA, με τις δύο πρώτες να εστιάζουν στα μικρότερα τμήματα τα οποία είναι προς μελέτη-ανάλυση του δείγματος και την τρίτη και τελευταία να επικεντρώνεται στην αλληλούχιση ολόκληρου του δείγματος με όλα δηλαδή τα τμήματά του. Τέτοιου τύπου μέθοδοι συνηθέστερα αξιοποιούνται με στόχο την ανίχνευση των γενετικής φύσης ασθενειών, όπως είναι π.χ. η αξιοποίηση των microarrays για τη συγκριτική αποτίμηση των καρκινικών αλλά και των μηκαρκινικών κυττάρων καθώς ο καθορισμός εκείνων των κυττάρων τα οποία και τίθενται ως άμεσα υπεύθυνα για την ανάπτυξη της ασθένειας.



Εικ. 5 - Η τεχνολογία microarrays
Πηγή: Βουλγαρίδης & Σαμαράς, 2012 (σελ. 30)

Πιο αναλυτικά, όμως και στο πλαίσιο και άλλων νεωτερικότερων μεθόδων κι πιο αποδοτικών συνάμα, που έως και σήμερα έχουν υλοποιηθεί, και παρά την ευρεία χρήση – αξιοποίηση της μεθόδου των μικροσυστοιχιών DNA ή DNA microarrays στη βάση πως ενείχε μιας μεγάλης ευκολίας ως προς τον χειρισμό, μιας τεράστιας διάχυσης της τεχνολογίας αλλά κι ενός πολύ μικρού κόστους, εντούτοις θα ήταν χρήσιμο να αναφερθούν και οι εξής πλέον συνηθέστεροι μέθοδοι μελέτης της γονιδιακής έκφρασης της τελευταίας δεκαετίας (Maier, Güell & Serrano 2009 · Vogel & Marcotte 2012):

1. *Ανάλυση μεταγραφώματος (transcriptome analysis)*: με σημείο αναφοράς την ποσοτικοποίηση της γονιδιακής έκφρασης και επίσης με σημείο αναφοράς πως το σύνολο των κυτταρικών λειτουργιών επιτυγχάνονται με τη συμβολή των πρωτεϊνών, το πιο λογικό θα ήταν να γίνει μια προσπάθεια καταμέτρησης όλων των συγκεντρώσεων των πρωτεϊνών εντός ενός κυττάρου. Συναφούς μορφής αναλύσεις - προσεγγίσεις ομαδοποιούνται βάσει της γενικής ορολογίας «πρωτεωμική ανάλυση». Πρέπει να σημειωθεί πως σύγχρονου τύπου μέθοδοι δίνουν τη δυνατότητα ποσοτικοποίησης 8000-10000 πρωτεϊνών σε ένα κυτταρικό δείγμα. Εντούτοις όλες αυτές οι πρωτεωμικές αναλύσεις δεν είναι και η πλέον προτιμητέα τεχνική, στην περίπτωση που τεθεί η ανάγκη ανάλυσης μιας γονιδιακής έκφρασης για μια πλειάδα λόγων, εκ των οποίων ορισμένοι είναι οι ακόλουθοι (Maier, Güell & Serrano 2009 · Vogel & Marcotte 2012): α). αφορά τα τεχνικής φύσης θέματα, μιας και οι συγκεντρώσεις των πρωτεϊνών εντός ενός κυτταρικού πληθυσμού τείνουν να διαφοροποιούνται μεταξύ τους ριζικά, καταλαμβάνοντας μάλιστα και πλείστες τάξεις μεγέθους. Ορισμένες εξ αυτών ενυπάρχουν σε πολλά εκατομμύρια αντίγραφα και κάποιες πάλι άλλες ενυπάρχουν σε μόνον λίγες δεκάδες αντίγραφα. β). σχετίζεται με τον παράγοντα της ποικιλομορφίας στις λειτουργίες που υλοποιούν, η οποία ποικιλομορφία τείνει να αντικατοπτρίζεται και στην κυτταρική χωροθέτησή τους καθώς και στα των χημικών τους ιδιοτήτων. Σε ό, τι αφορά στα ένζυμα πρέπει να σημειωθεί πως αυτά ενέχουν μιας διαλυτής εικόνας και με πολλή ευκολία απομονώνονται από το κυτταροδιάλυμα (cytosol), όμως αναφορικά με τις διαμεμβρανικές πρωτεΐνες καθώς και τους μεταφορείς αυτές είναι τοποθετημένες μέσα στις λιπιδικές μεμβράνες. Γενικότερα, δεν υπάρχει η δυνατότητα μιας πειραματικής διάταξης υπό μία καθόλα διακριτική ικανότητα, η οποία και να δίνει την ευκαιρία της ταυτοποίησης αλλά και της ποσοτικοποίησης των πρωτεϊνών σε ίχνη και ταυτόχρονα να μπορεί να λαμβάνει κι ένα ικανοποιητικής φύσης δείγμα από όλες τις θέσεις του κυττάρου. Με την προσπέλαση λοιπόν όλων αυτών των τεχνικών δυσκολιών, εκείνο που έχει αξία είναι να δοθεί σημασία μεγαλύτερη στη διαδικασία παραγωγής των πρωτεϊνών,

η οποία είναι μια εξόχως δύσκολη διαδικασία, που εμπεριέχει πλήθος σταδίων, τα οποία συνοπτικά είναι και τα εξής: καταρχάς είναι το στάδιο της ενεργοποίησης μεταγραφής όλων των γονιδίων τους, εν συνεχεία είναι το στάδιο επεξεργασίας-ανάλυσης του mRNA και ακολουθεί και η μετάφρασή του από τα ριβοσώματα, ένα άλλο στάδιο είναι των μετα-μεταφραστικών αλλαγών ή τροποποιήσεων και καταλήγει η όλη διαδικασία στο τελικό στάδιο αναδίπλωσης των αμινοξικών αλυσίδων σε μια πλήρως λειτουργική πρωτεΐνη. Από όλα αυτά τα προαναφερθέντα στάδια, πρέπει να τονιστεί ότι το 1^ο σε επίπεδο χρόνου δηλαδή η μεταγραφή των mRNA είναι κι εκείνο το στοιχείο, το οποίο διαθέτει εν τέλει και τα πιο πολλά θετικά χαρακτηριστικά ως προς την αξιοπιστία και την αποτελεσματικότητα των μετρήσεων. Το mRNA έχει την ιδιότητα να απομονώνεται με ευκολία και σε ποσότητες μάλιστα τέτοιες, οι οποίες ενδείκνυνται για μια εξαιρετικά αποτελεσματική ποσοτικοποίηση. Επίσης και η πλειονότητα των μεθόδων, που ενέχονται, έχουν τη δύναμη να ενδυναμώνουν κατά πολύ το δείγμα καθώς και να συμβάλουν στην ταυτοποίηση ακόμη και των μορίων mRNA, που βρίσκονται σε ελάχιστες συγκεντρώσεις, και σημειωτέον πως η μεταγραφή είναι και το στάδιο εκείνο στη διαδικασία της έκφρασης των γονιδίων, το οποίο περνά από πολύ αυστηρές ρυθμίσεις (Maier, Güell & Serrano 2009 · Vogel & Marcotte 2012). Με βάση λοιπόν όλη την παραπάνω ανάλυση, το σύνολο των μετρήσεων σε mRNA μπορεί κι εξασφαλίζει μια ομοιογένεια του δείγματος, επίσης είναι πολύ πιο εύκολο να δοκιμαστεί σε πειράματα, ενώ διαθέτει και μια ακρίβεια ως προς την ποσότητα. Ακόμη, τα επίπεδα του mRNA αντικατοπτρίζουν αμεσότερα όλες τις μεταβολές που επισυμβαίνουν στο πλαίσιο έκφρασης των γονιδίων, στη βάση πως ο μεσολαβών χρόνος από τότε που ένα γονίδιο κινητοποιείται σε παραγωγή της πρωτεΐνης που κωδικοποιεί, τότε αυτή η συνθήκη μπορεί να μην επιτρέπεται για την ανάλυση μιας ταχείας απόκρισης σε διάφορα ερεθίσματα. Ολοκληρώνοντας, να αναφερθεί πως μια ομάδα μελετών έχουν καταλήξει στο συμπέρασμα πως υπό σταθερές συνθήκες (steady state) όλα τα επίπεδα των mRNA δείχνουν να συνδέονται κατά πολύ με τα αντίστοιχα των πρωτεϊνών, και αν και ο βαθμός της συσχέτισής τους είναι

περιορισμένος για τα mRNA που βρίσκονται σε χαμηλή, και πάλι όμως η ποσοτικοποίηση των mRNA έχει τεθεί ως η πλέον πιο διαδομένη, άμεση και δη η πιο συστηματική μέθοδος για την επεξεργασία της γονιδιακής έκφρασης (Maier, Güell & Serrano 2009 · Vogel & Marcotte 2012).

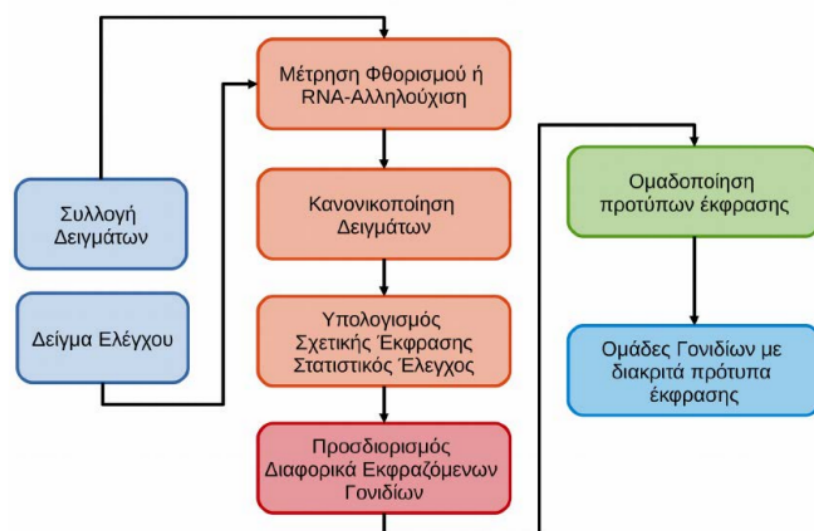
και

2. *Αλληλούχιση RNA (RNA Sequencing)*: η πολύ αυξημένη ζήτηση για την αλληλούχιση χαμηλού κόστους έχει οδηγήσει την επιστημονική κοινότητα τα τελευταία χρόνια στη διαμόρφωση κι ανάπτυξη της αλληλούχισης υψηλής απόδοσης ή αλλιώς της αλληλούχισης της επόμενης γενιάς (next generation sequencing, NGS). Ουσιαστικά πρόκειται για μια βιολογικής φύσης τεχνολογία της επανομαζόμενης ως «παραλληλοποίησης» της διεργασίας της αλληλούχισης, υπό μια πρακτική, η οποία και κάνει εφικτό τον προσδιορισμό της πρωτοταγούς διαδοχής βάσεων πολλών εκατομμυρίων ή και δεκάδων εκατομμυρίων αλληλουχιών σε παράλληλο χρόνο. Αναφορικά με τον παράγοντα τώρα της επεκτασιμότητας, και η ταχύτητα αλλά και η συσχέτιση μεταξύ κόστους και απόδοσης των NGS εφαρμογών δίνουν τη δυνατότητα στον ερευνητικό κόσμο να επεξεργαστεί τα βιολογικά συστήματα σε ένα τομέα, ο οποίος έως πρότινος δεν ήταν εφικτός. Προς το παρόν πάντως χρήσιμο θα ήταν να περιγραφεί ενδελεχώς η όλη διεργασία αποκομιδής των δεδομένων για ένα πείραμα έκφρασης των γονιδίων μέσω αξιοποίησης NGS. Αυτή λοιπόν η NGS ποσοτικοποίηση της γονιδιακής έκφρασης υλοποιείται μέσω της μαζικής αλληλούχισης mRNA, η οποία με τη σειρά της έχει την ιδιότητα απομόνωσής της από το δείγμα, κι ακολούθως μεταβάλλεται σε cDNA όπως γίνεται και στην περίπτωση των μικροσυστοιχιών. Σε ό, τι έχει να κάνει τώρα με την καταγραφή των διαφορών, που ενέχει η τεχνική της RNA, ή αλλιώς της λεγόμενης ως RNASeq, αυτές λοιπόν οι διαφορές αρχίζουν ως εξής: στην αρχή το δείγμα υποβάλλεται σε μια φάση κλασμάτωσης πριν την αλληλούχιση, στη βάση πως όλες οι προϋπάρχουσες τεχνολογίες στην πλειονότητά τους τείνουν να εξαγάγουν πολύ αξιόπιστα ευρήματα – αποτελέσματα σχετικά με αλληλουχίες, οι οποίες όμως δεν είναι μεγαλύτερες από 300-500 βάσεις. Ακολούθως, κι

