



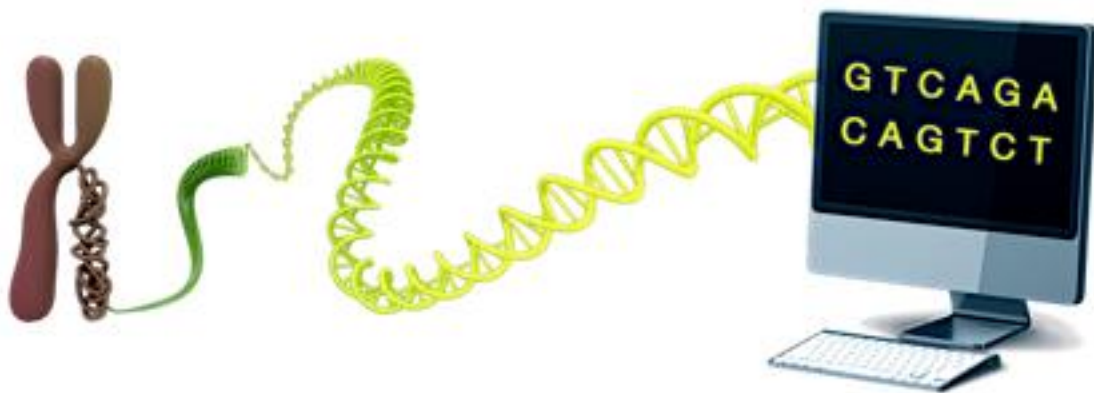
ΑΛΕΞΑΝΔΡΕΙΟ Τ.Ε.Ι. ΘΕΣΣΑΛΟΝΙΚΗΣ
ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



Τμήμα Μηχανικών
Πληροφορικής ΑΤΕΙΘ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Η Βιοπληροφορική και η χρήση της στην αλληλούχηση του
γενετικού υλικού του ανθρώπου



Του φοιτητή

Κοτρώτσιου Σωτήριου

Αρ. Μητρώου: 10/3641

Επιβλέπων καθηγητής

Αντωνίου Ευστάθιος

Θεσσαλονίκη 2017

Πρόλογος

Είναι αλήθεια ότι κατά τη συγγραφή μιας εργασίας ο πρόλογος αποτελεί το μοναδικό κομμάτι στο οποίο ο συγγραφέας έχει την ευχέρεια της χρήσης του πρώτου ενικού προσώπου. Δράττομαι λοιπόν της ευκαιρίας να απευθυνθώ στους αναγνώστες αυτής της πτυχιακής με τον ίδιο ακριβώς τρόπο που θα απευθυνόμουν σε αυτούς σε μια ζωντανή αναμεταξύ μας συζήτηση.

Για την αποπεράτωση των σπουδών μου στο Α.Τ.Ε.Ι. της Θεσσαλονίκης, ήταν απαραίτητη η ανάληψη ενός θέματος πτυχιακής εργασίας. Επιθυμούσα να αναλάβω ένα θέμα ξεχωριστό. Ένα θέμα το οποίο δεν ακουγόταν στις συζητήσεις ανάμεσα στους φοιτητές του τμήματός μου. Πιστεύω ακράδαντα ότι η εποχή στην οποία ζούμε αποτελεί την εποχή της ιδέας. Έτσι, μετά από ώριμη σκέψη και έρευνα, αποφάσισα να επιχειρήσω την εκπόνηση μιας πτυχιακής εργασίας το θέμα της οποίας θα άγγιζε ένα σημαντικό κομμάτι ενός τεράστιου πραγματικά κλάδου, αυτού της Βιοπληροφορικής. Όπως θα δούμε και κατά τη διάρκεια της ανάγνωσης της εργασίας, η Βιοπληροφορική είναι ένας αυτόνομος διεπιστημονικός κλάδος που έχει δημιουργηθεί με τη σύμπραξη πολλών υπαρχουσών επιστημών. Το επίπεδο ανάπτυξης της είναι αρκετά υψηλό και αυτό φαίνεται από την ίδρυση διαφόρων επιστημονικών περιοδικών, τη διεξαγωγή συνεδρίων καθώς και τη λειτουργία εξειδικευμένων προγραμμάτων σπουδών σε διάφορα πανεπιστήμια ανά τον κόσμο. Εμείς θα συζητήσουμε ένα από τα πολλά θέματα που θίγει η Βιοπληροφορική. Η συγκεκριμένη εργασία πραγματεύεται τη χρήση της Βιοπληροφορικής στην αλληλούχηση του γενετικού υλικού του ανθρώπου (DNA Sequencing). Με την ολοκλήρωση του προγράμματος αλληλούχησης ολόκληρου του ανθρωπίνου γονιδιώματος «ξεκλειδώθηκαν» γνώσεις οι οποίες έχουν συνεισφέρει τα μάλα στην κατανόηση της λειτουργίας του οργανισμού του ανθρώπου.

Πιστεύω ότι η συγκεκριμένη πτυχιακή θα βοηθήσει τον αναγνώστη να αποκομίσει χρήσιμες πληροφορίες τόσο για το γενετικό υλικό (του ανθρώπου) όσο και για τις δυνατότητες, υπολογιστικές και μη, που προκύπτουν απ' αυτό.

Περίληψη

Η αλληλούχιση του ανθρωπίνου γενετικού υλικού αποτελεί ένα από τα σημαντικότερα επιτεύγματα της σύγχρονης Μοριακής Βιολογίας. Μέσα από τη διαδικασία αυτή ο άνθρωπος έχει την ευκαιρία να ανακαλύψει τη γονιδιακή του ταυτότητα αφού το γενετικό υλικό είναι υπεύθυνο για τη μεταφορά της γενετικής πληροφορίας. Ο ρόλος της Βιοπληροφορικής στην αλληλούχιση του γενετικού υλικού, αλλά και γενικότερα στην έρευνα των βιολόγων, είναι απαραίτητος πλέον αφού χωρίς τις δυνατότητες που παρέχει (βιολογικές βάσεις δεδομένων, αλγόριθμοι αναζήτησης ή ευθυγράμμισης, μέθοδοι εντοπισμού ομοιοτήτων κ.ά.) θα ήταν αδύνατη η διαχείριση ενός τόσο μεγάλου όγκου βιολογικών πληροφοριών.

Abstract

Human DNA sequencing is one of the major achievements of modern Molecular Biology. Through this process, man has the opportunity to discover his genetic identity, since the DNA is responsible for transferring the genetic information. The role of Bioinformatics in the human DNA sequencing - and in biologists' research in general - is really necessary because without its potential (biological databases, search and alignment algorithms, similarity detection methods, etc.) it would be impossible for scientists to manage the huge amount of biological data.

Πίνακας περιεχομένων

Πρόλογος.....	1
Περίληψη.....	2
Abstract.....	3
Πίνακας περιεχομένων.....	4
Εισαγωγή.....	6
Κεφάλαιο 1^ο: Η επιστήμη της Βιοπληροφορικής.....	8
1.1 Εισαγωγή στη Βιοπληροφορική.....	8
1.2 Ιστορική Αναδρομή.....	10
Κεφάλαιο 2^ο: Η επιστήμη της Βιολογίας.....	20
2.1 Εισαγωγή στο γονιδίωμα.....	20
2.2 Η δομή του DNA.....	20
2.3 Οργάνωση του γενετικού υλικού των ευκαρυωτικών οργανισμών.....	23
2.4 Μελετώντας τα ανθρώπινα χρωμοσώματα – Καρυότυπος.....	24
2.5 Η αντιγραφή του DNA.....	25
2.6 Η έκφραση της γενετικής πληροφορίας.....	26
2.6.1 Η μεταγραφή του DNA.....	28
2.6.2 Ο γενετικός κώδικας.....	29
2.6.3 Μετάφραση.....	31
2.7 Γονιδιακή ρύθμιση.....	34
2.7.1 Η γονιδιακή ρύθμιση στους ευκαρυωτικούς οργανισμούς.....	35
Κεφάλαιο 3^ο: Η αλληλούχιση του γενετικού υλικού του ανθρώπου (DNA Sequencing).....	37
3.1 Εισαγωγή στην αλληλούχιση του γενετικού υλικού (DNA Sequencing).....	37
3.2 Η διαδικασία της αλληλούχισης του γονιδιώματος μέσα από την αλληλούχιση του γενετικού υλικού.....	38
3.3 Η συμβολή της Πληροφορικής στην επιστήμη της Βιοπληροφορικής.....	44
3.3.1 Αποθήκευση των αλληλουχιών σε ηλεκτρονικό υπολογιστή.....	45
3.3.2 Στοίχιση αλληλουχιών.....	46
3.3.3 Στοίχιση ζεύγους αλληλουχιών (Pair – Wise Alignment).....	46
3.3.3.1 Dot matrix analysis.....	47
3.3.3.2 Dynamic programming algorithm.....	47

3.3.4 Πολλαπλή στοίχιση αλληλουχιών (<i>Multiple Sequence Alignment</i>)	48
3.3.4.1 Προοδευτική πολλαπλή στοίχιση αλληλουχιών (<i>Progressive Multiple Sequence Alignment</i>).....	49
3.3.4.2 Επαναληπτική πολλαπλή στοίχιση αλληλουχιών (<i>Iterative Multiple Sequence Alignment</i>).....	50
3.3.5 Ολική στοίχιση αλληλουχιών – Αλγόριθμος <i>Needleman and Wunsch</i>	50
3.3.6 Τοπική στοίχιση αλληλουχιών – Αλγόριθμος <i>Smith and Waterman</i>	52
3.3.7 Η χρησιμότητα της στοίχισης των αλληλουχιών.....	53
3.4 Εφαρμογές του <i>DNA Sequencing</i>	54
3.4.1 Οι γονιδιακές μεταλλάξεις	55
3.4.1.1 Μεταβολικά νοσήματα από γονιδιακές μεταλλάξεις	56
3.4.2 Οι χρωμοσωμικές μεταλλάξεις	57
3.4.2.1 Οι αριθμητικές χρωμοσωμικές ανωμαλίες.....	57
3.4.2.2 Οι δομικές χρωμοσωμικές ανωμαλίες	60
3.4.3 Ο προγεννητικός έλεγχος	61
Κεφάλαιο 4^ο: Βάσεις δεδομένων νουκλεοτιδικών αλληλουχιών	63
4.1 Νουκλεοτιδικές βάσεις δεδομένων και αλγόριθμοι αναζήτησης	63
4.2 Ο αλγόριθμος <i>FASTA</i>	64
4.3 Ο αλγόριθμος <i>BLAST</i>	66
Κεφάλαιο 5^ο: Προβληματισμοί σχετικά με τις δυνατότητες της Βιοπληροφορικής	69
Βιβλιογραφία.....	71

Εισαγωγή

Σκοπός της πτυχιακής εργασίας είναι η γνωριμία με ένα νέο, ανερχόμενο, διεπιστημονικό τομέα, αυτόν της Βιοπληροφορικής. Μελετάμε σημαντικά στοιχεία του Βιολογικού κλάδου εξειδικεύοντας στη διαδικασία της αλληλούχισης του γενετικού υλικού του ανθρώπου (Human DNA Sequencing). Η πληροφορική αποτελεί αναπόσπαστο κομμάτι της Βιοπληροφορικής αφού αναλαμβάνει να αναζητήσει, να συλλέξει και να αναλύσει στοιχεία τα οποία υπάρχουν σε βιολογικές βάσεις δεδομένων. Εξειδικευμένοι αλγόριθμοι αναλαμβάνουν να ευθυγραμμίσουν αλληλουχίες (ή αλλιώς ακολουθίες) γενετικού υλικού εξάγοντας συμπεράσματα σχετικά με τη δομή και τη λειτουργία του γενετικού υλικού και κατ' επέκταση ενός ολόκληρου γονιδίου.

Στο 1^ο κεφάλαιο θα μελετήσουμε στοιχεία σχετικά με την επιστήμη της Βιοπληροφορικής (ορισμούς, προέλευση κ.ά.). Μέσα από μια ιστορική αναδρομή θα δούμε ανακαλύψεις οι οποίες οδήγησαν μοιραία στη δημιουργία της επιστήμης αυτής.

Στο 2^ο κεφάλαιο θα μελετήσουμε στοιχεία σχετικά με την επιστήμη της Βιολογίας. Θα ασχοληθούμε με την ανακάλυψη, τη δομή και την οργάνωση του γενετικού υλικού φτάνοντας στα χρωμοσώματα (καρυότυπος) του ατόμου. Η αντιγραφή του DNA καθώς και η έκφραση της γενετικής πληροφορίας (μέσω της μεταγραφής, του γενετικού κώδικα και της μετάφρασης) είναι πολύ σημαντικές για να κατανοήσουμε τη λειτουργία του γενετικού υλικού στον ανθρώπινο οργανισμό. Τέλος, θα αναφερθούμε στη σημαντικότητα της γονιδιακής ρύθμισης.

Το 3^ο κεφάλαιο πραγματεύεται τη διαδικασία της αλληλούχισης του γενετικού υλικού του ανθρώπου. Μέσα από την αλληλούχιση ολόκληρου του γονιδιώματος θα οδηγηθούμε στην αλληλούχιση του γενετικού υλικού. Με τη δυνατότητα της αποθήκευσης των αλληλουχιών σε ηλεκτρονικό υπολογιστή μέσω της πληροφορικής, μπορούμε να μελετήσουμε τις διάφορες μεθόδους που χρησιμοποιούνται για την ευθυγράμμιση δύο ή περισσότερων αλληλουχιών. Θα εξετάσουμε αλγορίθμους οι οποίοι χρησιμοποιούνται για ολική ή τοπική στοίχιση. Υπάρχει επίσης αναφορά στις

εφαρμογές του DNA sequencing μέσα από τα διάφορα παραδείγματα που σχετίζονται με γονιδιακές ή χρωμοσωμικές μεταλλάξεις. Το κεφάλαιο κλείνει με τη μελέτη της δυνατότητας του προγεννητικού ελέγχου.

Στο 4^ο κεφάλαιο θα αναφερθούμε στις νουκλεοτιδικές βάσεις δεδομένων. Παραθέτουμε τις κύριες νουκλεοτιδικές βάσεις δεδομένων ανά τον κόσμο ενώ παράλληλα μελετάμε τους κύριους αλγορίθμους αναζήτησης και ευθυγράμμισης αλληλουχιών στις βάσεις αυτές.

Στο 5^ο κεφάλαιο θα υπάρξουν κάποιοι προβληματισμοί σχετικά με τις δυνατότητες που παρέχει η επιστήμη της Βιοπληροφορικής.

Κεφάλαιο 1^ο: Η επιστήμη της Βιοπληροφορικής

1.1 Εισαγωγή στη Βιοπληροφορική

Η ραγδαία ανάπτυξη που παρατηρείται στον επιστημονικό χώρο δημιούργησε την ανάγκη «συνεργασίας» επιστημών που διέπονται από διαφορετικά αντικείμενα. Στην παρούσα κατάσταση είναι δύσκολο να μιλήσει κανείς για αυτοτελή επιστήμη. Παράδειγμα της συνεργασίας, ή αλλιώς της διεπιστημονικότητας, που αναφέρθηκε αποτελεί και η επιστήμη της βιοπληροφορικής. Ο κλάδος αυτός είναι ένας από τους πλέον αναπτυσσόμενους κλάδους σε παγκόσμιο επίπεδο. Η βιοπληροφορική, με μια απλή ματιά στα συνθετικά της λέξης, αποτελεί τη συνένωση των επιστημών της βιολογίας και της πληροφορικής. Πιο συγκεκριμένα, αποτελεί έναν επιστημονικό κλάδο ο οποίος έλκει την καταγωγή του από τη μοριακή βιολογία και ιδιαίτερα από τη μελέτη των βιολογικών αλληλουχιών και δομών (όπως το DNA Sequencing). Όμως, η διεπιστημονικότητα της βιοπληροφορικής δεν τελειώνει εκεί. Εάν κανείς επεκτείνει τη μελέτη του στο συγκεκριμένο κλάδο, θα καταλάβει ότι ο διεπιστημονικός αυτός κλάδος είναι αποτέλεσμα περισσότερων από δύο επιστημονικών πεδίων. Η βιοπληροφορική συνδυάζει την επιστήμη των υπολογιστών (αλγόριθμοι, μηχανική μάθηση κ.ά.), τη στατιστική, τα μαθηματικά και τη μηχανική (engineering) με στόχο την ανάλυση και την ερμηνεία βιολογικών δεδομένων.

Λόγω της διεπιστημονικότητας που διέπει τη βιοπληροφορική, υπάρχουν πολλές και αντικρουόμενες απόψεις σχετικά με τον ορισμό αλλά και το επιστημολογικό της καθεστώς. Μία άποψη λέει ότι η βιοπληροφορική αποτελεί απλώς την εφαρμογή κάποιων τεχνολογιών σε βιολογικά προβλήματα. Υπάρχει και μια άλλη προσέγγιση η οποία υποστηρίζει ότι η βιοπληροφορική είναι μια ξεχωριστή επιστήμη, η οποία ναι μεν εφαρμόζει υπολογιστικά και μαθηματικά εργαλεία σε προβλήματα βιολογικής φύσεως, αλλά κάνει και κάτι παραπάνω: προσπαθεί να παράγει γενικότερους νόμους που διέπουν αυτά τα βιολογικά συστήματα. Πολλές φορές, η βιοπληροφορική καλείται και υπολογιστική βιολογία (computational biology) διότι για

πολλούς αυτές οι δύο έννοιες είναι αδιάκριτες. Παρ' όλα αυτά, ένας λογικός διαχωρισμός που προκύπτει είναι ότι ο όρος βιοπληροφορική αναφέρεται στη χρήση αλγορίθμων και υπολογιστικών τεχνικών με σκοπό την απάντηση βιολογικών ερωτημάτων (π.χ. αναζήτηση μιας ακολουθίας σε μια βάση δεδομένων, χειρισμός μεγάλου όγκου ακολουθιών κλπ.), ενώ ο όρος υπολογιστική βιολογία αναφέρεται στα θεωρητικά αποτελέσματα στα οποία στηρίζεται κανείς για να αναπτύξει έναν αλγόριθμο ή έναν γενικό νόμο. Ας δούμε όμως μερικούς ορισμούς:

Το NCBI δίνει τον παρακάτω ορισμό:

(<http://www.ncbi.nlm.nih.gov/Class/MLACourse/Modules/MolBioReview/bioinformatics.html>)

«Bioinformatics is the field of science in which biology, computer science, and information technology merge into a single discipline. The ultimate goal of the field is to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology can be discerned. There are three important sub-disciplines within bioinformatics: the development of new algorithms and statistics with which to access relationships among members or large data sets; the analysis and interpretation of various types of data including nucleotide and amino acid sequences, protein domains, and protein structure; the development and implementations of tools that enable efficient access and management of different types of information»

Υπάρχει επίσης και ο αντίστοιχος ορισμός του Luscombe (Luscombe, Greenbaum, & Gerstein, 2001):

«Bioinformatics is conceptualizing biology in terms of macromolecules (in the sense of physical-chemistry) and then applying “informatics” techniques (derived from disciplines such as applied maths, computer science, and statistics) to understand and organize the information associated with these molecules, on a large scale»

Όπως και αυτός του Fredj Tekaia:

«The mathematical, statistical and computing methods that aim to solve biological problems using DNA and amino acid sequences and related information»

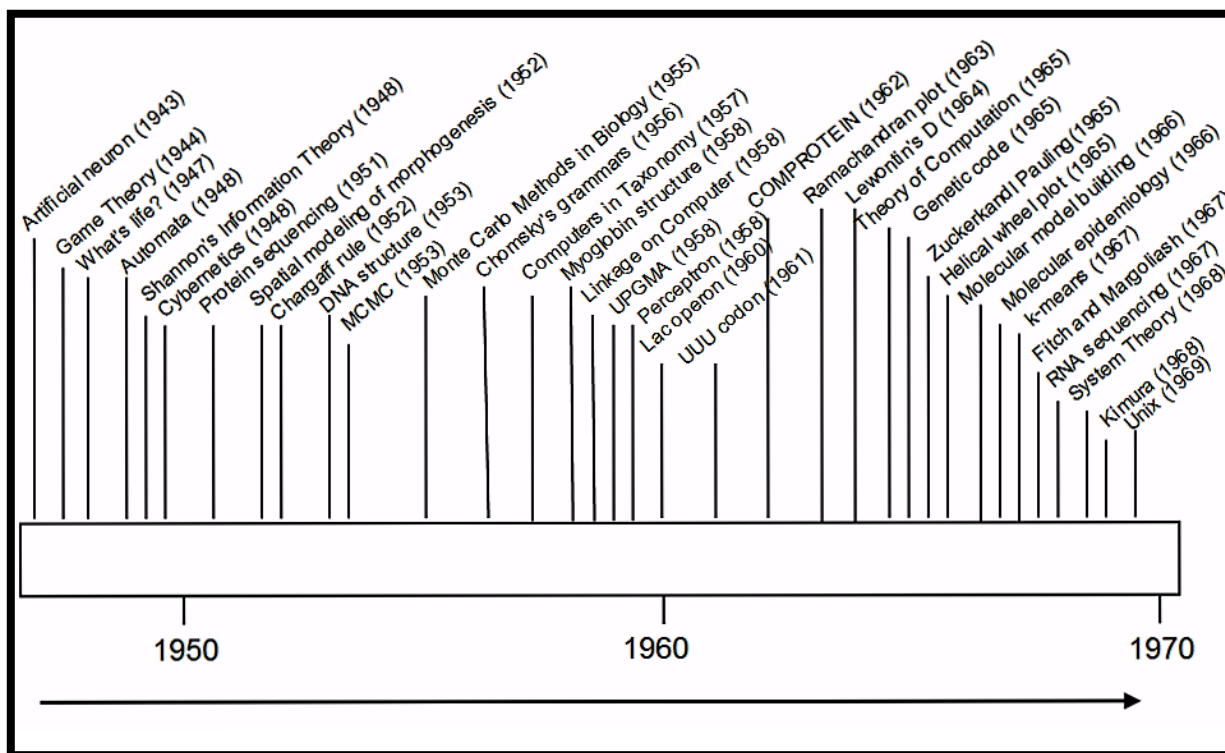
Για να μπορέσουμε να αντιληφθούμε το μέγεθος της επιστήμης της βιοπληροφορικής θα πρέπει να την μελετήσουμε από τα αρχικά της στάδια. Η ιστορική αναδρομή που ακολουθεί θα μας βοηθήσει να καταλάβουμε ακόμα καλύτερα τόσο το μέγεθος όσο και τη διεπιστημονικότητα του τομέα της βιοπληροφορικής.

1.2 Ιστορική Αναδρομή

Για να ανιχνεύσουμε τις απαρχές της σύγχρονης υπολογιστικής βιολογίας θα πρέπει να ανατρέξουμε στις απαρχές της ίδιας της σύγχρονης βιολογίας του 1950 και 1960. Οι πρώτες ενδείξεις για κάποια μορφή ψηφιακής πληροφορίας στις βιολογικές ακολουθίες πηγάζουν από τα πειράματα που εκτέλεσε ο Chargaff τα οποία έδειξαν ότι το ποσοστό αδενίνης είναι το ίδιο με το ποσοστό της θυμίνης και το ποσοστό γουανίνης είναι το ίδιο με το ποσοστό της κυτοσίνης σε κάθε μόριο DNA. Αργότερα, τα πειράματα αυτά χρησιμοποιήθηκαν από τους Watson και Crick για να προσδιορίσουν την τρισδιάστατη δομή του γενετικού υλικού (DNA) στο χώρο η οποία τους έδωσε και το Νόμπελ. Για αυτό τους το επίτευγμα, οι Watson και Crick χρησιμοποίησαν δεδομένα του Wilkins, ο οποίος βραβεύτηκε μαζί τους, καθώς και της Franklin, η οποία όμως είχε πεθάνει στο ενδιάμεσο. Ακολούθησαν οι πρωτοποριακές για την εποχή τους μελέτες των Jacob και Monod σχετικά με τη γονιδιακή ρύθμιση (το οπερόνιο της λακτόζης) οι οποίες έλαβαν χώρα τη δεκαετία του 1960. Όσον αφορά τις πρωτεΐνες, προσδιορίστηκαν οι πρώτες τρισδιάστατες δομές (ινσουλίνη και μυογλοβίνη) από τους Perutz και Kendrew οι οποίοι βραβεύτηκαν επίσης με Νόμπελ το 1962. Ακολουθεί μια σειρά αναπαράστασης παρόμοιων σημαντικών δομών (λυσοζύμη, παπαΐνη, ριβονουκλεάση κ.ο.κ.) οι οποίες ανοίγουν το δρόμο για τη μελέτη της λειτουργίας και της δομής των πρωτεϊνών σε ατομικό επίπεδο. Η εύρεση της πρωτοταγούς δομής των πρωτεϊνών έγινε το 1951 ενώ 16 χρόνια μετά βρέθηκε η πρωτοταγής δομή του RNA. Γίνεται λοιπόν σαφές ότι πολλά από τα προβλήματα με

τα οποία ασχολείται η βιοπληροφορική προέρχονται από την έκρηξη ανακαλύψεων που πραγματοποιήθηκαν τη δεκαετία του 1960 στον τομέα της μοριακής βιολογίας.

Την ίδια χρονική περίοδο, δηλαδή τις δεκαετίες 1950 και 1960, είχαμε την εμφάνιση πολλών θεμελιωδών στοιχείων της σύγχρονης πληροφορικής όπως είναι η θεωρία της πληροφορίας του Shannon, η μηχανή του Turing, η θεωρία των παιγνίων του von Neumann, η μελέτη των συμβολοσειρών, η θεωρία των συστημάτων κ.ά. Παράλληλα, αποκρυπτογραφήθηκε ο γενετικός κώδικας ο οποίος απετέλεσε ένα πολύ σημαντικό βήμα στην εξέλιξη όλων των βιοεπιστημών αφού έγινε αντικείμενο έντονης θεωρητικής και υπολογιστικής επεξεργασίας. Την ίδια εποχή έλαβε χώρα και η πρώτη προσπάθεια χρήσης βιολογικών αλληλουχιών για εξελικτικές μελέτες από τους Zuckerkandl και Pauling. Οι ίδιες αλληλουχίες χρησιμοποιήθηκαν και για την κατασκευή φυλογενετικών δένδρων από τους Fitch και Margoliash.

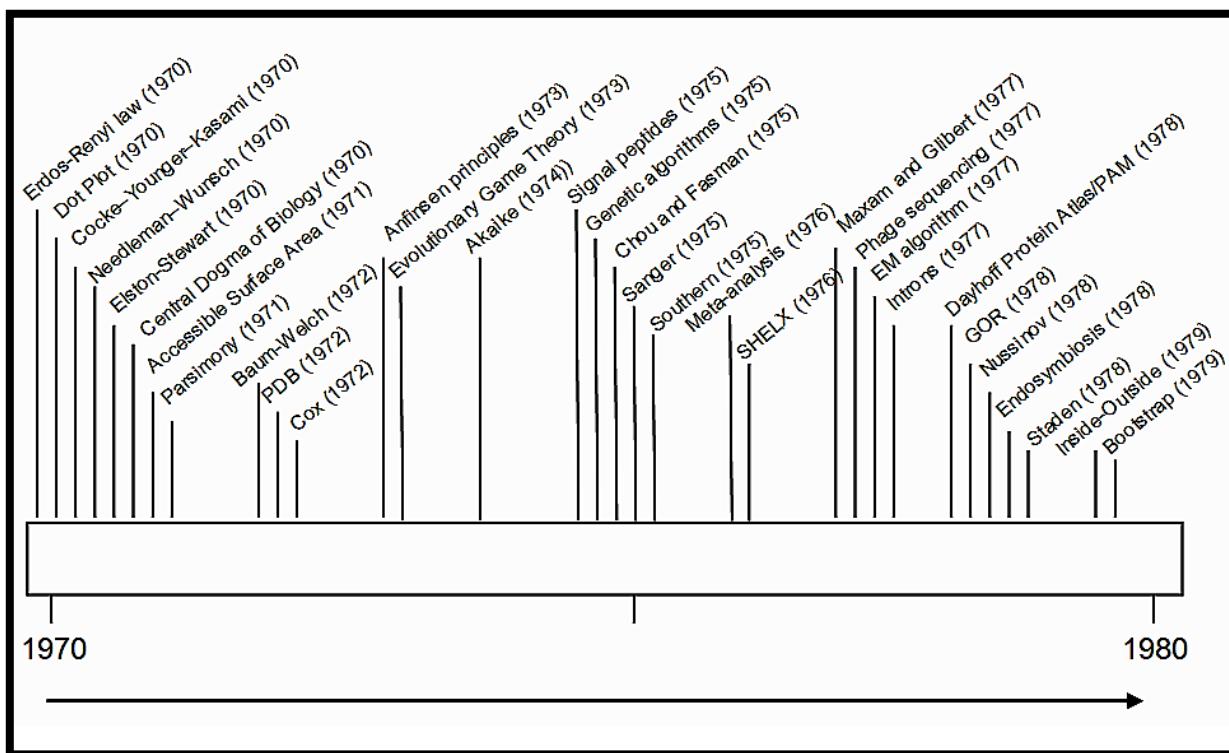


Εικόνα 1.1: Ανακαλύψεις σχετικές με τη βιοπληροφορική μέχρι και το τέλος της δεκαετίας του 1960.

