

Βιοπληροφορική Προσέγγιση

Δημήτριος Σ. Κανάκογλου
Ανδρομάχη Παμπάλου

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

«Μυοσκελετική Ογκολογία:
Διάγνωση – Θεραπεία – Έρευνα»



Δομή Παρουσίασης

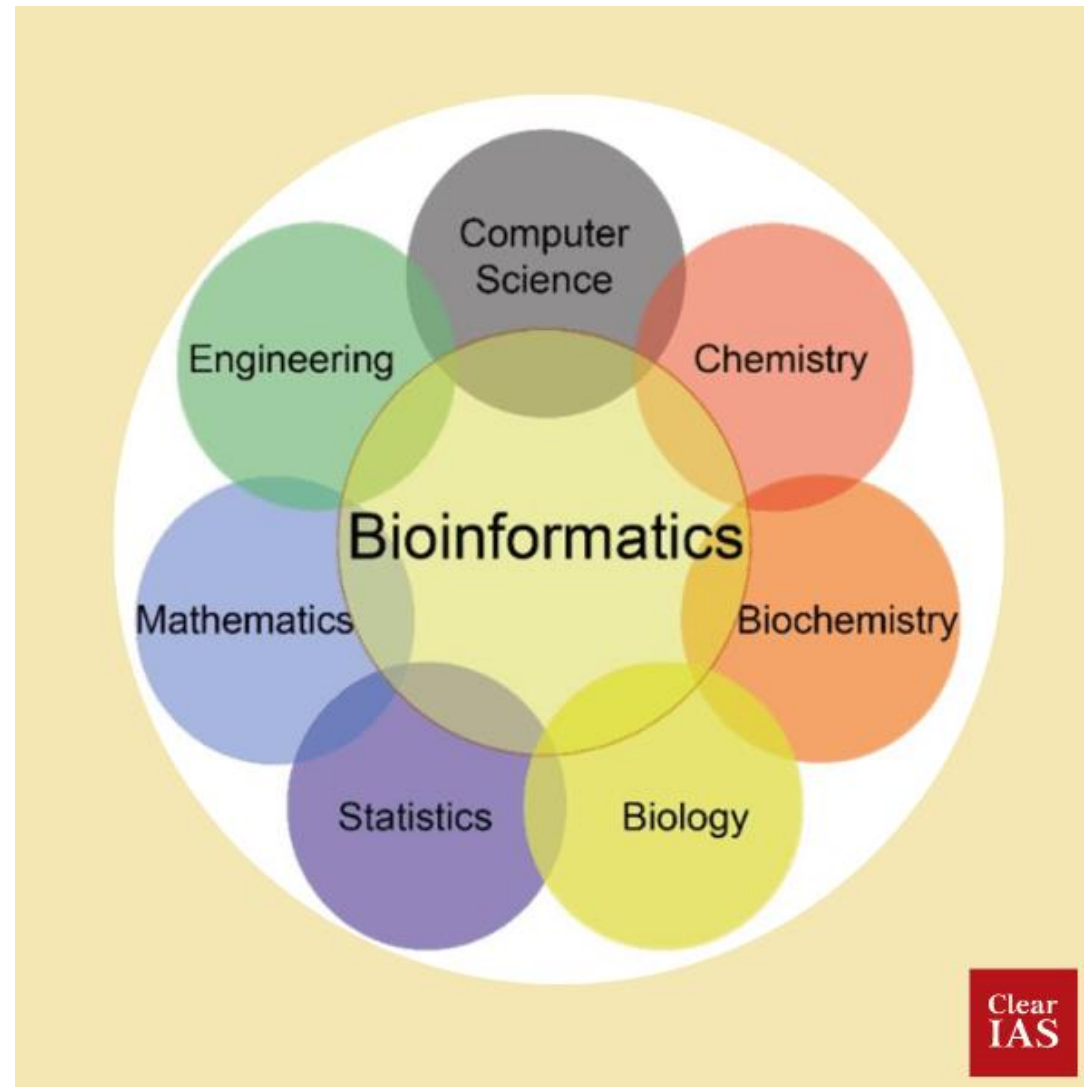
- Εισαγωγή στην Βιοπληροφορική Προσέγγιση
- Διεπιστημονικότητα της Σύγχρονης Έρευνας και ο Ρόλος της Βιοπληροφορικής
- Τομείς Εφαρμογής της Βιοπληροφορικής
 - Αλγόριθμοι Ευθυγράμμισης και Στοίχισης Ακολουθιών
 - Φυλογενετική
 - Τεχνητή Νοημοσύνη και Προβλεπτικές Μέθοδοι
 - Δομική Βιοπληροφορική
 - Γλώσσες Προγραμματισμού και Συστήματα
 - Βιολογία Συστημάτων
- Πρακτική Εφαρμογή της Βιοπληροφορικής στον Τομέα της Μυοσκελετικής Ογκολογίας
- Βιβλιογραφία και Χρήσιμοι Σύνδεσμοι

Εισαγωγή στην Βιοπληροφορική Προσέγγιση (I)

- Η βιοπληροφορική είναι ο επιστημονικός τομέας που συνδυάζει τη βιολογία, την επιστήμη των υπολογιστών (computer science) και την τεχνολογία της πληροφορικής (information technology) για την ανάλυση και ερμηνεία βιολογικών δεδομένων (όπως οι γενομικές αλληλουχίες). Πρόκειται για έναν διεπιστημονικό τομέα, ο οποίος περιλαμβάνει ένα ευρύ φάσμα θεμάτων και τεχνικών, όπως:
 - i. Ανάλυση αλληλουχιών: η μελέτη της δομής και της λειτουργίας των αλληλουχιών DNA, RNA και πρωτεϊνών.
 - ii. Ομικές τεχνολογίες: γονιδιωματική – μεταγραφωμική – πρωτεομική – επιγονιδιωματική.
 - iii. Βιολογία συστημάτων: η μελέτη βιολογικών συστημάτων και δικτύων σε μοριακό ή κυτταρικό επίπεδο και σε επίπεδο οργανισμών ή πληθυσμών.
 - iv. Ανακάλυψη και ανάπτυξη φαρμάκων: η χρήση της βιοπληροφορικής για τον εντοπισμό νέων φαρμακευτικών στόχων και για το σχεδιασμό και τη βελτιστοποίηση φαρμάκων.
 - v. Εξελικτική βιολογία: η χρήση της βιοπληροφορικής για τη μελέτη της εξέλιξης των οργανισμών και των γονιδιωμάτων τους.
- Ο τομέας της βιοπληροφορικής εξελίσσεται με ταχείς ρυθμούς, λόγω της ραγδαίας προόδου της τεχνολογίας, όπως των NGS (next-generation sequencing) τεχνολογιών και του AI (artificial intelligence), καθώς και της διαθεσιμότητας βιολογικών δεδομένων μεγάλης κλίμακας (big data).