αναλόγως με την εκάστοτε υιοθετηθείσα τεχνολογία το cDNA ενδυναμώνεται με τη συμβολή μιας αλυσιδωτής αντίδρασης πολυμεράσης (PCR), η οποία εν προκειμένω ενείχε κι άλλων τρόπων διεξαγωγής, και εν τέλει καταφέρνει να αλληλουχείται μαζικά και συνηθέστερα με την αρωγή της «σύνθεσης». Και αυτό ακριβώς πώς μεταφράζεται: αυτό λοιπόν δείχνει πως οι νέοι κλώνοι cDNA παράγονται εντός του εκμαγείου των κλώνων εκ του εκάστοτε δείγματος και με την όλη αυτήν τη διεργασία της σύνθεσης να αποτυπώνεται κιόλας ως νουκλεοτίδιο-νουκλεοτίδιο (Wang, Gerstein & Snyder 2009). Το αποτέλεσμα που εξάγεται είναι ένα αρχείο, το οποίο κι εμπεριέχει μια πλειάδα δεκάδων εκατομμυρίων αλληλουχιών μικρού μήκους σε ένα πλαίσιο από 100 έως και 500 βάσεις. Επιπρόσθετα μια άλλη διαφοροποίηση ως προς τις μεθοδολογίες, οι οποίες και στηρίζονται στην υβριδοποίηση, είναι εκείνη που συνδέεται με την ποσοτικοποίηση των αποτελεσμάτων. Αυτή πρέπει να σημειωθεί πως συνηθέστερα πραγματοποιείται με στόχο τη διεξαγωγή ενός πειράματος αλληλούχισης της επόμενης όμως γενιάς με την αρωγή των ακόλουθων βημάτων (Mortazavi et al. 2008): 1ο. Έλεγχος της ποιότητας των αλληλουχιών αλλά κι αποκλεισμός όλων όσα δεν καλύπτουν ορισμένα κριτήρια αξιοπιστίας. Είναι σημαντικό να τονιστεί ότι σε αυτό το πρώτο βήμα απορρίπτονται όλες εκείνες οι αλληλουχίες, οι οποίες δεν ενέχουν των απαραίτητων προϋποθέσεων σε επίπεδο ποιότητας. 2ο. Χαρτογράφηση των αλληλουχιών στο γονιδίωμα αναφοράς, το οποίο εν ολίγοις αναφέρεται σε μια κύρια διεργασία διαμέσω της οποίας ποσοτικοποιούνται τα αποτελέσματα. Για καθεμία από τα αναρίθμητα σε ποσότητα έως και τα εκατομμύρια των μικρών αλληλουχιών ή αλλιώς των επονομαζόμενων ως «αναγνώσεων αλληλουχιών» (sequence reads ή και reads) συνήθως εντοπίζεται η θέση του γονιδιώματος από την οποία μάλιστα και πηγάζει κι επονομάζεται ως γονιδίωμα αναφοράς (reference genome). Μετά λοιπόν από τη χαρτογράφηση όλων των reads, δίνεται και η δυνατότητα αναγνώρισης του αριθμού επακριβώς του «διαβάστηκε» για το εκάστοτε νουκλεοτίδιο του υπό επεξεργασία γονιδιώματος, 3ο. Ποσοτικοποίηση του αριθμού των αναγνώσεων ανά μετάγραφο. Με σημείο

αναφοράς το δεδομένο των θέσεων των γονιδίων/μεταγράφων μέσα στο γονιδίωμα αναφοράς είναι εφικτή η καταμέτρηση του αριθμού των αναγνώσεων, που επικαλύπτονται με κάθε γενετικό τόπο ένα-ένα. Εάν μάλιστα ληφθεί υπόψη και το μήκος της αντίστοιχης γονιδιωματικής περιοχής ή αλλιώς ληφθεί υπόψη η ακρίβεια των μεταγράφων mRNA που πηγάζουν από αυτήν), καθώς και το σύνολο του αριθμού των αναγνώσεων που εξήχθησαν από το πείραμα, τότε θα μπορεί να καταλήξει σε μια αριθμητική τιμή που είναι και δηλωτική της ποσότητας mRNA, η οποία και προϋπήρχε στο αρχικό δείγμα από τον εν λόγω συγκεκριμένο γενετικό τόπο (Geoffrey & Robert, 2011).

Ανεξαρτήτως όμως της όποιας τεχνικής έχει αξιοποιηθεί έως και στις μέρες μας σχετικά με την ανάλυση της γονιδιακής έκφρασης, οι φάσεις ή τα βήματά της είναι συγκεκριμένα. Θα ήταν λοιπόν πολύ χρήσιμο να αναφερθούν και τα προτεινόμενα στάδια επεξεργασίας – ανάλυσης ενός πειράματος γονιδιακής έκφρασης, τα οποία είναι και τα ακόλουθα (βλ. Εικ. 6): 1^{ov} η απομόνωση του mRNA, 2^{ov} η ποσοτικοποίησή του, 3^{ov} ο προσδιορισμός των διαφορικά εκφραζόμενων γονιδίων και 4^{ov} η κατηγοριοποίηση των γονιδίων ανάλογα με τα πρότυπα έκφρασής τους (Geoffrey & Robert, 2011).



Εικ. 6 – Σχηματική απεικόνιση των σταδίων ανάλυσης ενός πειράματος γονιδιακής έκφρασης από την αποκομιδή των πρωτογενών δεδομένων ως τη δημιουργία ομάδων γονιδίων με χαρακτηριστικά πρότυπα έκφρασης

Πηγή:

<file:///C:/Users/WORK2/Downloads/%CE%92%CE%99%CE%9F%CE%A0%CE%9B%CE%97%CE%A1%CE%9F%CE%A6%CE%9F%CE%A1%CE%99%CE%9A%CE%97%20-%20%CE%93%CE%9F%CE%9D%CE%99%CE%94%CE%99%CE%91%CE%9A%CE%95%CE%A3%20%CE%95%CE%9A%CE%A6%CE%A1%CE%91%CE%A3%CE%95%CE%99%CE%A3.pdf>

- **Υπολογιστική εξελικτική βιολογία** (*computational evolutionary biology*): η εξελικτική βιολογία αποτελεί τον κλάδο εκείνον της επιστήμης της Βιολογίας, η οποία και (δι-)ερευνά την προέλευση και καταγωγή των ειδών όπως επίσης (δι-)ερευνά και το σύνολο των εξελικτικών σχέσεων που ενυπάρχουν μεταξύ τους. Η επιστήμη της πληροφορικής κατάφερε τα μέγιστα σε ό, τι αφορούσε στην αναπτυξιακή πορεία του κλάδου αυτού, δίνοντας πολύ δυνατά εργαλεία στα χέρια των εξελικτικών βιολόγων. Ειδικότερα σε όλο αυτό συνέβαλε (Rodríguez, 2012, όπ. αναφ. στο Βουλγαρίδης & Σαμαράς, 2012): 1^ο η εμπέδωση της εξελικτικής πορείας της πλειονότητας εκ των οργανισμών υπό το πρίσμα των μεταβολών στο DNA τους, 2^ο επίσης βοήθησε η εξέλιξη των πολυσύνθετων υπολογιστικών συστημάτων αναφορικά με την ανάλυση πληθυσμών στους οποίους και πραγματοποιείται πρόβλεψη για τη μελλοντική εικόνα των πληθυσμών αυτών, 3^ο η ανίχνευση στοιχείων ενός ολοένα κι διογκούμενου αριθμού τύπων κι οργανισμών, 4^ο η συγκριτική αποτίμηση ολόκληρων γονιδιωμάτων, γεγονός που δίνει τη δυνατότητα ανάλυσης πολύ πιο πολύπλοκων εξελικτικών φαινομένων, όπως π.χ. είναι η οριζόντια μορφή μεταφορά των γονιδίων (*horizontal gene transfer*) ή και η επανάληψη των γονιδίων (*gene duplication*).

2.7 ΕΡΓΑΛΕΙΑ ΤΗΣ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ

Η πρόοδος του κλάδου της Βιοπληροφορικής προέκυψε από την ανάγκη των ερευνητών να εισαχθούν στην ανάλυση και την επεξεργασία βιοϊατρικών κατά κύριο λόγο δεδομένων, τα οποία συλλέγονταν από πολλά και διάφορα μέρη του κόσμου. Αυτή η ταχύτατη συγκέντρωση των δεδομένων ήταν η κατάληξη μιας σειράς μεγάλων τεχνολογικών αλμάτων, που με τη σειρά τους βοήθησαν στην παραγωγή μιας τεράστιας ποσότητας από δεδομένα βιολογικών αλληλουχιών (Zayed, 2019). Όλη αυτή η διεργασία οδήγησε στην ανάγκη ανάπτυξης ακόμη πιο πολλών αλγορίθμων και υπολογιστικών εργαλείων, που να επεξεργάζονται, να διαχειρίζονται καθώς και να αποθηκεύουν αυτά τα δεδομένα. Έτσι όλες αυτές οι εξελίξεις δημιούργησαν τον κλάδο, τον οποίο στις μέρες μας επονομάζουμε ως Βιοπληροφορική (Attwood et al., 2011 · Mohamadou, Halidou & Karpn, 2020).

Σε ό, τι αφορά πιο συγκεκριμένα στην αλματώδη εξέλιξη της υπολογιστικής βιολογίας αυτή οδήγησε στην αξιοποίηση μιας σειράς αλγοριθμικών και υπολογιστικών μέσων στην καθημερινή εργασία του βιολόγου επιστήμονα, όπου οι ενέργειές του κινούνται μεταξύ δύο τομέων στο αντικείμενο της Βιοπληροφορικής: 1^ο στην απλή επεξεργασία κι ανάλυση του λογισμικού, δηλαδή των διαφόρων αλγορίθμων στοίχισης ή και πολλαπλούς στοίχισης κλπ., 2^ο στη χρήση έτοιμων και προκαθορισμένων αλγορίθμων, προκειμένου να υλοποιήσουν σύνθετες αναλύσεις με στόχο την εμπειριστατωμένη απάντηση σε ένα βιολογικής φύσης ερώτημα. Στην περίπτωση αυτήν δεν γίνεται καμία χρήση ούτε του λογισμικού αλλά ούτε και των αλγορίθμων, και στην ανάλυση ενός ορισμένου προβλήματος αξιοποιώντας όμως εν προκειμένω και το λογισμικό αλλά και τους αλγόριθμους (Κίζυλη, 2017).

Ακόμη πρέπει να τονιστεί πως το κυρίαρχο μέσο-εργαλείο στην υπολογιστική ανάλυση αλληλουχιών καθώς και το βασικό μέλημα των μοριακών βιολόγων, που εργάζονταν σε εργαστήρια, ήταν η αναζήτηση σε βάσεις δεδομένων. Μια ομάδα όμως δεδομένων, που διαρκώς αυξάνονταν, είχε

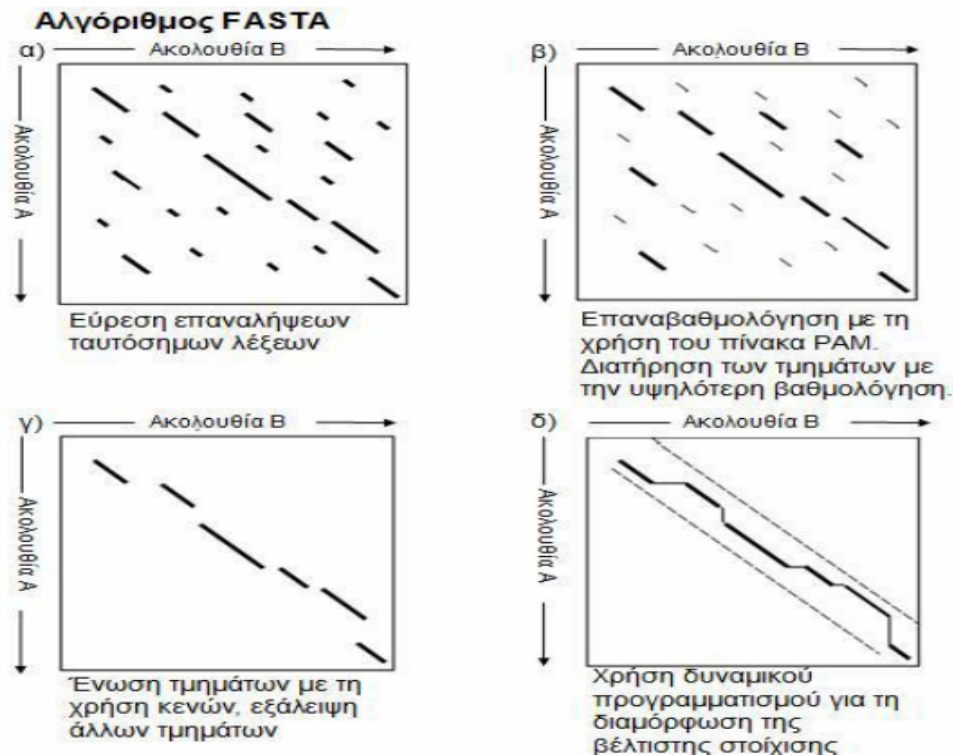
ως συνέπεια να δημιουργηθεί ένα μεγάλο πρόβλημα, στη βάση πως εξελικτικά και με ταχύτατους ρυθμούς ήταν σαφές πως ο αριθμός των καταχωρήσεων θα διογκωνόταν ακόμη περισσότερο. Οι επιστήμονες γνωρίζοντας αυτήν την εξέλιξη αποφάσισαν να ξεκινήσουν μία έρευνα για τον σχεδιασμό ταχύτατων και αποδοτικών ευριστικών αλγορίθμων, οι οποίοι εν προκειμένω θα διεξήγαγαν μιας συναφούς μορφής εργασίας όμως σε εξαιρετικά μειωμένο χρόνο. Σε ό, τι αφορά πιο συγκεκριμένα στην έννοια και τη χρησιμότητα του ευριστικού αλγορίθμου να σημειωθεί πως συνηθέστερα αποδίδει το ίδιο καλά με τον αυστηρά μαθηματικής φύσης αλγόριθμο και υλοποιεί τις διάφορες αναλύσεις με πιο γρήγορο όμως ρυθμό τις περισσότερες φορές. Οι σημαντικότεροι αλγόριθμοι αυτής της μορφής είναι οι κάτωθι δηλαδή ο Fasta και ο Blast, που αναλύονται κι ενδελεχώς (Κίζυλη, 2017 · Zayed, 2019).