Με όλα τα παραπάνω αντιλαμβάνεται κανείς το λόγο που οι πρώτες προσπάθειες υπολογιστικής αντιμετώπισης προβλημάτων βιολογικής φύσεως

εμφανίστηκαν τη δεκαετία του 1960 οδηγώντας σήμερα στους όρους υπολογιστική βιολογία και βιοπληροφορική.

Τη δεκαετία που ακολούθησε, η έρευνα συνεχίστηκε με αυξανόμενο ρυθμό. Την περίοδο αυτή έκανε την εμφάνισή του ο αλγόριθμος του Fitch για τη φειδωλή ανακατασκευή φυλογενετικών δένδρων με τη χρήση βιολογικών αλληλουχιών (μέθοδος της μέγιστης φειδωλότητας). Επίσης, με την αποκρυπτογράφηση του γενετικού κώδικα είχε πλέον διαλευκανθεί ο ρόλος του RNA στη μεταγραφή και τη μετάφραση, γεγονός που οδήγησε στη διατύπωση του κεντρικού δόγματος της βιολογίας από τον Crick το 1970. Την ίδια χρονική περίοδο εμφανίστηκαν οι πρώτες μεθοδολογίες αλληλούχισης νουκλεϊκών οξέων από τους Sanger και Maxam-Gilbert οι οποίες έδωσαν ώθηση στη μελέτη των γονιδιωμάτων ανοίγοντας το δρόμο για τις σύγχρονες μεθόδους αλληλούχισης.



Εικόνα 1.2: Ανακαλύψεις σχετικές με τη βιοπληροφορική μέχρι και το τέλος της δεκαετίας του 1970.

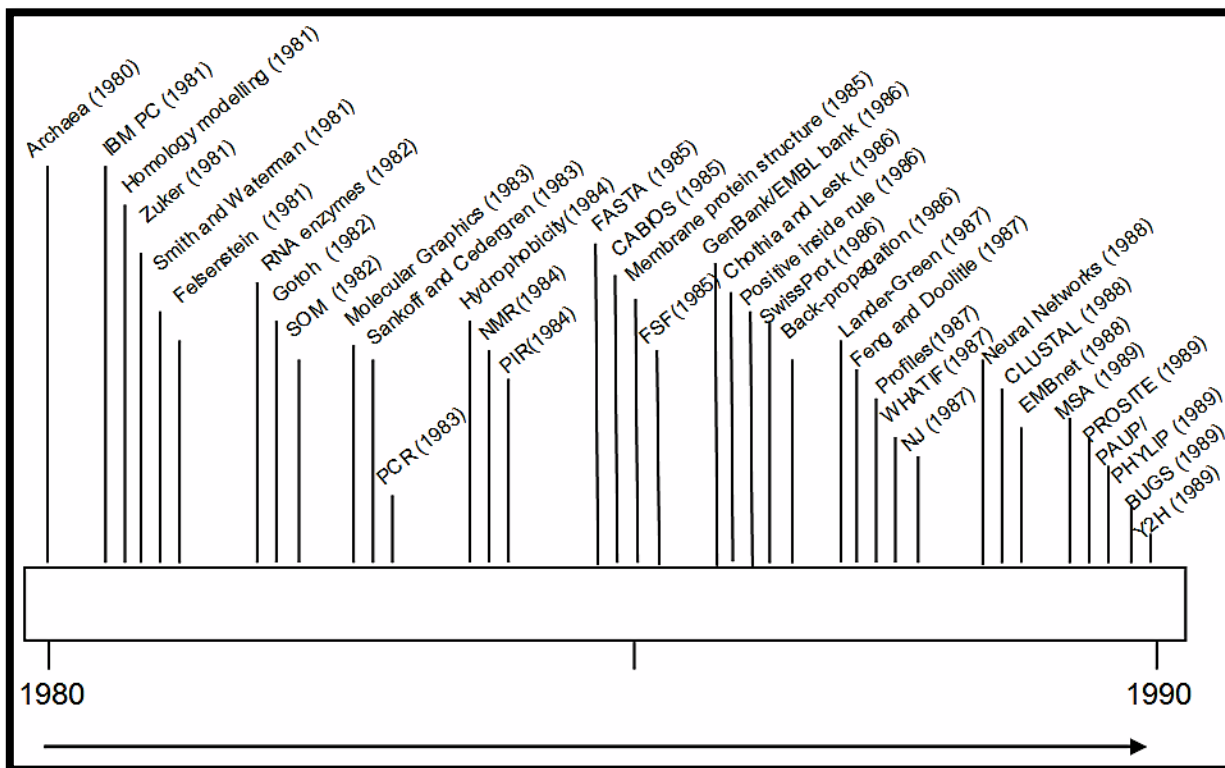
Μια από τις πιο σημαντικές αλγοριθμικές συνεισφορές στη βιοπληροφορική που έλαβαν χώρα τη δεκαετία του 1970, ήταν η εμφάνιση των αλγορίθμων δυναμικού

προγραμματισμού για τη στοίχιση βιολογικών αλληλουχιών (κυρίως πρωτεϊνών). Ο πρώτος αλγόριθμος παρουσιάστηκε το 1970 από τους Needleman και Wunsch και ήταν αλγόριθμος για ολική στοίχιση βιολογικών αλληλουχιών. Το 1970 έκανε επίσης την εμφάνισή του το διάγραμμα σημείων (dot-plot ή dot matrix) ενώ ακολούθησαν και άλλες προσεγγίσεις στη μεθοδολογία και τα στατιστικά της στοίχισης. Προς το τέλος της δεκαετίας εμφανίστηκαν και οι πρώτες βάσεις βιολογικών δεδομένων. Η RHB εμφανίστηκε το 1972 ενώ η Dayhoff παρουσίασε το 1978 την πρώτη συλλογή πρωτεϊνικών αλληλουχιών οι οποίες ήταν γνωστές εκείνα τα χρόνια. Τέλος, έκαναν την εμφάνισή τους και τα πρώτα προγράμματα H/Y για απλές αναλύσεις σε βιολογικές αλληλουχίες (μετάφραση μιας κωδικής αλληλουχίας, εύρεση προτύπων κ.ά.).

Η δεκαετία του 1980 ήταν η δεκαετία στην οποία το πεδίο της βιοπληροφορικής πήρε πλέον μια ξεκάθαρη μορφή, σαν ένας ξεχωριστός κλάδος θέτοντας τα δικά του προβλήματα και παρουσιάζοντας τα δικά του επιτεύγματα. Ξεκινούν οι σχετικές με το αντικείμενο δημοσιεύσεις στα υψηλού κύρους βιολογικά περιοδικά (Nature, Nucleic Acid Research, Science) ενώ παράλληλα εμφανίζονται τα πρώτα εξειδικευμένα περιοδικά (Computer Applications in Biosciences). Φυσικά, δεν πρέπει να ξεχνάμε ότι την εποχή αυτή είχε αρχίσει να γίνεται διαδεδομένη η χρήση υπολογιστικών συστημάτων (με παράλληλη ανάπτυξη hardware και software) με αποτέλεσμα πολλές ανακαλύψεις να επωφεληθούν από την πρόοδο αυτή.

Στο πεδίο της ανάλυσης αλληλουχιών μακρομορίων, η μελέτη των αλγορίθμων στοίχισης καθώς και των αποτελεσματικών υλοποιήσεών τους συνεχίστηκε με εντατικό ρυθμό. Έτσι, το 1981 έχουμε την ανακάλυψη του αλγορίθμου τοπικής στοίχισης από τους Smith και Waterman. Την ίδια περίοδο παρουσιάζονται οι πρώτες αποτελεσματικές υλοποιήσεις για γρήγορη στοίχιση και αναζήτηση ομοιότητας σε μια βάση δεδομένων (FASTA). Παράλληλα έγιναν οι πρώτες επεξεργασίες της πολλαπλής στοίχισης σε θεωρητικό επίπεδο, επινοήθηκε η ιεραρχική πολλαπλή στοίχιση και παρουσιάστηκε το CLUSTAL. Εκείνη την περίοδο επινοήθηκαν και τα προφίλ αλληλουχιών (sequence profiles) τα οποία χρησιμοποιήθηκαν για τη μελέτη πρωτεϊνικών οικογενειών. Τα προφίλ αυτά εφαρμόστηκαν σε πληθώρα

παραδειγμάτων με εντυπωσιακά αποτελέσματα και χρησιμοποιούνται μέχρι και σήμερα. Ακόμη, εμφανίζονται τα πρώτα βιβλία σχετικά με την υπολογιστική ανάλυση αλληλουχιών.

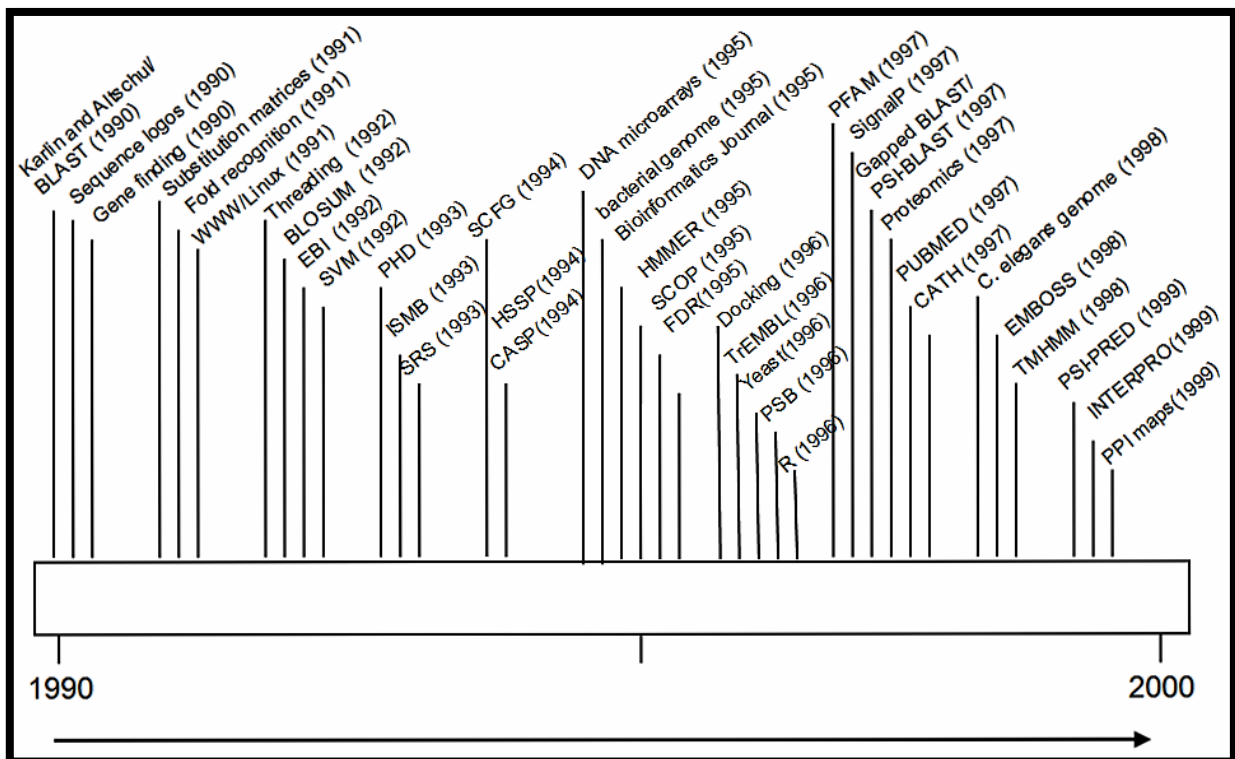


Εικόνα 1.3: Ανακαλύψεις σχετικές με τη βιοπληροφορική μέχρι και το τέλος της δεκαετίας του 1980.

Την ίδια εποχή, η πρόοδος στις μεθόδους αλληλούχισης του DNA, η εμφάνιση της PCR, καθώς και η συνεχής βελτίωση στις τεχνικές προσδιορισμού της τρισδιάστατης δομής των μακρομορίων, είχαν σαν αποτέλεσμα τη ραγδαία αύξηση του όγκου των δεδομένων προς επεξεργασία. Έτσι, ήταν αναγκαία η δημιουργία μεγαλύτερων και πιο οργανωμένων βιολογικών βάσεων δεδομένων. Το 1986 εμφανίστηκαν στο προσκήνιο οι δύο πιο γνωστές μέχρι σήμερα βάσεις δεδομένων νουκλεοτιδικών αλληλουχιών (GenBank και EMBL Data Library), ενώ ένα χρόνο μετά εμφανίστηκε η βάση δεδομένων των πρωτεϊνικών αλληλουχιών SwissProt. Παράλληλα έγιναν προσπάθειες για δημιουργία δικτύων με σκοπό την διευκόλυνση της υπολογιστικής έρευνας στην βιολογία (EMBnet και BIONET), ενώ παρουσιάστηκαν και οι πρώτοι κατάλογοι με σχετικό λογισμικό (LiMB).

Την ίδια χρονική περίοδο προτάθηκε ο αλγόριθμος του Fenselstein για την εκτίμηση φυλογενετικών δένδρων μέσω της μέγιστης πιθανοφάνειας, ανακάλυψη που έδωσε μεγάλη ώθηση στο συγκεκριμένο πεδίο. Παράλληλα, παρατηρείται μεγάλη πρόοδος στην εξελικτική μελέτη των γονιδιωμάτων καθώς πραγματοποιείται η μελέτη των φυλογενετικών δεικτών, όπως το rRNA, αλλά και η εξελικτική ιστορία των εσωνίων, των εξωνίων και της συρραφής.

Στη δεκαετία που ακολούθησε (1990), η έρευνα στον τομέα της υπολογιστικής βιολογίας πραγματικά εκτινάχθηκε. Σημάδι αυτής της δεκαετίας αποτέλεσε επίσης η εξάπλωση του παγκοσμίου ιστού και του διαδικτύου καθώς και η εξάπλωση των προσωπικών Η/Υ. Παράλληλα, ο όρος βιοπληροφορική αρχίζει να χρησιμοποιείται ευρέως από την επιστημονική κοινότητα, ενώ το πιο γνωστό περιοδικό στο χώρο, το «Computer Applications in the Biosciences» (CABIOS), αλλάζει όνομα το 1995 και αναγνωρίζεται πλέον ως «Bioinformatics».



Εικόνα 1.4: Ανακαλύψεις σχετικές με τη βιοπληροφορική μέχρι και το τέλος της δεκαετίας του 1990.

Το BLAST (Basic Local Alignment Search Tool) απετέλεσε την πιο σημαντική ανακάλυψη στον τομέα της στοίχισης αλληλουχιών. Η δημοσίευσή του πραγματοποιήθηκε το 1990 από επιστήμονες του NCBI. Το BLAST «πάτησε» πάνω στις ανακαλύψεις σχετικά με τη στατιστική κατανομή του σκορ (score) της τοπικής στοίχισης (το θεώρημα Karlin-Altschul) συμβάλλοντας καθοριστικά στον τρόπο που θα διεξαγόταν πλέον τόσο η αναζήτηση ομοιότητας σε βάσεις δεδομένων όσο και η στοίχιση, καθώς ήταν ταχύτερο από κάθε άλλο αλγόριθμο επιτρέποντας ταχείες αναζητήσεις, ενώ παράλληλα έδινε και μια εκτίμηση για τη στατιστική σημαντικότητα των στοιχίσεων. Η πρώτη του έκδοση δεν περιελάμβανε στοιχίσεις με κενά σε αντίθεση με τη δεύτερη η οποία μεταξύ άλλων περιελάμβανε και επιπλέον παραλλαγές όπως αυτή του PSI-BLAST. Έτσι, τα προγράμματα πολλαπλής στοίχισης αλληλουχιών έκαναν δυναμική εμφάνιση με εκδόσεις για μαζική χρήση σε Η/Υ ακόμα και για παραθυρικά περιβάλλοντα.

Την ίδια χρονική περίοδο εκτελέστηκαν οι πρώτες επιτυχημένες προσπάθειες αλληλούχισης ολόκληρων γονιδιωμάτων, στην αρχή βακτηρίων και στη συνέχεια ευκαρυωτικών οργανισμών, οι οποίες άνοιξαν νέους ορίζοντες στη συγκριτική γονιδιωματική αποτελώντας έναυσμα για την ανάπτυξη των πρώτων αλγορίθμων εύρεσης γονιδίων (gene finders). Εμφανίζονται επίσης οι μικροσυστοιχίες DNA για τη μέτρηση της γονιδιακής έκφρασης, τεχνολογία που είχε μεγάλη επίδραση στη βιοστατιστική, τη βιοπληροφορική καθώς και την ιατρική πληροφορική.

Οι εξελίξεις στη βιοπληροφορική πυροδότησαν την ανάπτυξη νέων βάσεων δεδομένων. Ανάμεσά τους βρίσκονται βάσεις με δομικές ταξινομήσεις των πρωτεϊνών (όπως η SCOP και η CATH), αλλά και βάσεις με ταξινομήσεις που βασίζονται σε χαρακτηριστικά πρότυπα (patterns) μιας ακολουθίας (η PROSITE, η PFAM και η INTERPRO). Επιπροσθέτως, το 1992 ιδρύεται στη Μεγάλη Βρετανία το EBI (European Bioinformatics Institute), το μεγαλύτερο ινστιτούτο βιοπληροφορικής της Ευρώπης, το οποίο αποτελεί κοινοπραξία των EMBL και Wellcome Trust. Στο ινστιτούτο αυτό στεγάστηκαν σε πρώτη φάση οι βάσεις δεδομένων EMBL, EMBL-Bank και SwissProt-TrEMBL. Αργότερα δημιουργήθηκαν ερευνητικές ομάδες οι οποίες συνέβαλαν στα διάφορα γονιδιωματικά προγράμματα της εποχής, ενώ λίγο

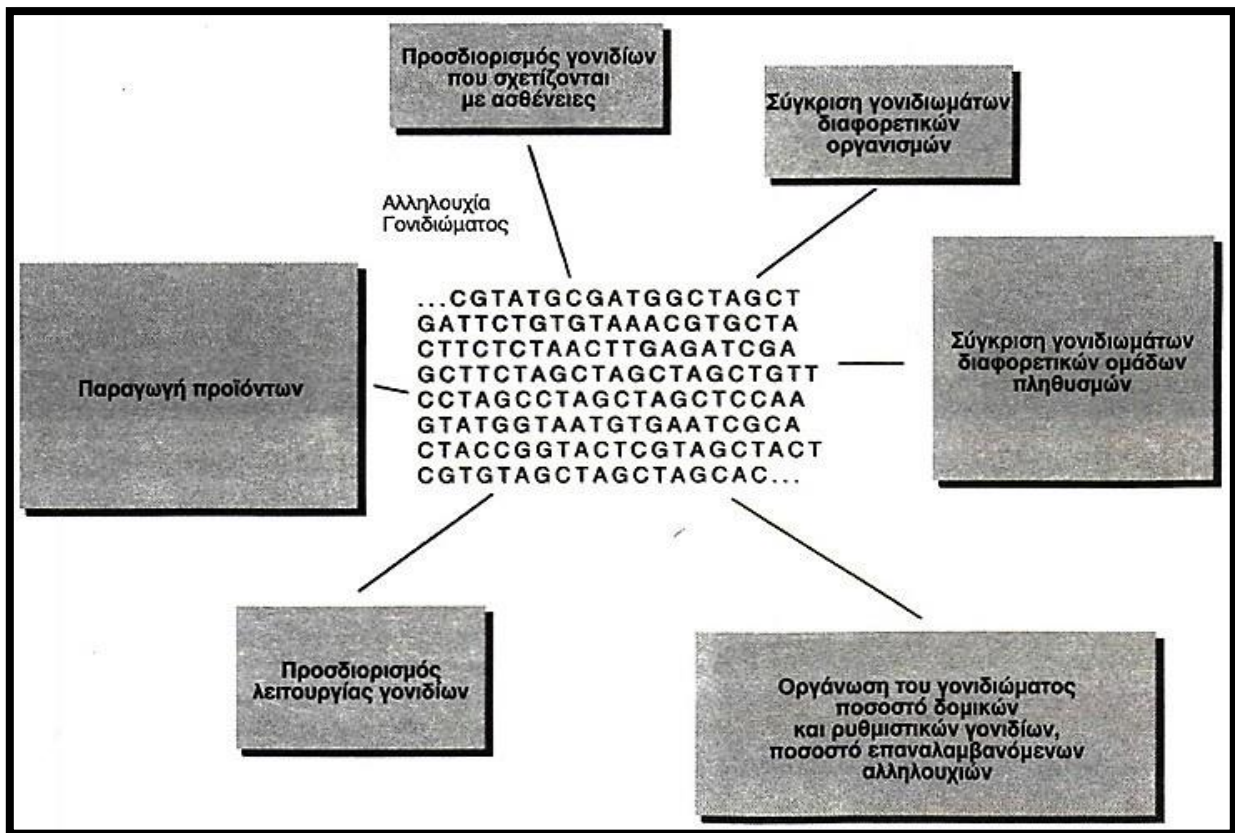
αργότερα λειτούργησε και η TrEMBL. Το 1993 ξεκίνησαν τα πρώτα συνέδρια ISMB και λίγα χρόνια αργότερα ιδρύθηκε η ISCB.

Τέλος, τη δεκαετία αυτή έκανε την εμφάνισή της μια μεθοδολογία που θα επικρατούσε τα επόμενα χρόνια στην ανάλυση των αλληλουχιών, το Hidden Markov Model (HMM). Το μοντέλο αυτό παρουσιάστηκε μετά από εργασίες που πραγματοποίησαν οι Krogh, Eddy, και Hughey και εφαρμόστηκε τόσο στις μεθόδους πρόγνωσης, όσο και στη μοντελοποίηση των πολλαπλών στοιχίσεων και στην αναζήτηση μακρινών ομοιοτήτων.

Μόλις στις αρχές του 21^{ου} αιώνα συναντάμε το πέρας ενός μεγάλου, παγκόσμιου εγχειρήματος το οποίο είχε ξεκινήσει περίπου την τελευταία δεκαετία του προηγούμενου αιώνα. Το εγχείρημα αυτό δεν είναι άλλο από την χαρτογράφηση ολόκληρου του ανθρωπίνου γονιδιώματος. Η προσπάθεια αυτή χρηματοδοτήθηκε από τις Ηνωμένες Πολιτείες της Αμερικής με το ποσό των 3 δισεκατομμυρίων δολαρίων και είχε διάρκεια 15 χρόνια περίπου (200 εκατομμύρια δολάρια για κάθε έτος έρευνας). Ο επιστημονικός στόχος ήταν η χαρτογράφηση των γονιδίων του ανθρώπου καθώς και η αλληλούχιση του γενετικού του υλικού. Η χαρτογράφηση θα αποκάλυπτε τη θέση και την απόσταση των τότε γνωστών εκατό χιλιάδων γονιδίων που βρίσκονταν σε κάθε ένα από τα ανθρώπινα κύτταρα. Η αλληλούχιση θα καθόριζε τη σειρά των τεσσάρων βάσεων των νουκλεοτιδίων – της αδενίνης (A), της θυμίνης (T), της γουανίνης (G) και της κυτοσίνης (C) – που συνθέτουν το μόριο του DNA. Το πρώτο κίνητρο που οδήγησε σε αυτήν την τεράστια έρευνα ήταν αυτό το οποίο δημιούργησε όλες τις βασικές επιστήμες, η ανάγκη του να γνωρίζουμε. Το δεύτερο κίνητρο ήταν ίσως ακόμη πιο σημαντικό, αφού αποτελούσε τον προσδιορισμό των τεσσάρων χιλιάδων γονιδίων περίπου για τα οποία υπήρχαν υποψίες πως θεωρούνταν υπεύθυνα για διάφορες κληρονομικές ασθένειες. Η έρευνα αυτή θα προετοίμαζε το έδαφος και για τη λεγόμενη «γονιδιακή θεραπεία».

Τελικά, όπως αποδεικνύεται από τα αποτελέσματα της έρευνας, το ανθρώπινο DNA είναι σε μεγάλο βαθμό «άχρηστο». Με τον όρο άχρηστο εννοούμε ότι το ενενήντα οκτώ τοις εκατό (98%) του γενετικού υλικού δεν κωδικοποιείται σε πρωτεΐνες. Το μισό από το άχρηστο DNA αποτελείται από επαναλαμβανόμενες

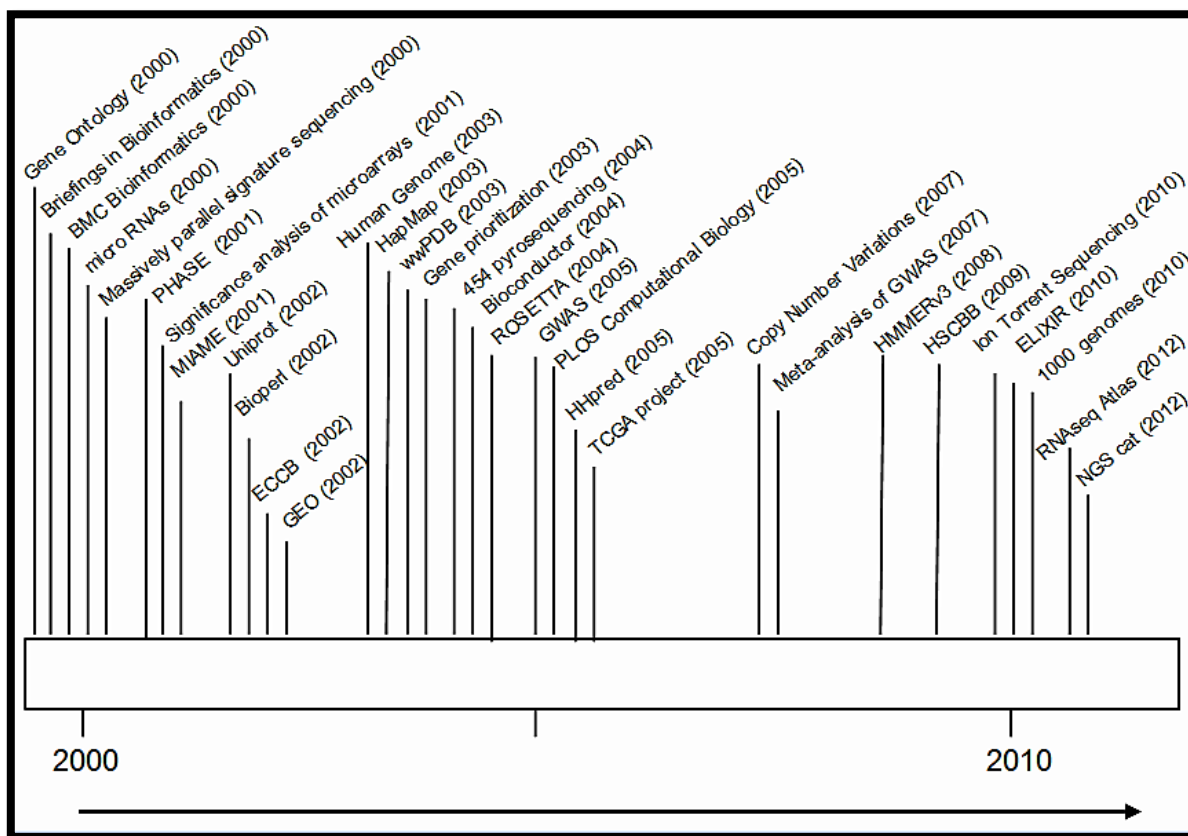
αλληλουχίες διαφόρων τύπων, οι περισσότερες από τις οποίες είναι παρασιτικά στοιχεία που έχουν κληρονομηθεί από το μακρινό εξελικτικό παρελθόν μας. Μόλις το δύο τοις εκατό (2%) του DNA αποτελείται από αλληλουχίες που κωδικοποιούνται σε πρωτεΐνες και λειτουργούν σαν γονίδια. Το συμπέρασμα που προκύπτει είναι ότι η επιστημονική κοινότητα έχει μπροστά της πολύ δρόμο ώστε να σχηματίσει μια ολοκληρωμένη άποψη επί του θέματος, αφού σε κάθε ανακάλυψη ο ρόλος του DNA μοιάζει να περιπλέκεται όλο και περισσότερο.