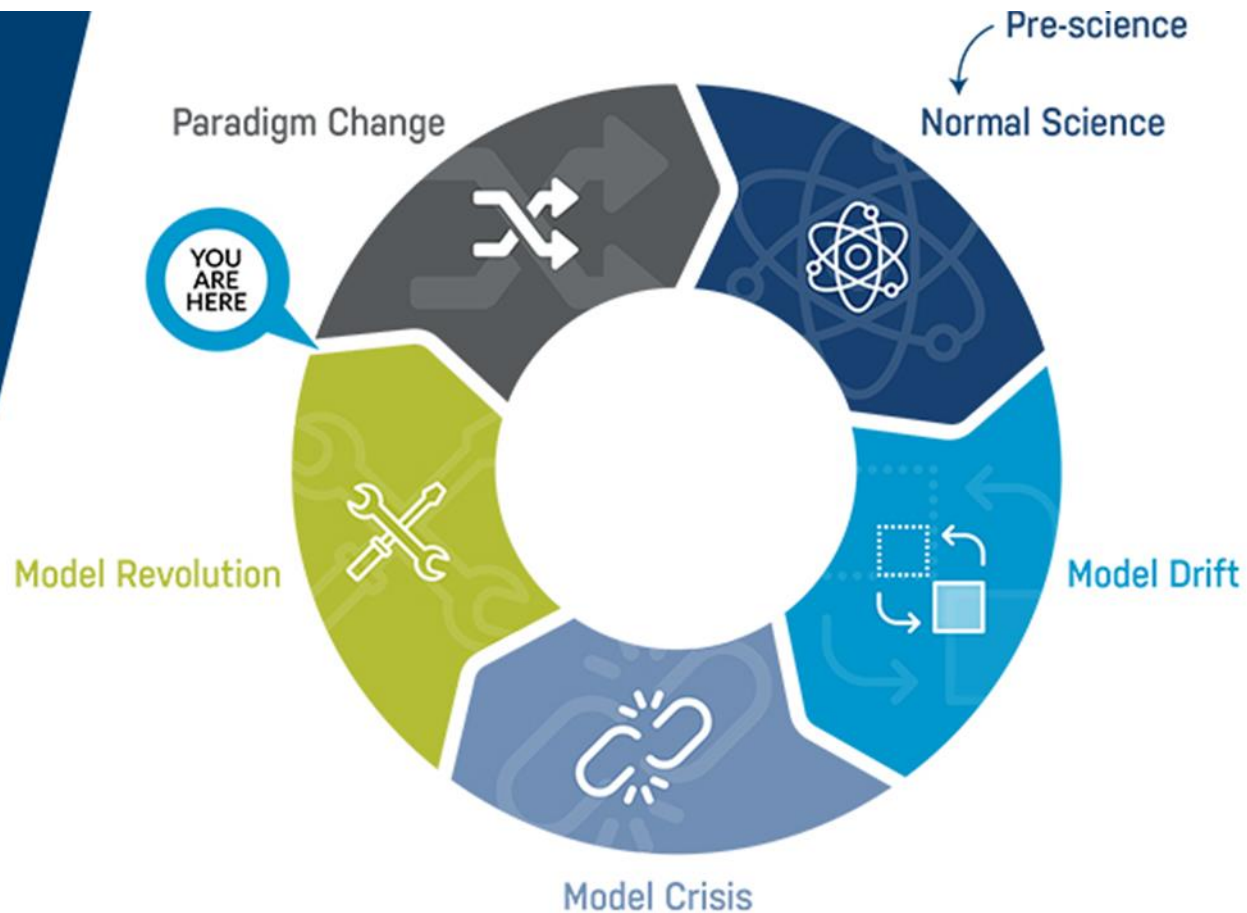
Εισαγωγή στην Βιοπληροφορική Προσέγγιση (II)



Διεπιστημονικότητα της Σύγχρονης Έρευνας και ο Ρόλος της Βιοπληροφορικής (I)

- Η σύγχρονη έρευνα αποκτά ολοένα και περισσότερο διεπιστημονικά χαρακτηριστικά, καθώς επιστήμονες από διαφορετικούς τομείς συνεργάζονται για να αντιμετωπίσουν πολύπλοκα προβλήματα που δεν μπορούν να επιλυθούν από κανέναν μεμονωμένο κλάδο. Η βιοπληροφορική είναι ένα χαρακτηριστικό παράδειγμα, καθώς συγκεντρώνει εμπειρογνωμοσύνη από τη βιολογία, την επιστήμη των υπολογιστών και την τεχνολογία της πληροφορικής για την ανάλυση και την ερμηνεία βιολογικών δεδομένων μεγάλης κλίμακας.
- Ένας από τους βασικούς τρόπους με τους οποίους η βιοπληροφορική συμβάλλει στη διεπιστημονική έρευνα, είναι η παροχή εργαλείων και μεθόδων για την ανάλυση και την οπτικοποίηση δεδομένων. Για παράδειγμα, βιοπληροφορικά λογισμικά μπορεί να χρησιμοποιηθούν για να αναλύσουν αλληλουχίες DNA, να αναγνωρίσουν γονίδια και μεταλλάξεις, να προβλέψουν τη δομή και τη λειτουργία των πρωτεϊνών και να μελετήσουν τις αλληλεπιδράσεις μεταξύ διαφορετικών μορίων σε επίπεδο συστήματος. Αυτό δίνει τη δυνατότητα σε ερευνητές από διαφορετικούς τομείς, όπως η γενετική, η βιοχημεία και η ιατρική, να κάνουν νέες ανακαλύψεις και να αναπτύξουν νέες θεραπείες.
- Μια άλλη σημαντική πτυχή της βιοπληροφορικής είναι η ανάπτυξη βάσεων δεδομένων που καθιστούν τα βιολογικά δεδομένα μεγάλης κλίμακας δημοσίως διαθέσιμα σε ερευνητές σε όλο τον κόσμο. Αυτό επιτρέπει σε επιστήμονες από διαφορετικούς κλάδους να έχουν πρόσβαση, να αναλύουν και να συγκρίνουν δεδομένα από διαφορετικούς οργανισμούς και να κάνουν νέες συνδέσεις και ανακαλύψεις που δεν θα ήταν δυνατές αν εργαζόντουσαν μεμονωμένα (π.χ., TCGA, GWAS).

Διεπιστημονικότητα της Σύγχρονης Έρευνας και ο Ρόλος της Βιοπληροφορικής (II)



- Τα πέντε στάδια της αλλαγής παραδείγματος:
 - i. Κανονική επιστήμη: Το κοινό (τρέχον) μοντέλο κατανόησης και επίλυσης προβλημάτων για οποιονδήποτε επιστημονικό κλάδο.
 - ii. Ολίσθηση του τρέχοντος μοντέλου: Ο επιστημονικός κλάδος αρχίζει να αντιμετωπίζει νέα ερωτήματα ή ανωμαλίες που το τρέχον μοντέλο δεν μπορεί να αντιμετωπίσει.
 - iii. Κρίση τρέχοντος μοντέλου: Το τρέχον μοντέλο δεν μπορεί πλέον να εφαρμοστεί σε όλα τα προβλήματα και οι ανεπίλυτες ανωμαλίες εμφανίζονται σε μεγαλύτερο βαθμό.
 - iv. Επιστημονική επανάσταση: Ένας ή περισσότεροι υποψήφιοι εμφανίζονται για να επιλύσουν την κρίση και ο κλάδος αγωνίζεται με ένα νέο μοντέλο.
 - v. Αλλαγή παραδείγματος: Το νέο μοντέλο καθιερώνεται από τους πρώτους ισχυρούς υποστηρικτές του και γίνεται «κανονική επιστήμη».