2.7.1 Ο αλγόριθμος FASTA

Αναφορικά με το πεδίο ανάλυσης των αλληλουχιών μακρομορίων, η όλη ερευνητική μελέτη για τους αλγορίθμους στοίχισης καθώς και για τις αποτελεσματικές εφαρμογές τους πρέπει να σημειωθεί ότι συνεχίστηκε με πολύ συστηματικό ρυθμό. Και σε όλη αυτήν την εξέλιξη πρωτεύοντα ρόλο διαδραμάτισε η ανάπτυξη του αλγορίθμου της τοπικής στοίχισης από τους Smith και Waterman κατά το έτος 1981, με τον πρώτο ευριστικής μορφής και τοπικής ευθυγράμμισης αλγόριθμο, που στόχο είχε την ταχύτατη στοίχιση κι ανίχνευση κοινού στοιχείου σε μια βάση δεδομένων, κι αυτός ήταν ο FASTA (Μπάγκος, 2015). Χρήσιμο θα ήταν στο σημείο αυτό να αποσαφηνιστεί η έννοια του «ευριστικού», κι εδώ σημαίνει ο αλγόριθμος εκείνος, που έχει τη δυνατότητα ταχύτατα και με περισσή ευκολία να βρει μία λύση, η οποία όμως δε θα είναι απαραίτητα και η καλύτερη. Η λέξη άλλωστε «FASTA» αποτελεί μια συντομογραφία του «FAST-ALL», και το γεγονός πως κυριάρχησε οφείλεται στο γεγονός πως αξιοποιήθηκε με άμεσο στόχο μια ταχεία ευθυγράμμιση και για τις νουκλεοτιδικές αλληλουχίες αλλά και για τις πρωτεϊνικές αλληλουχίες. Εκτός

από αυτόν λοιπόν τον αρχικό αλγόριθμο, πρέπει να επισημανθεί πως υπήρξαν και αρκετές παραλλαγές του, εκ των οποίων ο πρωτεύων μάλιστα στόχος του εν λόγω FASTA ήταν ο κατά προσέγγιση εντοπισμός της διαγωνίου περιμετρικά της οποίας αναπτυσσόταν και η στοίχιση. Με αυτόν τον τρόπο έγινε εφικτό και να περιοριστεί ριζικά και το μέγεθος της αναζήτησης, καθώς και οι όποιες εντοπιζόμενες ομοιότητες, που ενυπήρχαν εκτός της διαγωνίου, στην πλειονότητα των περιπτώσεων δεν συνηθιζόταν να λαμβάνονται και υπ'όψιν (Μπάγκος, 2015).

Στην Εικόνα όμως 7 αποτυπώνονται πολύ παραστατικά τα βασικά στάδια-βήματα ως εξής (<http://www.ebi.ac.uk/Tools/sss/fasta/nucleotide.html>): 1. Στην αρχή κατασκευάζεται ένα τύπου ευρετήριο, εντός του οποίου ενυπάρχουν οι θέσεις όλων των k-tuples - εν προκειμένω εννοούνται οι λέξεις μεγέθους k, το τύποις μέγεθος για τις νουκλεοτιδικές αλληλουχίες τίθεται από το 4 έως και το 6 – οι οποίες μάλιστα θέσεις διαπιστώνονται να υπάρχουν και στις δύο συγκρινόμενες αλληλουχίες. 2. Μέσω της διαφοροποίησης μεταξύ των θέσεων των k-tuples στις δύο αλληλουχίες ανιχνεύεται η διαγώνιος, εντός της οποίας μάλιστα ενυπάρχουν, ενώ στη συνέχεια ανιχνεύονται και όλες οι διαγώνιοι με τα πιο πολλά κιάλας k-tuples. 3. Μέσω της εισαγωγής των κενών - υπό το πρίσμα καταμέτρησης της αντίστοιχης ποινής – γίνεται κι εφικτή η σύμπτυξη των διαγωνίων αυτών, και 4. Με τη συμβολή της διεργασίας του εξόχως δυναμικού προγραμματισμού επιτυγχάνεται και η στο μέγιστο στοίχιση (<http://www.ebi.ac.uk/Tools/sss/fasta/nucleotide.html>).



Εικ. 7 – Διαγραμματική σχεδίαση (τα κύρια στάδια-βήματα) του αλγόριθμου Fasta
Πηγή: Κίζυλη, 2017 (σελ. 27)

2.7.2 Ο αλγόριθμος *BLAST*

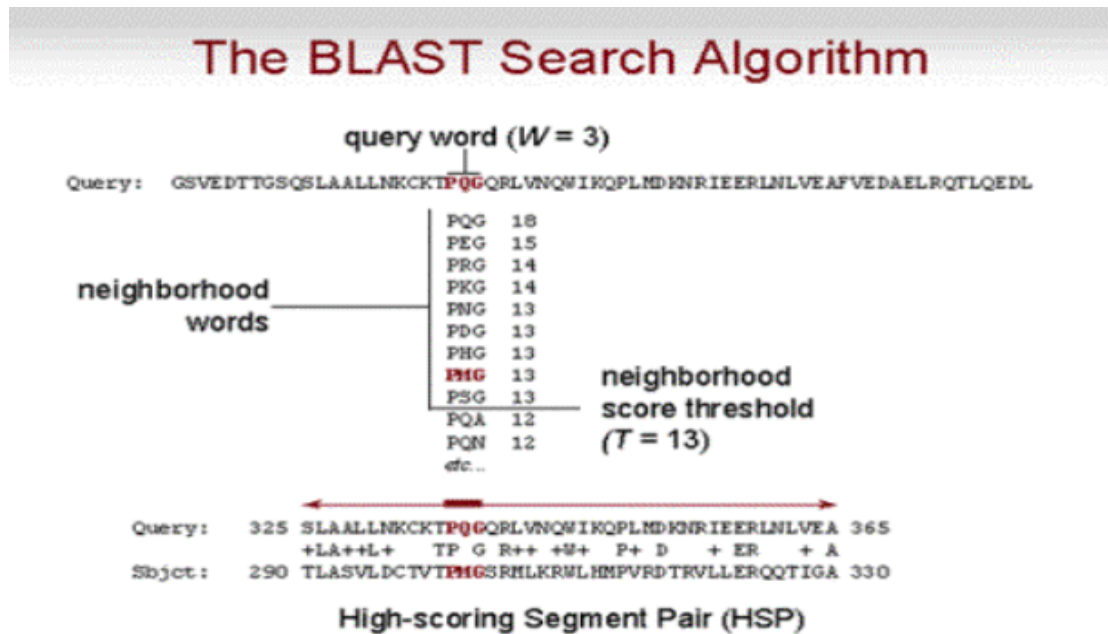
Ένα σπουδαίο επίτευγμα των τελευταίων δέκα περίπου ετών υπήρξε η εξέλιξη του ευριστικής μορφής αλγορίθμου BLAST⁶ (Basic Local Alignment Search Tool) από μια ομάδα επιστημόνων του NCBI κατά το έτος 1990. Αυτός ο αλγόριθμος βοηθά καταλυτικά στο ψάξιμο κοινών στοιχείων μεταξύ των αλληλουχιών σε βάσεις δεδομένων καθώς και στην στοίχιση, στη βάση πως τίθεται ως ο πλέον τάχιστος συγκριτικά με κάθε άλλον αλγόριθμο και επειδή δίνει τη δυνατότητα ταχέων αναζητήσεων υπό το πρίσμα να παρέχει πρωτόγνωρα μια εκτίμηση σε ό, τι αφορά στη στατιστική αξία των στοιχίσεων (Μπάγκος, 2015 · Hu, Chitnis, Monos & Dinh, 2021). Ο αλγόριθμος BLAST αποτελεί τον πλέον πιο

⁶ <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

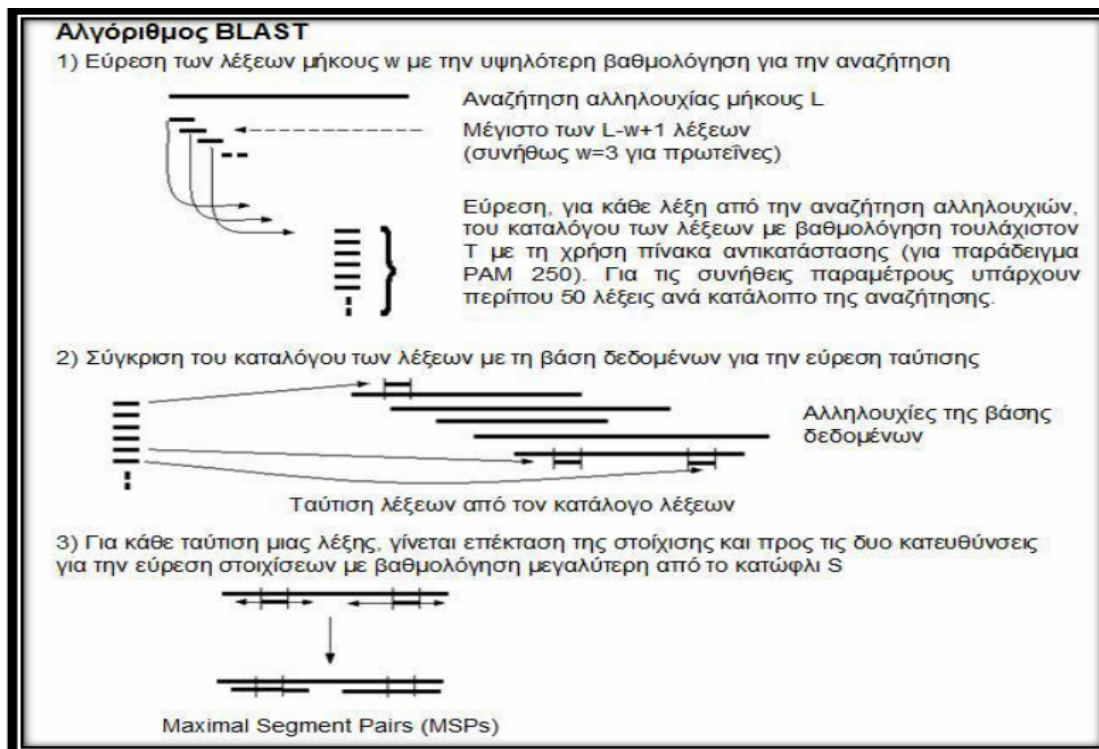
αξιοποιούμενο στο πεδίο της βιοπληροφορικής, για τον οποίο έχει παρατηρηθεί να έχουν εξελιχθεί και κάποιες παραλλαγές του που όμως φαίνεται να χρησιμοποιούνται για πολύ πιο εξειδικευμένα θέματα - προβλήματα (Attwood et al., 2011). Πιο συγκεκριμένα, σε αυτόν τον αλγόριθμο η λέξη BLAST απαρτίζεται από τα αρχικά των λέξεων Basic Local Alignment Search Tool, στην αρχή είχε αναπτυχθεί από το National Center for Biotechnology Information (NCBI) όντας ως ένας αλγόριθμος ευθυγράμμισης των αλληλουχιών βελτιστοποιημένης ταχύτητας, ο οποίος συνηθίζει να ανιχνεύει με στόχο τη βέλτιστη τοπικής φύσης στοίχιση στις διάφορες βάσεις των αλληλουχιών (Μπάγκος, 2015 · <https://blast.ncbi.nlm.nih.gov>).