Εικόνα 1.5: Εφαρμογές του προγράμματος του ανθρώπινου γονιδιώματος.

Με το άνωθεν πείραμα προέκυψε μια αλματώδης ανάπτυξη της γονιδιωματικής στις διάφορες μορφές που τη διέπουν (συγκριτική, λειτουργική και δομική γονιδιωματική). Παράλληλα εμφανίστηκαν οι τεχνικές αλληλούχησης νέας γενιάς, οι οποίες έδωσαν έναυσμα σε μελέτες σχετικές με τη γονιδιακή έκφραση, κάνοντας εύκολο τον εντοπισμό πολυμορφικών θέσεων. Οι ίδιες τεχνικές αναμένεται να επηρεάσουν και την προσωποποιημένη ιατρική. Φυσικά, ο τεράστιος όγκος δεδομένων που προέκυψε από τις σχετικές μελέτες, οδήγησε στην ανάγκη

δημιουργίας εξειδικευμένων βάσεων δεδομένων καθώς και νέων αλγορίθμων πρόγνωσης. Πραγματοποιήθηκαν επίσης συγχωνεύσεις μεγάλων βάσεων δεδομένων όπως αυτή των SwissProt και PIR (σχημάτισαν την UniProt). Επίσης, την ίδια περίοδο πραγματοποιείται ταχύτερη ανάπτυξη των διαδικτυακών εφαρμογών αλλά και του ελεύθερου λογισμικού βιοπληροφορικής. Τέλος, σχετικά με το αλγοριθμικό κομμάτι, εμφανίζονται τα Support Vector Machines (SVM's) σε προβλήματα πρόγνωσης, η χρήση των οποίων απεδείχθη σε ορισμένες περιπτώσεις πιο αποτελεσματική από τα νευρωνικά δίκτυα.



Εικόνα 1.6: Ανακαλύψεις σχετικές με τη βιοπληροφορική μέχρι και το τέλος της δεκαετίας του 2000.

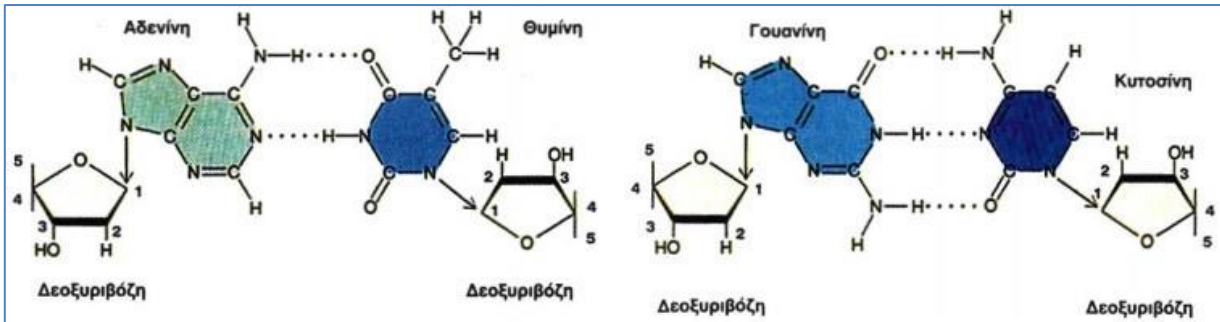
Κεφάλαιο 2^ο: Η επιστήμη της Βιολογίας

2.1 Εισαγωγή στο γονιδίωμα

Παρ' όλο που το DNA είχε εντοπιστεί στον πυρήνα των κυττάρων από το 1869, έπρεπε να περάσουν 75 χρόνια περίπου για να γνωστοποιηθεί ότι αποτελεί το γενετικό υλικό των οργανισμών. Αρχικά οι επιστήμονες πίστευαν ότι οι πρωτεΐνες αποτελούν τα μόρια τα οποία μεταφέρουν την γενετική πληροφορία επειδή παρουσιάζουν μεγαλύτερη ποικιλομορφία λόγω του ότι είναι αποτέλεσμα συνδυασμού είκοσι διαφορετικών αμινοξέων. Το DNA σε πρώτη φάση είχε αποκλειστεί διότι ήταν συνδυασμός τεσσάρων (μόνο) νουκλεοτιδίων. Πριν φτάσουμε στο 1944, στο έτος που αποδείχτηκε ότι το DNA είναι το γενετικό υλικό, ο Griffith είχε ήδη πραγματοποιήσει ένα πείραμα το 1928 το οποίο όμως τον οδήγησε σε ασαφή συμπεράσματα όσον αφορά το γενετικό υλικό. Το ίδιο πείραμα επανέλαβαν δεκαέξι χρόνια αργότερα, οι Avery, Mac-Leod και McCarty σε δοκιμαστικό σωλήνα (in vitro) και έφτασαν στο συμπέρασμα ότι τελικά το DNA είναι το γενετικό υλικό ενός οργανισμού. Η οριστική επιβεβαίωση για τον ισχυρισμό αυτό ήρθε το 1952 με πειράματα που εκτέλεσαν οι Hershey και Chase.

2.2 Η δομή του DNA

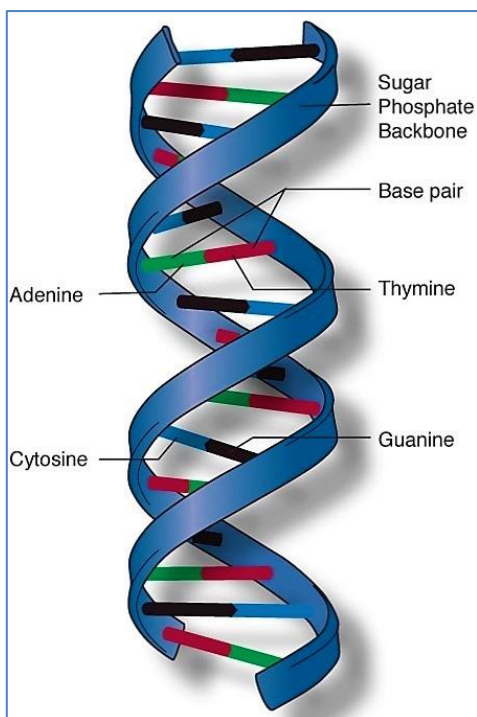
Το DNA, όπως και το RNA, είναι ένα μακρομόριο το οποίο αποτελείται από νουκλεοτίδια. Κάθε νουκλεοτίδιο αποτελείται με τη σειρά του από τρία συστατικά: μία πεντόζη, η οποία στο DNA ονομάζεται δεοξυριβόζη, μία φωσφορική ομάδα, η οποία είναι δεμένη με την πεντόζη, και μια αζωτούχο βάση. Οι αζωτούχες βάσεις είναι: η αδενίνη (A), η θυμίνη (T), η γουανίνη (G) και η κυτοσίνη (C). Ένα DNA νουκλεοτίδιο περιέχει μία μόνο εκ των τεσσάρων αζωτούχων βάσεων. Σε κάθε νουκλεοτίδιο η αζωτούχος βάση συνδέεται με τον 1' άνθρακα της πεντόζης (δεοξυριβόζη για το DNA) ενώ η φωσφορική ομάδα είναι συνδεδεμένη με τον άνθρακα στην θέση 5'.



Εικόνα 2.1: Οι τέσσερις αζωτούχες βάσεις του DNA. Η αδενίνη συνδέεται με τη θυμίνη με δύο δεσμούς υδρογόνου και η γουανίνη συνδέεται με την κυτοσίνη με τρεις δεσμούς υδρογόνου.

Μια πολυνουκλεοτιδική αλυσίδα σχηματίζεται από την ένωση πολλών νουκλεοτιδίων με ομοιοπολικό δεσμό. Αυτός ο δεσμός δημιουργείται μεταξύ του υδροξυλίου που βρίσκεται στον 3' άνθρακα της πεντόζης του πρώτου νουκλεοτιδίου και της φωσφορικής ομάδας που είναι δεμένη στον 5' άνθρακα της πεντόζης του ακόλουθου νουκλεοτιδίου. Ο δεσμός αυτός είναι γνωστός ως 3'-5' φωσφοδιεστερικός δεσμός. Τελικά, η πολυνουκλεοτιδική αλυσίδα αποτελείται από επανάληψη των μορίων φωσφορική ομάδα – πεντόζη – φωσφορική ομάδα – πεντόζη. Ο σκελετός δηλαδή της αλυσίδας παραμένει σταθερός. Το μόνο που αλλάζει κάθε φορά είναι η αζωτούχος βάση που συνδέεται με τον 1' άνθρακα. Ανεξάρτητα από τον αριθμό των νουκλεοτιδίων της αλυσίδας, το πρώτο νουκλεοτίδιο έχει πάντα ελεύθερη τη φωσφορική του ομάδα ενώ το τελευταίο έχει ελεύθερο το υδροξύλιό του. Έτσι, η πολυνουκλεοτιδική αλυσίδα έχει πραγματικό προσανατολισμό 5'-3'.

Με τα πειράματα που είχαν πραγματοποιηθεί στο παρελθόν, είμασταν σε θέση να γνωρίζουμε τόσο τις ιδιότητες όσο και τη χημική σύσταση του DNA. Επίσης ήταν γνωστό ότι σε κάθε μόριο ο αριθμός των νουκλεοτιδίων που είχαν ως βάση την αδενίνη ήταν ίσος με τον αριθμό των νουκλεοτιδίων που είχαν ως βάση την θυμίνη. Αυτό φυσικά ίσχυε και για τις αζωτούχες βάσεις κυτοσίνη και γουανίνη. Το πρόβλημα όμως ήταν ότι δεν υπήρχε κοινά αποδεκτή πρόταση για τη δομή του DNA στο χώρο. Το 1953 πραγματοποιήθηκε η μεγαλύτερη βιολογική ανακάλυψη του 20^{ου} αιώνα η οποία και καθιέρωσε την τρισδιάστατη δομή του DNA στο χώρο. Η ανακάλυψη αυτή ήταν το αποτέλεσμα της ερευνητικής εργασίας δύο ομάδων επιστημόνων: των Wilkins και Franklin καθώς και των Watson και Crick. Έχοντας σαν στήριγμα τα



αποτελέσματα της έρευνας των δύο ομάδων, οι Watson και Crick διατύπωσαν το μοντέλο της **διπλής έλικας του DNA**. Νέα στοιχεία προκύπτουν από αυτή τη διατύπωση.

Το DNA αποτελείται τελικά από δύο πολυνουκλεοτιδικές αλυσίδες οι οποίες σχηματίζουν μια δεξιόστροφη διπλή έλικα. Η έλικα αυτή είναι ένας σταθερός σκελετός που αποτελείται από μόρια φωσφορικής ομάδας – πεντόζης (δεοξυριβόζης) τα οποία ενώνονται με φωσφοδιεστερικό δεσμό. Ο σκελετός βρίσκεται στο εξωτερικό του μορίου και είναι υδρόφιλος. Στο εσωτερικό του σταθερού αυτού σκελετού

Εικόνα 2.2: Η διπλή έλικα του DNA.

βρίσκονται οι υδρόφοβες αζωτούχες βάσεις. Ο κανόνας της συμπληρωματικότητας προσδιορίζει τον τρόπο με τον οποίο συνδέονται οι βάσεις της μιας αλυσίδας με τις βάσεις της απέναντι αλυσίδας. Τη σύνδεση αυτή πραγματοποιούν δεσμοί υδρογόνου που καθιστούν τη δομή του μορίου δευτεροταγή. Οι δεσμοί που χρειάζονται για να συνδεθεί η αδενίνη με τη θυμίνη (και αντίστροφα) είναι δύο. Για τη σύνδεση μεταξύ γουανίνης και κυτοσίνης (και αντίστροφα) χρειάζονται τρεις. Η αλληλουχία της μιας αλυσίδας καθορίζει την αλληλουχία της άλλης, γεγονός που καθιστά τις δύο αλυσίδες αντιπαράλληλες.

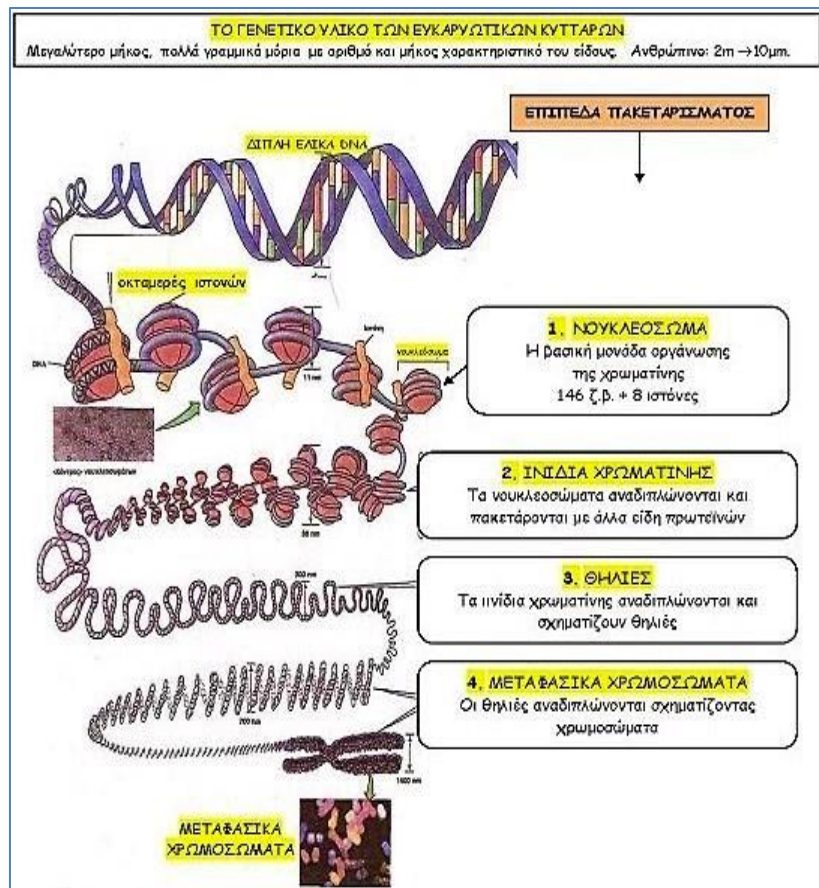
Τα παραπάνω στοιχεία προσδίδουν λειτουργίες στο DNA όπως αυτή της αποθήκευσης της γενετικής πληροφορίας η οποία επιτυγχάνεται με την οργάνωσή της σε λειτουργικές μονάδες που ονομάζονται γονίδια. Μέσω του αυτοδιπλασιασμού του, το DNA διατηρεί και μεταβιβάζει τη γενετική πληροφορία που είναι αποθηκευμένη. Η έκφραση της γενετικής πληροφορίας πραγματοποιείται με τον έλεγχο της σύνθεσης των πρωτεϊνών. Το γενετικό υλικό ενός κυττάρου αποτελεί το γονιδίωμά του. Υπάρχουν δύο ομάδες κυττάρων: τα απλοειδή και τα διπλοειδή κύτταρα. Στα απλοειδή κύτταρα, όπως είναι οι γαμέτες των διπλοειδών οργανισμών, υπάρχει ένα μόνο αντίγραφο του γονιδιώματος. Στα διπλοειδή κύτταρα, όπως είναι τα

σωματικά κύτταρα των ανώτερων ευκαρυωτικών οργανισμών, το γονιδίωμα υπάρχει σε δύο αντίγραφα. Στους ευκαρυωτικούς οργανισμούς το γενετικό υλικό κατανέμεται στον πυρήνα και στα μιτοχόνδρια των κυττάρων (όταν πρόκειται για φυτά κατανέμεται και στους χλωροπλάστες).

2.3 Οργάνωση του γενετικού υλικού των ευκαρυωτικών οργανισμών

Το συνολικό DNA που υπάρχει σε κάθε ευκαρυωτικό κύτταρο αποτελείται από

πολλά γραμμικά μόρια των οποίων ο αριθμός και το μήκος είναι χαρακτηριστικά για τα διάφορα είδη των οργανισμών. Τα μόρια του DNA πακετάρονται με πρωτεΐνες σχηματίζοντας ινίδια χρωματίνης. Τα ινίδια αυτά, μετά από ειδική επεξεργασία σε ηλεκτρονικό μικροσκόπιο, εμφανίζονται σαν κομπολόγια από χάντρες. Οι χάντρες είναι γνωστές ως νουκλεοσώματα. Κάθε



νουκλεόσωμα απαρτίζεται από DNA μήκους 146

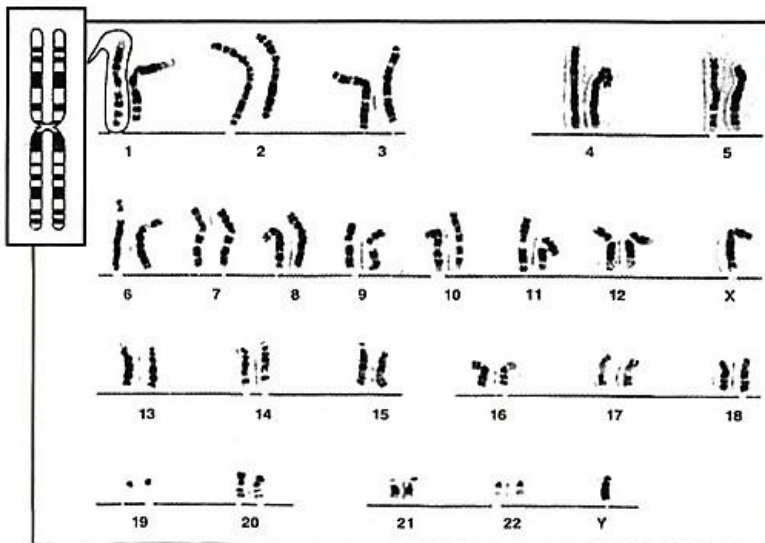
Εικόνα 2.3: Επίπεδα πακεταρίσματος του DNA στο μεταφασικό χρωμόσωμα.

ζευγών βάσεων και από οκτώ μόρια πρωτεϊνών, τις λεγόμενες ιστόνες. Το DNA τυλίγεται γύρω από το οκταμερές των ιστονών. Με την αναδίπλωση των νουκλεοσωμάτων επιτυγχάνεται πακετάρισμα μεγαλύτερου βαθμού σχηματίζοντας έτσι ένα δίκτυο ινιδίων χρωματίνης. Η κατάσταση αυτή ονομάζεται μεσόφαση. Στο τέλος της αντιγραφής, κάθε ινίδιο έχει διπλασιαστεί. Τα δύο αντίγραφα κάθε ινιδίου

συνδέονται με μια δομή γνωστή ως κεντρομερίδιο. Τα διπλασιασμένα χρωμοσώματα ονομάζονται αδελφές χρωματίδες για το χρονικό διάστημα που είναι συνδεδεμένα. Κατά το στάδιο της μετάφασης¹, οι αδελφές χρωματίδες βρίσκονται σε πλήρη συσπείρωση αναμένοντας την κυτταρική διαίρεση. Στο τέλος της κυτταρικής διαίρεσης προκύπτουν δύο νέα κύτταρα, γενετικά όμοια μεταξύ τους καθώς και με το αρχικό.

2.4 Μελετώντας τα ανθρώπινα χρωμοσώματα – Καρυότυπος

Η μελέτη των χρωμοσωμάτων του ανθρώπου είναι εφικτή μόνο σε κύτταρα τα οποία διαιρούνται (κύτταρα από ανθρώπινο ιστό). Τα χρωμοσώματα μελετώνται στο στάδιο της μετάφασης όπως αναφέραμε προηγουμένως. Για να γίνει αυτό, οι



Εικόνα 2.4: Καρυότυπος φυσιολογικού αρσενικού ατόμου, στον οποίο έχει προστεθεί απεικόνιση του 1^{ου} χρωμοσώματος (χρώση Giemsa).

επιστήμονες χρησιμοποιούν ουσίες οι οποίες σταματούν την κυτταρική διαίρεση. Κατόπιν, τα κύτταρα επωάζονται σε υποτονικό διάλυμα με σκοπό τη διάτρηση της κυτταρικής τους μεμβράνης. Τα χρωμοσώματα τώρα μπορούν να τοποθετηθούν σε αντικειμενοφόρο πλάκα.

Με τη βοήθεια ειδικών χρωστικών ουσιών μπορεί

κανείς να τα παρατηρήσει στο μικροσκόπιο. Τέλος, τα χρωμοσώματα ταξινομούνται σε ζεύγη κατά ελαττούμενο μέγεθος. Η απεικόνιση αυτή ονομάζεται καρυότυπος. Στον ανθρώπινο οργανισμό, είτε είναι αρσενικός είτε είναι θηλυκός, υπάρχουν 23 ζεύγη χρωμοσωμάτων μέσα στον πυρήνα των σωματικών κυττάρων. Τα 22 από τα 23 ζεύγη είναι μορφολογικά ίδια (αυτοσωμικά χρωμοσώματα) και στις δύο περιπτώσεις. Το 23^ο ζεύγος στα θηλυκά αποτελείται από δύο X χρωμοσώματα. Στα

¹ Ο υψηλός βαθμός συσπείρωσης στο στάδιο αυτό καθιστά τα χρωμοσώματα ευδιάκριτα με το οπτικό μικροσκόπιο.

αρσενικά το ζεύγος αυτό απαρτίζεται από ένα Y και ένα X χρωμόσωμα. Για το λόγο αυτό τα χρωμοσώματα X, Y ονομάζονται φυλετικά.

Όπως προαναφέραμε, στα μιτοχόνδρια υπάρχει γενετικό υλικό το οποίο κωδικοποιεί έναν μικρό αριθμό πρωτεϊνών σε συνεργασία με το γενετικό υλικό του πυρήνα του κυττάρου. Αυτή η ιδιότητα καθιστά τα μιτοχόνδρια ημιαυτόνομα. Στους ανώτερους οργανισμούς, όπως και στον άνθρωπο, τα μιτοχόνδρια προέρχονται από το ωάριο (μητρική προέλευση). Αυτή η γνώση κατευθύνει τους επιστήμονες σε θέματα που αφορούν μιτοχονδριακές ασθένειες.

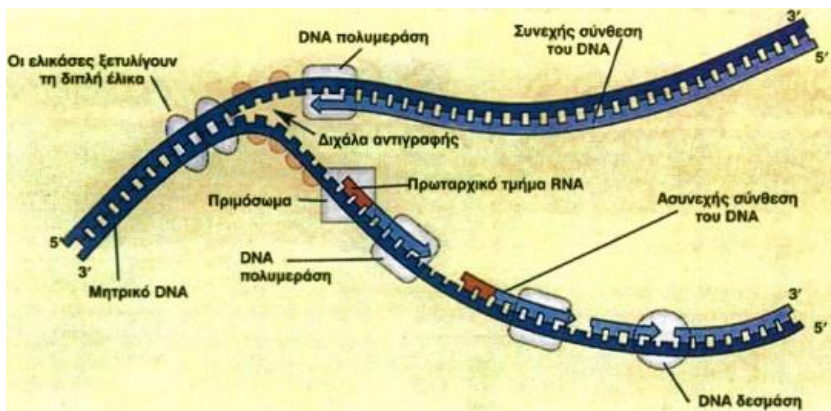
2.5 Η αντιγραφή του DNA

Με γνώμονα τον κανόνα της συμπληρωματικότητας των βάσεων του DNA, οι Watson και Crick φαντάστηκαν την διπλή έλικα να ξετυλίγεται και καθένα από τα δύο κομμάτια της να λειτουργεί σαν καλούπι για την δημιουργία μια νέας συμπληρωματικής αλυσίδας. Ο μηχανισμός αυτός ονομάστηκε ημισυντηρητικός. Με μια πρώτη ματιά, η αντιγραφή του γενετικού υλικού φαντάζει απλή. Παρ' όλα αυτά αποτελεί μια ιδιαίτερα περίπλοκη διαδικασία λόγω των ενζύμων και των πρωτεϊνών που χρησιμοποιούνται από τα κύτταρα.

Η αντιγραφή ξεκινάει από καθορισμένα σημεία, τις λεγόμενες θέσεις έναρξης της αντιγραφής. Θέσεις έναρξης αντιγραφής αποτελούν εκατοντάδες σημεία σε όλο το μήκος του DNA. Για να ξεκινήσει η διαδικασία της αντιγραφής του DNA πρέπει οι δύο αλυσίδες να ξετυλιχθούν στις θέσεις έναρξης αντιγραφής. Οι DNA ελικάσες είναι τα ένζυμα εκείνα που αναλαμβάνουν να σπάσουν τους δεσμούς υδρογόνου μεταξύ των δύο αλυσίδων (συγκεκριμένα μεταξύ των αζωτούχων βάσεων των δύο αλυσίδων). Όταν επιτευχθεί το άνοιγμα της έλικας δημιουργείται μια «θηλιά²» αυξανόμενη και προς τις δύο κατευθύνσεις. Μέσα στο κύτταρο υπάρχει ένα ειδικό σύμπλοκο, αποτελούμενο από πολλά ένζυμα, το οποίο ονομάζεται πριμόσωμα. Το πριμόσωμα συνθέτει στις θέσεις έναρξης της αντιγραφής μικρά τμήματα RNA, τα επονομαζόμενα πρωταρχικά τμήματα. Τα τμήματα αυτά είναι συμπληρωματικά προς τις μητρικές

² Οι θηλιές που δημιουργούνται κατά την έναρξη της αντιγραφής σε ένα μόριο DNA είναι ορατές με το ηλεκτρονικό μικροσκόπιο.

αλυσίδες. Μέσα στο πριμόσωμα υπάρχουν τα ένζυμα DNA πολυμεράσες τα οποία επιμηκύνουν τα πρωταρχικά τμήματα, τοποθετώντας συμπληρωματικά δεοξυριβονουκλεοτίδια απέναντι από τις μητρικές αλυσίδες. Κατά τη διάρκεια της δημιουργίας δεσμών υδρογόνου μεταξύ των συμπληρωματικών αζωτούχων βάσεων των δεοξυριβονουκλεοτιδίων, σχηματίζονται νέα μόρια DNA. Τυχόν λάθη κατά τη διάρκεια της αντιγραφής διορθώνονται από τις DNA πολυμεράσες με απίστευτη ακρίβεια. Λάθη που δεν επιδιορθώνονται από τις πολυμεράσες, επιδιορθώνονται σε μεγάλο ποσοστό από άλλα ειδικά για αυτή τη λειτουργία ένζυμα. Οι DNA πολυμεράσες αναλαμβάνουν επίσης την αντικατάσταση των πρωταρχικών τμημάτων RNA με τμήματα DNA. Επειδή η αντιγραφή πραγματοποιείται με προσανατολισμό 5'-3', οι πολυμεράσες τοποθετούν τα νουκλεοτίδια στο ελεύθερο 3' άκρο της δεοξυριβόζης του τελευταίου νουκλεοτιδίου κάθε αναπτυσσόμενης αλυσίδας. Για να εφαρμοστεί αυτός ο κανόνας, κάθε τμήμα DNA που γίνεται η αντιγραφή φέρει μία συνεχή και μία ασυνεχή αλυσίδα. Η DNA δεσμάση αναλαμβάνει να συνδέσει τα κομμάτια της ασυνεχούς αλυσίδας



Εικόνα 2.5: Η αντιγραφή του DNA.

μεταξύ τους. Τα κομμάτια που προκύπτουν από τις διάφορες θέσεις αντιγραφής συνδέονται επίσης με τη βοήθεια του ίδιου ενζύμου. Τελικά, κάθε νέα έλικα που παράγεται έχει τις δύο αλυσίδες της αντιπαράλληλες.