Τομείς Εφαρμογής της Βιοπληροφορικής

Αλγόριθμοι Ευθυγράμμισης και Στοίχισης Ακολουθιών (I)

- Στη βιοπληροφορική, η ευθυγράμμιση ακολουθιών είναι η διαδικασία διάταξης δύο ή περισσότερων ακολουθιών DNA/ RNA ή αμινοξέων σε μια συγκεκριμένη σειρά, έτσι ώστε οι παρόμοιες περιοχές να τοποθετούνται η μια κάτω από την άλλη. Διαφορετικοί αλγόριθμοι μπορούν να χρησιμοποιηθούν για την ευθυγράμμιση ακολουθιών και ο καθένας έχει τα δικά του πλεονεκτήματα και αδυναμίες. Μερικοί από τους πιο συχνά χρησιμοποιούμενους αλγορίθμους ευθυγράμμισης ακολουθιών είναι οι εξής:
 - i. Ο αλγόριθμος Needleman-Wunsch: Πρόκειται για έναν αλγόριθμο ολικής στοίχισης που συγκρίνει δύο αλληλουχίες στο σύνολό τους και βρίσκει την καλύτερη ολική στοίχιση. Βασίζεται στον δυναμικό προγραμματισμό και χρησιμοποιείται για τη σύγκριση αλληλουχιών διαφορετικού μήκους και για την εύρεση συντηρημένων περιοχών.
 - ii. Αλγόριθμος Smith-Waterman: Πρόκειται για έναν αλγόριθμο τοπικής στοίχισης που συγκρίνει δύο αλληλουχίες και βρίσκει την καλύτερη στοίχιση για μια συγκεκριμένη περιοχή ενδιαφέροντος. Βασίζεται επίσης στον δυναμικό προγραμματισμό και χρησιμοποιείται για την εύρεση ομοιοτήτων εντός αλληλουχιών ίδιου μήκους.
 - iii. BLAST (Basic Local Alignment Search Tool): Πρόκειται για έναν ευρέως χρησιμοποιούμενο αλγόριθμο για την αναζήτηση ομοιοτήτων. Βασίζεται στην ευρετική προσέγγιση, χρησιμοποιεί έναν πίνακα βαθμολόγησης για τη σύγκριση των ακολουθιών και χρησιμοποιείται για τον γρήγορο εντοπισμό παρόμοιων ακολουθιών σε μεγάλες βάσεις δεδομένων.
 - iv. Clustal: Πρόκειται για έναν αλγόριθμο ευθυγράμμισης πολλαπλών ακολουθιών που ευθυγραμμίζει πολλές ακολουθίες ταυτόχρονα. Βασίζεται σε μια προοδευτική μέθοδο και χρησιμοποιείται για την ευθυγράμμιση αλληλουχιών από συγγενείς οργανισμούς.
 - v. T-Coffee: Πρόκειται για έναν αλγόριθμο ευθυγράμμισης πολλαπλών ακολουθιών που συνδυάζει τα πλεονεκτήματα διαφορετικών μεθόδων, βασίζεται σε αλγορίθμους συναίνεσης και χρησιμοποιείται για την ευθυγράμμιση ακολουθιών από οργανισμούς που έχουν μακρινή συγγένεια.
 - vi. MUSCLE (Multiple Sequence Comparison by Log-Expectation): Πρόκειται για έναν αλγόριθμο ευθυγράμμισης πολλαπλών ακολουθιών που βασίζεται σε μια προοδευτική προσέγγιση και χρησιμοποιεί μια ποικιλία ευρετικών μεθόδων για τη βελτίωση της ακρίβειας της ευθυγράμμισης. Χρησιμοποιείται για την ευθυγράμμιση μεγάλου αριθμού αλληλουχιών από απομακρυσμένους οργανισμούς με μακρινή συγγένεια.
- Αυτά είναι μερικά μόνο παραδείγματα από τους πολλούς διαφορετικούς αλγορίθμους στοίχισης ακολουθιών που είναι διαθέσιμοι στα χέρια των ερευνητών. Η επιλογή του αλγορίθμου εξαρτάται από το συγκεκριμένο ερευνητικό ερώτημα, τον τύπο του ερωτήματος, το μέγεθος των αλληλουχιών που συγκρίνονται, καθώς και τους διαθέσιμους υπολογιστικούς πόρους.



Τομείς Εφαρμογής της Βιοπληροφορικής

Αλγόριθμοι Ευθυγράμμισης και Στοίχισης Ακολουθιών (II)

- Το BLAST είναι ένα από τα πιο ευρέως χρησιμοποιούμενα εργαλεία για τη σύγκριση αλληλουχιών. Είναι ένας εξαιρετικά γρήγορος αλγόριθμος που επιτρέπει στους ερευνητές να συγκρίνουν μια αλληλουχία ενδιαφέροντος σε μια μεγάλη βάση δεδομένων, προκειμένου να εντοπίσουν παρόμοιες αλληλουχίες.
 - ❑ Χρησιμοποιείται ευρέως σε διάφορους τομείς, όπως στην ανακάλυψη γονιδίων, στη λειτουργική γονιδιωματική, στη φυλογενετική κ.ο.κ.
 - ❑ Χρησιμοποιείται επίσης σε πολλούς άλλους τομείς, όπως στην ιατρική, στη γεωπονία κ.ο.κ.
- Ο Clustal είναι ένας ευρέως χρησιμοποιούμενος αλγόριθμος πολλαπλή στοίχισης.
 - ❑ Χρησιμοποιείται για την ευθυγράμμιση αλληλουχιών συγγενών οργανισμών και λόγω αυτής της λειτουργίας του έχει χρησιμοποιηθεί σε πολλούς ερευνητικούς τομείς όπως στη φυλογενετική, στην πρόβλεψη πρωτεϊνικών δομών, στη λειτουργική γονιδιωματική κ.ο.κ.
- Οι δύο αυτοί αλγόριθμοι είναι διαθέσιμοι ως εργαλεία γραμμής εντολών και μπορούν εύκολα να ενσωματωθούν σε αυτοματοποιημένες βιοπληροφορικές ροές.

CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix ...

JD Thompson, [DG Higgins](#), [TJ Gibson](#) - Nucleic acids research, 1994 - academic.oup.com

... alignments are now carried out using the 'progressive' approach of Feng and Doolittle (1). In ... to the progressive multiple alignment method which greatly improve the sensitivity without ...

☆ Save 📄 Cite Cited by 71716 Related articles All 48 versions

Basic local alignment search tool

[SF Altschul](#), [W Gish](#), [W Miller](#), [EW Myers](#)... - Journal of molecular ..., 1990 - Elsevier

... A new approach to rapid sequence comparison, basic local alignment search tool (BLAST), directly approximates alignments that optimize a measure of local similarity, the maximal ...