Σε ό, τι αφορά στα προς εκτέλεση αρχικά στάδια-βήματα, πρέπει να τονιστεί πως εντοπίζονται κάποια κοινά στοιχεία με τα αντίστοιχα στάδια-βήματα εκτέλεσης του FASTA, εντούτοις ο ίδιος ο BLAST παρατηρείται να ενέχει καλύτερης ταχύτητας στη βάση πως δεν προτιμά να στοιχίζει αλληλουχίες (σημειώνοντας όμως και τις όποιες σχετικές παραμέτρους), οι οποίες μάλιστα διαπιστώνεται να υπάρχουν μέσα στη βάση των δεδομένων υπό το πλαίσιο πως δεν εμφανίζουν κανένα σπουδαίο και υπολογίσιμο κοινό στοιχείο. Εν προκειμένω, τα βασικά στάδια - βήματα του αλγόριθμου BLAST (βλ. Εικ. 8 και 9) είναι και τα ακόλουθα (<https://blast.ncbi.nlm.nih.gov>): 1. Στην αρχή διαμορφώνεται ένας τύπου κατάλογος, που εμπεριέχει όλες τις λέξεις (words), οι οποίες θα μπορούσαν να συνταιριάζουν με την όποια λέξη μιας άγνωστης αλληλουχίας. Όλες αυτές οι λέξεις είναι σημαντικό να ξεπερνούν την τιμή του οριζόμενου ορίου, 2. Το αμέσως επόμενο στάδιο-βήμα είναι η ανίχνευση όλων αυτών των λέξεων με την αρωγή του αλγόριθμου εντός των αλληλουχιών, οι οποίες διαπιστώνεται να ενυπάρχουν μέσα στη βάση των δεδομένων. Κάθε φορά που ανιχνεύεται μία λέξη, τότε ακριβώς και ξεκινά μία διεργασία εξάπλωσης του «ευρήματος» και προς τις δύο κατευθύνσεις της εν λόγω αλληλουχίας. Αυτή η εξάπλωση εξακολουθεί να διενεργείται μάλιστα για όσο χρόνο το score αυξάνεται, 3. Όλες οι περιοχές, οι οποίες ενέχουν και το πιο αυξημένο score, τίθενται και ως οι υποψήφιες περιοχές εντός των οποίων

ανιχνεύεται ένα κοινό στοιχείο (high scoring pairs). Εντός των ευρημάτων-αποτελεσμάτων καταγράφονται και όλες οι περιοχές των high scoring pairs όπου και το score τείνει να ξεπερνά και το όριο, εν τέλει όμως επιλέγονται οι τοπικής φύσης μόνον στοιχίσεις, οι οποίες κι ενέχουν και της πιο αυξημένης στατιστικής αξίας (<https://blast.ncbi.nlm.nih.gov> · Hu, Chitnis, Monos & Dinh, 2021).



Εικ. 8 – Σχηματική αποτύπωση του αλγόριθμου Blast
 Πηγή: Κίζυλη, 2017 (σελ. 29)



Εικ. 9 – Τα βασικά στάδια-βήματα του Blast
Πηγή: Κοτρώτσιος, 2017

Σε μια συγκριτική αποτίμηση μεταξύ του BLAST και του FASTA: ο αλγόριθμος BLAST υπολογίζει τη στατιστική αξία βάσει των παραμέτρων σπουδαιότητας της κατανομής, οι οποίες έχουν εκτιμηθεί από προσομοιώσεις, οι οποίες με τη σειρά τους εφαρμόστηκαν από πριν, ενώ και οι τιμές τους έχουν ήδη αποθηκευτεί. Στον αντίποδα ο FASTA συνηθίζει να καταμετρά όλες αυτές τις παραμέτρους από όλες τις άλλες αλληλουχίες που ενυπάρχουν μέσα στη βάση δεδομένων, εντός της οποίας και διεξάγεται μία αναζήτηση. Εν ολίγοις, ο μεν BLAST διαπιστώνεται να είναι και ο πιο τάχιστος συγκριτικά με τον δε FASTA. Αναφορικά με τις πρώτες εκδόσεις του BLAST, πρέπει να σημειωθεί ότι δεν εμφάνισαν τη δυνατότητα εισαγωγής των κενών στη διάρκεια της διεργασίας της εξεύρεσης ομοιοτήτων ανάμεσα στις διάφορες αλληλουχίες. Παρόλα αυτά με την έλευση της δεύτερης έκδοσης είναι σημαντικό πως προστέθηκε και η δυνατότητα εισαγωγής των κενών κατά τη διάρκεια της ευθυγράμμισης. Φυσικό βέβαια ήταν να ανακύψουν και άλλες νεότερες εκδόσεις του BLAST, εκ των

οποίων η καθεμιά τους να μπορεί να αξιοποιηθεί για έναν ανεξάρτητο σκοπό. Εν κατακλείδι, και ο BLAST αλλά και ο FASTA δείχνουν να μπορούν να παράξουν σημαντικά αποτελέσματα και κυρίως σχεδόν παραπλήσια με εκείνα, που παράγονται εκ της χρήσης-αξιοποίησης της τεχνικής του δυναμικού προγραμματισμού. Τέλος, σε ό, τι αφορά και στην επιλογή του πακέτου δυνατοτήτων ανάμεσα σε BLAST και σε FASTA, αυτό έχει σχέση με επιστημονικής φύσης κριτήρια, όπως για παράδειγμα είναι ο επακριβής στόχος της έρευνας, η απαιτούμενη ταχύτητα, οι τύποι των ακολουθιών που υπόκεινται σε σύγκριση μεταξύ τους (ή νουκλεοτιδικές ή πρωτεϊνικές), κλπ. (Μπάγκος, 2015 · <https://blast.ncbi.nlm.nih.gov> · <http://www.ebi.ac.uk/Tools/sss/fasta/nucleotide.html>).

Και ο BLAST και ο FASTA υπάγονται στην κατηγορία κατά ζεύγη στοίχισης.

2.7.3 To Clustal

Το Clustal όντας το πλέον γνωστό εργαλείο πολλαπλής στοίχισης των βιολογικών ακολουθιών είχε ως βασικό του στόχο τον εντοπισμό συγγένειας αλλά κι εξέλιξής τους. Το Clustal έχει τη δυνατότητα να επιτελέσει πολλαπλής φύσης στοίχισεις έως και πάνω από τον αριθμό των 100 νουκλεοτιδικών ή πρωτεϊνικών ακολουθιών, κι ενέχει δύο (2) κύριων τύπων της ClustalW που συμπεριλαμβάνει μία διεπαφή γραμμής εντολών και της ClustalX που αποτυπώνεται εντός ενός καθόλα γραφικού περιβάλλοντος (Καλλά, 2012).

2.8 ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ

Η Βιοπληροφορική βάσει των τεχνικών, που συνήθως αξιοποιεί, δίνει τη δυνατότητα δημιουργίας και παροχής νέων γνώσεων, οι οποίες γνώσεις κρίνονται ως πολύ σημαντικές και σε άλλες επιστήμες και αντικείμενα, όπως είναι για παράδειγμα η (εξατομικευμένη) Ιατρική (βλ. Εικ. 10) αλλά και η Φαρμακολογία σχετικά με τον ορθολογικό σχεδιασμό φαρμάκων (βλ. Εικ. 11).

Αναλυτικότερα (Dogra, Butner, Chuang, Caserta, Goel, Brinker & Wang, 2019 · Zayed, 2019)⁷:

1^ο στην Ιατρική η Βιοπληροφορική τείνει να ενασχολείται με την αντιπαραβολή του γενετικού υλικού των καρκινικών κυττάρων με το αντίστοιχο των υγιών κυττάρων, με την αξιοποίηση των ακολουθιακών τεχνικών, την ταυτοποίηση των γενετικών μεταβολών που προκύπτουν κατά την μετατροπή των υγιών κυττάρων σε καρκινικά⁸. Μάλιστα και η έρευνα των Dogra, Butner, Chuang, Caserta, Goel, Brinker & Wang (2019) εστίασε στην αντιμετώπιση του καρκίνου, ο οποίος αποτελεί ένα εκ των κορυφαίων προβλημάτων της υγειονομικής περίθαλψης παγκοσμίως με όλον τον ερευνητικό κόσμο να εξακολουθεί να διερευνά συστηματικά την εν λόγω νόσο, στοχεύοντας στην εύρεση καλύτερων φαρμάκων καθώς και πιο στοχευμένων πρακτικών χορήγησης των φαρμάκων στους ασθενείς. Η αναγκαιότητα λοιπόν για μια πιο ευθεία αντιμετώπιση του καρκίνου και των καρκινικών εν προκειμένω κυττάρων καθώς και των διαφόρων μεταβολικών νόσων και κληρονομικών ασθενειών έχει στρέψει την προσοχή των ερευνητών προς την εφαρμογή της νανοτεχνολογίας στην ιατρική, καθώς αυτός ο τομέας της νανοϊατρικής του καρκίνου έχει σημειώσει σημαντική πρόοδο, παρόλα αυτά ακόμη υπάρχουν (βιοφυσικά) εμπόδια σε ό, τι αφορά στην κλινική του μετάφραση. Όμως και η έρευνα Zayed (2019) επικεντρώθηκε στον σπουδαίο ρόλο, που διαδραματίζουν τα μαθηματικά στην εξέλιξη της ιατρικής, μέσω αξιοποίησης των τεχνικών των

⁷ <https://docplayer.gr/47186459-Eisagogi-sti-viopliforiki.html>

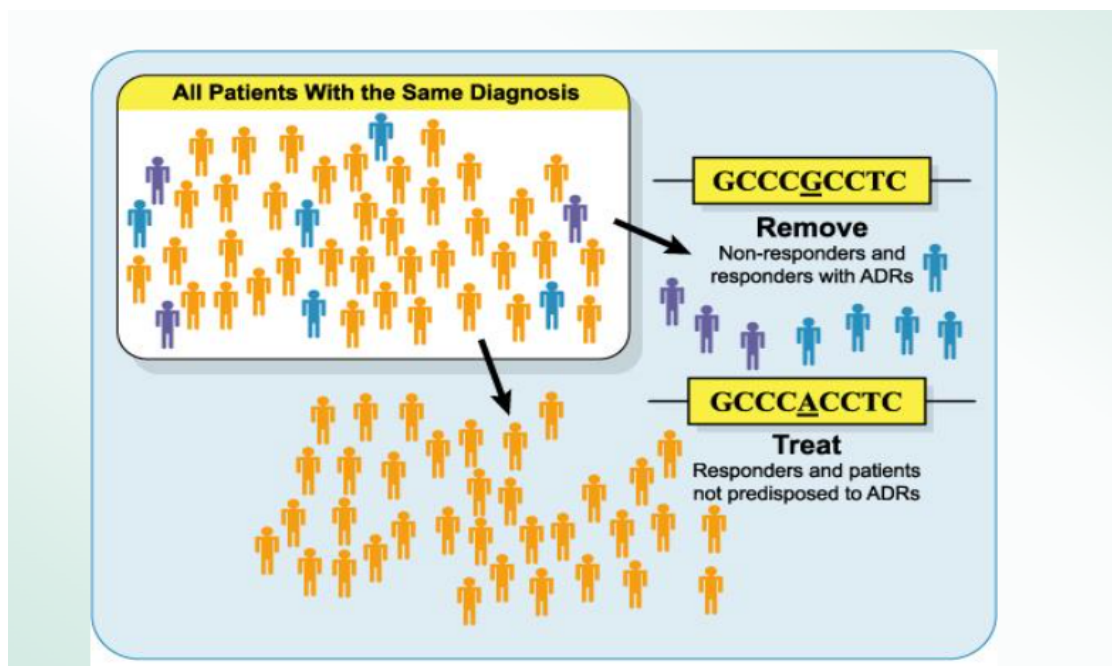
⁸ <https://helios-eie.ekt.gr/EIE/bitstream/10442/474/1/M01.027.01.pdf>

πιθανοτήτων και των στατιστικών για την επικύρωση της αποτελεσματικότητας πιο καινοτόμων εναλλακτικών νέων φαρμάκων ή διαδικασιών ή και για την εκτίμηση του ποσοστού επιβίωσης ασθενών με καρκίνο που υποβάλλονται σε κάποιες θεραπευτικές αγωγές. Κάτι ανάλογο επεδίωξε κι έπραξε και η έρευνα των Mohamadou, Halidou & Karpn (2020), καθώς ενέσκηψε στην πρόσφατη πανδημία COVID-19 υπό την αρωγή των μαθηματικών μοντέλων και της τεχνητής νοημοσύνης (AI) όμως συνδυαστικά, προσπαθώντας να ανακαλύψει τρόπους και να ανταποκριθεί στην έκτακτη ανάγκη για μια τάχιστα ανίχνευση, κατανόηση και θεραπεία της εν λόγω απρόσμενης ασθένειας. Πιο συγκεκριμένα, προκειμένου η πανδημία αυτή να τεθεί υπό τον έλεγχο της επιστημονικής κοινότητας και σιγά-σιγά με την πάροδο του χρόνου να εξλειφθεί τέθηκε η ανάγκη να κατανοηθεί η εξελικτική πορεία και κυρίως η παθογένεια του εν λόγω ιού, μία διαδικασία στην οποία συνέβαλε ριζικά η επιστήμη της Βιοπληροφορικής.