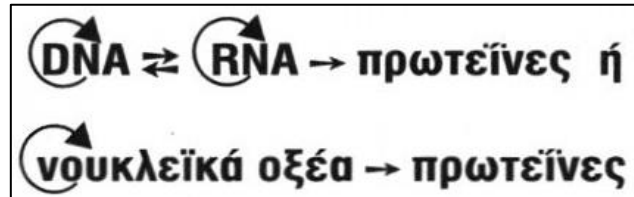
2.6 Η έκφραση της γενετικής πληροφορίας

Το γενετικό υλικό ενός οργανισμού αποτελεί έναν μοριακό «σκληρό δίσκο» μέσα στον οποίο βρίσκονται αποθηκευμένες ακριβείς οδηγίες, καθοριστικές για τη δομή και τη λειτουργία του οργανισμού. Παράλληλα περιέχονται πληροφορίες για τον αυτοδιπλασιασμό του DNA, καθώς και για τη μεταβίβαση των γενετικών οδηγιών ενός κυττάρου στα θυγατρικά του. Για την έκφραση της γενετικής πληροφορίας πρέπει να

πραγματοποιηθούν κάποιες ενέργειες. Η διαδικασία της αντιγραφής του DNA που περιγράψαμε διαιωνίζει τη γενετική πληροφορία. Με τη διαδικασία της μεταγραφής, μεταφέρεται η πληροφορία από τα γονίδια (τμήματα του DNA με συγκεκριμένη ακολουθία) στο RNA. Ακολουθεί η διαδικασία της μετάφρασης μέσω της οποίας το RNA μεταφέρει την πληροφορία στις πρωτεΐνες, οι οποίες καθορίζουν τη δομή και τη λειτουργία των κυττάρων. Οι παραπάνω ενέργειες αποτελούν τη λεγόμενη γονιδιακή έκφραση.



Για αρκετό καιρό οι επιστήμονες πίστευαν πως η ροή της γενετικής πληροφορίας γινόταν προς μία μόνο κατεύθυνση, όπως περιγράφεται στην παραπάνω εικόνα. Με την ανακάλυψη όμως ιών που έχουν σαν γενετικό υλικό RNA, το κεντρικό αυτό δόγμα σταμάτησε να υφίσταται. Σήμερα λοιπόν ισχύει το δόγμα που περιγράφεται στην παρακάτω εικόνα.



Σε κάθε ομάδα κυττάρων εκφράζονται διαφορετικά γονίδια τα οποία διακρίνονται σε δύο κατηγορίες:

- Στα γονίδια που μεταγράφονται σε mRNA και αργότερα μεταφράζονται σε πρωτεΐνες.
- Στα γονίδια που μεταγράφονται και παράγουν tRNA, rRNA και snRNA.

Και τα τέσσερα είδη μορίων (mRNA, tRNA, rRNA και snRNA) παράγονται κατά τη διαδικασία της μεταγραφής. Ακολουθεί η περιγραφή του κάθε μορίου:

Αγγελιαφόρο RNA (mRNA): Μεταφέρει την πληροφορία του DNA με σκοπό τη δημιουργία μιας πολυπεπτιδικής αλυσίδας.

Ριβοσωμικό RNA (rRNA): Συνδέεται με πρωτεΐνες σχηματίζοντας το ριβόσωμα το οποίο είναι απαραίτητο για την πρωτεϊνοσύνθεση.

Μεταφορικό RNA (tRNA): Συνδέεται με ένα συγκεκριμένο αμινοξύ το οποίο και μεταφέρει στη θέση όπου πραγματοποιείται η πρωτεϊνοσύνθεση.

Μικρό πυρηνικό RNA (snRNA): Μικρά μόρια που συνδέονται με πρωτεΐνες σχηματίζοντας μικρά ριβονουκλεοπρωτεϊνικά σωματίδια. Ευθύνονται για την κατάλυση της «ωρίμανσης» του αγγελιαφόρου RNA (mRNA).

2.6.1 Η μεταγραφή του DNA

Η διαδικασία της μεταγραφής καταλύεται από ένα ένζυμο που ονομάζεται RNA πολυμεράση. Οι μεταγραφικοί παράγοντες είναι οι πρωτεΐνες που βοηθούν την RNA πολυμεράση να προσδεθεί στις ειδικές περιοχές του DNA (υποκινητές) ώστε να ξεκινήσει σωστά η φάση της μεταγραφής. Οι υποκινητές βρίσκονται πάντα πριν από την αρχή κάθε γονιδίου. Με την έναρξη της μεταγραφής ενός γονιδίου, η RNA πολυμεράση προκαλεί τοπικό ξετύλιγμα της διπλής έλικας του DNA στον υποκινητή με τον οποίο συνδέθηκε. Αμέσως μετά, τοποθετεί τα ριβονουκλεοτίδια (νουκλεοτίδια του RNA) ακριβώς απέναντι από τα δεοξυριβονουκλεοτίδια μιας DNA αλυσίδας ακολουθώντας πάντα τον κανόνα της συμπληρωματικότητας των βάσεων (όπως συμβαίνει και στην αντιγραφή). Στην προκειμένη περίπτωση απέναντι από την αζωτούχο βάση αδενίνη (A) θα τοποθετηθεί το ριβονουκλεοτίδιο που φέρει την συμπληρωματική βάση ουρακίλη (U) διότι δεν υπάρχει ριβονουκλεοτίδιο που να έχει ως αζωτούχο βάση τη θυμίνη (T). Όπως και η αντιγραφή έτσι και η μεταγραφή έχει προσανατολισμό 5'-3'. Ειδικές αλληλουχίες που βρίσκονται στο τέλος του γονιδίου σταματούν τη σύνθεση RNA επιτρέποντας την απελευθέρωσή του. Οι αλληλουχίες αυτές ονομάζονται αλληλουχίες λήξης της μεταγραφής.

Το RNA που συντίθεται αποτελεί συμπλήρωμα της μεταγραφόμενης αλυσίδας της διπλής έλικας η οποία ονομάζεται μη κωδική. Το άλλο μισό της έλικας αποτελεί την κωδική αλυσίδα.

Το mRNA που παράγεται κατά τη μεταγραφή ενός γονιδίου υφίσταται μία πολύπλοκη διαδικασία ωρίμανσης αφού δεν είναι έτοιμο να μεταφραστεί ακόμα. Αυτό

το mRNA ονομάζεται πρόδρομο και περιέχει εξώνια³ και εσώνια⁴. Η ωρίμανση του προδρόμου mRNA πραγματοποιείται με τη βοήθεια του snRNA το οποίο αναλαμβάνει να αποκόψει τα εσώνια από το mRNA ράβοντας παράλληλα τα εξώνια μεταξύ τους. Τελικά το ώριμο mRNA αποτελείται μόνο από εξώνια έχοντας το 5' και το 3' άκρο του αμετάφραστα. Το mRNA είναι πλέον έτοιμο να μεταφερθεί από τον πυρήνα στο κυτταρόπλασμα και συγκεκριμένα στα ριβοσώματα όπου λαμβάνει χώρα η πρωτεϊνοσύνθεση.

2.6.2 Ο γενετικός κώδικας

Με το πέρας της διαδικασίας της μεταγραφής, οι πληροφορίες που βρίσκονταν στα γονίδια μεταφέρονται στο mRNA με γνώμονα πάντα τον κανόνα της συμπληρωματικότητας των νουκλεοτιδικών βάσεων. Επόμενο στάδιο για τη μεταφορά της γενετική πληροφορίας είναι η μετάφραση, γνωστή και ως πρωτεϊνοσύνθεση, η οποία πραγματοποιείται με τη βοήθεια του γενετικού κώδικα.

Γενετικός κώδικας							
		Δεύτερο γράμμα					
		U	C	A	G		
Πρώτο γράμμα	U	UUU } φαινυλαλανίνη UUC } UUA } λευκίνη UUG }	UCU } UCC } σερίνη UCA } UCG }	UAU } τυροσίνη UAC } UAA } λήξη UAG } λήξη	UGU } κυστεΐνη UGC } UGA } λήξη UGG } τρυπτοφάνη	U	Τρίτο γράμμα
	C	CUU } λευκίνη CUC } CUA } CUG }	CCU } CCC } προλίνη CCA } CCG }	CAU } ιστιδίνη CAC } CAA } γλουταμίνη CAG }	CGU } CGC } αργινίνη CGA } CGG }	C	
	A	AUU } ισολευκίνη AUC } AUA } AUG } μεθειονίνη έναρξη	ACU } ACC } θρεονίνη ACA } ACG }	AAU } ασπαραγγίνη AAC } AAA } λυσίνη AAG }	AGU } σερίνη AGC } AGA } αργινίνη AGG }	A	
	G	GUU } βαλίνη GUC } GUA } GUG }	GCU } GCC } αλανίνη GCA } GCG }	GAU } ασπαρτικό οξύ GAC } GAA } γλουταμινικό GAG } οξύ	GGU } GGC } γλυκίνη GGA } GGG }	G	

Εικόνα 2.6: Ο γενετικός κώδικας.

³ Οι αλληλουχίες που μεταφράζονται σε αμινοξέα.

⁴ Οι αλληλουχίες που δεν μεταφράζονται σε αμινοξέα και αποτελούν ενδιάμεσες αλληλουχίες μεταξύ των εξωνίων.

Ο γενετικός κώδικας αποτελεί έναν κώδικα αντιστοίχισης νουκλεοτιδίων mRNA με αμινοξέα πρωτεϊνών. Δηλαδή, η αλληλουχία των αζωτούχων βάσεων του mRNA καθορίζει την αλληλουχία των αμινοξέων στις πρωτεΐνες με βάση τον κώδικα αυτό. Γι' αυτό η πρωτεϊνοσύνθεση είναι πραγματικά μια «μετάφραση» από τη γλώσσα των βάσεων στη γλώσσα των αμινοξέων.

Όπως προαναφέραμε, οι πρωτεΐνες είναι αποτέλεσμα συνδυασμού είκοσι διαφορετικών αμινοξέων. Το RNA είναι συνδυασμός τεσσάρων διαφορετικών νουκλεοτιδίων (όπως και το DNA). Για να μπορέσουν να κωδικοποιηθούν και τα είκοσι αμινοξέα που αποτελούν την πρωτεΐνη, έπρεπε τα νουκλεοτίδια να συνδυαστούν ανά τρία⁵ ώστε να αντιστοιχιστούν με ένα αμινοξύ.

Ο γενετικός κώδικας φέρει τα παρακάτω χαρακτηριστικά:

1. Ο γενετικός κώδικας ονομάζεται κώδικας τριπλέτας. Αυτό σημαίνει ότι μια τριάδα νουκλεοτιδίων, το κωδικόνιο, κωδικοποιεί ένα αμινοξύ.
2. Ο γενετικός κώδικας είναι συνεχής. Το mRNA διαβάζεται συνεχώς ανά τρία νουκλεοτίδια χωρίς να παραλείπεται κάποιο νουκλεοτίδιο.
3. Κάθε νουκλεοτίδιο ανήκει σε ένα μόνο κωδικόνιο. Αυτό καθιστά τον γενετικό κώδικα μη επικαλυπτόμενο.
4. Όλοι οι οργανισμοί έχουν τον ίδιο γενετικό κώδικα αφού αυτός είναι σχεδόν καθολικός.
5. Ο γενετικός κώδικας χαρακτηρίζεται ως εκφυλισμένος. Εάν εξαιρέσουμε δύο αμινοξέα (τη μεθειονίνη και την τρυπτοφάνη), τα υπόλοιπα 18 μπορούν να κωδικοποιηθούν από δύο μέχρι και έξι διαφορετικά κωδικόνια. Τα διαφορετικά κωδικόνια που κωδικοποιούν το ίδιο αμινοξύ ονομάζονται συνώνυμα.
6. Ο γενετικός κώδικας περιέχει κωδικόνιο έναρξης και κωδικόνιο λήξης. Σε όλους τους οργανισμούς, το κωδικόνιο έναρξης είναι το AUG. Για τη λήξη, δηλαδή για τον τερματισμό της σύνθεσης της πολυπεπτιδικής

⁵ Συνδυάζοντας τρία νουκλεοτίδια έχουμε παραπάνω συνδυασμούς από αυτούς που χρειαζόμαστε για να εκφράσουμε ένα αμινοξύ ($4^3 = 64$ πιθανοί συνδυασμοί).

αλυσίδας, το mRNA πρέπει να συναντήσει ένα από τα κωδικόνια UGA, UAG και UAA.

Το τμήμα ενός γονιδίου και του mRNA του που κωδικοποιεί μια πολυπεπτιδική αλυσίδα αρχίζει με το κωδικόνιο έναρξης και τελειώνει με το κωδικόνιο λήξης.

2.6.3 Μετάφραση

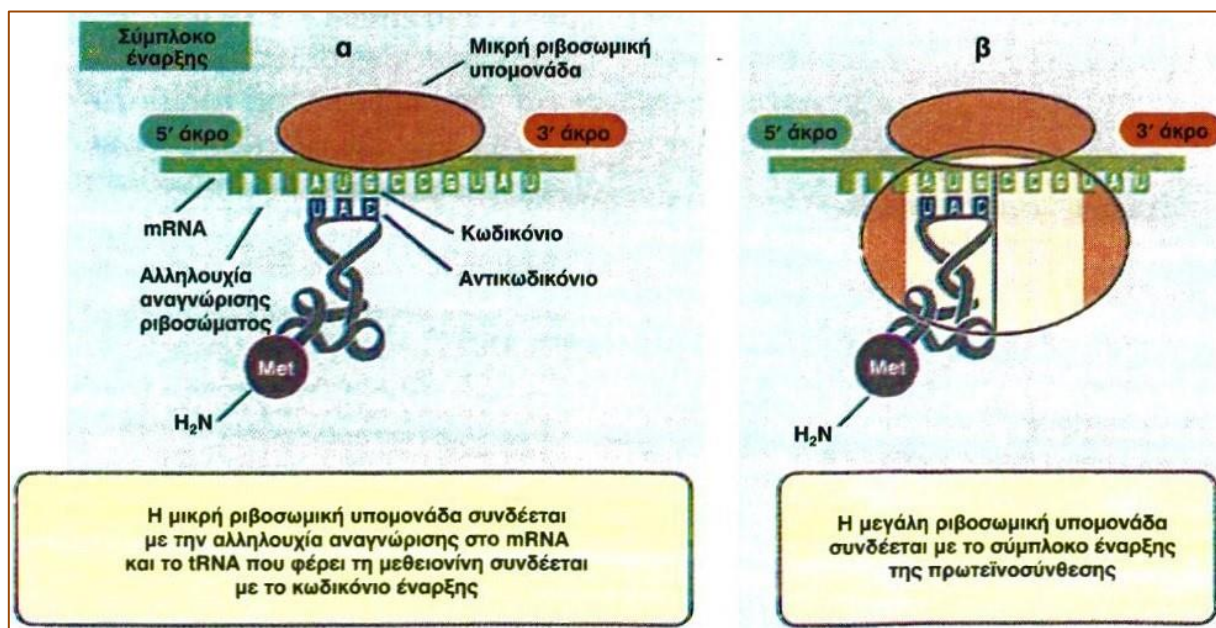
Όπως σχολιάσαμε προηγουμένως, η μετάφραση του mRNA είναι η αντιστοίχιση των κωδικονίων σε αμινοξέα καθώς και η διαδοχική σύνδεση των αμινοξέων αυτών σε πολυπεπτιδική αλυσίδα. Η διαδικασία αυτή εκτελείται στα ριβοσώματα με την βοήθεια των tRNA και τη συμμετοχή αρκετών πρωτεϊνών και ενέργειας. Τα ριβοσώματα μπορούν να χρησιμοποιηθούν ως θέση μετάφρασης για οποιοδήποτε mRNA. Με την ιδιότητα αυτή, οι επιστήμονες χρησιμοποιούν βακτήρια ώστε να παράγουν ανθρώπινες πρωτεΐνες.

Το ριβόσωμα αποτελείται από δύο υπομονάδες, τη μικρή και τη μεγάλη υπομονάδα. Στη μικρή υπομονάδα υπάρχει μια θέση πρόσδεσης του mRNA ενώ στη μεγάλη υπομονάδα υπάρχουν δύο θέσεις εισδοχής των tRNA. Κάθε μόριο tRNA φέρει μια ειδική τριπλέτα νουκλεοτιδίων η οποία ονομάζεται αντικωδικόνιο. Το αντικωδικόνιο προσδένεται με το αντίστοιχο κωδικόνιο του mRNA ακολουθώντας τον κανόνα της συμπληρωματικότητας. Επιπροσθέτως, κάθε μόριο tRNA διαθέτει μια ειδική θέση σύνδεσης με ένα συγκεκριμένο αμινοξύ.

Η διαδικασία της πρωτεϊνοσύνθεσης πραγματοποιείται σε τρία στάδια: την έναρξη, την επιμήκυνση και τη λήξη. Παρακάτω αναλύουμε τι συμβαίνει στο κάθε στάδιο.

Έναρξη: Κατά τη φάση της έναρξης το mRNA προσδένεται, μέσω μιας αλληλουχίας που υπάρχει στην 5' αμετάφραστη περιοχή του, με το ριβοσωμικό RNA (rRNA) της μικρής υπομονάδας του ριβοσώματος ακολουθώντας πάντα τον κανόνα της συμπληρωματικότητας των βάσεων. Το κωδικόνιο έναρξης του mRNA είναι πάντα το AUG και σ' αυτό προσδένεται το tRNA το οποίο φέρει το αμινοξύ

μεθειονίνη⁶. Το σύμπλοκο που δημιουργείται μετά την πρόσδεση του mRNA στη μικρή ριβοσωμική υπομονάδα και του tRNA που μεταφέρει τη μεθειονίνη καλείται σύμπλοκο έναρξης της πρωτεϊνοσύνθεσης. Ακολούθως, η μεγάλη ριβοσωμική υπομονάδα συνδέεται με τη μικρή.

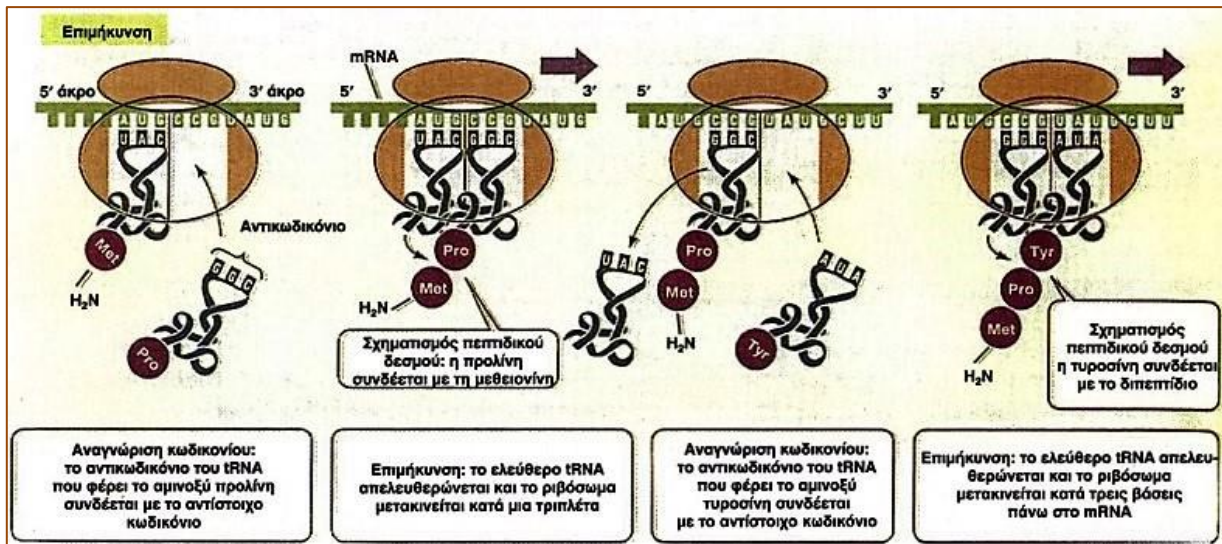


Εικόνα 2.7: Η πρωτεϊνοσύνθεση αρχίζει με το σχηματισμό ενός συμπλόκου έναρξης.

Επιμήκυνση: Στη φάση αυτή, ένα δεύτερο μόριο tRNA με το δικό του αντικωδικόνιο που είναι συμπληρωματικό του δεύτερου κωδικονίου του mRNA τοποθετείται στην κατάλληλη εισδοχή της μεγάλης ριβοσωμικής υπομονάδας, μεταφέροντας το δεύτερο αμινοξύ. Το πρώτο (μεθειονίνη) και το δεύτερο αμινοξύ συνδέονται με πεπτιδικό δεσμό. Αμέσως μετά, το πρώτο tRNA αποσυνδέεται από το ριβόσωμα και απελευθερώνεται στο κυτταρόπλασμα όπου εκεί θα συνδεθεί ξανά με μεθειονίνη, έτοιμο για μια επόμενη χρήση. Αυτή τη στιγμή, το ριβόσωμα και το mRNA έχουν ένα tRNA πάνω στο οποίο έχουν δεθεί δύο αμινοξέα. Με τον τρόπο αυτό επιτυγχάνεται η σταδιακή επιμήκυνση της πολυπεπτιδικής αλυσίδας. Το ριβόσωμα κινείται κατά μήκος του mRNA κλιμακωτά κατά ένα κωδικόνιο. Ένα τρίτο tRNA είναι έτοιμο να προσδεθεί μεταφέροντας το δικό του αμινοξύ. Πεπτιδικός δεσμός

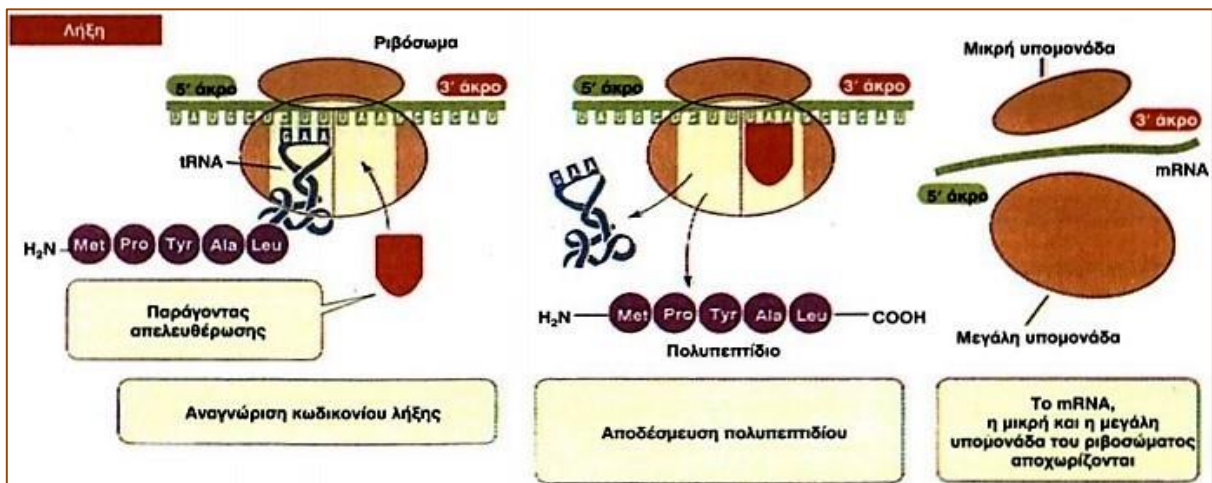
⁶ Δεν έχουν όλες οι πρωτεΐνες του οργανισμού μεθειονίνη ως πρώτο αμινοξύ. Αυτό συμβαίνει γιατί, σε πολλές πρωτεΐνες, μετά τη σύνθεσή τους απομακρύνονται ορισμένα αμινοξέα από το αρχικό αμινικό υγρό.

σχηματίζεται ανάμεσα στο δεύτερο και το τρίτο αμινοξύ. Η πολυπεπτιδική αλυσίδα επιμηκύνεται όσο νέα tRNA μεταφέρουν αμινοξέα τα οποία συνδέονται μεταξύ τους.



Εικόνα 2.8: Επιμήκυνση της πρωτεϊνοσύνθεσης.

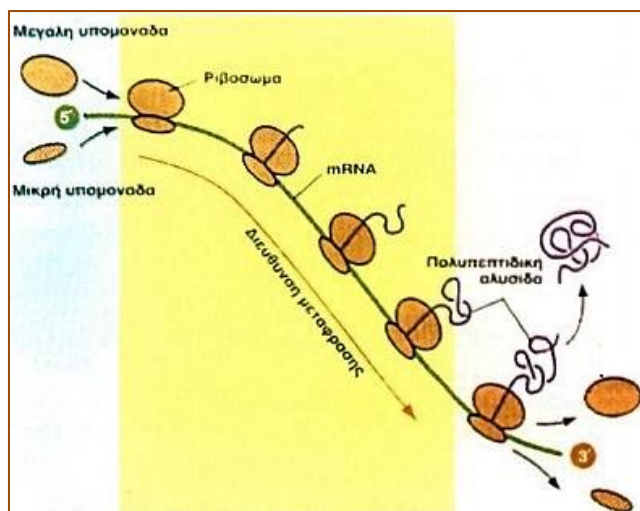
Λήξη: Η πολυπεπτιδική αλυσίδα σταματά να αναπτύσσεται όταν προκύψει ένα από τα τρία κωδικόνια λήξης (UGA, UAG ή UAA) διότι δεν υπάρχουν tRNA που να αντιστοιχούν σε αυτά. Τελικά, το τελευταίο tRNA απομακρύνεται από το ριβόσωμα επιτρέποντας την απελευθέρωση της πολυπεπτιδικής αλυσίδας.



Εικόνα 2.9: Λήξη της πρωτεϊνοσύνθεσης.

Από ένα γονίδιο μπορούν να μεταγραφούν πολλά μόρια mRNA. Επίσης, ένα mRNA μπορεί να μεταφράζεται από πολλά ριβοσώματα ταυτόχρονα, με το καθένα να

βρίσκεται σε διαφορετικό σημείο κατά μήκος του μορίου. Αμέσως μετά τη μετάφραση των πρώτων κωδικονίων από το ριβόσωμα, η θέση έναρξης του mRNA είναι ελεύθερη και ένα νέο ριβόσωμα μπορεί να προσδεθεί. Το σύμπλεγμα των ριβοσωμάτων με το mRNA καλείται πολύσωμα. Τα χαρακτηριστικά αυτά καθιστούν την πρωτεϊνοσύνθεση μια «οικονομική διαδικασία». Ένα κύτταρο μπορεί να παράγει μεγάλα ποσά μιας πρωτεΐνης από ένα ή από δύο αντίγραφα ενός γονιδίου.



Εικόνα 2.10: Ένα πολύσωμα αποτελείται από πολλά ριβοσώματα τα οποία προσδένονται κατά μήκος ενός μορίου mRNA και το μεταφράζουν.

2.7 Γονιδιακή ρύθμιση

Γονιδιακή έκφραση ονομάζεται όλη η διαδικασία με την οποία ένα γονίδιο ενεργοποιείται με σκοπό την παραγωγή μιας πρωτεΐνης. Φυσικά, σε κάθε κύτταρο δεν παράγονται όλες οι πρωτεΐνες σε κάθε χρονική στιγμή. Επίσης, οι πρωτεΐνες ενός κυττάρου δεν παράγονται σε ίσες ποσότητες διότι το κύτταρο χρειάζεται κάθε πρωτεΐνη σε συγκεκριμένη ποσότητα. Είναι απαραίτητη η ύπαρξη και η σωστή λειτουργία ενός προγράμματος (αλγόριθμος) ρύθμισης της γονιδιακής έκφρασης ώστε να παρέχονται οδηγίες σχετικές με το είδος αλλά και την ποσότητα των πρωτεϊνών που πρέπει να παραχθούν σε μια συγκεκριμένη χρονική στιγμή. Αυτό συμβαίνει διότι εάν όλα τα γονίδια λειτουργούσαν με τον ίδιο ρυθμό, ορισμένες πρωτεΐνες θα παράγονταν σε πολύ μεγαλύτερες ποσότητες από αυτές που χρειάζονται ενώ άλλες σε ανεπαρκείς ποσότητες.