☆ Save 📄 Cite Cited by 103718 Related articles All 80 versions



Τομείς Εφαρμογής της Βιοπληροφορικής

Αλγόριθμοι Ευθυγράμμισης και Στοίχισης Ακολουθιών (III)

- Παράδειγμα χρήσης του αλγόριθμου Clustal σε δεδομένα ακολουθιών 16s βακτηρίων.

```
Rhodopirellula -----CG-----TGGCTTGACGGGCGGTGTGTACAAGGCTCAG
Rhodococcus -----
Aquifex ACGCTCCCGCCTGGGGATGGGCCCCGCGGCCATCAGGTAGTTGGTGGGGTAACGGCCAC
Thermus CTTTGCCCGCTTCCGGATGGGCCCCGCGTCCCATCAGCTAGTTGGTGGGGTAATGGCCAC
Thermotoga GTTTGCCCGCCGAGGAGGGGCCCGCGGC-CATCAGGTAGTTGGTGGGGTAACGGCCAC
Microbacterium GATTTTTCGGTGGGGATGGGCTCGCGGCCTATCAGCTTGTGGCGAGGTAAATGGCTCAC
Arthrobacter GATTTTTGGTCAGGGATGAACTCGCGGCCTATCAGCTTGTGGTGAAGTAATGGCTCAC
Cellulosimicrobium AGTTTTTCGGTCAGGGATGGGCTCGCGGCCTATCAGCTTGTGGTGGGGTATGGCCTAC
Corynebacterium CTTTATGCGGTGTGGGATGAGCCTGCGGCCTATCAGCTTGTGGTGGGGTAATGGCCTAC
Nocardia GATTTATCGGTGCAGGATGGGCCCCGCGGCCTATCAGCTTGTGGTGGGGTAATGGCCTAC
IHBB GATTTATCGGTGCAGGATGGGCCCCGCGGCCTATCAGCTTGTGGTGGGGTAATGGCCTAC
cqsV23 GATTTATCGGTGCAGGATGGGCCCCGCGGCCTATCAGCTTGTGGTGGGGTAATGGCCTAC
Mycobacterium GCTTTTGCGGTGTGGGATGGGCCCCGCGGCCTATCAGCTTGTGGTGGGGTATGGCCTAC
Gordonia GCTTTTGCGGTTAGGATGGGCCCCGCGGCCTATCAGCTTGTGGTGGGGTAATGGCCTAC
MTR-11 GCTTTTGCGGTTAGGATGGGCCCCGCGGCCTATCAGCTTGTGGTGGGGTAATGGCCTAC
Fusobacterium CTATATGCGCTGTGAGAGAGCTTTCGCTCCCATTAGCTAGTTGGAGAGGTAAACGGCTCAC
Chlamydia GACCTTTTCGGTTAAGGGAGAGTCTATGTGATATCAGCTAGTTGGTGGGGTAAAAGCCTAC
Dehalococcoides CGTAAGGCGCTTGGTGAAGGGCTTTCGCTCCGATTAGCTAGTTGGTGGGGTAACGGCCTAC
Synechococcus CTTTTTTGGCCTGGGGATGAGCTCGCGGTGGATTAGCTAGTTGGTGGGGTAATGGCCTAC
Brevibacillus -----
Paenibacillus AAGCTGTGGCTTACAGATGGGCTGCGGCGCATTAGCTAGTTGGTGGGGTAACGGCTCAC
Bacillus CGGCTACCACTTACAGATGGACCCGCGGCGCATTAGCTAGTTGGTGAAGTGACGGCTCAC
Lysinibacillus CGGCTGTCGCTATAGGATGGGCCCCGCGGCGCATTAGCTAGTTGGTGAAGTAACGGCTCAC
Porphyromonas TATTTATAGCTGTAAGATAGGCATGCGTCCCATTAGCTGGTTGGTGAAGTAACGGCTCAC
Treponema ACGGCTCCGCTTACAGATGGGCCCCGCGTCCCATTAGCTAGTTGGTGAAGTAAAAGGCCAC
Chlorobium GATTTATCGCCAAAAGATGAGCCTATATCCCATCAGGTAGTTGGTAGGGTAAAAGCCTAC
Fibrobacter GATTTATCGGTATGGGATGAGCTTTCGCTCCGATTAGCTAGTTGGCGGGGAACGGCCAC
Sulfurospirillum ACTCCGGTGCCAAGGGATGGGGCTGTATCCTATCAGCTTGTGGTGAAGTAATGGCTCAC
Nitrosomonas GACCTTGCCTAAAGGAGCGGCCGATGTCTGATTAGCTAGTTGGTGGGGTAAAAGCCTAC
Alcaligenes GACCTTGCCTAATTGGAGCGGCCGATATCGGATTAGCTAGTTGGTGGGGTAACGGCTCAC
Shewanella GGCCTTCCGCGATTGGATGAACCTAGGTGGGATTAGCTAGTTGGTGAAGTAATGGCTCAC
Serratia GGCCTCTGCCATCAGATGTGCCAGATGGGATTAGCTAGTAGGTGGGGTAATGGCTCAC
Pantoea -----
```

[HTML] [Deciphering the biodesulfurization potential of two novel *Rhodococcus* isolates from a unique Greek environment](#)

PD Glekas, O Martzoukou, ME Mastrodima... - AIMS ..., 2022 - aimspress.com

Sustainable **biodesulfurization** (BDS) processes require the use of microbial biocatalysts that display high activity against the recalcitrant heterocyclic sulfur compounds and can ...

Τομείς Εφαρμογής της Βιοπληροφορικής

Φυλογενετική (I)

- Η φυλογενετική είναι ένα πεδίο της βιοπληροφορικής που ασχολείται με τη μελέτη των εξελικτικών σχέσεων μεταξύ διαφορετικών οργανισμών.
- Ο στόχος της φυλογενετικής είναι να συμπεράνει την εξελικτική ιστορία διαφορετικών ειδών, να κατανοήσει πώς σχετίζονται μεταξύ τους και πώς έχουν εξελιχθεί με την πάροδο του χρόνου.
- Η φυλογενετική χρησιμοποιεί μια ποικιλία εργαλείων και τεχνικών, όπως η ευθυγράμμιση αλληλουχιών, η εύρεση γονιδίων, η συναρμολόγηση γονιδιώματος (*de novo*) και η κατασκευή φυλογενετικών δέντρων.
 - Ο πιο κοινός τύπος δεδομένων που χρησιμοποιείται στη φυλογενετική είναι οι αλληλουχίες DNA και πρωτεϊνών, αλλά μπορούν επίσης να χρησιμοποιηθούν και άλλοι τύποι δεδομένων.
 - Τα φυλογενετικά δέντρα χρησιμοποιούνται για να αναπαραστήσουν τις εξελικτικές σχέσεις μεταξύ διαφορετικών οργανισμών. Αυτά τα δέντρα κατασκευάζονται χρησιμοποιώντας διαφορετικούς αλγόριθμους, όπως maximum parsimony, maximum likelihood και Bayesian inference, που λαμβάνουν υπόψη τις ομοιότητες και τις διαφορές μεταξύ διαφορετικών ακολουθιών.
 - Το δέντρο που προκύπτει είναι ένα διάγραμμα διακλάδωσης που δείχνει τις εξελικτικές σχέσεις μεταξύ διαφορετικών ειδών, με τα πιο στενά συγγενικά είδη να διακλαδώνονται κοντινότερα.
- Η φυλογενετική έχει πολλές εφαρμογές στη βιολογία, την ιατρική και τη γεωργία:

- Μπορεί να χρησιμοποιηθεί για την κατανόηση της εξέλιξης των παθογόνων μικροοργανισμών
- Για το σχεδιασμό εμβολίων και φαρμάκων που στοχεύουν συγκεκριμένα παθογόνα στελέχη.
- Για την κατανόηση της εξέλιξης των καλλιεργούμενων φυτών και για το σχεδιασμό νέων ποικιλιών καλλιεργειών που είναι πιο ανθεκτικές σε ασθένειες και παράσιτα.