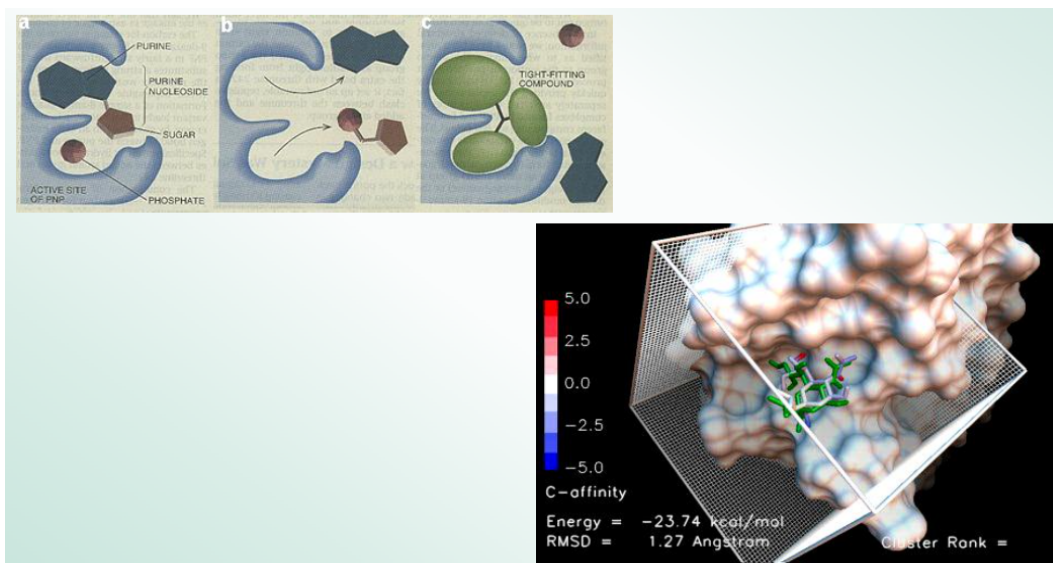
και

2^ο στη Φαρμακολογία η Βιοπληροφορική βοηθά σε επίπεδο διαμόρφωσης μιας ορθολογικής οργάνωσης των φαρμάκων εκείνων, που τίθενται για την αντιμετώπιση μολυσματικών ασθενειών (Κίζυλη, 2017). Αυτό βέβαια, εάν επιτυγχανόταν, τότε θα έλυνε και το πρόβλημα θνησιμότητας των μικρών παιδιών, μιας και οι μολυσματικές ασθένειες ταλαιπωρούν πολύ κόσμο όντας μία εκ των σημαντικότερων αιτίων θανάτου παγκοσμίως. Και σε αυτό κύρια αιτία είναι το αυξημένο κόστος των φαρμάκων. Όμως με την ταχύτατη ανάπτυξη της τεχνολογίας και δη της πληροφορικής με τις υπολογιστικές της τεχνικές είναι πλέον εφικτή η πρόγνωση-πρόβλεψη «ομοιότητας των φαρμάκων», που εν ολίγοις επρόκειτο για ανίχνευση και ανάλυση μορίων για τα οποία υπάρχει μεγάλη πιθανότητα να εξλειφθούν στην τελευταία φάση ανίχνευσης και ανάλυσης. Αυτή λοιπόν η διεργασία της ανίχνευσης των «όμοιων φαρμάκων» είναι δυνατό να επιτευχθεί με τη συμβολή του γενετικού αλγορίθμου καθώς και μέσω άλλων πρακτικών που στηρίζονται στο νευρωνικό δίκτυο. Μάλιστα είναι πολύ ελπιδοφόρο πως η πλειοψηφία εκ των

προγραμματιστών κάνουν φιλότιμη προσπάθεια, προκειμένου να σχεδιάσουν αποδοτικής φύσης αλγόριθμους, οι οποίοι θα έχουν τη δυνατότητα πρόγνωσης-πρόβλεψης των πρωτεϊνικών δομών καθώς και της αλληλεπίδρασης που διεξάγεται μεταξύ αυτών των δομών και των μικρότερων μορίων. Βέβαια αυτό το εγχείρημα των προγραμματιστών είναι εξόχως απαιτητικό σε ό, τι αφορά τους υπολογισμούς και έως και στις μέρες μας δεν έχει γίνει εφικτό να υλοποιηθεί, στη βάση πως απουσιάζει η αναγκαία υπολογιστική δύναμη (Κίζυλη, 2017 · Dogra, Butner, Chuang, Caserta, Goel, Brinker & Wang, 2019).



Εικ. 10 – Εξατομικευμένη Ιατρική
Πηγή: MacLeod & Evans, 2001



Εικ. 11 – Ορθολογικός σχεδιασμός φαρμάκων

Πηγή: <https://docplayer.gr/47186459-Eisagogi-sti-viopliroriki.html> (16)

Γενικότερα στις εφαρμογές της Βιοπληροφορικής έχουν συμπεριληφθεί έως και σήμερα οι κάτωθι (Μπάγκος, 2015 · Dogra, Butner, Chuang, Caserta, Goel, Brinker & Wang, 2019 · Zayed, 2019):

- η ανακάλυψη πολλών νέων φαρμάκων, στη βάση πως η επιστήμη της Βιοπληροφορικής έδωσε τη δυνατότητα στους επιστήμονες να καταφέρουν να παράξουν ειδικά φάρμακα περισσότερο οικονομικά και δη προσिता καθώς επίσης να το επιτύχουν αυτό και μέσα σε έναν πολύ μικρό χρόνο, με την αξιοποίηση της μοριακής μοντελοποίησης (molecular simulation), εν προκειμένω μέσω των υπολογιστικών μεθόδων, που μιμούνται και αναλύουν – επεξεργάζονται την όλη συμπεριφορά των μορίων,
- η προσωποποιημένη ιατρική, όντας ένα μοντέλο της ιατρικής φροντίδας, το οποίο καταφέρνει και προσαρμόζει τη θεραπεία στο ξεχωριστό γονιδίωμα του εκάστοτε ανθρώπινου οργανισμού,
- η προληπτική ιατρική, η οποία επικεντρώνεται στο θέμα της υγείας και του κάθε ατόμου ξεχωριστά αλλά και των ομάδων. Αξιοποιεί πλείστες ερευνητικές μεθόδους, όπως είναι η Βιοστατιστική, η Βιοπληροφορική και η Επιδημιολογία, με στόχο της να μπορέσει να αναγνωρίσει τις

κύριες αιτίες των διαφόρων εκδηλωθεισών νόσων αλλά και να μπορέσει να αξιοποιήσει τον εντοπισμό τους επιδιώκοντας τη δημιουργία προγραμμάτων σχεδιασμένων να μπορούν να αποφεύγουν τη νόσηση, την αναπηρία και τον θάνατο από μια ασθένεια,

και

- η νεωτερικής φύσης γονιδιωματική θεραπευτική αγωγή, η οποία εν ολίγοις σημαίνει μια αντικατάσταση των ελαττωματικών γονιδίων από άλλα λειτουργικά, στα κύτταρα ενός ασθενούς. Παρόλα αυτά ως πρακτική δυστυχώς δεν έχει υλοποιηθεί ακόμη έως και στις μέρες μας, στη βάση πως η διαμόρφωση μιας εξατομικευμένης θεραπείας για το γονιδίωμα ενός ατόμου τείνει να αποτελεί μια πολύ περίπλοκη πρόκληση για την επιστημονική κοινότητα.

ΚΕΦΑΛΑΙΟ 3 – ΤΑ ΜΑΘΗΜΑΤΙΚΑ

Η επιστήμη της Βιοπληροφορικής υιοθετεί και υλοποιεί μεθόδους της Πληροφορικής, οι οποίες με τη σειρά τους προέρχονται από πολλά επιστημονικά πεδία, όπως είναι τα εφαρμοσμένα μαθηματικά (όπως π.χ. οι διαφορικές εξισώσεις κλπ.)⁹, η Πληροφορική και η Στατιστική, προκειμένου να αντιληφθεί και συνάμα να καταφέρει να οργανώσει όλο αυτό το πλήθος των πληροφοριών που συσχετίζονται με τη μοριακή βιολογία σε μία ευρεία κλίμακα βέβαια. Πιο συγκεκριμένα η Βιοπληροφορική κάνει χρήση των υπολογιστικών μεθόδων με στόχο να μελετήσει, να επεξεργαστεί και να οργανώσει όλα τα βιολογικής μορφής δεδομένα-ευρήματα που είναι και διαθέσιμα, εν ολίγοις συλλέγει, αποθηκεύει και εκτιμά-αναλύει-επεξεργάζεται τα βιολογικά δεδομένα (Cohen, 2004 · Hack & Kendall, 2006 · Jungck & Weisstein, 2013).

Σε ό, τι αφορά όμως ειδικά την επιστήμη των Μαθηματικών στη συνέχεια της παρούσας εργασίας πιο συγκεκριμένα θα μελετηθούν με σύντομης μορφής παραδείγματα οι κλάδοι της Γραμμικής Άλγεβρας (βλ. Υποεν. 3.1), των

⁹ <https://math.uoi.gr>

Διαφορικών Εξισώσεων (βλ. Υποεν. 3.2), της Στατιστικής (βλ. Υποεν. 3.3) και των Διακριτών Μαθηματικών (βλ. Υποεν. 3.4), στη βάση πως οι επιστήμονες αξιοποιούν διάφορες πιθανότητες αλλά και στατιστικές, προκειμένου να αναγνωρίσουν διαφορετικά εκφραζόμενα γονίδια, ακολούθως να ταξινομήσουν δείγματα σύμφωνα με τα προφίλ έκφρασής τους καθώς και να φτάσουν να επικυρώσουν πειραματικά αποτελέσματα.

3.1 ΓΡΑΜΜΙΚΗ ΑΛΓΕΒΡΑ

Η γραμμική άλγεβρα έχει υιοθετηθεί και εφαρμόζεται εκτενώς στη βιοπληροφορική με στόχο την ανάλυση δεδομένων και τη μοντελοποίηση. Οι επιστήμονες μάλιστα αξιοποιούν διάφορες τεχνικές, όπως είναι για παράδειγμα η ανάλυση κύριου συστατικού (PCA) αλλά και η αποσύνθεση μονής τιμής (SVD) προκειμένου να επεξεργαστούν μεγάλα σύνολα δεδομένων αλλά και να αναγνωρίσουν μοτίβα στα δεδομένα. Ως εκ τούτου η *Γραμμική άλγεβρα* έχει τεθεί ως ένας από τους σπουδαιότερους κλάδους των Μαθηματικών, με πρωταρχικό αντικείμενο μελέτης του την επίλυση των συστημάτων γραμμικών εξισώσεων, τα διάφορα μητρώα, τις ορίζουσες αλλά και την ανάλυση των διανυσματικών περιοχών, των γραμμικής φύσης απεικονίσεων, των χαρακτηριστικών καθώς και των πρακτικών τους. Παρά το γεγονός πως σε θεωρητικό επίπεδο έχει τεθεί ως ένας αυτόνομος κλάδος των Μαθηματικών, εντούτοις η Γραμμική άλγεβρα, επειδή ακριβώς τείνει να παρέχει τα κυριότερα εργαλεία σχετικά με τη συστηματοποίηση των δυσκολιών, γι'αυτό και έχει τεθεί ως μια βασική προϋπόθεση σε ποικίλα επιστημονικά πεδία στα Εφαρμοσμένα μαθηματικά, στην πλειονότητα των Φυσικών επιστημών όπως και στις Επιστήμες των Υπολογιστών και της Μηχανικής, στην Επιστήμη της Οικονομίας κλπ. (<https://www.dit.uoi.gr/lessons/?id=1>).

Πιο συγκεκριμένα, η Γραμμική άλγεβρα δίνει έμφαση στους ακόλουθους στόχους(επιδιώξεις)

(<https://pithos.oceanos.grnet.gr/public/xVNSXsxtNMnK7OyY7FxxS5>):

1. Ανάλυση για το ποιο ακριβώς και με ποιες συνθήκες ισχύει μία λύση, ενδεχόμενο υπάρχει (μοναδική λύση, καμία λύση, απειρία λύσεων),
2. Διατύπωση και καταμέτρηση της προτεινόμενης λύσης, και αν αυτή η λύση είναι και η μόνη ενδεχομένως,
3. Παρουσίαση του συνόλου των προτεινόμενων τρόπων λύσης, σε περιπτώσεις που όλες αυτές είναι άπειρες.

Στο σημείο αυτό θα ήταν χρήσιμο να τονιστεί πως δεν έχει και τόσο ενδιαφέρον να επιλυθεί και μόνον ένα ορισμένο γραμμικό σύστημα, αλλά πολύ περισσότερο έχει σημασία να φτιαχτεί ένα πιο γενικό πλαίσιο, εντός του οποίου και θα γίνει εφικτό να επιτευχθούν όλοι οι προαναφερόμενοι στόχοι σε πιο γενικά μοντέλα της μορφής. Ακόμη θα ήταν εξίσου χρήσιμο να επισημανθεί πως σχεδόν όλες οι εξισώσεις είναι πολλές έως κι ενός ποσοστού των 104 περίπου (<https://pithos.oceanos.grnet.gr/public/xVNSXsxtNMnK7OyY7FxxS5>).