Τα κύτταρα ενός πολυκύτταρου οργανισμού (όπως ο άνθρωπος) διαφέρουν τόσο στη δομή όσο και στη λειτουργία τους. Η ζωή ξεκινάει, όταν ένα γονιμοποιημένο ωάριο διαιρείται με μίτωση, παράγοντας έτσι τρισεκατομμύρια κύτταρα τα οποία φέρουν τα ίδια γονίδια. Τα ανθρώπινα κύτταρα, από τα αρχικά κίβλας στάδια της

εμβρυογένεσης, ακολουθούν μια διαδικασία μέσω της οποίας εξειδικεύονται στο να εκτελούν επιμέρους λειτουργίες. Η διαδικασία αυτή είναι γνωστή ως κυτταρική διαφοροποίηση. Για παράδειγμα, τα κύτταρα ενός πολύπλοκου πολυκύτταρου οργανισμού, όπως αυτά του μυϊκού και του νευρικού συστήματος (καθώς και όλα τα υπόλοιπα), έχουν όλα το ίδιο γενετικό υλικό, οπότε και τα ίδια γονίδια. Διαφέρουν όμως στη μορφή και στη λειτουργία τους. Όλα τα κύτταρα υπακούουν στις ίδιες γενετικές οδηγίες. Έχουν όμως αναπτύξει μηχανισμούς μέσω των οποίων εκφράζουν τη γενετική τους πληροφορία επιλεκτικά, ακολουθώντας τις απαραίτητες οδηγίες για τις διάφορες χρονικές στιγμές. Κάθε κυτταρικός τύπος εκτελεί εξειδικευμένες λειτουργίες οι οποίες πρέπει να συνυπάρχουν σε πλήρη συντονισμό με τις λειτουργίες των υπολοίπων κυττάρων. Πρέπει δηλαδή όλα τα κύτταρα να ακολουθούν πιστά τις λειτουργίες που προβλέπει γι' αυτά ο αλγόριθμος της ζωής. Σε περίπτωση που τα κύτταρα «βγουν» από το αυστηρό πρόγραμμα που τους επιβάλλει ο αλγόριθμος, λαμβάνει χώρα η μεταστροφή τους σε καρκινικά.

2.7.1 Η γονιδιακή ρύθμιση στους ευκαρυωτικούς οργανισμούς

Η ρύθμιση της έκφρασης των γονιδίων στα ευκαρυωτικά κύτταρα πραγματοποιείται σε πολλά επίπεδα με ιδιαίτερα πολύπλοκους μηχανισμούς, αποτελώντας αντικείμενο συνεχούς μελέτης. Η γονιδιακή έκφραση ρυθμίζεται στα ακόλουθα επίπεδα:

- **Στο επίπεδο της μεταγραφής:** Συγκεκριμένοι μηχανισμοί ελέγχουν ποια γονίδια θα μεταγραφούν ή/και με ποια ταχύτητα θα εκτελεστεί η μεταγραφή. Κάθε γονίδιο έχει το δικό του υποκινητή και μεταγράφεται αυτόνομα.
- **Στο επίπεδο μετά τη μεταγραφή:** Περιλαμβάνονται οι μηχανισμοί που χρησιμοποιούνται για την ωρίμανση του πρόδρομου mRNA και παράλληλα καθορίζεται η ταχύτητα με την οποία το mRNA ελευθερώνεται στο κυτταρόπλασμα.
- **Στο επίπεδο της μετάφρασης:** Το χρονικό διάστημα ζωής των mRNA μορίων στο κυτταρόπλασμα ποικίλλει για τα διάφορα είδη RNA, διότι

αυτά αποικοδομούνται μετά από κάποια χρονική περίοδο. Επίσης, ποικίλλει και η ικανότητα πρόσδεσης του mRNA στα ριβοσώματα.

- **Στο επίπεδο μετά τη μετάφραση:** Η πρωτεΐνη που παράχθηκε στο πέρας της πρωτεϊνοσύνθεσης, ίσως κριθεί απαραίτητο να υποστεί τροποποιήσεις, ώστε να γίνει βιολογικά λειτουργική.

Κεφάλαιο 3^ο: Η αλληλούχιση του γενετικού υλικού του ανθρώπου (DNA Sequencing)

3.1 Εισαγωγή στην αλληλούχιση του γενετικού υλικού (DNA Sequencing)

Με τον όρο αλληλούχιση του γενετικού υλικού εννοείται η διαδικασία προσδιορισμού της ακριβούς σειράς των νουκλεοτιδίων, ή αλλιώς των αζωτούχων βάσεων, που απαρτίζουν το DNA ενός οργανισμού. Το ανθρώπινο γονιδίωμα είναι κατασκευασμένο από περισσότερα από τρία δισεκατομμύρια από αυτά τα γενετικά «γράμματα». Τη σήμερον ημέρα, η αλληλούχιση του γενετικού υλικού πραγματοποιείται σε μεγάλη κλίμακα από μηχανήματα υψηλής τεχνολογίας. Τα μηχανήματα αυτά είναι ικανά να διαβάζουν μια ακολουθία γενετικού υλικού όπως ο άνθρωπος είναι ικανός να διαβάζει μια πρόταση. Η αλληλούχιση ενός γονιδιώματος συνήθως αποκαλείται και αποκωδικοποίηση, αφού η αλληλουχία των βάσεων που προκύπτει αποτελεί μια μεγάλη σειρά από γράμματα μιας μυστήριας «γλώσσας». Όταν διαβάζουμε ένα κείμενο, το νόημα δεν βρίσκεται μόνο στην ακολουθία των γραμμάτων. Βρίσκεται στις λέξεις που δημιουργούν τα γράμματα αυτά καθώς και στη γραμματική της εκάστοτε γλώσσας (σημεία στίξης, παράγραφοι κλπ). Παρομοίως, το ανθρώπινο γονιδίωμα είναι κάτι περισσότερο από μια αλληλουχία. Έτσι, αντιλαμβανόμαστε ότι η αλληλούχιση αυτή κάθε αυτή, δε μας αποκαλύπτει από μόνη της τα γενετικά μυστικά ενός ολόκληρου είδους. Απομένει πολλή δουλειά από τους επιστήμονες. Θα πρέπει να μεταφράσουν αυτές τις σειρές γραμμάτων σε μια κατανοητή διαδικασία λειτουργίας του γονιδιώματος: τι κάνουν τα γονίδια που απαρτίζουν το γονιδίωμα, πώς σχετίζονται τα διάφορα γονίδια μεταξύ τους καθώς και πώς συντονίζονται τα διάφορα κομμάτια του γονιδιώματος τα οποία βοηθούν τόσο στην ανάπτυξη όσο και στη συντήρηση του ανθρωπίνου οργανισμού.

Η γνώση των αλληλουχιών DNA έχει πλέον καταστεί απαραίτητη για τη βασική βιολογική έρευνα. Φυσικά η γνώση αυτή δεν συνεισφέρει μόνο σ' αυτόν τον τομέα. Η

χρήση της είναι εξίσου απαραίτητη και σε πολλούς εφαρμοσμένους τομείς όπως είναι η ιατρική διάγνωση, η βιοτεχνολογία, η ιατροδικαστική βιολογία και η ιολογία. Η μεγάλη ταχύτητα αλληλούχισης, η οποία έχει επιτευχθεί με τις σύγχρονες τεχνολογίες αλληλούχισης DNA, έχει συμβάλει στην αλληλούχιση ολόκληρων γονιδιωμάτων πολλών τύπων και ειδών ζωής, συμπεριλαμβανομένου και αυτού του ανθρωπίνου γονιδιώματος. Ποια είναι όμως η διαδικασία της λεγόμενης αλληλούχισης του γενετικού υλικού και κατ' επέκταση του γονιδιώματος;

3.2 Η διαδικασία της αλληλούχισης του γονιδιώματος μέσα από την αλληλούχιση του γενετικού υλικού

Η διαδικασία της αλληλούχισης ολόκληρου του ανθρωπίνου γονιδιώματος δεν πραγματοποιήθηκε σε μία φάση διότι οι μέθοδοι αλληλούχισης του γενετικού υλικού περιορίζονται σε μικρά κομμάτια DNA. Επομένως, οι επιστήμονες έπρεπε να σπάσουν το γονιδίωμα σε μικρά κομμάτια (γενετικού υλικού), να αλληλουχίσουν τα κομμάτια αυτά και στη συνέχεια να τα επανασυναρμολογήσουν στη σωστή σειρά ώστε να φθάσουν τελικά στην αλληλούχιση του συνόλου του γονιδιώματος. Μεγάλο κομμάτι της διαδικασίας της αλληλούχισης απετέλεσε η επανακατασκευή αυτού του βιολογικού παζλ.

Υπάρχουν δύο προσεγγίσεις για την κοπή του γονιδιώματος σε κομμάτια και για την μετέπειτα επανασυναρμολόγησή του. Η μία στρατηγική είναι γνωστή ως “*clone-by-clone*”. Κατά την προσέγγιση αυτή, δημιουργείται ένα χάρτης για κάθε χρωμόσωμα του γονιδιώματος πριν χωριστεί το DNA σε κομμάτια έτοιμα για αλληλούχιση. Το γονιδίωμα σπάει σε μεγάλα τμήματα με μήκος περίπου 150.000 ζεύγη βάσεων (*BP*). Η θέση αυτών των κομματιών στο χρωμόσωμα χαρτογραφείται για να γίνει εφικτή η διαδικασία επανακατασκευής του αρχικού κλώνου στη σωστή σειρά (μετά το πέρας της αλληλούχισης). Τα κομμάτια αυτά τοποθετούνται μέσα σε *BACs* (*Bacterial Artificial Chromosomes*) τα οποία τελικά θα τοποθετηθούν σε βακτηριακά κύτταρα ώστε να αναπτυχθούν. Τα τμήματα του γενετικού υλικού αντιγράφονται κάθε φορά που τα βακτήρια αναπαράγονται (αυξάνονται σε αριθμό) παράγοντας έτσι πολλά όμοια αντίγραφα DNA. Στη συνέχεια, το DNA στους

επιμέρους βακτηριακούς κλώνους κόβεται σε ακόμη μικρότερα κομμάτια. Κάθε κομμάτι έχει μήκος 500 ζεύγη βάσεων για να είναι διαχειρίσιμο κατά τη διαδικασία της αλληλούχισης. Ακολούθως, τα κομμάτια τοποθετούνται μέσα σε έναν φορέα του οποίου η αλληλουχία είναι γνωστή (όπως αυτή των βακτηρίων). Αμέσως μετά, υπόκεινται αλληλούχιση, ξεκινώντας με τη γνωστή αλληλουχία του φορέα και μελετώντας την άγνωστη αλληλουχία DNA που υπάρχει. Συνεχίζοντας της αλληλούχιση, τα μικρά κομμάτια του γενετικού υλικού ενώνονται δημιουργώντας τα μεγάλα κομμάτια που είχαν τοποθετηθεί μέσα στα BACs. Αυτή η επανένωση πραγματοποιείται με τη βοήθεια λογισμικού που βρίσκεται σε ηλεκτρονικούς υπολογιστές οι οποίοι γνωρίζουν τα κοινά τμήματα (κοινές αλληλουχίες στα διάφορα χρωμοσώματα). Τελικά, ακολουθώντας την αρχική χαρτογράφηση, τα μεγάλα κομμάτια επανασυναρμολογούνται σε χρωμοσώματα σηματοδοτώντας την ολοκλήρωση της αλληλούχισης του γονιδίου.

Η δεύτερη στρατηγική είναι γνωστή ως *“whole-genome shotgun”*. Η μέθοδος αυτή περιλαμβάνει τυχαία διάσπαση αλληλουχιών DNA σε πολλά μικρά τμήματα τα οποία επανασυναρμολογούνται στην αρχική αλληλουχία κοιτώντας πάντα τα κοινά σημεία (βάσει αλληλουχιών). Το γονιδίωμα ενός μεγάλου οργανισμού, όπως είναι ο άνθρωπος, είναι ιδιαίτερα δύσκολο να κλωνοποιηθεί, να αλληλουχιθεί και να επανασυναρμολογηθεί εξαιτίας του μεγέθους του και της πολυπλοκότητας της κατασκευής του. Έτσι, η μέθοδος *“clone-by-clone”*, παρά την μεθοδικότητα και την αξιοπιστία που προσφέρει, καθίσταται χρονοβόρα για περιπτώσεις σαν αυτή του ανθρωπίνου γονιδιώματος. Με την εμφάνιση της «φθηνής» αλληλούχισης καθώς και με την εξέλιξη των προγραμμάτων των ηλεκτρονικών υπολογιστών, οι ερευνητές βασίζονται στην μέθοδο *“whole-genome shotgun”* για να αντιμετωπίσουν μεγαλύτερα και πιο πολύπλοκα γονίδια. Η μέθοδος αυτή χρησιμοποιήθηκε επίσημα από τον Fred Sanger και τους συνεργάτες του για την αλληλούχιση μικρών γονιδιωμάτων όπως είναι αυτά των ιών και των βακτηρίων. Με τη χρήση της, παρακάμπτουμε την χρονοβόρα χαρτογράφηση και την κλωνοποίηση που συναντάμε στη μέθοδο *“clone-by-clone”*. Αρχικά, ολόκληρο το γονιδίωμα διασπάται σε μικρά τμήματα DNA τα οποία είναι έτοιμα για αλληλούχιση. Τα τμήματα αυτά φέρουν ποικίλα μεγέθη τα οποία κυμαίνονται από 2-20 kilobases (δηλαδή 2.000-20.000 BP) έως 200-300 kilobases

(200.000-300.000 BP). Τα τμήματα αυτά υπόκεινται σε αλληλούχιση με σκοπό τον προσδιορισμό της σειράς των βάσεων του γενετικού υλικού. Στη συνέχεια, συναρμολογούνται με τη βοήθεια προγραμμάτων ηλεκτρονικών υπολογιστών τα οποία μπορούν να βρουν τις κοινές αλληλουχίες. Μπορούμε να φανταστούμε την εν λόγω μέθοδο σαν να τεμαχίζουμε πολλαπλά αντίγραφα ενός βιβλίου (στην περίπτωση μας αυτό είναι το γονιδίωμα), να ανακατεύουμε τα κομμάτια αυτά μεταξύ τους και μετά να συναρμολογούμε πάλι το αρχικό κείμενο (γονιδίωμα) βρίσκοντας κομμάτια κειμένου που είναι κοινά (επικαλυπτόμενα), ενώνοντας τα κομμάτια αυτά ώστε να δημιουργηθεί το πρωτότυπο βιβλίο.

Η μέθοδος “*whole-genome shotgun*” χρησιμοποιήθηκε επίσης από τον Craig Venter, ιδρυτή της ιδιωτικής εταιρίας Celera Genomics, για την αλληλούχιση του γονιδιώματος του ανθρώπου. Ο Venter ήθελε να πραγματοποιήσει την διαδικασία της αλληλούχισης ταχύτερα από τη δημόσια χρηματοδοτούμενη προσπάθεια και πίστευε ότι η μέθοδος αυτή ήταν ο καλύτερος τρόπος για να πετύχει το σκοπό του. Για τη συναρμολόγηση της αλληλουχίας, ο Venter χρησιμοποίησε τα δημοσίως διαθέσιμα στοιχεία της μεθόδου “*clone-by-clone*” από το Human Genome Project. Σήμερα, με την εξέλιξη της τεχνολογίας, η μέθοδος “*whole-genome shotgun*” χρησιμοποιείται για να βελτιωθεί η ακρίβεια των υπαρχόντων αλληλουχιών γονιδιωμάτων (όπως του ανθρωπίνου). Η αποσφαλμάτωση και το γέμισμα των κενών είναι μερικές από τις χρήσεις της μεθόδου για τις αλληλουχίες που δεν συναρμολογήθηκαν σωστά (παρά τη χρήση της “*clone-by-clone*”).

Κάθε μία προσέγγιση έχει τα θετικά και τα αρνητικά της. Για παράδειγμα, η μέθοδος “*clone-by-clone*” είναι αξιόπιστη αλλά αργή, και το σημείο της χαρτογράφησης μπορεί να αποδειχτεί χρονοβόρο. Αντίθετα, η μέθοδος “*whole-genome shotgun*” είναι δυνητικά γρηγορότερη αλλά είναι εξαιρετικά δύσκολο να συναρμολογηθούν τόσα μικρά κομμάτια αλληλούχισης μονομιάς. Παρακάτω θα εξετάσουμε πώς ακριβώς πραγματοποιείται η αλληλούχιση του γενετικού υλικού μέσω της οποίας οδηγούμαστε τελικά στην αλληλούχιση του γονιδιώματος (εφόσον το επιθυμούμε).

Στη διαδικασία της αλληλούχισης του DNA χρησιμοποιείται μια τεχνική γνωστή ως ηλεκτροφόρηση μέσω της οποίας διαχωρίζουμε κομμάτια DNA τα οποία διαφέρουν σε μήκος κατά μία μόνο βάση. Στην ηλεκτροφόρηση, το γενετικό υλικό το οποίο πρόκειται να υποστεί αλληλούχιση τοποθετείται στην άκρη ενός τζελ (μιας γέλης), μιας «πλάκας» η υφή της οποίας μοιάζει με ζελατίνη. Αργότερα, τοποθετούνται ηλεκτρόδια σε κάθε άκρο αυτού του τζελ δημιουργώντας ρεύμα (κύκλωμα) κάνοντας τα μόρια του γενετικού υλικού να κινούνται μέσα στο τζελ. Τα μικρότερα μόρια κινούνται γρηγορότερα μέσα στο υλικό με αποτέλεσμα τα μόρια του DNA να διαχωρίζονται σε ζώνες ανάλογα με το μέγεθός τους. Το πρόβλημα είναι ότι με τη διαδικασία της ηλεκτροφόρησης μπορούμε να διαχωρίσουμε περίπου 500 βάσεις σε σαφείς ζώνες. Μέχρι και το πέρας του 1980, τα στοιχεία που προέκυπταν από το τζελ ηλεκτροφόρησης έπρεπε στο τέλος, μετά από μεγάλη προσπάθεια, να διαβαστούν από κάποιον. Κάθε κομμάτι του γενετικού υλικού συναπτόταν σε μία ραδιενεργή πλακέτα και αμέσως μετά κατασκευαζόταν μια εικόνα του τζελ με τη βοήθεια ακτίνων X η οποία βοηθούσε στο να γίνουν ευδιάκριτες οι θέσεις των ζωνών του DNA. Μέσω της ανάλυσης των γραμμών και των στηλών των ζωνών του γενετικού υλικού, ο επιστήμονας μπορούσε τελικά να προσδιορίσει την αλληλουχία του DNA. Όπως γίνεται αντιληπτό, η διαδικασία αυτή ήταν αργή, επίπονη και συχνά γεμάτη με σφάλματα.

Σήμερα, τα εγχειρήματα αλληλούχισης DNA ευρείας κλίμακας που εκτελούνται, θα ήταν αδύνατο να υλοποιηθούν χωρίς τα μηχανήματα αυτόματης αλληλούχισης DNA. Τα μηχανήματα αυτά άρχισαν να είναι εμπορικά διαθέσιμα στα τέλη του 1980 και έχουν κάνει τη διαδικασία αλληλούχισης πολύ πιο γρήγορη και πολύ πιο αξιόπιστη. Για να καταλάβουμε την ταχύτητα αλληλούχισης αρκεί να σκεφτούμε ότι ένας άνθρωπος μπορεί να παράγει μια ολοκληρωμένη αλληλουχία αποτελούμενη από 20.000-50.000 βάσεις σε διάστημα ίσο του ενός χρόνου. Πλέον, ένα μηχάνημα μπορεί να εκτελέσει την ίδια διαδικασία σε διάστημα λίγων ωρών. Τα περισσότερα μηχανήματα αυτόματης αλληλούχισης είναι σχεδιασμένα με τρόπο βασισμένο στην πρωτότυπη, χειροκίνητη διαδικασία αλληλούχισης. Για τη χρήση του μηχανήματος, ένας τεχνικός τοποθετεί τζελ ανάμεσα σε δύο γυάλινες πλάκες οι οποίες απέχουν μεταξύ τους λιγότερο από μισό χιλιοστό. Αμέσως μετά, το γενετικό υλικό τοποθετείται

σε κάθε μια από τις διαδρομές που έχουν σχηματιστεί. Οι διαδρομές αυτές έχουν μήκος περίπου 30 εκατοστά. Καθώς τα κομμάτια του DNA κινούνται μέσα στο τζελ, το μηχάνημα διαβάζει τη σειρά των βάσεων του DNA και αποθηκεύει την πληροφορία αυτή στη μνήμη του υπολογιστή του.

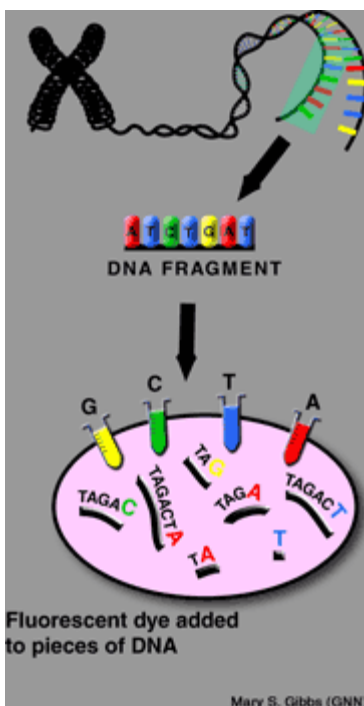
Σε κάποιες ακόμη πιο νέες μηχανές, γνωστές και ως *capillary*⁷ sequencers, το γενετικό υλικό κινείται πλέον σε μια συστοιχία από πολύ λεπτούς, γυάλινους σωλήνες γεμάτους με τζελ. Όπως και στα μηχανήματα με τις γυάλινες πλάκες, οι *capillary sequencers* διαβάζουν της αλληλουχία των βάσεων καθώς το γενετικό υλικό κινείται μέσα στο τζελ. Η ταχύτητα με την οποία εκτελούν τη διαδικασία αλληλούχισης είναι διπλάσια από την ταχύτητα των μηχανημάτων που φέρουν γυάλινες πλάκες. Επίσης, είναι εντελώς αυτόνομα αφού η διαδικασία τοποθέτησης του γενετικού υλικού μέσα στους σωλήνες γίνεται με τη χρήση ρομποτικού βραχίονα. Ακόμη, οι σωλήνες του μηχανήματος γεμίζουν αυτόματα με τζελ ενώ η ίδια η μηχανή αυτοκαθαρίζεται ώστε να είναι έτοιμη για την επόμενη αλληλούχιση. Αυτές οι ιδιότητες καθιστούν την ανθρώπινη παρέμβαση ελάχιστη, αφού είναι απαραίτητη μόνο στις περιπτώσεις γεμίσματος της μηχανής με τζελ, νερό ή οποιοδήποτε άλλο διάλυμα κρίνεται απαραίτητο για αυτήν ώστε να λειτουργήσει. Αξίζει να σημειωθεί ότι στα μεγάλα έργα αλληλούχισης γενετικού υλικού χρησιμοποιούνται συνδυαστικά και τα δύο προαναφερθέντα μηχανήματα. Πώς όμως μια μηχανή αλληλούχισης γνωρίζει πότε μια βάση αποτελεί αδενίνη, κυτοσίνη, γουανίνη ή θυμίνη;



Εικόνα 3.1: Οι σωλήνες *Capillary*.

Οι μηχανές αλληλούχισης δε μπορούν να διαβάσουν τις αζωτούχες βάσεις άμεσα. Έτσι, οι επιστήμονες πρέπει πρώτα να εκτελέσουν ένα περίπλοκο σύνολο διαδικασιών ώστε το γενετικό υλικό να προετοιμαστεί για αλληλούχιση. Πριν υποστεί αλληλούχιση, ένα κομμάτι DNA αντιγράφεται αρκετές φορές, ενώ στη συνέχεια χωρίζεται σε τέσσερις ομάδες (δέσμες) με σκοπό να προετοιμαστεί για περαιτέρω αντιγραφή. Στο δεύτερο πέρασμα, μικρή ποσότητα από μια χημικά τροποποιημένη αζωτούχα βάση προστίθεται σε κάθε ομάδα. Δηλαδή, στη μία ομάδα τοποθετούμε μια

⁷ Ονομάζονται έτσι διότι το μέγεθός τους είναι όμοιο με το μέγεθος μιας ανθρώπινης τρίχας.



Εικόνα 3.2: Οι ομάδες DNA του παραδείγματος.

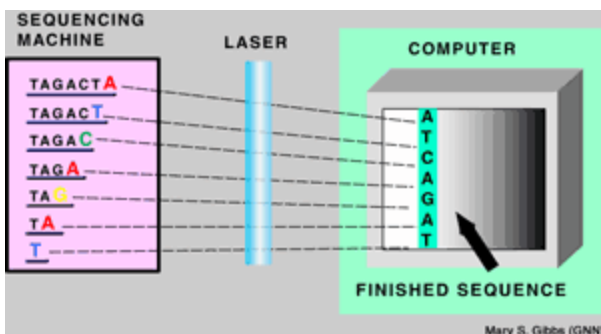
χημικά τροποποιημένη θυμίνη, στην άλλη μια χημικά τροποποιημένη αδενίνη και ούτω καθεξής. Η ενσωμάτωση μιας εκ των χημικά τροποποιημένων αζωτούχων βάσεων σε ένα μόριο DNA έχει σαν αποτέλεσμα τη διακοπή της ανάπτυξης της αλυσίδας των βάσεων. Τελικά, μια ομάδα DNA θα περιέχει τα κομμάτια του γενετικού υλικού που τελειώνουν σε T, μια άλλη θα περιέχει τα κομμάτια που τελειώνουν σε A, η επόμενη θα περιέχει τα κομμάτια που τελειώνουν σε G και η τέταρτη θα περιέχει τα κομμάτια που τελειώνουν σε C. Στη δεύτερη φάση αντιγραφής, προστίθεται μια διαφορετική φθορίζουσα χρωστική σε κάθε ομάδα DNA. Έτσι, κάθε κομμάτι γενετικού υλικού που τελειώνει σε T έχει ξεχωριστό χρώμα από τα κομμάτια που τελειώνουν με μία εκ των υπολοίπων τριών βάσεων. Το παρακάτω

παραδείγμα θα μας βοηθήσει να καταλάβουμε τη διαδικασία καλύτερα.

Ας υποθέσουμε ότι εκτελούμε την παραπάνω διαδικασία στη αλληλουχία TAGACT. Στο τέλος της δεύτερης φάσης της αντιγραφής κάθε ομάδα θα περιέχει τα ακόλουθα κομμάτια γενετικού υλικού στα οποία θα δώσουμε τα ακόλουθα χρώματα (http://www.genomenetwork.org/resources/whats_a_genome/Chp2_2.shtml):

- T → μπλε χρώμα, TAGACT → μπλε χρώμα.
- TA → κόκκινο χρώμα, TAGA → κόκκινο χρώμα.
- TAG → κίτρινο χρώμα.
- TAGAC → πράσινο χρώμα.

Σε ένα σωληνάριο μιας μηχανής αλληλούχισης μεταβαίνει γενετικό υλικό και από τις τέσσερις ομάδες. Επειδή, όπως προαναφέραμε, τα μικρότερα μόρια DNA μετακινούνται μέσα στο τζελ πιο γρήγορα, τα κομμάτια γενετικού υλικού που διασχίζουν το τζελ είναι όλο και πιο μεγάλα σε μήκος. Για την ακρίβεια είναι μεγαλύτερα κατά μία αζωτούχα βάση. Στο συγκεκριμένο παράδειγμα, το πρώτο κομμάτι που διασχίζει το τζελ είναι μια θυμίνη (T → μπλε). Το επόμενο κομμάτι είναι η



Εικόνα 3.3: Η διαδικασία της ανίχνευσης των βάσεων.

αλληλουχία TA → κόκκινο. Αμέσως μετά ακολουθεί η αλληλουχία TAG → κίτρινο. Η επόμενη ακολουθία είναι η TAGA → κόκκινο και ούτω καθεξής. Καθώς τα τμήματα γενετικού υλικού αναδύονται από το τζελ, περνάνε από ένα λέιζερ το οποίο κάνει τα μόρια που φέρουν χρωστική να φθορίζουν. Ένας ανιχνευτής διαβάζει το χρώμα του κάθε φθορισμού ενώ αμέσως μετά, ένα ειδικό λογισμικό αναλαμβάνει να ταιριάξει το χρώμα με την αντίστοιχη αζωτούχα βάση. Με τον τρόπο αυτό, η αλληλουχία μεγαλώνει βάση-βάση. Κάθε ακολουθία που παράγεται από τη μηχανή αλληλούχισης και αποτελείται από περίπου 500 βάσεις, ονομάζεται “read”.