Rhodospirillum	-----CG-----TGGCTTGACGGGCGGTGTGTACAAGGCTCAG
Rhodococcus	
Aquifex	ACGGCTCCGCCCTGGGGAATGGCCCGGGCCCATCAGGTAGTGTGGTGGGATACGGCCAC
Thermus	CTTTGGCCGCTCCGGATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Thermotoga	GTTCGGCCCGGGGAGGGCCCGGGCCGCGCATAGAGTAGTGTGGTGGGATACGGCCAC
Microbacterium	GATTTTCGGTGGGGAATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Aerobacter	GATTTTTCGGTGGGGAATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Cellulomicrobium	AGTTTTCGGTGGGGAATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Cornebacterium	CTTTATCGGTGGGGAATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Nocardia	GATTTATCGGTGGGGAATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
IHB	GATTTATCGGTGGGGAATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
eqs23	GATTTATCGGTGGGGAATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Mycobacterium	GCTTTTGGGTGGGGAATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Gordonia	GCTTTTGGGTGGGGAATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
MFR-11	GCTTTTGGGTGGGGAATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Fusobacterium	CTTTATCGGTGGGGAATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Chlamydia	GACCTTCGGTAAAGGGAATCTATGTGATACAGCTAGTGTGGTGGGATACGGCCAC
Dentalococcus	GCTAAAGGCGCTGGTGGGGAATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Symbioboscus	CTTTTTCGGTGGGGAATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Brevibacillus	
Paenibacillus	AAAGCTGGCTTACAGATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Bacillus	CGGCTACCACTTACAGATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Lysinibacillus	CGGCTTCGGTAAAGGGAATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Porphyromonas	GATTTATCGGTGGGGAATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Treponema	ACGGCTCCGCTTCAGGATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Chlorobium	GATTTATCGGCAAGGGAATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Fibrobacter	GATTTATCGGTAAGGGAATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Sulfurospirillum	ACTCCGGTCCCAAGGGAATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Nitrosomonas	GACCTTCGGTAAAGGGAATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Alcaligenes	GACCTTCGGTAAAGGGAATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Shewanella	GGCCTTCGGGATTCAGGATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Serratia	GGCCTTCGGGATTCAGGATGGCCCGGGCCCATCAGCTAGTGTGGTGGGATACGGCCAC
Pantoea	

Τομείς Εφαρμογής της Βιοπληροφορικής

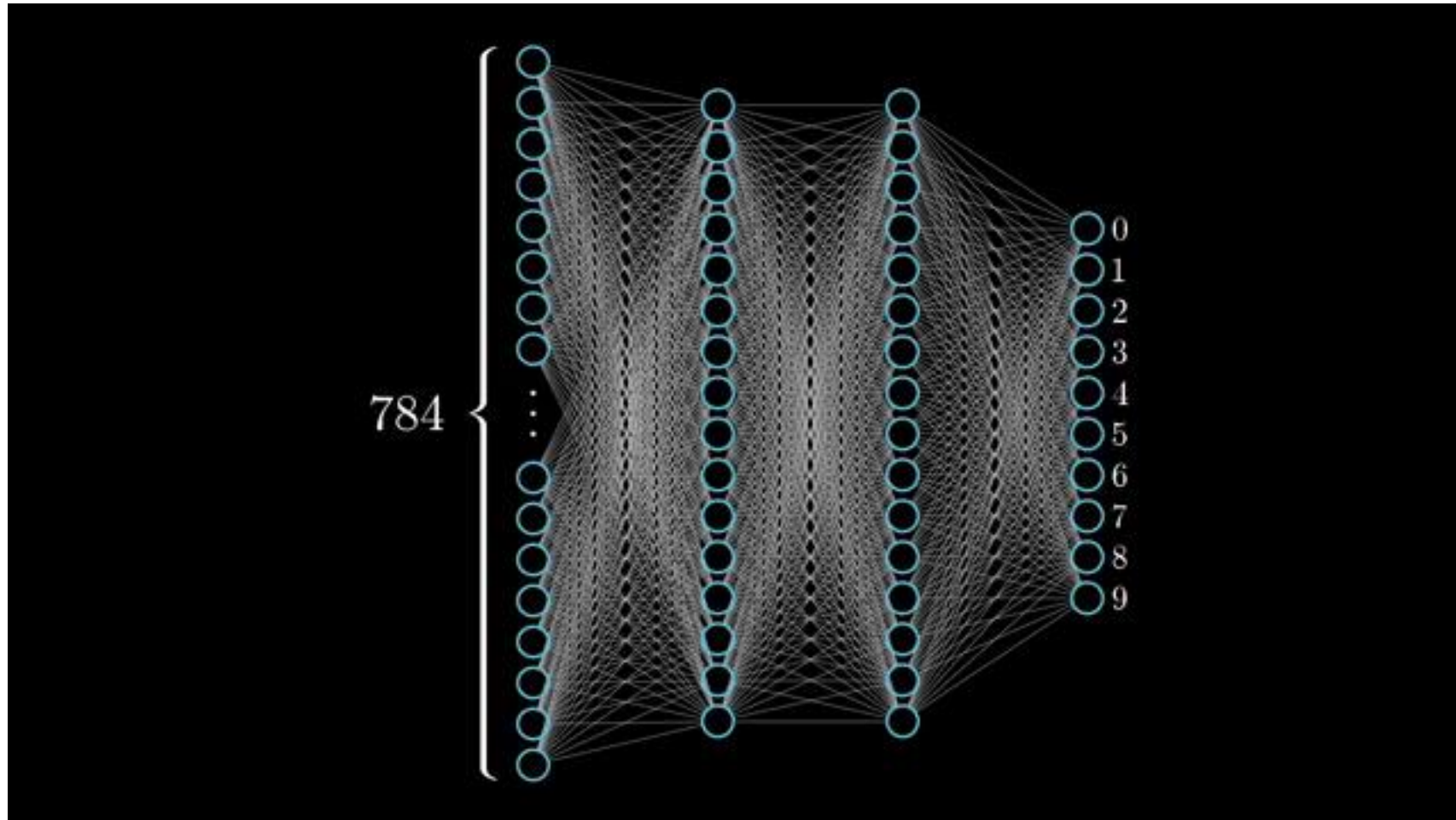
Τεχνητή Νοημοσύνη και Προβλεπτικές Μέθοδοι (I)