3.2 ΔΙΑΦΟΡΙΚΕΣ ΕΞΙΣΩΣΕΙΣ

Ως *Διαφορική εξίσωση* ορίζεται μία μαθηματικής φύσης εξίσωση, η οποία και συνδέει τις τιμές μιας άγνωστης συνάρτησης, μιας ή και πιο πολλών συχνά μεταβλητών καθώς και των παραγώγων της σε πρώτο, δεύτερο ή και πιο ανώτερο βαθμό. Οι Διαφορικές εξισώσεις διαδραματίζουν εν προκειμένω έναν βασικότατο ρόλο στην επιστήμη της Φυσικής, με πολυάριθμες εφαρμογές σε οικονομία, σε βιολογία κλπ. (Σγαρδέλης, 2005). Οι Διαφορικές εξισώσεις επεξεργάζονται στα Μαθηματικά με ποικίλες μεθόδους θεώρησης, οι οποίες θεωρήσεις και συνηθέστερα ενασχολούνται με τις λύσεις τους, εν προκειμένω με τις συναρτήσεις που καλύπτουν και την εκάστοτε εξίσωση. Συνήθως οι πιο απλές Διαφορικές εξισώσεις μόνον επιδέχονται κι επίλυσης, που πηγάζουν από τους αναλυτικούς τύπους (Σγαρδέλης, 2005).

3.3 ΣΤΑΤΙΣΤΙΚΗ

Η *Στατιστική* αποτελεί μια συστηματική μαθηματική επιστήμη, η οποία συνήθως εξάγει έγκυρες γνώσεις υπό το πρίσμα αξιοποίησης εμπειρικών δεδομένων παρατήρησης ή και διαφόρων πειραμάτων. Το πρωταρχικό αντικείμενο της έρευνας της Στατιστικής είναι η συλλογή/συγκέντρωση, η ταξινόμηση/κατηγοριοποίηση, η επεξεργασία, η παρουσίαση, η ανάλυση και η ερμηνεία πολλών δεδομένων, προκειμένου να εξαχθούν ασφαλή κι έγκυρα συμπεράσματα (ευρήματα) με στόχο τη λήψη σωστών αποφάσεων. Ουσιαστικά επρόκειτο για μια επιστήμη, της οποίας οι εφαρμογές ενέχουν ενός ευρέος πεδίου σε τομείς όπως είναι η Διοίκηση, οι Θετικές ή και οι Κοινωνικές επιστήμες κλπ. Η Στατιστική στις μέρες μας υπάγεται στα Εφαρμοσμένα μαθηματικά, με όλους όσοι ενασχολούνται με το στατιστικό αυτό ερευνητικό πεδίο ανάλυσης να επονομάζονται ή στατιστικοί ή και στατιστικολόγοι (Σγαρδέλης, 2015).

Ειδικότερα, η Στατιστική όντας μια κατηγορία των Μαθηματικών ουσιαστικά παρέχει και δύο δυνατότητες: 1^ο την περιγραφή των αριθμητικών συνόλων δεδομένων της έρευνας και 2^ο την επεξεργασία τους. Απόρροια αυτών των δύο δυνατοτήτων είναι και η κύρια κατηγοριοποίησή της σε 1. *Περιγραφική Στατιστική*, όπου περιγράφονται τα στατιστικά στοιχεία ύστερα από τη συγκέντρωση και την ταξινόμηση σε ομάδες όλων των στατιστικών δεδομένων, τα οποία ακολούθως και καταγράφονται σε μια μορφή μελέτης σε πίνακες ή διαγράμματα ή «πίτες» με τις τιμές ή τα ποσοστά τους, και σε 2. *Αναλυτική στατιστική*, η οποία θεωρείται πιο περίπλοκη, διεξάγεται με ποικίλες τεχνικές ο καθορισμός ποσοστού της εμπιστοσύνης για την εξαγωγή έγκυρων συμπερασμάτων, μέσω όμως ενός μικρού δείγματος στοιχείων από ένα γενικότερο σύνολο (Σγαρδέλης, 2005).

3.4 ΔΙΑΚΡΙΤΑ ΜΑΘΗΜΑΤΙΚΑ

Τα *Διακριτά Μαθηματικά* ενασχολούνται με την (κατα-)μέτρηση διακριτών εν προκειμένω ξέχωρων αντικειμένων, όπως είναι π.χ. τα pixels σε μία οθόνη, ή οι χαρακτήρες ενός κωδικού ή password κλπ. Καλύτερα, τα Διακριτά Μαθηματικά μελετούν τις μαθηματικές δομές που εκ φύσεως είναι διακριτές και δεν απαρτίζονται από την έννοια της συνέχειας. Όλα αυτά τα αντικείμενα είναι κατά βάση αριθμήσιμα σύνολα, όπως π.χ. αποτελούν οι ακέραιοι αριθμοί, τα γραφήματα κλπ. Πρέπει να τονιστεί ότι υφίσταται μια στενή σχέση ανάμεσα στα Διακριτά Μαθηματικά και την επιστήμη των Υπολογιστών, με βασικό ιδιαίτερο χαρακτηριστικό πως στους Υπολογιστές τα διακριτά αντικείμενα σκιαγραφούνται με μεγάλη ευκολία, και συνηθέστερα οι αλγόριθμοι επικεντρώνονται στον χειρισμό και την οργάνωση των διακριτών οντοτήτων (π.χ. των ακέραιων αριθμών κλπ.). Εντούτοις τα Διακριτά μαθηματικά αξιοποιούνται σε εφαρμόσιμο επίπεδο για την εξάλειψη πάσης φύσεως προβλημάτων από πολλά επιστημονικά πεδία όντας αναγκαία για τη επιλογή ενεργειών σε μη συνεχείς συγκυρίες (Κολουτζάκης & Παπαχριστόπουλος, 2015 · Σγαρδέλης, 2005 · Rosen, 2014 · <https://eclass.upatras.gr/courses/CULTURE159>).

ΚΕΦΑΛΑΙΟ 4 - Η ΣΥΜΒΟΛΗ-ΣΥΝΕΙΣΦΟΡΑ ΤΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΣΤΗ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ

Η επιστήμη της *Βιοπληροφορικής* -όπως πάρα πολλές φορές αναφέρθηκε στη διάρκεια σύνθεσης της παρούσας εργασίας- αποτελεί έναν ξεκάθαρα διεπιστημονικό κλάδο. Όμως παρά το γεγονός πως ένας κοινώς αποδεκτός ορισμός δεν υπήρξε, μια φιλότιμη προσπάθεια περί της εννοιολογικής της αποσαφήνισης έγινε, όταν ετέθη ως ένας επιστημονικός χώρος, εντός του οποίου διεξάγεται μία συνεργασία της Βιολογίας με τις επιστήμες της

Πληροφορικής, της Στατιστικής και των Μαθηματικών, όπου κι εξερευνώνται καινοτόμες μέθοδοι, προκειμένου να προσεγγίσουν και να αναλύσουν ενδελεχώς τα διάφορα βιολογικά ζητήματα-προβλήματα, όπως επίσης και την ιδέα των κύριων αρχών της Βιολογίας (Κοσσιδά, 2008 · Conference, 2006 · Jungck & Weisstein, 2013). Ειδικότερα, η Βιοπληροφορική επρόκειτο για έναν γνωστικό τομέα με οριοθετημένο αλλά κι ευρύ συνάμα πεδίο εφαρμοστικότητας κι αλληλεπίδρασης με τη σύγχρονη μοριακή, πληθυσμιακή καθώς και με τη περιβαλλοντική βιολογία. Στις μέρες μας μάλιστα η Βιοπληροφορική έχει τεθεί σε παγκόσμιο επίπεδο ως ένας εκ των πλέον εξελισσόμενων κλάδων, ενώ έχει δείξει και πολύ σπουδαία επιτεύγματα με εξίσου σπουδαίες επενδυτικές πρακτικές. Εν ολίγοις η Βιοπληροφορική έχει καταλάβει μια πρωτεύουσα θέση στις τωρινές σύγχρονες εξελίξεις των Επιστημών Ζωής, με αντιπροσωπευτικότερο παράδειγμα προγράμματος την «Αποκωδικοποίηση» των Γονιδιωμάτων όπως και του Ανθρώπου. Επίσης ένα άλλο χαρακτηριστικό παράδειγμα σε όλη την ιστορική πορεία της επιστήμης της Βιοπληροφορικής είναι η τύπου «στρατολόγηση» των ειδικών από πολλά και διάφορα γνωστικά αντικείμενα, όπως π.χ. από την επιστήμη της Βιολογίας, της Πληροφορικής, των Μαθηματικών κλπ., υπό το πρίσμα της αξιοποίησης όμως των μαθηματικών μεθόδων και όλων των υπολογιστικών τεχνικών σε ό, τι αφορά στη ακριβή περιγραφή και τη μελέτη των βιολογικών μοντέλων (He & Petoukhov, 2011). Μάλιστα η σταδιακά αυξανόμενη πολυπλοκότητα των επεξεργαζόμενων δεδομένων καθώς και η ποικιλότητά τους κάνει επιβεβλημένη τη συμβολή και κυρίως τη συνεργασία των ειδικών σε όλα αυτά τα επιστημονικά πεδία, και μάλιστα επαρκώς καταρτισμένων σε αρχές και σε μέθοδο της σύγχρονης Βιοπληροφορικής, προκειμένου να γίνει εφικτό να ανταποκριθούν και στις υψηλές απαιτήσεις του εν λόγω ερευνητικού πεδίου αλλά και του παραγωγικού (Αγγελίδης, 2015 · Klyuchko, 2017).

Η τεχνολογία της e-NIOS μάλιστα ως μια εξαιρετικά «ευφυής» -omic πλατφόρμα, η επονομασία της οποίας πηγάζει από το ακρωνύμιο e-Noesis Inspired Operational Systems δείχνοντας την ενσωμάτωση της ηλεκτρονικής

νόησης σε πολύ οργανωμένα μοντέλα αυτοματοποιημένης μορφής ανάλυσης όλων των δεδομένων, εμπεριέχει ενσωματωμένα μια πλειάδα από νεωτερικές ιδιότητες. Όλες αυτές λοιπόν οι ιδιότητες δίνουν τη δυνατότητα για προτυποποίηση κι αυτοματοποίηση της όλης διεργασίας ανάλυσης από την έναρξη έως και τη λήξη, προκειμένου να αναδυθούν αξιοποιήσιμα αποτελέσματα-ευρήματα με στόχο στις Βιοιατρικές και τις Βιοτεχνολογικές εφαρμογές. Μάλιστα στα τελευταία χρόνια έχει διαπιστωθεί μια πολύ μεγάλη εξέλιξη των εργαστηριακών πρακτικών σε ό, τι αφορά στην (κατα-)μέτρηση των μοριακών παραγόντων από βιολογικά δείγματα όπως είναι το DNA ή και το RNA). Όλη αυτή λοιπόν η εξέλιξη όμως δεν είχε και τα αναμενόμενα αποτελέσματα, στη βάση πως η όλη εκδηλωθείσα πολυπλοκότητα έδειχνε να ξεπερνά κατά πολύ τους προσδοκώμενους στόχους σχετικά με ταχύτατες θεραπείες των πιο πολλών τουλάχιστον ασθενειών. Η πλειονότητα των δεδομένων συσσωρεύτηκε γρήγορα μεν, όμως δεν έγινε εφικτό να επεξεργαστεί κιάλας με τον ίδιο αποτελεσματικό ρυθμό με έμφαση στην καρδιά του προβλήματος, που ήταν η ενδογενής βιολογική πολυπλοκότητα. Η παραπάνω έτσι τεχνολογία της e-NIOS ήρθε για να επιλύσει το πρόβλημα της ιεράρχησης των βιοδεικτών ανάμεσα σε ένα πλήθος παραγόντων, οι οποίοι και τείνουν να λειτουργούν αλλά και να επηρεάζουν συνάμα, και μάλιστα πολύπλοκα, εκ των οποίων κι ελάχιστοι διαθέτουν μια αιτιακή σχέση με το βιολογικό φαινόμενο όπως για παράδειγμα είναι η εμφάνιση-πρόκληση μιας ασθένειας. Πολύ πιο επιδεινωμένη είναι η περίπτωση αναζήτησης απόλυτα γενικεύσιμων γονιδιακών παραγόντων μιας και κρίνεται ως απόλυτα μάταιη, στη βάση πως δεν επηρεάζουν απόλυτα κι αυστηρά με επαναλαμβανόμενο τρόπο. Σε αυτές τις περιπτώσεις λοιπόν η λύση έχει σχέση με την εύρεση πολλαπλών δυναμικών συνδυασμών, τους οποίους μάλιστα η μεθοδολογία μιας εταιρείας οφείλει και να ψάχνει αλλά και να το δείχνει εμπράκτως. Αυτή η υπολογιστικής μορφής πλατφόρμα της e-NIOS είναι και η μόνη, η οποία έχει καταφέρει να ενσωματώσει τη βιολογική γνώση στη γονιδιωματική ανάλυση και οργανώνει αυτοματοποιημένα το σύνολο της

διεργασίας, με έμφαση στην τελική απόδοση ελάχιστων αξιοσήμαντων βιοδεικτών και εξόχως δυναμικών συνδυαστικών πρακτικών τους. Αυτή η τακτική της εταιρείας είναι η απόρροια πολλών ετών και δείχνει μια διεπιστημονική τεχνογνωσία σε πολλαπλούς τομείς, όπως της μοριακής βιολογίας, της στατιστικής κλπ.¹⁰.