Κατά το πέρας της διαδικασίας αλληλούχισης, εξέρχεται από τη μηχανή αυτό που οι επιστήμονες ονομάζουν “raw” sequence (ακατέργαστη αλληλουχία). Στην “raw” sequence, τα “reads” είναι όλα μπερδεμένα μεταξύ τους όπως είναι τα κομμάτια ενός παζλ του οποίου το κουτί μόλις ανοίχτηκε. Αναπόφευκτα, η “raw” sequence περιέχει κάποια κενά καθώς και κάποιες ασάφειες. Ακολουθεί η διαδικασία τερματισμού, η οποία αναλαμβάνει να συναρμολογήσει τις μικρές αλληλουχίες σε μια μεγάλη, συνεχή αλληλουχία χωρίς κενά και λάθη. Η διαδικασία τερματισμού περιλαμβάνει τόσο τη συναρμολόγηση, όπου τα επιμέρους “reads” ενώνονται με τη σωστή σειρά σε μια συνεχή αλληλουχία, όσο και την εκτέλεση ενός διπλού ελέγχου ο οποίος αποσκοπεί στην εξάλειψη λαθών και κενών. Η διαδικασία τερματισμού συνήθως διαρκεί περισσότερο και από την ίδια την αλληλούχιση του γενετικού υλικού.

3.3 Η συμβολή της Πληροφορικής στην επιστήμη της Βιοπληροφορικής

Η χρήση των ηλεκτρονικών υπολογιστών και της Πληροφορικής δίνει στους επιστήμονες πληθώρα δυνατοτήτων όσον αφορά τη διαχείριση των αλληλουχιών. Η σημαντικότερη εξ αυτών είναι η στοίχιση (alignment) ή αλλιώς σύγκριση (ψάχνοντας την ομοιότητα) των διαφόρων βιολογικών αλληλουχιών. Πριν αναφερθούμε σε αυτή

τη διαδικασία όμως, θα πρέπει να βρούμε έναν τρόπο ώστε να συνδέσουμε το κομμάτι της βιολογίας με αυτό της πληροφορικής.

3.3.1 Αποθήκευση των αλληλουχιών σε ηλεκτρονικό υπολογιστή

Η πληροφορία μιας αλληλουχίας αποθηκεύεται στον ηλεκτρονικό υπολογιστή σαν ένα σύνολο από απλές γραμμές αποτελούμενες από χαρακτήρες οι οποίες ονομάζονται συμβολοσειρές (strings). Οι συμβολοσειρές αυτές είναι παρόμοιες με τις συμβολοσειρές που εμφανίζονται στο τερματικό (οθόνη) ενός υπολογιστή (όπως εμφανίζεται για παράδειγμα ένα κείμενο). Κάθε χαρακτήρας αποθηκεύεται σε δυαδικό κώδικα στη μικρότερη μονάδα μνήμης, το *byte*. Κάθε *byte* αποτελείται από 8 *bit*, με κάθε *bit* να μπορεί να πάρει τιμή 0 ή 1, παράγοντας έτσι 256 πιθανούς συνδυασμούς. Καθένας από τους συνδυασμούς έχει το δικό του ισοδύναμο στον πίνακα ASCII. Για παράδειγμα, κάποιες ASCII τιμές χαρακτηρίζονται ως *keyboard characters* ενώ άλλες ως *special control characters*, όπως είναι ο *end-of-file character* ο οποίος σηματοδοτεί το τέλος ενός αρχείου κειμένου. Ένα αρχείο αποτελούμενο μόνο από χαρακτήρες ASCII ονομάζεται αρχείο ASCII (ASCII file). Για λόγους ευκολίας, όλες οι δυαδικές τιμές μπορούν να γραφούν σε δεκαεξαδική μορφή, η οποία αντιστοιχεί στη δική μας δεκαδική μορφή με τους αριθμούς από 0 έως 9 συν τα γράμματα του λατινικού αλφαβήτου από A έως F. Συνεπώς, το δεκαεξαδικό *0F* αντιστοιχεί στο δυαδικό *0000 1111* και το δεκαδικό *15*, και το *FF* αντιστοιχεί στο δυαδικό *1111 1111* και το δεκαδικό *255*. Συμπερασματικά, μια αλληλουχία DNA αποθηκεύεται και διαβάζεται σαν μια σειρά *8-bit* λέξεων στην παραπάνω δυαδική μορφή.

Προτού χρησιμοποιηθεί ένα αρχείο αλληλουχίας (sequence file) σε ένα πρόγραμμα ανάλυσης αλληλουχιών, είναι αναγκαίο να βεβαιωθούμε ότι το αρχείο αυτό περιέχει μόνο χαρακτήρες αλληλουχίας χωρίς επιπλέον ειδικούς χαρακτήρες (special characters) που χρησιμοποιούνται από τους διάφορους επεξεργαστές κειμένου (text editors). Ένας περιττός χαρακτήρας μπορεί να δημιουργήσει πρόβλημα στη συνοχή ενός sequence file. Για τον λόγο αυτόν, οι διάφοροι επεξεργαστές κειμένου παρέχουν τη δυνατότητα αποθήκευσης αρχείων τα οποία αποτελούνται μόνο από ASCII χαρακτήρες. Μόνο τέτοιου είδους αρχεία είναι κατάλληλα για επεξεργασία από τα διάφορα προγράμματα ανάλυσης αλληλουχιών. Επίσης, τα

αρχεία αλληλούχισης που δεν περιέχουν ASCII χαρακτήρες υπάρχει περίπτωση να μην μπορούν να μεταφερθούν σωστά από μηχανή σε μηχανή (υπολογιστή) προκαλώντας απρόβλεπτη συμπεριφορά του λογισμικού επικοινωνίας μεταξύ των δύο μηχανών. Κάποια από τα λογισμικά επικοινωνίας μπορούν να ρυθμιστούν κατάλληλα ώστε να αγνοούν τα όποια λάθη. Ένα τέτοιο παράδειγμα αποτελεί το πρωτόκολλο FTP.

Έχοντας εξετάσει τον τρόπο σύνδεσης της βιολογίας με την πληροφορική μπορούμε πλέον να ερευνήσουμε την διαδικασία στοίχισης των διαφόρων βιολογικών αλληλουχιών.

3.3.2 Στοίχιση αλληλουχιών

Με τον όρο Στοίχιση Αλληλουχιών εννοείται η διαδικασία σύγκρισης δύο (pair-wise alignment) ή περισσότερων (multiple sequence alignment) αλληλουχιών στις οποίες αναζητάμε μια ακολουθία από ατομικούς χαρακτήρες ή από πρότυπα χαρακτήρων οι οποίοι εντοπίζονται με την ίδια σειρά στις συγκρινόμενες αλληλουχίες. Για να γίνει πιο κατανοητό, φανταστείτε μια σελίδα όπου οι συγκρινόμενες αλληλουχίες είναι γραμμένες η μία κάτω από την άλλη σε ξεχωριστές γραμμές (rows). Πανομοιότυποι ή παρόμοιοι χαρακτήρες τοποθετούνται στην ίδια στήλη (column) ενώ οι ανόμοιοι χαρακτήρες μπορούν είτε να τοποθετηθούν στη στήλη που ήδη βρίσκονται, υποδεικνύοντας την αναντιστοιχία (mismatch), ή να τοποθετηθούν απέναντι από ένα κενό (gap) που υπάρχει στην άλλη αλληλουχία. Σε μια βέλτιστη ευθυγράμμιση (στοίχιση), οι ανόμοιοι χαρακτήρες και τα κενά τοποθετούνται με τέτοιο τρόπο ώστε να προκύψουν όσο το δυνατόν περισσότερα ταιριάσματα χαρακτήρων στις στήλες μεταξύ των δύο αλληλουχιών. Η διαδικασία της ευθυγράμμισης στηρίζεται σε διάφορους πίνακες (score matrices) οι οποίοι βαθμολογούν τις ομοιότητες (matches), τις διαφορές και τα κενά μεταξύ δύο διαδοχικών συμβόλων. Οι αλληλουχίες που ευθυγραμμίζονται άμεσα με αυτόν τον τρόπο μπορούμε να πούμε ότι είναι όμοιες.

3.3.3 Στοίχιση ζεύγους αλληλουχιών (Pair – Wise Alignment)

Η στοίχιση δύο αλληλουχιών πραγματοποιείται με τις ακόλουθες μεθόδους:

- Dot matrix analysis.
- The dynamic programming (DP) algorithm.
- Word or k-tuple method. Η μέθοδος αυτή χρησιμοποιείται κυρίως από προγράμματα τύπου FASTA και BLAST τα οποία σχετίζονται με βάσεις δεδομένων. Θα αναφερθούμε σε αυτή στο επόμενο κεφάλαιο.

3.3.3.1 Dot matrix analysis

Η μέθοδος η οποία πρέπει να χρησιμοποιείται πρώτα για την ευθυγράμμιση δύο αλληλουχιών, εκτός και αν οι αλληλουχίες είναι γνωστό ότι είναι όμοιες, είναι η μέθοδος dot matrix. Η μέθοδος αυτή αναπαριστά κάθε πιθανή στοίχιση σαν μια διαγώνιο στον πίνακα που δημιουργείται μεταξύ των συγκρινόμενων αλληλουχιών. Η ίδια μέθοδος μπορεί εύκολα να εντοπίσει τις όποιες εισαγωγές/διαγραφές καθώς και τις άμεσες ή τις ανεστραμμένες επαναλήψεις στις διάφορες αλληλουχίες.

Για να εκτελέσουμε τη διαδικασία, η μία αλληλουχία (ας την ονομάσουμε αλληλουχία X) γράφεται οριζόντια στην κορυφή (είτε αυτό είναι οθόνη είτε είναι χαρτί) ενώ η δεύτερη (ας την ονομάσουμε αλληλουχία Y) γράφεται κάθετα στο αριστερό μέρος. Ξεκινάμε με τον πρώτο χαρακτήρα της αλληλουχίας Y και κινούμαστε οριζόντια στην ίδια γραμμή τοποθετώντας μια τελεία στις στήλες στις οποίες ο χαρακτήρας της Y είναι ίδιος με αυτόν της αλληλουχίας X (θυμίζει τους δισδιάστατους πίνακες της πληροφορικής). Συνεχίζουμε με τον δεύτερο χαρακτήρα της αλληλουχίας Y τον οποίο συγκρίνουμε με ολόκληρη την αλληλουχία X τοποθετώντας ξανά τελεία στις περιπτώσεις ομοιότητας. Η ίδια διαδικασία πραγματοποιείται μέχρι να γεμίσει η οθόνη, ή το χαρτί, τελείες οι οποίες αναπαριστούν όλα τα πιθανά matches μεταξύ των δύο αλληλουχιών. Κάθε περιοχή (region) στην οποία παρατηρείται ομοιότητα είναι εμφανής λόγω της διαγωνίου που έχει σχηματιστεί από τις τελείες. Οι τελείες που δεν βρίσκονται πάνω στη διαγώνιο που έχει σχηματιστεί υποδηλώνουν τυχαίες ομοιότητες οι οποίες συνήθως δεν είναι σημαντικές.

3.3.3.2 Dynamic programming algorithm

Ένας αλγόριθμος δυναμικού προγραμματισμού είναι μια εξαιρετικά σημαντική υπολογιστική μέθοδος ευθυγράμμισης αλληλουχιών διότι μας παρέχει την βέλτιστη ευθυγράμμιση μεταξύ δύο αλληλουχιών. Υπάρχουν διάφορα προγράμματα καθώς και

websites τα οποία βασίζονται στον δυναμικό προγραμματισμό για να εκτελέσουν μια τέτοιου είδους διαδικασία.

Η μέθοδος συγκρίνει κάθε ζευγάρι χαρακτήρων ανάμεσα σε δύο αλληλουχίες και παράγει τη βέλτιστη ευθυγράμμιση. Ο αλγόριθμος διαχειρίζεται τα matches, τα mismatches και τα gaps ώστε να πετύχει τον μεγαλύτερο αριθμό ομοιοτήτων μεταξύ των αλληλουχιών. Βασικό χαρακτηριστικό του αλγορίθμου είναι ότι «σπάει» ένα μεγάλο πρόβλημα (το οποίο θέλει πολλούς υπολογισμούς για να λυθεί) σε μικρότερα προβλήματα των οποίων η λύση επιτυγχάνεται ευκολότερα. Το άθροισμα αυτών των μικρών προβλημάτων δίνει τελικά τη λύση του μεγάλου. Η όλη διαδικασία σε γενικές γραμμές είναι η εξής: οι δύο αλληλουχίες, έστω A με n χαρακτήρες και B με m χαρακτήρες, τοποθετούνται σε έναν πίνακα $Table\ n*m$ με στοιχεία $Table(i,j)$. Κάθε στοιχείο του πίνακα είναι η τιμή του score για την καλύτερη στοίχιση μέχρι το στοιχείο A_i και B_j . Ουσιαστικά, χρησιμοποιούμε την μέθοδο dot matrix τοποθετώντας αριθμητικές τιμές αντί για τελείες. Γνωρίζοντας τα στοιχεία $Table(i-1,j-1)$, $Table(i-1,j)$ και $Table(i,j-1)$ μπορούμε κάθε φορά να υπολογίζουμε το στοιχείο $Table(i,j)$. Δηλαδή, γνωρίζοντας αυτά τα 3 γειτονικά στοιχεία μπορούμε να υπολογίζουμε κάθε φορά το στοιχείο $Table(i,j)$. Σημαντικό ρόλο στην εύρεση του εκάστοτε στοιχείου παίζουν επίσης τόσο η επιλογή του συστήματος του score όσο και η συνάρτηση ποινής για ένα πιθανό κενό. Προφανώς, κάθε μη διαγώνια μετάβαση σημαίνει εισαγωγή κενού σε μια από τις δύο αλληλουχίες. Έχοντας υπολογίσει όλα τα στοιχεία του πίνακα, μπορούμε κινούμενοι προς τα πίσω να βρούμε την καλύτερη δυνατή στοίχιση (με βάση το τελικό score) των δύο αλληλουχιών.

3.3.4 Πολλαπλή στοίχιση αλληλουχιών (Multiple Sequence Alignment)

Έχοντας καταστήσει εφικτή την ευθυγράμμιση δύο αλληλουχιών, εύλογα αναρωτιέται κανείς εάν θα μπορούσε να εκτελέσει κάποια διαδικασία για στοίχιση περισσότερων των δύο αλληλουχιών. Η διαδικασία αυτή επιτυγχάνεται ξανά με τη μέθοδο του δυναμικού προγραμματισμού. Η πολλαπλή στοίχιση χρησιμοποιείται τις περισσότερες φορές για ευθυγράμμιση πρωτεϊνικών αλληλουχιών. Παρ' όλα αυτά βρίσκει εφαρμογή και σε περιπτώσεις νουκλεοτιδικών αλληλουχιών. Όπως είδαμε προηγουμένως, στην περίπτωση των δύο αλληλουχιών, τόσο η στοίχιση όσο και η

οπτικοποίηση αποτελούν σχετικά εύκολα βήματα. Όσο όμως μεγαλώνει ο αριθμός των αλληλουχιών τόσο δυσκολεύουν τα πράγματα. Για παράδειγμα, στην περίπτωση ευθυγράμμισης 3 αλληλουχιών η πολλαπλή στοίχιση αντιστοιχεί στην εύρεση του βέλτιστου μονοπατιού σε έναν τρισδιάστατο πίνακα (κύβο) πλέον, όπου οι έδρες του αποτελούν τους πίνακες που αντιστοιχούν στις κατά ζεύγη στοιχίσεις των αλληλουχιών. Προφανώς, η προσθήκη όλο και περισσότερων αλληλουχιών στη διαδικασία της πολλαπλής στοίχισης απαιτεί και περισσότερες διαστάσεις στον πίνακα. Επίσης, η συνάρτηση του score, η οποία θα κρίνει ποια είναι η βέλτιστη στοίχιση με βάση πάντα τα matches, τα mismatches και τα gaps, θα πρέπει να προσαρμοστεί στην εκάστοτε περίπτωση. Ο πιο γνωστός τρόπος πολλαπλής στοίχισης αλληλουχιών είναι η στοίχιση κατά ζεύγη με όλους τους πιθανούς συνδυασμούς ζευγών. Και σε αυτήν την περίπτωση φυσικά βασιζόμαστε στον πίνακα που προκύπτει από την dot matrix analysis.

3.3.4.1 Προοδευτική πολλαπλή στοίχιση αλληλουχιών (Progressive Multiple Sequence Alignment)

Η πιο γνωστή ευριστική (heuristic) μέθοδος η οποία χρησιμοποιείται για πολλαπλή στοίχιση είναι η μέθοδος progressive multiple alignment. Σύμφωνα με αυτή τη μέθοδο, η στοίχιση των αλληλουχιών γίνεται προοδευτικά ξεκινώντας από δύο αλληλουχίες (αυτές που παρουσιάζουν τη μεγαλύτερη ομοιότητα) μεταξύ των οποίων γίνεται προσπάθεια να επιτευχθεί η βέλτιστη στοίχιση. Στη συνέχεια προσθέτουμε σταδιακά στη στοίχιση μία-μία τις υπόλοιπες αλληλουχίες. Αν και έχουν παρουσιαστεί πολλές παραλλαγές της μεθόδου από τις αρχές της δεκαετίας του 1980, η γενική ιδέα της προοδευτικής πολλαπλής στοίχισης όπως διατυπώθηκε από τους Feng και Doolittle το 1987 (Feng & Doolittle, 1987) περιλαμβάνει τα παρακάτω βήματα:

- Αρχικές κατά ζεύγη στοιχίσεις όλων των ακολουθιών.
- Δημιουργία πίνακα αποστάσεων και ενός δένδρου οδηγού (guide tree), όπου τα φύλλα θα είναι οι αλληλουχίες, με βάση τις κατά ζεύγη στοιχίσεις.
- Προοδευτική στοίχιση των πιο όμοιων αλληλουχιών μεταξύ τους, μέχρι τέλους.

Όπως αναφέραμε προηγουμένως τα βήματα αυτά έχουν πραγματοποιηθεί με διάφορες κατά καιρούς παραλλαγές. Για παράδειγμα, οι αρχικές στοιχίσεις μεταξύ των αλληλουχιών ή οι ευθυγραμμίσεις μεταξύ των αποτελεσμάτων των στοιχίσεων (προοδευτική στοιχίση) μπορούν να γίνουν με την αλγοριθμική βοήθεια του δυναμικού προγραμματισμού. Το δένδρο οδηγός θα μπορούσε να δημιουργηθεί με τη βοήθεια αλγορίθμων μηχανικής μάθησης οι οποίοι βοηθούν στην ομαδοποίηση (clustering) με βάση την ομοιότητα.

3.3.4.2 Επαναληπτική πολλαπλή στοιχίση αλληλουχιών (Iterative Multiple Sequence Alignment)

Η βασική ιδέα της επαναληπτικής πολλαπλής στοιχίσης είναι ουσιαστικά η βελτίωση της προοδευτικής πολλαπλής στοιχίσης που προαναφέραμε. Στην προοδευτική πολλαπλή στοιχίση είναι πιθανό να προκύψουν κάποια σφάλματα (π.χ. χαμηλό score) στις αρχικές συγκρίσεις ειδικά εάν οι αλληλουχίες που συγκρίνονται είναι μακρινά συγγενικές. Αυτά τα πιθανά λάθη θα πολλαπλασιάζονται κατά τη διάρκεια της διαδικασίας της προοδευτικής πολλαπλής στοιχίσης. Η επαναληπτική πολλαπλή στοιχίση αναλαμβάνει να διορθώσει αυτά τα λάθη επιστρέφοντας σε προηγούμενες στοιχίσεις ζευγών ή πολλαπλές υποστοιχίσεις επαναλαμβάνοντας τη διαδικασία προκειμένου να πετύχει το καλύτερο συνολικό score στοιχίσης.

3.3.5 Ολική στοιχίση αλληλουχιών – Αλγόριθμος Needleman and Wunsch

Στην ολική στοιχίση (global alignment), η στοιχίση γίνεται σε όλο το πεδίο της αλληλουχίας και συμπεριλαμβάνει όσο το δυνατόν περισσότερα ταιριαστά ζεύγη ανάμεσα στις συγκρινόμενες αλληλουχίες. Οι αλληλουχίες αυτές μπορεί να είναι είτε νουκλεοτιδικές είτε πρωτεϊνικές. Η ολική στοιχίση βοηθάει τους επιστήμονες να βγάλουν συμπεράσματα σχετικά με πιθανή εξελικτική ή λειτουργική σχέση μεταξύ δύο ή περισσότερων αλληλουχιών. Ο πιο γνωστός αλγόριθμος για ολική στοιχίση είναι αυτός των Needleman and Wunsch (Needleman & Wunsch, 1970). Σύμφωνα με τον αλγόριθμο αυτό, το score κάθε κελιού υπολογίζεται με τον αναδρομικό τύπο:

$$F(i, j) = \max\{F(i - 1, j - 1) + s(x_i, y_j), F(i - 1, j) - d, F(i, j - 1) - d\} \quad (3.1)$$

Η τιμή για το κάτω δεξιά στοιχείο του πίνακα είναι εξ ορισμού το score για την καλύτερη δυνατή στοίχιση, ενώ για την αρχικοποίηση της πρώτης στήλης και της πρώτης γραμμής, έχουμε $F(i,0)=-id$ και $F(0,j)=-jd$ αντίστοιχα. Εάν ξεκινήσουμε μια αναδρομή (recursion) από το κάτω δεξιά στοιχείο του πίνακα ακολουθώντας κάθε φορά τα μέγιστα εντοπίζουμε σιγά-σιγά τη βέλτιστη διαδρομή, δηλαδή τη βέλτιστη στοίχιση. Με το ακόλουθο παράδειγμα θα ξεκαθαρίσουμε λίγο καλύτερα τα πράγματα.

Παράδειγμα 3.1 (Waterman, 1995)

Έστω λοιπόν ότι έχουμε τις ακόλουθες αλληλουχίες DNA: $y=CAGTATCGCA$ και $x=AAGTTAGCAG$. Ας εξετάσουμε την καλύτερη ολική στοίχιση που μπορούμε να πετύχουμε με:

$$s(x_i, y_j) = \begin{cases} 1, & \text{αν } x_i = y_j \\ -1, & \text{αν } x_i \neq y_j \end{cases} \quad (3.2)$$

και $d=1$, τότε με τη συμπλήρωση του πίνακα προκύπτει:

	-	A	A	G	T	T	A	G	C	A	G
-	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10
C	-1	-1	-2	-3	-4	-5	-6	-7	-6	-7	-8
A	-2	0	0	-1	-2	-3	-4	-5	-6	-5	-6
G	-3	-1	-1	1	0	-1	-2	-3	-4	-5	-4
T	-4	-2	-2	0	2	1	0	-1	-2	-3	-4
A	-5	-3	-1	-1	1	1	2	1	0	-1	-2
T	-6	-4	-2	-2	0	2	1	1	0	-1	-2
C	-7	-5	-3	-3	-1	1	1	0	2	1	0
G	-8	-6	-4	-2	-2	0	0	2	1	1	2
C	-9	-7	-5	-3	-3	-1	-1	1	3	2	1
A	-10	-8	-6	-4	-4	-2	0	0	2	4	3

Η ολική στοίχιση λοιπόν είναι:

A A G T – T A G C A G

C A G T A T C G C A –

το score της οποίας είναι ίσο με 3 (η τιμή του κελιού που βρίσκεται κάτω δεξιά στον πίνακα).

3.3.6 Τοπική στοίχιση αλληλουχιών – Αλγόριθμος Smith and Waterman

Στην τοπική στοίχιση (local alignment) αναζητούμε περιοχές στις οποίες παρατηρείται τοπική ομοιότητα. Η στοίχιση αυτή σταματά στις ταυτόσημες περιοχές ή στις περιοχές που έχουν μεγάλη ομοιότητα. Μεγαλύτερη προτεραιότητα δίνεται στην εύρεση περιοχών ομοιότητας (patterns) παρά στην εύρεση ομοιότητας ατομικών χαρακτήρων. Ο πιο γνωστός αλγόριθμος για τοπική στοίχιση είναι αυτός των Smith and Waterman (Smith & Waterman, 1981) και χρησιμοποιεί τον ακόλουθο αναδρομικό τύπο:

$$F(i, j) = \max\{F(i - 1, j - 1) + s(x_i, y_j), F(i - 1, j) - d, F(i, j - 1) - d, 0\} \quad (3.3)$$

με $F(i, 0) = 0$ και $F(0, j) = 0$

Ο αλγόριθμος αυτός είναι ίδιος με τον προηγούμενο που εξετάσαμε. Η μόνη διαφορά είναι ότι όταν μια στοίχιση δίνει αρνητικό score τερματίζεται και αρχίζει μια νέα. Έτσι, η αρχικοποίηση στον πίνακα είναι διαφορετική για να μπορούμε να εντοπίσουμε ομοιότητες σε οποιοδήποτε σημείο εκτός της κύριας διαγωνίου.

Παράδειγμα 3.2 (Waterman, 1995)

Εάν εφαρμόσουμε τη μέθοδό μας ακριβώς με τα στοιχεία που χρησιμοποιήσαμε και στο παράδειγμα 3.1 ο πίνακας θα γίνει ως εξής:

	-	A	A	G	T	T	A	G	C	A	G
-	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	1	0	0
A	0	1	1	0	0	0	1	0	0	2	1
G	0	0	0	2	1	0	0	2	1	1	3
T	0	0	0	1	3	2	1	1	1	0	2
A	0	1	1	0	2	2	3	2	1	2	1

T	0	0	0	0	1	3	2	2	1	1	1
C	0	0	0	0	0	2	2	1	3	2	1
G	0	0	0	1	0	1	1	3	2	2	3
C	0	0	0	0	0	0	0	2	4	3	2
A	0	1	1	0	0	0	1	1	3	5	4

Η τοπική στοίχιση λοιπών είναι:

A G T A T C G C A
A G T – T A G C A

το score της οποίας είναι ίσο με 5 (το κελί με τη μεγαλύτερη τιμή στον πίνακα).

3.3.7 Η χρησιμότητα της στοίχισης των αλληλουχιών

Η διαδικασία της στοίχισης των διαφόρων αλληλουχιών είναι πολύ χρήσιμη διότι μέσω αυτής οι επιστήμονες μπορούν να αντλήσουν πληροφορίες σχετικά με τη λειτουργία, τη δομή αλλά και την εξέλιξη των διαφόρων οργανισμών (ειδών). Είναι πολύ σημαντική η επίτευξη της βέλτιστης (optimal) στοίχισης ώστε να ανακαλυφθούν οι πληροφορίες αυτές. Οι αλληλουχίες που εμφανίζουν τη βέλτιστη ομοιότητα είναι πολύ πιθανό να έχουν την ίδια λειτουργία. Στην περίπτωση του DNA θα μπορούσαμε να πούμε ότι είναι πιθανό να έχουν τον ίδιο ρυθμιστικό ρόλο. Υπάρχουν και περιπτώσεις όπου δυο αλληλουχίες από διαφορετικούς οργανισμούς είναι όμοιες. Αυτό ενδέχεται να σημαίνει ότι υπήρξε μια κοινή αλληλουχία πρόγονος για τις δύο αυτές αλληλουχίες. Μπορούμε λοιπόν να εντοπίσουμε τις διάφορες αλλαγές μεταξύ των απογόνων αλληλουχιών σε σχέση με την πρόγονό τους.

Με την έλευση της αλληλούχισης ολόκληρων γονιδιωμάτων καθώς και τις συγκρίσεις αλληλουχίας ευρείας κλίμακας προκύπτουν νέα δεδομένα. Η ομοιότητα μεταξύ των αλληλουχιών ενδέχεται να κρύβει πιθανή ή καμία συγγένεια. Μια ενδεχόμενη γονιδιακή εξέλιξη πιστεύεται ότι συμβαίνει μέσω του διπλασιασμού ενός γονιδίου, δημιουργώντας δηλαδή δύο διαδοχικά αντίγραφα του γονιδίου, τον οποίο διαδέχεται μια μετάλλαξη στα αντίγραφα αυτά (αναλυτικότερη μελέτη των μεταλλάξεων στη συνέχεια του κεφαλαίου). Σε κάποιες σπάνιες περιπτώσεις, οι μεταλλάξεις στο ένα από τα δύο αντίγραφα προκαλούν αλλαγή στη λειτουργία του. Τα

δύο αντίγραφα, θα εξελιχθούν με ξεχωριστούς τρόπους. Έτσι, παρ' όλο που ο διαχωρισμός που προκύπτει στη λειτουργία θα δημιουργήσει δύο νέους απογόνους (οικογένειες) αλληλουχιών, οι αλληλουχίες μεταξύ των δύο οικογενειών θα είναι όμοιες λόγω του κοινού προγόνου.