- Η τεχνητή νοημοσύνη στη βιοπληροφορική αναφέρεται στη χρήση τεχνικών και μεθόδων για την πραγματοποίηση προβλέψεων σχετικά με βιολογικά συστήματα. Οι τεχνικές αυτές περιλαμβάνουν τη μηχανική μάθηση, τη βαθιά μάθηση και τα νευρωνικά δίκτυα και μπορούν να εφαρμοστούν σε ένα ευρύ φάσμα, όπως στην ευθυγράμμιση ακολουθιών, στο σχολιασμός γονιδίων, στην πρόβλεψη πρωτεϊνικών δομών και στην ανακάλυψη φαρμάκων.
 - i. Τεχνητή νοημοσύνη: Η τεχνητή νοημοσύνη (artificial intelligence) είναι ένας ευρύς τομέας που περιλαμβάνει πολλές διαφορετικές τεχνικές, όπως η μηχανική μάθηση και τα νευρωνικά δίκτυα. Οι μέθοδοι αυτοί μπορούν να χρησιμοποιηθούν για την ανάλυση μεγάλων βιολογικών δεδομένων, όπως τα γονιδιωματικά δεδομένα, και για την πραγματοποίηση προβλέψεων σχετικά με τα βιολογικά συστήματα.
 - ii. Βαθιά μάθηση: Η βαθιά μάθηση (deep learning) είναι ένα υποσύνολο της μηχανικής μάθησης που χρησιμοποιεί νευρωνικά δίκτυα πολλαπλών επιπέδων για να «μαθαίνει» από δεδομένα. Τα μοντέλα βαθιάς μάθησης μπορούν να χρησιμοποιηθούν για την ανάλυση βιολογικών δεδομένων, όπως εικόνες και ακολουθίες.
 - iii. Computer vision: Το CV είναι ένας τομέας που ασχολείται με την ανάλυση εικόνων (MRI, μικροσκοπίου, ακτινογραφίες) και βίντεο. Οι μέθοδοι υπολογιστικής όρασης μπορούν να χρησιμοποιηθούν για την ανάλυση εικόνων βιολογικών δειγμάτων, όπως εικόνες μικροσκοπίας. Για παράδειγμα, οι μέθοδοι υπολογιστικής όρασης μπορούν να χρησιμοποιηθούν για την ανίχνευση, ταξινόμηση και καταμέτρηση κυττάρων σε εικόνες από φωτογραφικό μικροσκόπιο.
 - iv. Hidden Markov Models (HMM): Τα μαρκοβιανά μοντέλα είναι μαθηματικά μοντέλα που χρησιμοποιούνται ευρέως στη βιοπληροφορική για την ανάλυση ακολουθιών και συμπεριλαμβάνονται συχνά στη μηχανική μάθηση.

Τομείς Εφαρμογής της Βιοπληροφορικής

Τεχνητή Νοημοσύνη και Προβλεπτικές Μέθοδοι (II)

- Παράδειγμα αναγνώρισης χαρακτήρων από μοντέλο βαθιάς μάθησης (Deep Learning).



Τομείς Εφαρμογής της Βιοπληροφορικής

Δομική Βιοπληροφορική (I)

- Η δομική βιοπληροφορική είναι πεδίο της βιοπληροφορικής το οποίο ασχολείται με τη μελέτη της τρισδιάστατης δομής των βιολογικών μακρομορίων -όπως οι πρωτεΐνες και τα νουκλεϊκά οξέα- καθώς και με τον τρόπο που αλληλεπιδρούν μεταξύ τους και με άλλα μακρομόρια. Στόχος της δομικής βιοπληροφορικής είναι η κατανόηση των μοριακών μηχανισμών των βιολογικών διεργασιών σε ατομικό επίπεδο.
 - ❑ Η δομική βιοπληροφορική χρησιμοποιεί μια ποικιλία εργαλείων και τεχνικών, όπως την ευθυγράμμιση ακολουθιών, τη μοντελοποίηση ομολογίας και τη μοριακή δυναμική. Ο συνηθέστερος τύπος δεδομένων που χρησιμοποιούνται στη δομική βιοπληροφορική είναι τα δεδομένα κρυσταλλογραφίας ακτίνων-X και πυρηνικού μαγνητικού συντονισμού (Nuclear Magnetic Resonance - NMR), αλλά μπορούν επίσης να χρησιμοποιηθούν και άλλοι τύποι δεδομένων, όπως π.χ. δεδομένα κρυσταλλογραφίας μικροσκοπίας (cryo-EM).
 - ❑ Μια από τις βασικές εφαρμογές της δομικής βιοπληροφορικής είναι η ανακάλυψη και η ανάπτυξη φαρμάκων. Με την κατανόηση της δομής των πρωτεϊνικών στόχων και του τρόπου με τον οποίο αλληλεπιδρούν με άλλα μόρια, είναι δυνατόν να σχεδιαστούν νέα φάρμακα που στοχεύουν ειδικά αυτές τις πρωτεΐνες και αναστέλλουν ή εντείνουν τη λειτουργία τους.
 - ❑ Η δομική βιοπληροφορική διαδραματίζει επίσης σημαντικό ρόλο στην κατανόηση των μηχανισμών των ενζύμων, των αλληλεπιδράσεων μεταξύ πρωτεϊνών και των αλληλεπιδράσεων πρωτεΐνης-νουκλεϊκών οξέων.

Τομείς Εφαρμογής της Βιοπληροφορικής

Δομική Βιοπληροφορική (II)

- Η αμινοξική ακολουθία καθορίζει τη δομή των πρωτεϊνών και η δομή υπαγορεύει τη βιοχημική λειτουργία.
- Πρωτεΐνες που μοιράζονται δομικά στοιχεία και ομοιότητες καλούνται ομόλογες.
- Οι πρωτεΐνες που μοιράζονται παρόμοιες αμινοξικές ακολουθίες συνήθως επιτελούν παρόμοιες βιοχημικές λειτουργίες.
 - Αυτό μπορεί να ισχύει και όταν βρίσκονται σε οργανισμούς που έχουν μακρινή συγγένεια.
- Επί του παρόντος, ο προσδιορισμός της λειτουργίας μιας νεοανακαλυφθείσας πρωτεΐνης αρχίζει συνήθως με την αναζήτηση πρωτεϊνών που έχουν εντοπιστεί προηγουμένως και είναι παρόμοιες ως προς την πολυπεπτιδική τους αλυσίδα.