Πολύ σημαντικό στο σημείο αυτό να τονιστεί -στη βάση της συμβολής και της συνεισφοράς των Μαθηματικών στη Βιοπληροφορική- πως ένα πλήθος από μαθηματικά μοντέλα έχει συμβάλει καταλυτικά στην επεξήγηση κι «ερμηνεία» πολλών νευρολογικών διαταραχών, και πιο συγκεκριμένα φάνηκε αυτό καθώς κατόρθωσαν έλληνες ερευνητές να επεξηγήσουν πολλές δυσλειτουργίες των μιτοχονδρίων, που είναι υπεύθυνες για τις νευρολογικές διαταραχές σε επίπεδο κοινών ασθενειών όπως είναι για παράδειγμα η νόσος του Αλτσχάιμερ κλπ. Μερικά από αυτά λοιπόν τα μαθηματικά μοντέλα όμως έχουν χρησιμεύσει καταλυτικά και σε άλλους διάφορους τομείς, όπως στην προσπάθεια να καταπολεμηθεί η νόσος του καρκίνου που οδήγησε στην αξιοποίηση της μεθόδου της νανοτεχνολογίας στον τομέα της ιατρικής, όπως αυτό διεφάνη στην έρευνα των Dogra, Butner, Chuang, Caserta, Goel, Brinker & Wang (2019). Επίσης και στην έρευνα Zayed (2019) δόθηκε εστίαση στην αξιοποίηση της μεθόδου των πιθανοτήτων και των στατιστικών για την επικύρωση υιοθέτησης της αποτελεσματικότητας νέων φαρμάκων ή για την εκτίμηση του ποσοστού επιβίωσης των ασθενών με καρκίνο που υποβάλλονται σε ορισμένες θεραπείες. Δε θα πρέπει όμως να παραλειφθεί και να μην αναφερθεί η συμβολή – συνεισφορά των μαθηματικών μοντέλων και στην πρόσφατη κι επίκαιρη πανδημία COVID-19, η ανίχνευση της οποίας πραγματοποιήθηκε με την αρωγή των μαθηματικών μοντέλων και της τεχνητής νοημοσύνης (AI) όντας τα δύο κύρια όπλα της πανδημίας αυτής της εποχής μας (Mohamadou, Halidou & Kapen, 2020). Στην έρευνα των Mohamadou et al (2020) επισημάνθηκε ότι η πλειονότητα εκ των μαθηματικών μοντέλων στηρίχτηκαν

¹⁰ <https://startupper.gr>

ειδικά στα μοντέλα Susceptible-Exposed-Infected-Removed (SEIR) και Susceptible-infected-Recovered (SIR), ενώ οι περισσότερες από τις υλοποιήσεις AI ήταν Convolutional Neural Network (CNN) σε ακτίνες X και εικόνες CT. Εν συνεχεία και η έρευνα των Shendure, Balasubramanian, Church, Gilbert, Rogers, Schloss & Waterston (2017) ανέδειξε τη σημαντικότητα προσδιορισμού της αλληλουχίας του DNA που στις μέρες μας έχει επεξεργαστεί ενδελεχώς, συμπεριλαμβάνοντας και τον «μετρητή» για ένα αυξημένο φάσμα των μοριακών φαινομένων. Όμως και μια άλλη έρευνα των He & Petoukhov (2011) συμπέρανε πως η πρόσφατη πρόοδος, που παρατηρήθηκε στον τομέα της βιοπληροφορικής, ουσιαστικά αποτέλεσε και την αφορμή για πλείστες γονιδιακές ακολουθίες. Εν προκειμένω τι ακριβώς μας λένε αυτές οι ακολουθίες; Διέπονται από κανόνες; Έχει φανεί λοιπόν πως πρέπει να αντιληφθούμε όλοι μας τη ζωή υπό το πρίσμα του γονιδίου, προκειμένου να κατανοήσουμε και να εμπεδώσουμε όλες αυτές τις ακολουθίες. Τα μαθηματικά είναι μια επιστήμη που μπορεί να καταφέρει αυτή τη σύνδεση ανάμεσα στις ακολουθίες, και οι οποίες παίζουν έναν πολύ ουσιαστικό ρόλο στην περιγραφή των περιπλοκοτήτων των βιολογικών δομών, των ακολουθιών αλλά και των διαδικασιών.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Η παρούσα εργασία ως κύριο σκοπό της είχε να μελετήσει βιβλιογραφικά την επιστήμη της Βιοπληροφορικής, με έμφαση στη συμβολή – συνεισφορά των μαθηματικών σε αυτήν την επιστήμη της Βιοπληροφορικής. Από τη Βιβλιογραφική ανασκόπηση, που πραγματοποιήθηκε, διαπιστώθηκε πως η επιστήμη της Βιοπληροφορικής αρχικά στράφηκε στην υλοποίηση διαφόρων προγραμμάτων σύγκρισης της ομοιότητας ακολουθίας με κύρια επιδίωξη την ανίχνευση εκείνων των ακολουθιών, οι οποίες και θα μπορούσαν να ευθύνονται για την εκδήλωση των γενετικών ασθενειών. Όμως η κύρια κινητοποίηση στον κλάδο της Βιοπληροφορικής σημειώθηκε με την αποκρυπτογράφηση του ανθρώπινου γονιδιώματος. Αυτός λοιπόν ο κλάδος εξαρτήθηκε άμεσα από την αποκρυπτογράφηση των γονιδιωμάτων και εν προκειμένω η ολοκληρωμένη χαρτογράφηση του ανθρώπινου γονιδιώματος προσέφερε και την απαιτούμενη ώθηση, προκειμένου η Βιοπληροφορική να τεθεί ως ένας εκ των πιο ενεργών επιστημονικών κλάδων. Στις μέρες μας μάλιστα η επιστήμη της Βιοπληροφορικής ενασχολείται πρωτίστως με τη μελέτη των γονιδιωμάτων (γονιδιωματική) αλλά και των πρωτεϊνών (πρωτεωμική), έχοντας καταφέρει πολύ σημαντικά επιτεύγματα. Εν κατακλείδι, η Βιοπληροφορική διεφάνη πως είναι ένας πολύ αναπτυσσόμενος διεπιστημονικός κλάδος, που ενέχει μιας στενής επαφής με τις επιστήμες της βιολογίας και των μαθηματικών καθώς και των υπολογιστών. Κι αυτό τεκμηριώνεται, καθώς η Βιοπληροφορική κάνοντας χρήση των υπαρχόντων πειραματικών δεδομένων μπόρεσε να αξιοποιήσει μεθόδους της στατιστικής, των μαθηματικών και της πληροφορικής, εισήγαγε εν ολίγοις τη μαθηματικοποίηση και την ποσοτικοποίηση των βιολογικών φαινομένων στην έρευνά της.

ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ

Ελληνόγλωσσες

Αγγελίδης, Π. (2015). *Opencourse Βιοπληροφορική*. Πανεπιστήμιο Δυτικής Μακεδονίας, Τμήμα Μηχανικών Πληροφορικής και Τηλεπικοινωνιών. Κοζάνη. <https://eclass.uowm.gr/courses/ICTE102> [Ανακτήθηκε από διαδίκτυο στις 29-04-2023].

Βουλγαρίδης, Α., Σαμαράς, Π. (2012). *ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ ΑΝΑΛΥΣΗ ΓΟΝΙΔΙΩΜΑΤΩΝ*. (Πτυχιακή/Διπλωματική εργασία). ΑΠΘ ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ. Θεσσαλονίκη. <http://ikee.lib.auth.gr/record/131117/files/%CE%A0%CF%84%CF%85%CF%87%CE%B9%CE%B1%CE%BA%CE%AE.pdf> [Ανακτήθηκε από διαδίκτυο στις 24-03-2023].

Δελβενιώτη, Δ. (2017). «*Διδακτικές προσεγγίσεις εννοιών της Βιοπληροφορικής στο Λύκειο: διεπιστημονική διερεύνηση*». (Διπλωματική εργασία). ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ ΣΧΟΛΗ ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ ΤΜΗΜΑ ΕΠΙΣΤΗΜΩΝ ΤΗΣ ΠΡΟΣΧΟΛΙΚΗΣ ΑΓΩΓΗΣ ΚΑΙ ΤΟΥ ΕΚΠΑΙΔΕΥΤΙΚΟΥ ΣΧΕΔΙΑΣΜΟΥ. Ρόδος. <https://hellanicus.lib.aegean.gr/bitstream/handle/11610/21096/%CE%94%CE%B9%CE%B4%CE%B1%CE%BA%CF%84%CE%B9%CE%BA%CE%AD%CF%82%20%CF%80%CF%81%CE%BF%CF%83%CE%B5%CE%B3%CE%B3%CE%AF%CF%83%CE%B5%CE%B9%CF%82%20%CE%B5%CE%BD%CE%BD%CE%BF%CE%B9%CF%8E%CE%BD%20%CF%84%CE%B7%CF%82%20%CE%92%CE%B9%CE%BF%CF%80%CE%BB%CE%B7%CF%81%CE%BF%CF%86%CE%BF%CF%81%CE%B9%CE%BA%CE%AE%CF%82%20%CF%83%CF%84%CE%BF%20%CE%9B%CF%8D%CE%BA%CE%B5%CE%B9%CE%BF%20%CE%B4%CE%B9%CE%B5%CF%80%CE%B9%CF%83%CF%84%CE%B7%CE%BC%CE%BF%CE%BD%CE%B9%CE%BA%CE%A>

[E%20%CE%B4%CE%B9%CE%B5%CF%81%CE%B5%CF%8D%CE%BD%CE%B7%CF%83%CE%B7.pdf?sequence=1&isAllowed=y](#) [Ανακτήθηκε από διαδίκτυο στις 29-04-2023].

Καλλά, Μ. Π. (2012). *ΕΞΟΡΥΞΗ ΓΝΩΣΗΣ ΑΠΟ ΙΑΤΡΟΒΙΟΛΟΓΙΚΑ ΔΕΔΟΜΕΝΑ (Biomedical data mining)*. (Διπλωματική Εργασία). Πανεπιστήμιο Πατρών Σχολή Θετικών Επιστημών Τμήμα Μαθηματικών ΜΠΣ « Μαθηματικά και Σύγχρονες Εφαρμογές». Πάτρα.
<https://nemertes.library.upatras.gr/server/api/core/bitstreams/c637bf76-d556-4259-b185-662915c83648/content> [Ανακτήθηκε από διαδίκτυο στις 27-08-2023].

Καναβάκης, Ε., Ξαϊδάρη, Α. (2001). Η χαρτογράφηση του ανθρώπινου γονιδιώματος. *Αρχεία Ελληνικής Ιατρικής*, 18(5): 474-484.
<https://www.mednet.gr/archives/2001-5/pdf/475.pdf> [Ανακτήθηκε στις 27-05-2023].

Κελεφιώτης, Δ. (2002). Ένας γενναίος νέος κόσμος, σαν όλους τους προηγούμενους. *ΕΠΙΣΤΗΜΗ ΚΑΙ ΚΟΙΝΩΝΙΑ: Επιθεώρηση Πολιτικής και Ηθικής Θεωρίας* (8-9), σσ. 119-146.

Κίζυλη, Θ. Γ. (2017). *Βιοπληροφορική ανάλυση του αναπνευστικού συστήματος*. (Διπλωματική εργασία). Εθνικό Μετσόβιο Πολυτεχνείο. Σχολή Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών. Αθήνα.
<https://dspace.lib.ntua.gr/xmlui/bitstream/handle/123456789/47160/%CE%92%CE%B9%CE%BF%CF%80%CE%BB%CE%B7%CF%81%CE%BF%CF%86%CE%BF%CF%81%CE%B9%CE%BA%CE%AE%20%CE%91%CE%BD%CE%AC%CE%BB%CF%85%CF%83%CE%B7%20%CE%91%CE%BD%CE%B1%CF%80%CE%BD%CE%B5%CF%85%CF%83%CF%84%CE%B9%CE%BA%CE%BF%CF%8D%20%CE%A3%CF%85%CF%83%CF%84%CE%AE%CE%BC%CE%B1%CF%84%CE%BF%CF%82.pdf?sequence=1> [Ανακτήθηκε από διαδίκτυο στις 24-03-2023].

Κολουντζάκης, Μ., Παπαχριστόπουλος, Χ. (2015). *Διακριτά Μαθηματικά*. Αθήνα: Κάλλιπος.