3.4 Εφαρμογές του DNA Sequencing

Η αλληλούχιση του γενετικού υλικού του ανθρώπου έχει βοηθήσει σημαντικά τους επιστήμονες τόσο στη μελέτη όσο και στην προσπάθεια αντιμετώπισης διαφόρων ασθενειών που σχετίζονται άμεσα με το ανθρώπινο DNA. Οι ασθένειες αυτές παρατηρούνται λόγω των διαφόρων αλλαγών που προκύπτουν στο γενετικό υλικό του ανθρώπου.

Το ανθρώπινο γενετικό υλικό μπορεί να υποστεί αλλαγές με πληθώρα διαφορετικών τρόπων. Το γεγονός αυτό δεν προκαλεί έκπληξη εάν σκεφθούμε για παράδειγμα με πόσους πιθανούς τρόπους μπορούν να αλλάξουν οι λέξεις που υπάρχουν σε αυτήν την πτυχιακή εργασία. Λίγες διακριτικές αλλαγές και η συγκεκριμένη πτυχιακή εργασία ίσως να μην παρήγαγε κανένα απολύτως νόημα. Έτσι λοιπόν, οι διάφορες αλλαγές στην αλληλουχία του γενετικού υλικού ονομάζονται μεταλλάξεις. Οι μεταλλάξεις δημιουργούν συνήθως διαφορετικό φαινότυπο χωρίς όμως αυτό να είναι πάντα απαραίτητο. Αυτό εξαρτάται από τον τρόπο επίδρασης της εκάστοτε αλλαγής στην πρωτεΐνη.

Υπάρχουν δύο μεγάλες κατηγορίες μεταλλάξεων:

- Οι γονιδιακές μεταλλάξεις.
- Οι χρωμοσωμικές μεταλλάξεις.

Η έκταση της αλλαγής στην ακολουθία και στον αριθμό των βάσεων στο γονιδίωμα του ανθρώπου είναι αυτή η οποία προκαλεί τον παραπάνω διαχωρισμό. Αν η αλλαγή αυτή προξενείται σε μικρό αριθμό βάσεων τότε αποτελεί γονιδιακή μετάλλαξη. Εάν πάλι η αλλαγή αυτή πραγματοποιείται σε μεγαλύτερο τμήμα του χρωμοσώματος ονομάζεται χρωμοσωμική μετάλλαξη. Οι μεταλλάξεις ευθύνονται για πολλές

περιπτώσεις καρκίνου καθώς και για διάφορες κληρονομικές ασθένειες. Παρακάτω θα μελετήσουμε λεπτομερώς τις δύο κατηγορίες των μεταλλάξεων.

3.4.1 Οι γονιδιακές μεταλλάξεις

Οι κύριες κατηγορίες γονιδιακών μεταλλάξεων είναι τρεις:

- Αντικατάσταση.
- Προσθήκη.
- Έλλειψη.

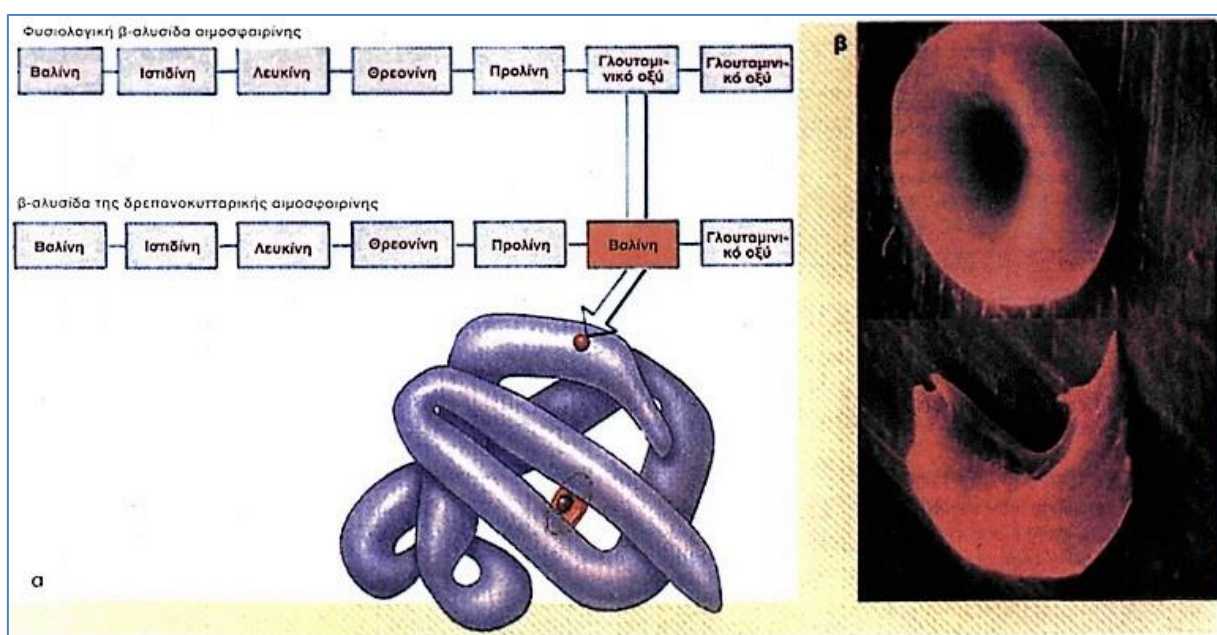
Η αντικατάσταση μιας και μόνο βάσης του DNA από τα δισεκατομμύρια των βάσεων μπορεί να επιφέρει ποικίλα αποτελέσματα στην πρωτεΐνη που παράγεται από το αντίστοιχο γονίδιο. Η αντικατάσταση αυτή μπορεί να δημιουργήσει μια τριπλέτα η οποία θα κωδικοποιεί ένα διαφορετικό αμινοξύ και κατ' επέκταση μια διαφορετική πρωτεΐνη. Σε άλλες περιπτώσεις, η αντικατάσταση μπορεί να μετατρέψει ένα κωδικόνιο, το οποίο είναι υπεύθυνο για την κωδικοποίηση ενός αμινοξέος, σε κωδικόνιο λήξης προκαλώντας τον τερματισμό της σύνθεσης της πολυπεπτιδικής αλυσίδας. Αυτό έχει ως απότοκο την καταστροφή της λειτουργικότητας της πρωτεΐνης.

Εξίσου σημαντική κατηγορία γονιδιακών μεταλλάξεων αποτελούν η προσθήκη ή η έλλειψη βάσεων. Διάφορες αλλαγές που προκύπτουν στον αριθμό των βάσεων του γενετικού υλικού μπορούν να επιφέρουν μεταλλαγμένους φαινότυπους. Η έλλειψη ή η προσθήκη διαδοχικών βάσεων σε αριθμό πολλαπλάσιο του τρία δημιουργεί έλλειψη ή προσθήκη ενός ή περισσότερων αμινοξέων στην πολυπεπτιδική αλυσίδα. Αυτό έχει ως συνέπεια την αλλαγή της λειτουργικότητας της πολυπεπτιδικής αλυσίδας.

Η πρώτη γενετική ασθένεια η οποία προέκυψε λόγω γονιδιακής μετάλλαξης ήταν η δρεπανοκυτταρική αναιμία. Παρ' όλο που η αιμοσφαιρίνη (HbA) στα ενήλικα άτομα αποτελείται από τέσσερις πολυπεπτιδικές αλυσίδες, δύο α και δύο β, ο Linus Pauling⁸ και οι συνεργάτες του παρατήρησαν ότι η αιμοσφαιρίνη κάποιου ενήλικα ο

⁸ Η ανακάλυψη έγινε το 1949 χωρίς όμως να διευκρινιστεί η αιτία που προκαλούσε τη διαφορά στην αιμοσφαιρίνη.

ο οποίος έπασχε από δρεπανοκυτταρική αναιμία είναι διαφορετική. Φτάνοντας στο σήμερα, τα πράγματα έχουν γίνει σαφώς πιο ξεκάθαρα. Η διαφορά εντοπίζεται στο έκτο αμινοξύ της β πολυπεπτιδικής αλυσίδας. Η βαλίνη αντικαθιστά το γλουταμινικό οξύ με συνέπεια την μεταλλαγμένη αιμοσφαιρίνη (HbS). Αυτό συμβαίνει διότι στην κωδική αλυσίδα του DNA αλλάζει μια βάση με αποτέλεσμα το κωδικόνιο (GTG) που κωδικοποιεί τη βαλίνη να αντικαθιστά το φυσιολογικό κωδικόνιο (GAG) που κωδικοποιεί το γλουταμινικό οξύ. Έτσι, λόγω της αλλαγής στη στερεοδιάταξη της αιμοσφαιρίνης, τα ερυθρά κύτταρα παίρνουν ένα χαρακτηριστικό δρεπανοειδές⁹ σχήμα. Η μη φυσιολογική κυκλοφορία που προκύπτει στο αίμα μπορεί να προκαλέσει



Εικόνα 3.4: (α) Η αντικατάσταση του γλουταμινικού οξέος με βαλίνη. (β) Φυσιολογικό ερυθροκύτταρο επάνω και δρεπανοειδές ερυθροκύτταρο κάτω.

ζημιές σε διάφορα όργανα όπως στους πνεύμονες ή στο σπλήνα.

3.4.1.1 Μεταβολικά νοσήματα από γονιδιακές μεταλλάξεις

Έχουν περιγραφεί περίπου διακόσιες διαταραχές του ανθρωπίνου μεταβολισμού. Οι διαταραχές αυτές αφορούν κυρίως τη λειτουργικότητα των ενζύμων. Οι διάφορες μεταλλάξεις που προκύπτουν σε γονίδια τα οποία κωδικοποιούν κάποια από τα ένζυμα είναι υπεύθυνες για πληθώρα μεταβολικών νοσημάτων (ασθενειών).

⁹ Σε συνθήκες έλλειψης οξυγόνου.

Τα νοσήματα αυτά είναι πιθανό να κληρονομηθούν στις επόμενες γενιές. Για παράδειγμα, τέτοιου είδους ασθένειες αποτελούν η φαινυλκετονουρία και ο αλφισμός.

Η φαινυλκετονουρία (PKU = Phenyl Keton Urea) είναι μια νόσος η οποία προκαλείται λόγω έλλειψης ενός ενζύμου το οποίο, στα φυσιολογικά άτομα, είναι υπεύθυνο για τη μετατροπή του αμινοξέος φαινυλαλανίνη σε τυροσίνη. Αυτό φυσικά προκαλεί συσσώρευση της φαινυλαλανίνης. Τα άτομα που νοσούν από αυτήν την ασθένεια αντιμετωπίζουν μη φυσιολογική ανάπτυξη και λειτουργία των κυττάρων του εγκεφάλου. Αυτό φυσικά επιφέρει διανοητική καθυστέρηση. Αν ο εντοπισμός της ασθένειας επιτευχθεί εγκαίρως, δηλαδή αν ανιχνευθεί κατά τη νεογνική ηλικία του ανθρώπου, η εμφάνιση των συμπτωμάτων αυτής μπορεί να αποφευχθεί με τη χρήση κατάλληλου διαιτολογίου¹⁰ εφ' όρου ζωής.

Ο αλφισμός αποτελεί νόσημα που προκαλείται λόγω της έλλειψης ενός ενζύμου το οποίο είναι απαραίτητο ώστε να σχηματιστεί η χρωστική ονόματι μελανίνη. Τα πάσχοντα από αλφισμό άτομα αντιμετωπίζουν έλλειψη της χρωστικής αυτής στην ίριδα του οφθαλμού, στα μαλλιά και στο δέρμα.

3.4.2 Οι χρωμοσωμικές μεταλλάξεις

Για να προκύψει μια οποιαδήποτε χρωμοσωμική ανωμαλία (μετάλλαξη), θα πρέπει οι αλλαγές που θα προκύψουν στο ανθρώπινο γονιδίωμα να είναι μεγαλύτερες σε έκταση από αυτές των γονιδιακών μεταλλάξεων. Οι χρωμοσωμικές ανωμαλίες όμως χωρίζονται σε δυο κατηγορίες. Έτσι λοιπόν προκύπτουν:

- Οι αριθμητικές χρωμοσωμικές ανωμαλίες.
- Οι δομικές χρωμοσωμικές ανωμαλίες.

Παρακάτω θα αναλύσουμε την κάθε μια κατηγορία ξεχωριστά με παραδείγματα για μια πιο ολοκληρωμένη εικόνα.

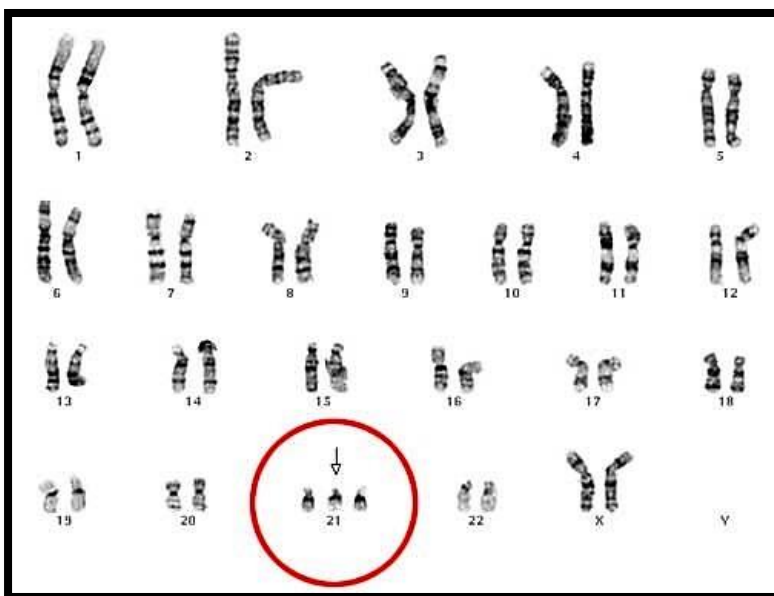
3.4.2.1 Οι αριθμητικές χρωμοσωμικές ανωμαλίες

Αριθμητικές χρωμοσωμικές ανωμαλίες αποκαλούνται οι αλλαγές που προκύπτουν στον αριθμό των χρωμοσωμάτων του ανθρώπου. Οι αλλαγές αυτές είναι

¹⁰ Διαιτολόγιο με περιορισμένη ποσότητα φαινυλαλανίνης.

αποτέλεσμα λαθών που συμβαίνουν κατά τη διαδικασία της μειωτικής διαίρεσης. Αν κατά τη μείωση δεν υπάρξει φυσιολογικός διαχωρισμός των ομόλογων χρωμοσωμάτων ή των αδελφών χρωματίδων¹¹ τότε προκύπτουν γαμέτες οι οποίοι φέρουν αριθμό χρωμοσωμάτων μεγαλύτερο ή μικρότερο του φυσιολογικού. Η μετέπειτα γονιμοποίηση των μη φυσιολογικών γαμετών με φυσιολογικό γαμέτη έχει ως απότοκο την παρασκευή ζυγωτού το οποίο έχει σαν γνώρισμα «λανθασμένη» ποσότητα DNA. Ο ανθρώπινος οργανισμός που προκύπτει και χαρακτηρίζεται από έλλειψη ή περίσσεια μικρού αριθμού χρωμοσωμάτων καλείται ανευπλοειδής. Εάν υπάρχει απουσία ενός μόνο χρωμοσώματος τότε το φαινόμενο αποκαλείται μονοσωμία¹². Από την άλλη πλευρά, η ύπαρξη ενός επιπλέον χρωμοσώματος ονομάζεται τρισωμία.

Είναι γεγονός ότι οι αριθμητικές χρωμοσωμικές ανωμαλίες δημιουργούνται τόσο στα αυτοσωμικά όσο και στα φυλετικά χρωμοσώματα. Μια πασίγνωστη εμφάνιση αριθμητικής, αυτοσωμικής χρωμοσωμικής ανωμαλίας είναι το σύνδρομο Down¹³.



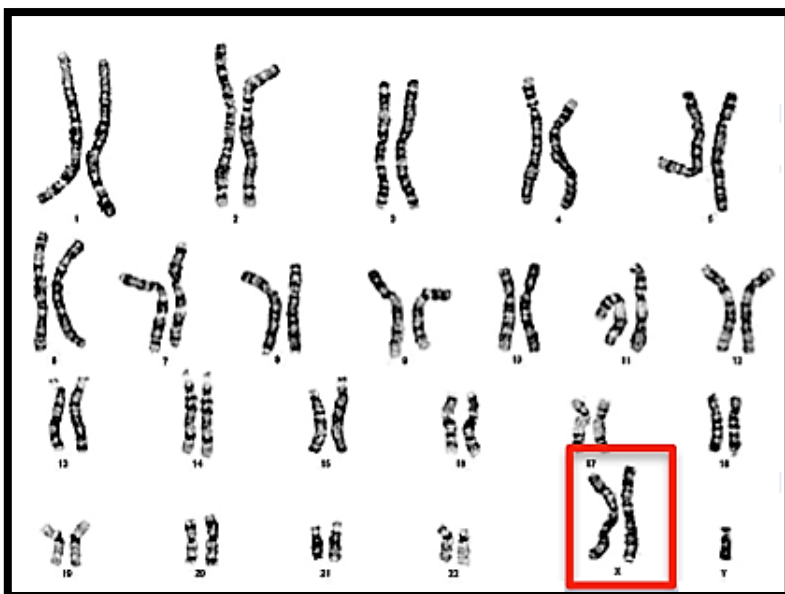
Εικόνα 3.5: Καρυότυπος ατόμου με σύνδρομο Down.

Οι άνθρωποι με σύνδρομο Down αντιμετωπίζουν προβλήματα όπως διανοητική καθυστέρηση, χαρακτηριστικές δυσμορφίες στο πρόσωπο και εμφανή καθυστέρηση στην ανάπτυξη. Στα πάσχοντα άτομα, σχεδόν στο εκατό τοις εκατό των περιπτώσεων, παρατηρείται ένα επιπλέον χρωμόσωμα 21 στον καρυότυπό τους. Το επιπλέον χρωμόσωμα απορρέει από το μη διαχωρισμό των χρωμοσωμάτων του 21^{ου} ζεύγους κατά το σχηματισμό γαμετών στη διαδικασία της μειωτικής διαίρεσης. Στις

¹¹ Το φαινόμενο αυτό ονομάζεται μη – διαχωρισμός.

¹² Η μονοσωμία συνήθως οδηγεί στο θάνατο του ατόμου διότι τα χρωμοσώματα με τα γονίδια που περιέχουν πρέπει να υπάρχουν σε δύο δόσεις ώστε να εξασφαλίζεται η σωστή ανάπτυξη του ζυγωτού. Εξαιρούνται τα φυλετικά χρωμοσώματα.

¹³ Σύνδρομο Down αποτελεί η τρισωμία στο χρωμόσωμα 21.



περισσότερες περιπτώσεις δημιουργείται wάριο με δύο χρωμοσώματα στη θέση 21. Ελάχιστες είναι οι αντίστοιχες περιπτώσεις για το σπερματοζwάριο. Συμπερασματικά, κομβικό ρόλο στη γέννηση ενός παιδιού με σύνδρομο Down παίζει η ηλικία της μητέρας.

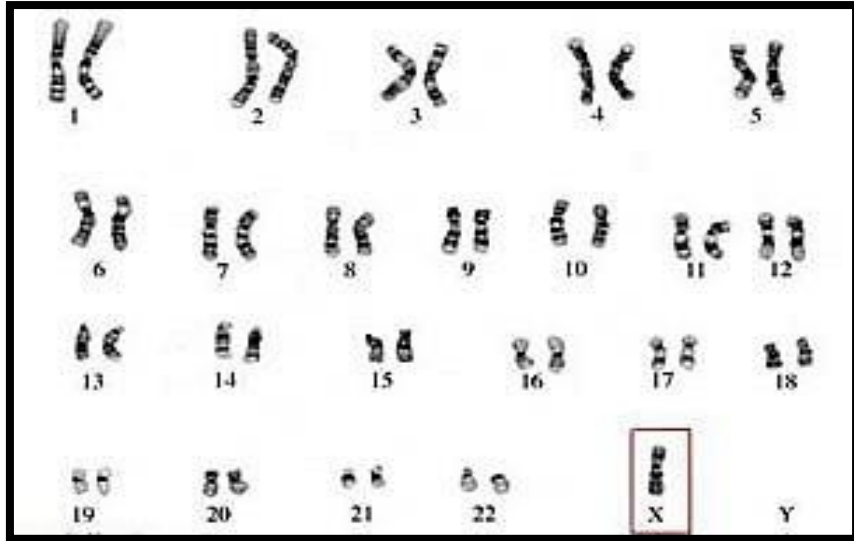
Εικόνα 3.6: Καρυότυπος ατόμου με σύνδρομο Klinefelter.

Πέρα από το σύνδρομο Down υπάρχουν και άλλες σχετικά συχνές αριθμητικές, αυτοσωμικές χρωμοσωμικές ανωμαλίες όπως είναι η τρισωμία 13 και 18. Τα πάσχοντα άτομα χαρακτηρίζονται από βαρύτερα του συνδρόμου Down συμπτώματα. Αυτό συμβαίνει λόγω μεγέθους των χρωμοσωμάτων που βρίσκονται στις θέσεις αυτές (13 και 18). Το γεγονός ότι είναι μεγαλύτερα από αυτά της θέσεις 21 καθιστά λογικό το να περιέχουν περισσότερα γονίδια. Πέρα όμως από τις αριθμητικές, χρωμοσωμικές ανωμαλίες στα αυτοσωμικά, υπάρχουν και οι αριθμητικές χρωμοσωμικές ανωμαλίες στα φυλετικά χρωμοσώματα. Δύο απ' τα διάφορα σύνδρομα της κατηγορίας αυτής εξετάζονται παρακάτω.

Αριθμητική χρωμοσωμική ανωμαλία στα φυλετικά χρωμοσώματα αποτελεί το σύνδρομο Klinefelter. Τα άτομα που πάσχουν απ' αυτό το σύνδρομο, ενώ εμφανίζουν φυσιολογικό αριθμό αυτοσωμικών χρωμοσωμάτων (44 στον αριθμό), έχουν τρία φυλετικά χρωμοσώματα (XXY) αντί του φυσιολογικού ζεύγους XY. Τα συγκεκριμένα άτομα φέρουν εξωτερικά γνωρίσματα αρσενικού ατόμου αλλά πάσχουν από πρόβλημα στείρωσης. Τα συμπτώματα του συνδρόμου εμφανίζονται μετά την εφηβεία.

Το σύνδρομο Turner αποτελεί ακόμα ένα παράδειγμα της κατηγορίας που προαναφέραμε. Και σε αυτήν την περίπτωση, τα πάσχοντα άτομα εμφανίζουν

φυσιολογικό αριθμό αυτοσωμικών χρωμοσωμάτων (44 στον αριθμό), αλλά έχουν μόνο ένα χρωμόσωμα X από το ζεύγος των χρωμοσωμάτων που καθορίζουν το φύλο. Αυτό το παράδειγμα αποτελεί την μοναδική



Εικόνα 3.7: Καρυότυπος ατόμου με σύνδρομο Turner.

μονοσωμία που έχει εντοπιστεί στον άνθρωπο. Τα εν λόγω άτομα, παρ' όλο που έχουν φαινότυπο θηλυκού ατόμου, δεν εμφανίζουν δευτερογενή γνωρίσματα του θηλυκού φύλου ενώ αντιμετωπίζουν και πρόβλημα στειρότητας.

3.4.2.2 Οι δομικές χρωμοσωμικές ανωμαλίες

Δομικές χρωμοσωμικές ανωμαλίες ονομάζονται οι αλλαγές (μεταλλάξεις) που συμβαίνουν στη δομή ενός ή περισσότερων χρωμοσωμάτων του ανθρώπου. Με τις δομικές χρωμοσωμικές αλλαγές παρατηρείται συνήθως τροποποίηση του φαινοτύπου του ανθρώπου. Η δημιουργία δομικών¹⁴ χρωμοσωμικών ανωμαλιών είναι αποτέλεσμα διαφόρων μηχανισμών κατά τη διάρκεια του κυτταρικού κύκλου. Οι ανωμαλίες αυτές επιφέρουν αλλαγή στην ποσότητα ή στη διάταξη της γενετικής πληροφορίας στα χρωμοσώματα. Ανάλογα με τον τύπο της αλλαγής διακρίνουμε τα παρακάτω είδη δομικών χρωμοσωμικών ανωμαλιών:

- Την έλλειψη.
- Τον διπλασιασμό.
- Την αναστροφή.
- Τη μετατόπιση.

¹⁴ Οι δομικές χρωμοσωμικές ανωμαλίες μπορούν να προκληθούν και από τη δράση μεταλλαξογόνων παραγόντων όπως οι ακτινοβολίες.

Έλλειψη παρατηρείται όταν υπάρχει απώλεια γενετικού υλικού. Παράδειγμα έλλειψης αποτελεί το σύνδρομο φωνή της γάτας (cri-du-chat) το οποίο είναι αποτέλεσμα έλλειψης ενός τμήματος από το χρωμόσωμα στη θέση 5. Το συγκεκριμένο σύνδρομο πήρε την ονομασία του από το γεγονός ότι το κλάμα των νεογέννητων που πάσχουν απ' αυτό, μοιάζει με το κλάμα της γάτας. Τα ίδια άτομα εμφανίζουν και διανοητική καθυστέρηση.

Διπλασιασμό έχουμε όταν υπάρχει επανάληψη ενός χρωμοσωμικού τμήματος σε ένα οποιοδήποτε χρωμόσωμα.

Η αναστροφή είναι αποτέλεσμα θραύσεων σε δύο διαφορετικά σημεία ενός χρωμοσώματος τα οποία ενώνονται με το υπόλοιπο χρωμόσωμα αφού πρώτα υποστούν αναστροφή. Αυτό φυσικά έχει ως συνέπεια την αλλαγή της διάταξης των γονιδίων στο χρωμόσωμα αυτό.

Η μετατόπιση προκαλείται ύστερα από θραύση ενός τμήματος του χρωμοσώματος το οποίο στη συνέχεια ενώνεται με ένα άλλο μη ομόλογο χρωμόσωμα. Στις αμοιβαίες μετατοπίσεις προκύπτει ανταλλαγή χρωμοσωμικών τμημάτων ανάμεσα σε μη ομόλογα χρωμοσώματα. Παρ' όλα αυτά δεν έχουμε απώλεια γενετικού υλικού με τα πάσχοντα άτομα να έχουν, συνήθως, φυσιολογικό φαινότυπο. Στα άτομα αυτά όμως ελλοχεύει ο κίνδυνος απόκτησης απογόνων οι οποίοι θα χαρακτηρίζονται από χρωμοσωμικές ανωμαλίες. Αυτό συμβαίνει διότι κατά το χρωμοσωμικό ζευγάρωμα στη διαδικασία της μειωτικής διαίρεσης προκύπτουν και μη φυσιολογικοί γαμέτες.

3.4.3 Ο προγεννητικός έλεγχος

Η χρήση της αλληλούχισης του γενετικού υλικού κρίνεται απαραίτητη και στη διαδικασία του προγεννητικού ελέγχου στον ανθρώπινο οργανισμό. Ο έλεγχος αυτός μπορεί να πραγματοποιηθεί είτε με τη λήψη χοριακών λαχνών, είτε με τη μέθοδο της αμνιοπαρακέντησης.

Η μέθοδος λήψης χοριακών λαχνών πραγματοποιείται συνήθως μεταξύ της 9^{ης} και της 12^{ης} εβδομάδας κύησης. Η διαδικασία αυτή περιλαμβάνει τη λήψη εμβρυϊκών

κυττάρων από τις προεκβολές (λάχνες) του χόριου¹⁵. Τα ληφθέντα κύτταρα είναι χρήσιμα για τον έλεγχο των χρωμοσωμάτων (καρυότυπος) καθώς και για βιοχημικές αναλύσεις ή ανάλυση του γενετικού υλικού.

Η μέθοδος της αμνιοπαρακέντησης πραγματοποιείται μεταξύ της 12^{ης} και της 16^{ης} εβδομάδας κύησης και αποτελεί έναν αξιόπιστο και ασφαλή τρόπο διάγνωσης γενετικών ανωμαλιών. Κατά τη διαδικασία αυτή λαμβάνεται μικρή ποσότητα αμνιακού υγρού, μέσα στο οποίο υπάρχουν εμβρυϊκά κύτταρα, από τον αμνιακό σάκο με τη βοήθεια μιας ειδικής βελόνας. Τα κύτταρα αυτά μπορούμε αργότερα να τα χρησιμοποιήσουμε τόσο για την ανάλυση του γενετικού υλικού όσο και για τη βιοχημική ανάλυση ορισμένων ενζύμων και πρωτεϊνών. Επιπρόσθετα, τα ίδια εμβρυϊκά κύτταρα μπορούμε να τα χρησιμοποιήσουμε, ύστερα από καλλιέργεια, για μελέτη του καρυότυπου μέσω του οποίου υπάρχει η δυνατότητα εντοπισμού διαφόρων χρωμοσωμικών ανωμαλιών.