Τομείς Εφαρμογής της Βιοπληροφορικής

Δομική Βιοπληροφορική (III)



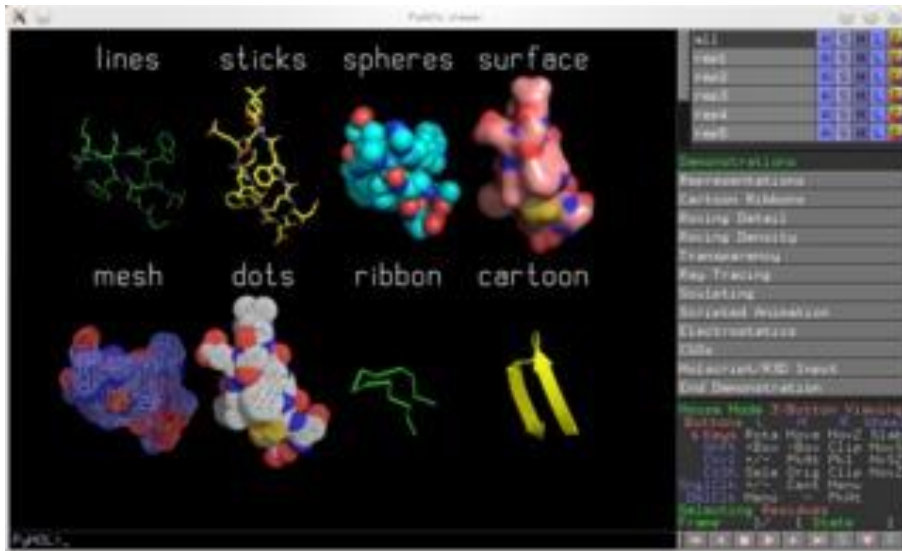
- Η μεγαλύτερη και πιο ενημερωμένη βάση δεδομένων με δεδομένα τρισδιάστατης δομής βιολογικών μορίων είναι η Protein Data Bank.
- Αναπτύσσεται με ρυθμό σχεδόν 10% ετησίως.
- Διαχειρίζεται τα Big Data ως παγκόσμιο δημόσιο αγαθό.
- Συνεισέφερε δεδομένα σε σχεδόν ένα εκατομμύριο δημοσιευμένες ερευνητικές εργασίες.

Molecular Type	X-ray	EM	NMR	Multiple methods	Neutron	Other	Total
Protein/Oligosaccharide	8,979	1,635	32	7	1	0	10,654
Protein/NA	8,032	2,919	281	6	0	0	11,238
Protein (only)	152,462	9,288	12,112	189	72	32	174,155
Other	163	8	31	0	0	0	202
Oligosaccharide (only)	11	0	6	1	0	4	22
Nucleic acid (only)	2,579	77	1,433	12	2	1	4,104
Total	172,226	13,927	13,895	215	75	37	200,375

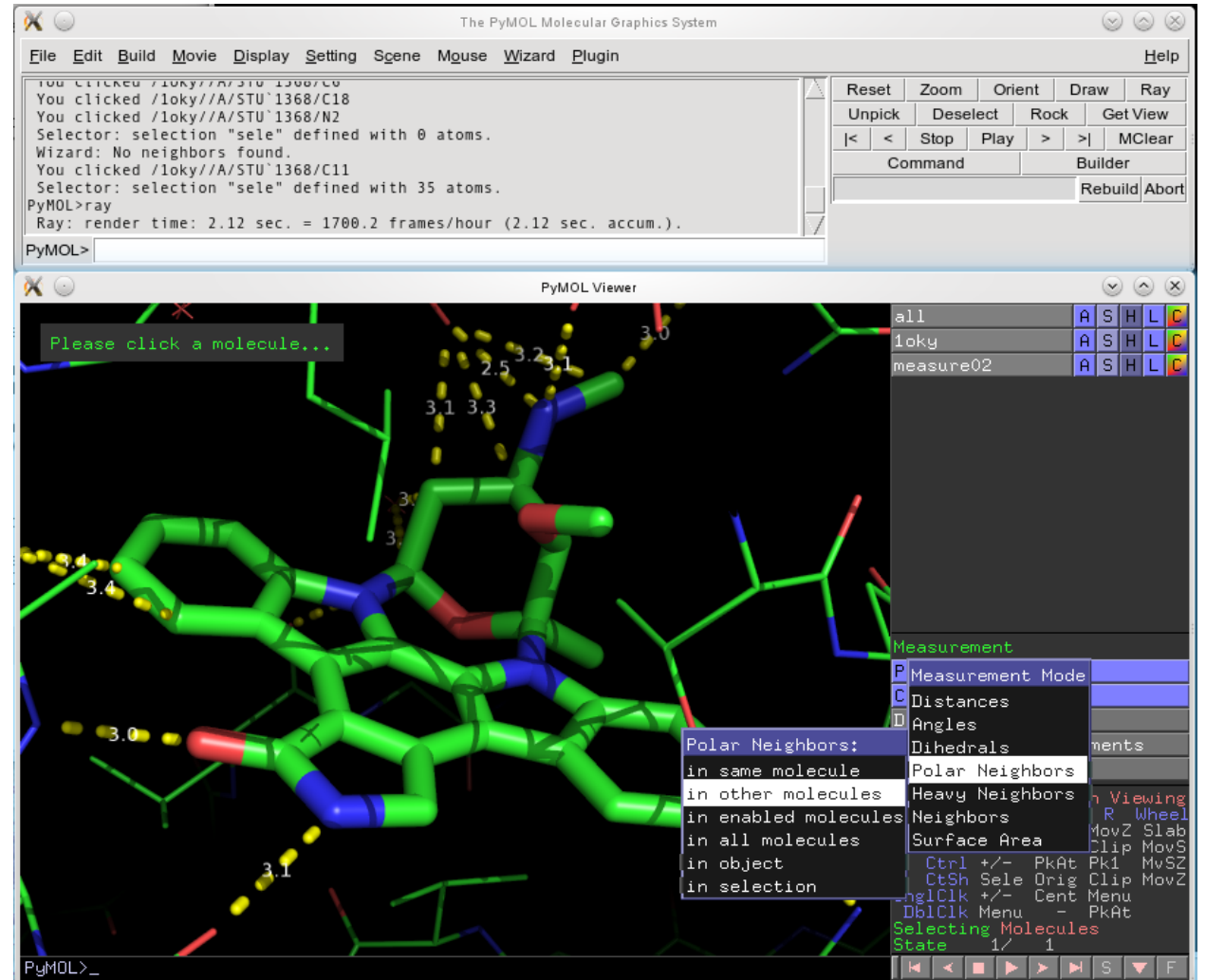
Τομείς Εφαρμογής της Βιοπληροφορικής

Δομική Βιοπληροφορική (IV)

- Η δομική απεικόνιση επιτρέπει στην αναγνώριση βιοχημικών λειτουργιών που ειδάλλως θα ήταν άγνωστες.