- Κοσσιδά, Σ. (2008). *Βιοπληροφορική. Δυνατότητες, προοπτικές*. Αθήνα: ΙΙΒΕΑΑ.
- Μπάγκος, Π. (2015). *Δομική Βιοπληροφορική* [Κεφάλαιο]. Στο Μπάγκος, Π. 2015. *Βιοπληροφορική* [Προπτυχιακό εγχειρίδιο]. Κάλλιπος, Ανοικτές Ακαδημαϊκές εκδόσεις. <https://repository.kallipos.gr/handle/11419/5025> [Ανακτήθηκε από διαδίκτυο στις 24-03-2023].
- Παπαϊωάννου Στέλλα-Συλβάνα (2007). *Βιοπληροφορική-Βάσεις Δεδομένων με γενετικά στοιχεία από μελέτες ψαριών – Αλεξάνδρειο Τεχνολογικό εκπαιδευτικό ίδρυμα Θεσσαλονίκης*.
- Σαχίνη, Ε., Μάλλιου, Ν., Χούσος, Ν., & Καραϊσκος, Δ. (2012). Ελληνικές Επιστημονικές Δημοσιεύσεις 2000-2010 - Τομέας Βιοεπιστημών. <https://metrics.ekt.gr/publications/40> [Ανακτήθηκε από διαδίκτυο στις 23-03-2023].
- Σγαρδέλης, Σ. Π. (2005). *Μαθηματικά Μοντέλα στη Βιολογία Σημειώσεις για το μάθημα Στατιστική – Μαθηματικά στη Βιολογία*. Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης Τμήμα Βιολογίας Τομέας Οικολογίας. Θεσσαλονίκη. <http://users.auth.gr/sgardeli/Mathjune05V1.pdf> [Ανακτήθηκε από διαδίκτυο στις 23-03-2023].
- Φραγκουδάκη, Ε. (2007). *Η νομική μεταχείριση των Εφαρμογών της βιογενετικής - ιδίως από πλευράς ιδιωτικού δικαίου*. Αθήνα: Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών (ΕΚΠΑ). Σχολή Νομικών, Οικονομικών και Πολιτικών Επιστημών. Τμήμα Νομικής. Τομέας Α' Ιδιωτικού Δικαίου, σ. 14.

Ξενογλωσσες

- Attwood, T. K., Gisel, A., Eriksson, N. E., & Bongcam-Rudloff, E. (2011). Concepts, historical milestones and the central place of bioinformatics in modern biology: a European perspective. *In Bioinformatics-trends and methodologies*. InTech.
- Bagos, P. G., Nikolaou, E. P., Liakopoulos, T. D. & Tsirigos, K. D. (2010). Combined prediction of Tat and Sec signal peptides with hidden Markov models. *Bioinformatics*, 26(22), 2811-2817.
- Baxevanis, A.D. & Ouellette, B.F.F. (2004). *ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ. Ένας Πρακτικός Οδηγός για την Ανάλυση Γονιδίων και Πρωτεϊνών*. μτφρ. Ε. Μουδριανάκη, Σ. Χαμόδρακα. Αθήνα: Παρισιάνος.
- Cohen, J.E. (2004). *Τα μαθηματικά είναι το επόμενο μικροσκόπιο της βιολογίας, μόνο καλύτερα. Η βιολογία είναι η επόμενη φυσική των μαθηματικών, μόνο καλύτερη*. <https://doi.org/10.1371/journal.pbio.0020439> [Ανακτήθηκε από διαδίκτυο στις 29-04-2023].
- Conference, I. B. S. I. R. (2006). *Statistical Advances in Biosciences and Bioinformatics*. Allied Publishers.
- Dogra, P., Butner, J. D., Chuang, Y. L., Caserta, S., Goel, S., Brinker, C. J. & Wang, Z. (2019). Mathematical modeling in cancer nanomedicine: a review. *Biomedical microdevices*, 21, 1-23. doi: [10.1007/s10544-019-0380-2](https://doi.org/10.1007/s10544-019-0380-2) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6449316> [Ανακτήθηκε στις 27-05-2023].
- Geoffrey, C. M., & Robert, H. E. (2011). *ΤΟ ΚΥΤΤΑΡΟ - ΜΙΑ ΜΟΡΙΑΚΗ ΠΡΟΣΕΓΓΙΣΗ* (5^η Αγγλική, 1^η Ελληνική εκδ., Τόμ. Α'). (Κ. Κωνσταντίνα, επιμ., Β. Βασιλική, Κ. Αδαμαντία, Λ. Λαοκρατία, Χ. Α. Ν., & Χ. Ουρανία, μτφρ.) Αλεξανδρούπολη: ΑΚΑΔΗΜΑΙΚΕΣ ΕΚΔΟΣΕΙΣ, Ι. Μπάσδρα και ΣΙΑ Ο.Ε., σσ. 3, 279.

- Hack, C. J., & Kendall, G. S. (2006). Bioinformatics: Current practice and future challenges for life science education. *Biochemistry and Molecular Biology Education*, 33(2), 82–85. <https://doi.org/10.1002/bmb.2005.494033022424> [Ανακτήθηκε από διαδίκτυο στις 03-05-2023].
- He, M. & Petoukhov, S. V. (2011). *Mathematics of Bioinformatics: Theory, Methods and Applications*. https://www.researchgate.net/publication/267003196_Mathematics_of_Bioinformatics_Theory_Practice_and_Applications [Ανακτήθηκε από διαδίκτυο στις 03-05-2023].
- Hu, T., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Human Immunology*, 82(11), 801-811. DOI: 10.1016/j.humimm.2021.02.012 <https://pubmed.ncbi.nlm.nih.gov/33745759> [Ανακτήθηκε στις 27-05-2023].
- Jungck, J. R. & Weisstein, A. E. (2013). Mathematics and evolutionary biology make bioinformatics education comprehensible. *Briefings in Bioinformatics*, 14(5), 599–609. <https://doi.org/10.1093/bib/bbt046> [Ανακτήθηκε από διαδίκτυο στις 03-05-2023].
- Klyuchko, O. M. (2017). ON THE MATHEMATICAL METHODS IN BIOLOGY AND MEDICINE. *Biotechnologia Acta*. <https://doi.org/10.15407/biotech10.03.031> [Ανακτήθηκε από διαδίκτυο στις 03-05-2023].
- Lesk, A. M. (2021). *Εισαγωγή στη Βιοπληροφορική*. (επιμ. Β. Α. Οικονομίδου). Αθήνα: Utopia.
- Luscombe, N. M., Greenbaum, D., & Gerstein, M. (2001). What is bioinformatics? A proposed definition and overview of the field. *Methods of information in medicine*, 40(4), 346-358.

- Machluf, Y., & Yarden, A. (2013). Integrating bioinformatics into senior high school: design principles and implications. *Briefings in bioinformatics*, 14(5), 648-660.
- Maier, Tobias, Marc Güell, and Luis Serrano. (2009). "Correlation of mRNA and Protein in Complex Biological Samples." *FEBS Letters* 583 (24): 3966–73. DOI: [10.1016/j.febslet.2009.10.036](https://pubmed.ncbi.nlm.nih.gov/19850042) <https://pubmed.ncbi.nlm.nih.gov/19850042> [Ανακτήθηκε στις 27-05-2023].
- Mortazavi, A., Brian A. W., McCue, K., Schaeffer, L. & Wold, B. (2008). "Mapping and Quantifying Mammalian Transcriptomes by RNA-Seq." *Nature Methods* 5 (7). Nature Publishing Group. <https://www.nature.com/articles/nmeth.1226> [Ανακτήθηκε στις 27-05-2023].
- Mohamadou, Y., Halidou, A., & Kapen, P. T. (2020). A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19. *Applied Intelligence*, 50(11), 3913-3925. DOI: [10.1007/s10489-020-01770-9](https://www.researchgate.net/publication/342711012_A_review_of_mathematical_modeling_artificial_intelligence_and_datasets_used_in_the_study_prediction_and_management_of_COVID-19) https://www.researchgate.net/publication/342711012_A_review_of_mathematical_modeling_artificial_intelligence_and_datasets_used_in_the_study_prediction_and_management_of_COVID-19 [Ανακτήθηκε στις 27-05-2023].
- Rosen, K. H. (2014). *Διακριτά Μαθηματικά και εφαρμογές*. Αθήνα: Τζιόλα.
- Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., & Waterston, R. H. (2017). DNA sequencing at 40: past, present and future. *Nature*, 550(7676), 345-353. DOI: [10.1038/nature24286](https://pubmed.ncbi.nlm.nih.gov/29019985) <https://pubmed.ncbi.nlm.nih.gov/29019985> [Ανακτήθηκε στις 27-05-2023].
- Vogel, C. & Marcotte, E. M. (2012). "Insights into the Regulation of Protein Abundance from Proteomic and Transcriptomic Analyses." *Nature Reviews. Genetics* 13 (4). Nature Publishing Group: 227–32. <https://www.nature.com/articles/nrg3185> [Ανακτήθηκε στις 27-05-2023].

Watson, J. D., Caudy, A. A., Myers, R. M., Witkowski, J. A., και Γονιδιώματα, Γ., & Παρουσίαση, Μ. Σ. (2007). *Ανασυνδυασμένο DNA. Γονίδια και γονιδιώματα-Μια συνοπτική παρουσίαση*. 3η αγγλική έκδοση-1η ελληνική έκδοση. Επιστημονικές εκδόσεις Ι. Μπάσδρα και σία ΟΕ.
<https://docplayer.gr/34476501-Anasyndyasmeno-dna-gonidia-kai-gonidiomat-a-mia-synoptiki-paroysiasi-richard-m-myers-jan-a-witkowski-james-d-watson-amy-a.html> [Ανακτήθηκε από διαδίκτυο στις 24-03-2023].

Zhong, W., Gerstein, M. & Snyder, M. (2009). "RNA-Seq: A Revolutionary Tool for Transcriptomics." *Nature Reviews. Genetics* 10 (1). Nature Publishing Group: 57–63. <https://www.nature.com/articles/nrg2484> [Ανακτήθηκε στις 27-05-2023].

Zayed, A. I. (2019). A new perspective on the role of mathematics in medicine. *Journal of Advanced Research*, 17, 49-54.
<https://doi.org/10.1016/j.jare.2019.01.016>
<https://www.sciencedirect.com/science/article/pii/S2090123219300165>
[Ανακτήθηκε στις 27-05-2023].

Διαδικτυακές

<https://www.dit.uoi.gr/lessons/?id=1> [Ανακτήθηκε από διαδίκτυο στις 23-03-2023].

<https://pithos.oceanos.grnet.gr/public/xVNSXsxtNMnK7OyY7FxxS5> [Ανακτήθηκε από διαδίκτυο στις 23-03-2023].

<https://eclass.upatras.gr/courses/CULTURE159> [Ανακτήθηκε από διαδίκτυο στις 23-03-2023].

https://eclass.uth.gr/modules/document/file.php/DIB_U_195/chapter1final.pdf
[Ανακτήθηκε από διαδίκτυο στις 23-03-2023].

<https://www.ncbi.nlm.nih.gov/Class/MLACourse/Modules/MolBioReview/bioinformatics.html> [Ανακτήθηκε από διαδίκτυο στις 23-03-2023].

http://jupiter.chem.uoa.gr/chem/talks/talks_2017/double_helix.pdf [Ανακτήθηκε από διαδίκτυο στις 24-03-2023].

<https://blast.ncbi.nlm.nih.gov/Blast.cgi> [Ανακτήθηκε από διαδίκτυο στις 24-03-2023].

<https://helios-eie.ekt.gr/EIE/bitstream/10442/474/1/M01.027.01.pdf> [Ανακτήθηκε από διαδίκτυο στις 24-03-2023].

<https://www.youtube.com/watch?v=BrEs7zFVrMI> [Ανακτήθηκε από διαδίκτυο στις 29-04-2023].

<https://eclass.uowm.gr/modules/document/file.php/ICTE102/%CE%94%CE%99%CE%91%CE%9B%CE%95%CE%9E%CE%95%CE%99%CE%A3/%CE%94%CE%B9%CE%AC%CE%BB%CE%B5%CE%BE%CE%B7%2003.pdf> [Ανακτήθηκε από διαδίκτυο στις 29-04-2023].

<http://www.ebi.ac.uk/Tools/sss/fasta/nucleotide.html> [Ανακτήθηκε από διαδίκτυο στις 29-04-2023].

<https://blast.ncbi.nlm.nih.gov/Blast.cgi> [Ανακτήθηκε από διαδίκτυο στις 29-04-2023].

<https://math.uoi.gr/index.php/2016-03-05-09-59-57/2016-03-05-10-00-48/2016-03-05-10-34-16> [Ανακτήθηκε από διαδίκτυο στις 30-04-2023].

<https://docplayer.gr/47186459-Eisagogi-sti-viopliroriki.html> [Ανακτήθηκε από διαδίκτυο στις 30-04-2023].

<https://startupper.gr/interviews/78845/enios-mia-exypni-omic-platforma> [Ανακτήθηκε από διαδίκτυο στις 30-04-2023].

<https://el.lambdageeks.com/difference-between-genomics-and-proteomics>

[Ανακτήθηκε από διαδίκτυο στις 30-04-2023].

<https://opencourses.auth.gr/modules/document/file.php/OCRS146/%CE%A0%CE%B1%CF%81%CE%BF%CF%85%CF%83%CE%B9%CE%AC%CF%83%CE%B5%CE%B9%CF%82/%CE%95%CE%BD%CF%8C%CF%84%CE%B7%CF%84%CE%B1%2004%3A%20%CE%A4%CE%BF%20%CE%B3%CE%BF%CE%BD%CE%B9%CE%B4%CE%AF%CF%89%CE%BC%CE%B1%20%CF%84%CE%BF%CF%85%20%CE%B1%CE%BD%CE%B8%CF%81%CF%8E%CF%80%CE%BF%CF%85.pdf>

[Ανακτήθηκε στις 27-05-2023].