Έπειτα από σύγκριση των δύο μεθόδων, οι επιστήμονες κατέληξαν ότι με τη λήψη χοριακών λαχνών μας δίνεται η δυνατότητα μιας πιο έγκαιρης διάγνωσης. Απ' την άλλη μεριά, με την αμνιοπαρακέντηση παρέχεται η δυνατότητα παρασκευής χρωμοσωμάτων καλύτερης ποιότητας από αυτών της μεθόδου λήψης χοριακών λαχνών.

¹⁵ Χόριο ονομάζεται η εμβρυϊκή μεμβράνη που συμμετέχει στο σχηματισμό του πλακούντα.

Κεφάλαιο 4^ο: Βάσεις δεδομένων νουκλεοτιδικών αλληλουχιών

4.1 Νουκλεοτιδικές βάσεις δεδομένων και αλγόριθμοι αναζήτησης

Τη σήμερον ημέρα, η διαδικασία συλλογής αλληλουχιών γενετικού υλικού με τις μεθόδους της Μοριακής Βιολογίας αποτελεί ρουτίνα για τους επιστήμονες. Επανάσταση στη διαδικασία αυτή όμως απετέλεσε η ψηφιοποίησή τους. Η δυνατότητα απεικόνισης των διαφόρων βιολογικών αλληλουχιών σε μορφή αναγνώσιμη από τους ηλεκτρονικούς υπολογιστές προσέφερε εντυπωσιακές δυνατότητες ανάλυσης, επεξεργασίας και αποθήκευσής τους. Αυτό είχε σαν απότοκο τη δημιουργία διαφόρων βιολογικών βάσεων δεδομένων και στην περίπτωση μας *Βάσεων Δεδομένων Νουκλεοτιδικών Αλληλουχιών*.

Οι πρώτες βάσεις δεδομένων που αναπτύχθηκαν και ονομάζονται *Πρωτογενείς* παρείχαν τον μηχανισμό για την πρόσβαση σε επιστημονικά και πειραματικά δεδομένα που αφορούσαν αλληλουχίες γενετικού υλικού. Εκτός αυτού, παρείχαν και το χώρο για την κατάθεση νέων επιστημονικών ευρημάτων, εφόσον αυτά υπήρχαν, σχετικών πάντα με τις αλληλουχίες. Ακολούθησαν νέου τύπου βάσεις δεδομένων, οι *Δευτερογενείς*, οι οποίες στηρίζονταν στις Πρωτογενείς βάσεις που αναφέραμε προηγουμένως. Αυτές περιείχαν επιπλέον σχολιασμό πάνω στα επιστημονικά δεδομένα ο οποίος δεν είχε απαραίτητα πειραματική στήριξη.

Οι τρεις πρώτες βάσεις νουκλεοτιδικών αλληλουχιών αναπτύχθηκαν σε τρεις διαφορετικές περιοχές του πλανήτη. Στην Αμερική έχουμε την *GenBank* η οποία είναι μια βάση ελεύθερα διαθέσιμη στην επιστημονική κοινότητα και βρίσκεται υπό την αιγίδα του *National Institutes of Health* των Ηνωμένων Πολιτειών της Αμερικής. Στην Ευρώπη έχει δημιουργηθεί η *EMBL Nucleotide Sequence Database* η οποία βρίσκεται υπό την αιγίδα του *European Molecular Biology Laboratory*. Η έδρα της βρίσκεται στο Cambridge του Ηνωμένου Βασιλείου και τη συντήρησή της έχει αναλάβει το Ευρωπαϊκό Ινστιτούτο Βιοπληροφορικής. Τέλος, στην Ιαπωνία έχουμε την *DNA Data Bank of Japan* η οποία ιδρύθηκε το 1986 στο Εθνικό Ινστιτούτο

Γενετικής και βρίσκεται υπό την αιγίδα του Υπουργείου Παιδείας, Επιστημών και Αθλητισμού της Ιαπωνίας. Ακολουθούν οι τρεις μεγάλες βάσεις δεδομένων με τις ιστοσελίδες τους:

- NCBI – GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>).
- EMBL – Bank (<http://www.ebi.ac.uk/>).
- DDBJ (<http://www.ddbj.nig.ac.jp/>).

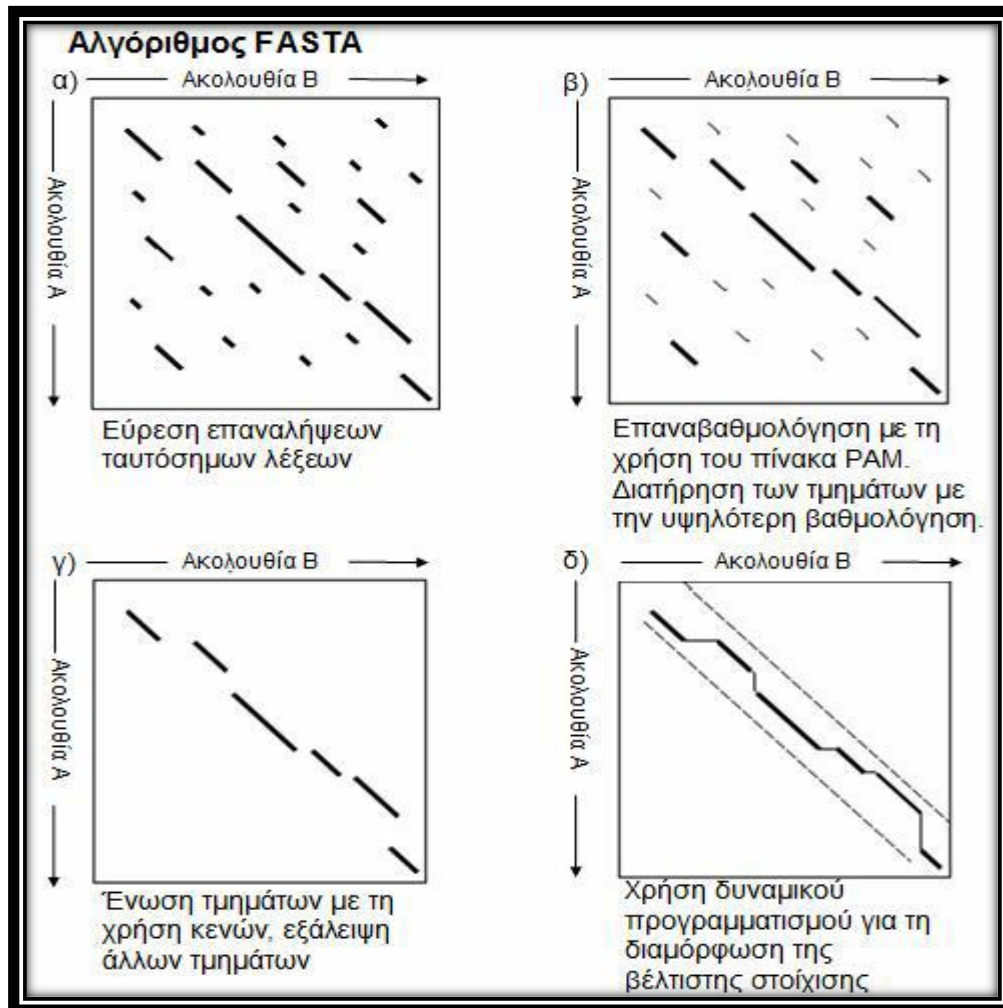
Όπως είναι λογικό, προέκυψε συνεργασία των τριών αυτών ερευνητικών κέντρων στο όνομα της επιστημονικής έρευνας. Η συνεργασία αυτή έγινε στο πλαίσιο της *International Nucleotide Sequence Database Collaboration* (<http://www.insdc.org/>). Πλέον, η πρόσβαση είναι ελεύθερη στις ιστοσελίδες και στα εργαλεία αναζήτησης και στοίχισης αλληλουχιών.

Στα υποκεφάλαια που ακολουθούν θα γνωρίσουμε τους δύο επικρατέστερους αλγόριθμους οι οποίοι χρησιμοποιούνται τόσο για αναζήτηση όσο και για στοίχιση αλληλουχιών με αλληλουχίες που υπάρχουν στις διάφορες βάσεις δεδομένων.

4.2 Ο αλγόριθμος FASTA

Ο αλγόριθμος FASTA (<http://www.ebi.ac.uk/Tools/sss/fasta/nucleotide.html>) ανήκει στην κατηγορία των ευριστικών αλγορίθμων και αποτελεί αλγόριθμο τοπικής ευθυγράμμισης. Ευριστικοί χαρακτηρίζονται οι αλγόριθμοι οι οποίοι μπορούν να βρουν μια λύση γρήγορα και εύκολα χωρίς όμως να σημαίνει ότι η λύση αυτή είναι απαραίτητα και η βέλτιστη. Η λέξη FASTA είναι συντομογραφία του *FAST-ALL* και η επικράτησή του οφείλεται στο γεγονός ότι χρησιμοποιείται για ταχεία ευθυγράμμιση τόσο για νουκλεοτιδικές όσο και για πρωτεϊνικές αλληλουχίες. Εκτός από τον αρχικό αλγόριθμο, έχουν υπάρξει και διάφορες παραλλαγές του. Ο κύριος στόχος του FASTA είναι ο εντοπισμός, κατά προσέγγιση, της διαγωνίου γύρω από την οποία εκτυλίσσεται η στοίχιση. Με τον τρόπο αυτό περιορίζει σημαντικά το εύρος αναζήτησης αφού οποιεσδήποτε ομοιότητες εκτός της διαγωνίου δε λαμβάνονται υπόψιν. Όπως θα δούμε και στην εικόνα 4.1 τα κύρια βήματα που πραγματοποιούνται είναι τα ακόλουθα:

- Αρχικά, δημιουργείται ένα ευρετήριο το οποίο περιέχει τις θέσεις όλων των k-tuples (δηλαδή των λέξεων με μέγεθος k, τυπικό μέγεθος για νουκλεοτιδικές αλληλουχίες είναι από 4 έως 6) που υπάρχουν και στις δύο συγκρινόμενες αλληλουχίες.
- Από τη διαφορά των θέσεων των k-tuples στις δύο αλληλουχίες εντοπίζεται η διαγώνιος στην οποία βρίσκονται ενώ ακολούθως εντοπίζονται όλες οι διαγώνιοι με τα περισσότερα k-tuples.
- Με την εισαγωγή κενών (πάντα με τον υπολογισμό της αντίστοιχης ποινής) επιτυγχάνεται συνένωση αυτών των διαγωνίων.
- Τελικά, μέσω της διαδικασίας του δυναμικού προγραμματισμού διαμορφώνεται η βέλτιστη στοίχιση.

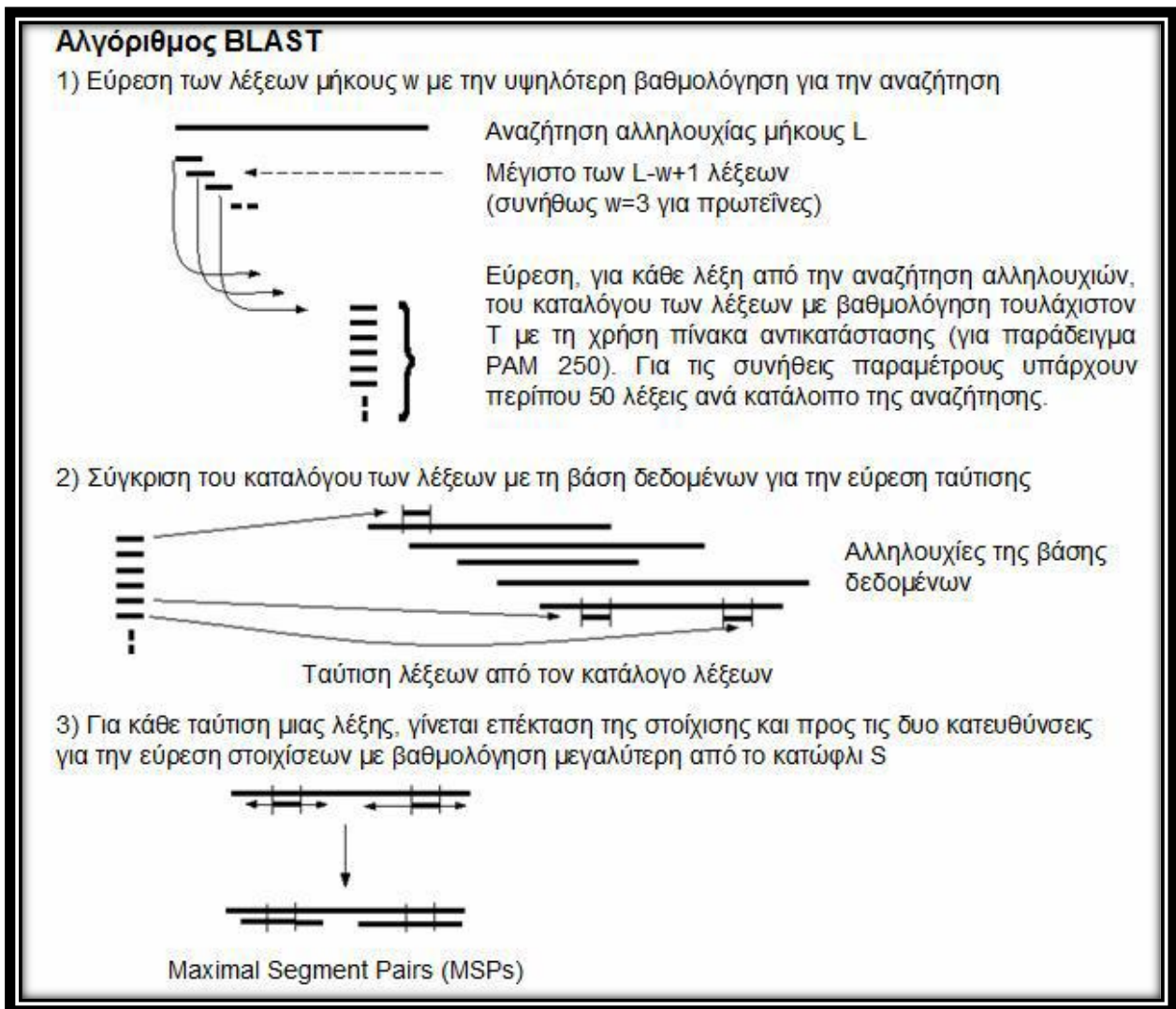


Εικόνα 4.1: Τα κύρια βήματα του FASTA.

4.3 Ο αλγόριθμος BLAST

Ο αλγόριθμος BLAST (<https://blast.ncbi.nlm.nih.gov>) ανήκει επίσης στην κατηγορία των ευριστικών αλγορίθμων. Η λέξη BLAST αποτελείται από τα αρχικά των λέξεων *Basic Local Alignment Search Tool*. Αρχικά είχε αναπτυχθεί από το National Center for Biotechnology Information (NCBI) και αποτελεί έναν αλγόριθμο ευθυγράμμισης αλληλουχιών βελτιστοποιημένης ταχύτητας ο οποίος ψάχνει την καλύτερη δυνατή τοπική στοίχιση στις διάφορες βάσεις αλληλουχιών. Τα αρχικά βήματα που εκτελούνται έχουν κάποιες ομοιότητες με τα βήματα εκτέλεσης του FASTA αλλά ο ίδιος ο BLAST είναι γρηγορότερος αφού αποφεύγει να στοιχίσει αλληλουχίες (προϋπολογίζοντας σχετικές παραμέτρους) οι οποίες βρίσκονται στη βάση δεδομένων και κρίνει ότι δεν παρουσιάζουν σημαντική ομοιότητα. Έτσι, σύμφωνα με την εικόνα 4.2 τα κύρια βήματα του αλγορίθμου είναι τα εξής:

- Αρχικά, κατασκευάζεται ένας κατάλογος με όλες τις λέξεις (words) οι οποίες θα ταίριαζαν με κάποια λέξη της άγνωστης αλληλουχίας. Οι λέξεις αυτές θα πρέπει να ξεπερνούν την τιμή του προκαθορισμένου κατωφλίου.
- Επόμενο βήμα αποτελεί η αναζήτηση των λέξεων αυτών (μέσω του αλγορίθμου) στις αλληλουχίες που βρίσκονται στη βάση δεδομένων. Κάθε φορά που εντοπίζεται κάποια λέξη εκκινείται μια διαδικασία επέκτασης του «ευρήματος» και προς τις δύο κατευθύνσεις της αλληλουχίας. Η επέκταση αυτή συνεχίζεται όσο το score εξακολουθεί να αυξάνεται.
- Οι περιοχές οι οποίες φέρουν το μεγαλύτερο score αποτελούν υποψήφια περιοχές στις οποίες εντοπίζεται ομοιότητα (high scoring pairs).
- Στα αποτελέσματα αναφέρονται οι περιοχές των high scoring pairs στις οποίες το score υπερβαίνει το κατώφλι. Τελικά, επιλέγονται μόνο οι τοπικές στοίχισεις οι οποίες εμφανίζουν την υψηλότερη στατιστική σημαντικότητα.



Εικόνα 4.2: Τα κύρια βήματα του BLAST.

Ο BLAST υπολογίζει τη στατιστική σημαντικότητα με βάση τις παραμέτρους της κατανομής οι οποίες έχουν υπολογιστεί από προσομοιώσεις που έχουν εκτελεστεί από πριν και οι τιμές τους έχουν αποθηκευτεί. Από την άλλη, ο FASTA υπολογίζει τις παραμέτρους αυτές από όλες τις άλλες αλληλουχίες που βρίσκονται στη βάση δεδομένων στην οποία γίνεται η αναζήτηση. Έτσι, ο BLAST καθίσταται γρηγορότερος από τον FASTA.

Οι πρώτες εκδόσεις του BLAST δεν παρείχαν την δυνατότητα εισαγωγής κενών κατά τη διαδικασία εξεύρεσης ομοιοτήτων μεταξύ των διαφόρων αλληλουχιών. Με τον ερχομό της δεύτερης έκδοσης προστέθηκε και η δυνατότητα εισαγωγής κενών

κατά την ευθυγράμμιση. Φυσικά, προέκυψαν και νέες εκδόσεις του BLAST με την κάθε μία από αυτές να χρησιμοποιείται για ξεχωριστό σκοπό.

Συμπερασματικά, τόσο ο BLAST όσο και ο FASTA παράγουν αποτελέσματα σχεδόν παραπλήσια με αυτά που παράγονται από τη χρήση της μεθόδου δυναμικού προγραμματισμού. Το ποιο πακέτο δυνατοτήτων θα επιλεγθεί (μεταξύ των BLAST και FASTA) επαφίεται σε επιστημονικά κριτήρια όπως είναι ο σκοπός της έρευνας, η ταχύτητα που απαιτείται, τα είδη των ακολουθιών που συγκρίνονται (νουκλεοτιδικές ή πρωτεϊνικές), οι ποινές για τα κενά κ.ά.

Κεφάλαιο 5^ο: Προβληματισμοί σχετικά με τις δυνατότητες της Βιοπληροφορικής

Έχοντας μελετήσει κάποιες από τις δυνατότητες που παρέχουν η Βιοπληροφορική και η διαδικασία της αλληλούχισης του γενετικού υλικού αντιλαμβάνεται κανείς το μέγεθος της δύναμης που πηγάζει από αυτήν την επιστήμη. Οι επιστήμονες κρατούν στα χέρια τους ένα πανίσχυρο «όπλο» η χρήση του οποίου μπορεί να αλλάξει ριζικά τον τρόπο αντιμετώπισης των διαφόρων προβλημάτων υγείας. Μέσα από τη συνεχή ανάγνωση και μελέτη του ανθρωπίνου γονιδιώματος έρχονται σταδιακά στο φως νέα ευρήματα τα οποία ίσως οδηγήσουν σε νέες εφαρμογές. Μια μελλοντική πρακτική εφαρμογή θα μπορούσε να είναι η ανάπτυξη αποτελεσματικών στρατηγικών τόσο για την πρόληψη όσο και για τη θεραπεία νοσημάτων σε ατομικό επίπεδο (η σε μικρές πληθυσμιακές ομάδες). Θα μπορούσε να ξεκινήσει δηλαδή η στοχευμένη θεραπεία στον εκάστοτε ασθενή δίχως να υπάρχει μια γενική πατέντα η οποία θα χρησιμοποιείται σε όλους τους ανθρώπους που πάσχουν από την ίδια ασθένεια. Ο κάθε οργανισμός άλλωστε ενδέχεται να ανταποκριθεί διαφορετικά από τους υπόλοιπους σε μια πιθανή θεραπεία.

Μέσα από τις δυνατότητες της αλληλούχισης του γενετικού υλικού προκύπτουν φυσικά και κάποια ηθικά ζητήματα. Είδαμε προηγουμένως ότι μέσα από τον προγεννητικό έλεγχο υπάρχει η δυνατότητα εντοπισμού πιθανών γενετικών ανωμαλιών. Στην περίπτωση διάγνωσης τέτοιων ανωμαλιών, δίνεται η δυνατότητα διακοπής της κύησης. Μία άποψη υποστηρίζει ότι σε περιπτώσεις όπου το έμβρυο διαγιγνώσκεται με ένα σύνδρομο η κατάσταση του οποίου μέσα από διαρκείς μελέτες καλείται μη αναστρέψιμη τότε είναι καλύτερα να διακόπτεται η κύηση. Με τον τρόπο αυτό αποφεύγεται ίσως μια «βασανιστική» και γεμάτη δυσκολίες για το έμβρυο ζωή. Σε ορισμένες περιπτώσεις η συνέχεια της κύησης εγκυμονεί κινδύνους και για την ίδια τη μητέρα. Μια άλλη άποψη υποστηρίζει ότι η διαδικασία διακοπής της κύησης ίσως ξεπερνά τα ανθρωπιστικά όρια. Με μία τέτοιου είδους κίνηση αφαιρούνται ουσιαστικά

το δικαίωμα για ζωή και για όνειρα ενός ατόμου (αγέννητου τη στιγμή εκείνη) παρά τη θέλησή του. Σε καταστάσεις σαν αυτές, οι αποφάσεις λαμβάνονται εξαιρετικά δύσκολα. Η ενημέρωση σε τέτοιου είδους ζητήματα είναι καλύτερα να πηγάζει από την ίδια την κοινωνία μέσω προγραμμάτων συμβουλευτικής γενετικής. Φυσικά, το ιατρικό απόρρητο θα πρέπει να ισχύει τόσο στην περίπτωση γενετικών εξετάσεων (και των αποτελεσμάτων τους) όσο και στην περίπτωση χρήσης σχετικής θεραπείας.

Συμπερασματικά αντιλαμβάνεται κανείς ότι με τη συνεχή έρευνα σε οποιαδήποτε επιστήμη η δίψα για γνώση παραμένει άσβεστη. Η γνώση είναι δύναμη. Τη δύναμη αυτή ο άνθρωπος οφείλει να τη διαχειρίζεται σωστά και με σοφία στοχεύοντας πάντα στο συνολικό όφελος. Οι ισορροπίες είναι λεπτές και μπορεί κανείς να τις υπερβεί πολύ εύκολα ειδικά σε περιπτώσεις όπου ο απώτερος σκοπός είναι το χρηματικό κέρδος. Η ανθρωπιά αποτελεί το κλειδί ώστε η εξέλιξη να έχει μονάχα θετικό αντίκτυπο στην κοινωνία που ζούμε.

Βιβλιογραφία

[Ελληνική]

Αλεπόρου – Μαρίνου Β., Αργυροκαστρίτης Α., Κομητοπούλου Α., Πιαλόγλου Π., Σγουρίτσα Β., Βιολογία (Θετικής κατεύθυνσης Γ' τάξης Γενικού Λυκείου), Ινστιτούτο Τεχνολογίας Υπολογιστών και Εκδόσεων «Διόφαντος», Ελλάδα, 2013.

Γιαλούρης Π., Μποσινάκου Κ., Σιδέρης Δ., Βιοχημεία (Τεχνολογικής κατεύθυνσης Γ' τάξης Γενικού Λυκείου), Οργανισμός Εκδόσεων Διδακτικών Βιβλίων, Αθήνα, Ελλάδα.

Καναβάκης Ε., Ξαϊδάρα Α., Η χαρτογράφηση του ανθρώπινου γονιδιώματος, Εργαστήριο Ιατρικής Γενετικής, Πανεπιστήμιο Αθηνών, Νοσοκομείο Παιδων «Αγία Σοφία», Αρχεία Ελληνικής Ιατρικής 2001, Αθήνα, Ελλάδα.

Μπάγκος Παντελής Γ., Βιοπληροφορική, Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών, Εθνικό Μετσόβιο Πολυτεχνείο, Ελλάδα, 2015.

Νανά Ευτυχία Σ., Διπλωματική Εργασία: Η χρήση της Βιοπληροφορικής στη μελέτη της νόσου του Καρκίνου, Αθήνα, Ελλάδα, 2009.

Σούμπαση – Γρίβα Β., Χατζησεβαστού – Λουκίδου Χ., Συμβουλευτική Γενετική, Έκδοση: C.City Publish, Θεσσαλονίκη, Ελλάδα, 2014.

[Ξενόγλωσση]

Alberts Bruce, Bray Dennis, Hopkin Karen, Johnson Alexander, Lewis Julian, Raff Martin, Roberts Keith, Walfer Peter, Essential cell biology: second edition, Garland Science Taylor & Francis Group.

Altschul, S. F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990), Basic local alignment search tool, J Mol Biol.

Durbin Richard, Sean Eddy R., Krogh Anders, Graeme Mitchison, Biological sequence analysis: Probabilistic models of proteins and nucleic acids, Cambridge University Press, 1998.

Feng, D. F., & Doolittle, R. F. (1987), Progressive sequence alignment as a prerequisite to correct phylogenetic trees, Journal of molecular evolution.

Hagen Joel B., The origins of bioinformatics, Nature Reviews | Genetics, Macmillan Magazines Ltd, 2000.

Hochreiter Sepp, Bioinformatics I: Sequence Analysis and Phylogenetics, Sepp Hochreiter, 2008.

Jones Neil C., Pevzner Pavel A., An Introduction to Bioinformatics Algorithms, A Bradford Book, The MIT Press, Cambridge, Massachusetts, London, England, 2004.

Luscombe, N. M., Greenbaum, D., & Gerstein, M. (2001), What is Bioinformatics? A proposed definition and overview of the field, Methods Inf Med.

Mount David W., Bioinformatics: Sequence and Genome Analysis, Cold Spring Harbor Laboratory Press, New York, 2001.

Needleman, S. B., & Wunsch, C. D. (1970), A general method applicable to the search for similarities in the amino acid sequence of two proteins, J Mol Biol.

Pearson, W.R., Lipman, D.J. (1988), Improved tools for biological sequence comparison, Proc Natl Acad Sci.

Posada David, Bioinformatics for DNA Sequence Analysis, Humana Pres, a part of Springer Science + Business Media, LLC 2009.

Russell Peter J., iGenetics: A Mendelian Approach, Pearson/Benjamin Cummings, 2006.

Smith, T. F., & Waterman, M. S. (1981), Identification of common molecular subsequences, J Mol Biol.

Waterman, M. S. (1995), *Introduction to Computational Biology*, Chapman and Hall, London.

Watson James D., *The double helix: first Edition*, 1968.

[Διαδικτυακές Πηγές]

<https://www.genome.gov/25019999/>

http://www.genomenewsnetwork.org/resources/whats_a_genome/Chp2_2.shtml

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.134.8387&rep=rep1&type=pdf>

<https://sites.google.com/site/edubioinformatics/istorikh-anadromh>

<http://www.nature.com/nature/journal/v431/n7011/full/nature03001.html>

<https://repository.kallipos.gr/>

<https://www.britannica.com/event/Human-Genome-Project>

<http://www.encyclopedia.com/science-and-technology/biology-and-genetics/genetics-and-genetic-engineering/human-genome-project>

<http://www.yourgenome.org/facts/what-is-clone-by-clone-sequencing>

<http://www.yourgenome.org/facts/what-is-shotgun-sequencing>