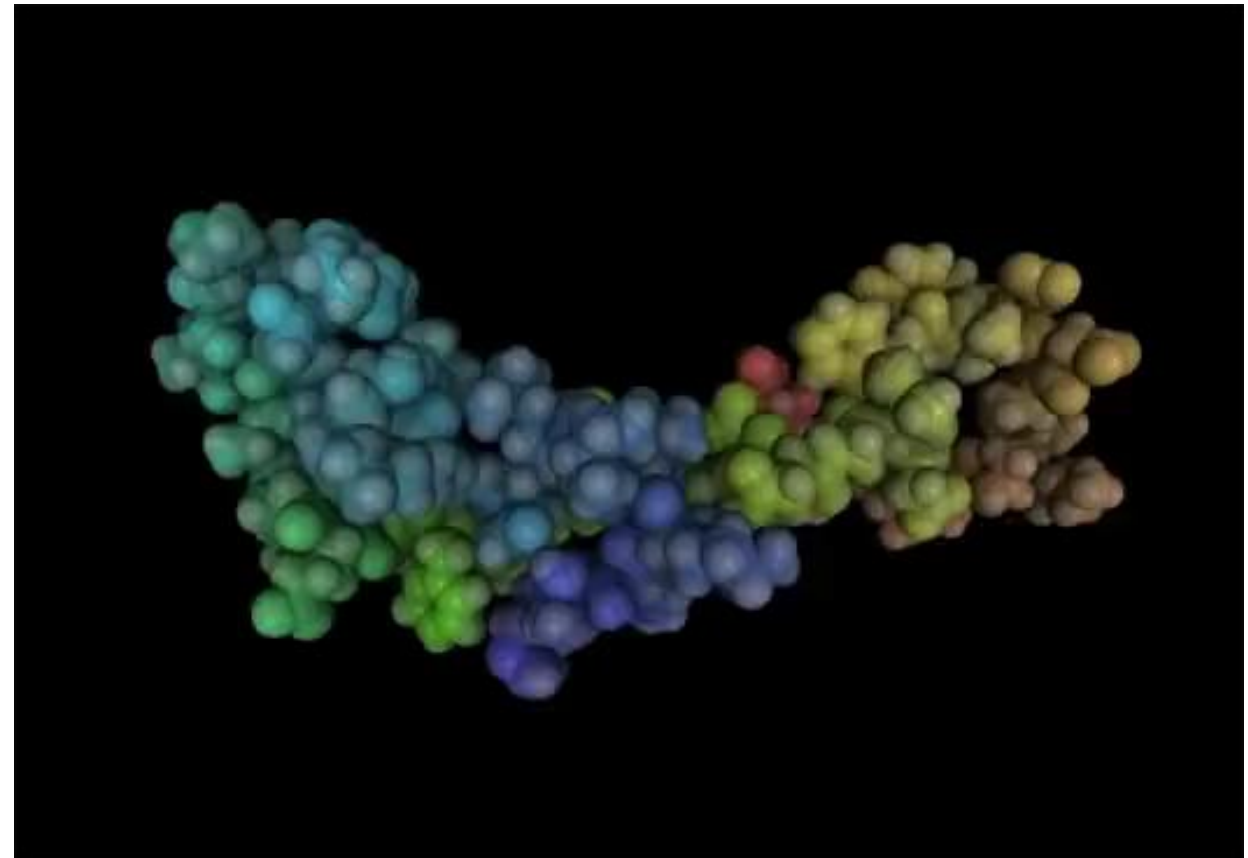
<https://www.pymol.org/pymol.html?>



Τομείς Εφαρμογής της Βιοπληροφορικής

Δομική Βιοπληροφορική (V)

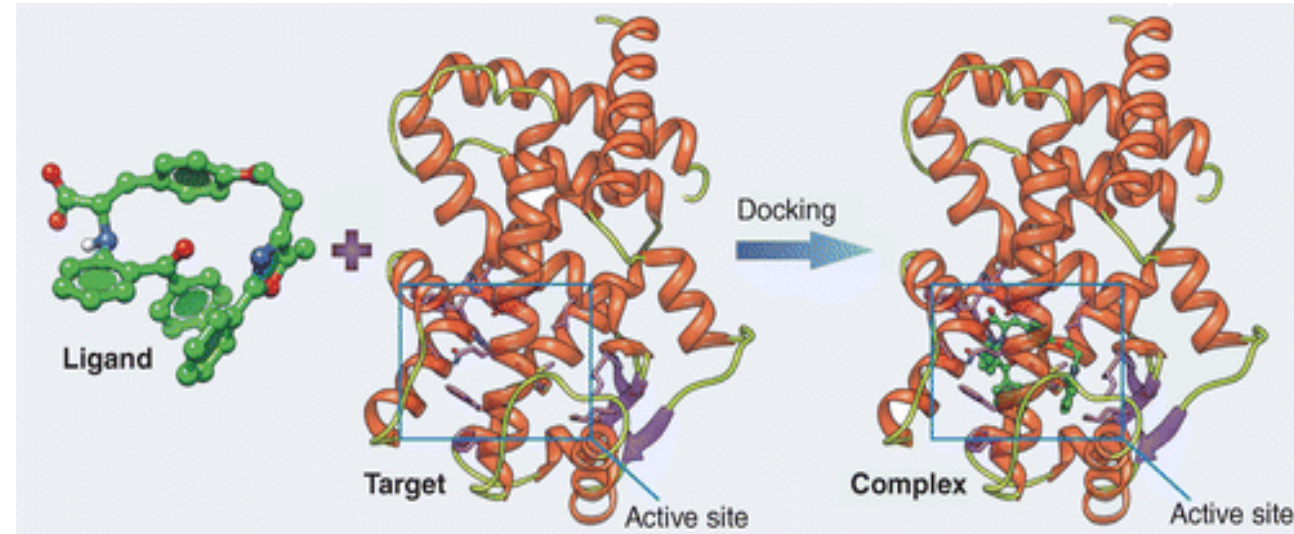
- Χρησιμοποιώντας τεχνικές δομικής βιοπληροφορικής, μπορούμε να κατανοήσουμε την τρισδιάστατη δομή των πρωτεϊνικών στόχων και τον τρόπο με τον οποίο αυτές αλληλεπιδρούν με άλλα μόρια. Αυτές οι πληροφορίες μπορούν στη συνέχεια να χρησιμοποιηθούν για τον σχεδιασμό νέων φαρμάκων που στοχεύουν ειδικά αυτές τις πρωτεΐνες και αναστέλλουν τη λειτουργία τους, προκειμένου να θεραπεύσουν ένα ευρύ φάσμα ασθενειών.



Τομείς Εφαρμογής της Βιοπληροφορικής

Δομική Βιοπληροφορική (VI)

- Η μοριακή πρόσδεση (ή μοριακός ελλιμενισμός / molecular docking) είναι μια υπολογιστική τεχνική που χρησιμοποιείται στη βιοπληροφορική για την πρόβλεψη των αλληλεπιδράσεων μεταξύ δύο μορίων, όπως μια πρωτεΐνη και ένα φάρμακο. Στόχος της μοριακής πρόσδεσης είναι η πρόβλεψη της τρισδιάστατης δομής του συμπλόκου που σχηματίζεται από τα δύο μόρια και η κατανόηση των αλληλεπιδράσεων μεταξύ τους σε ατομικό επίπεδο.
- Η διαδικασία της μοριακής πρόσδεσης μπορεί να χωριστεί σε δύο βασικά βήματα:
 - i. Αναζήτηση: Το βήμα αυτό περιλαμβάνει τη δημιουργία ενός μεγάλου αριθμού πιθανών συμπλόκων μεταξύ της πρωτεΐνης και του μοριακού στόχου. Αυτό συνήθως γίνεται με την τυχαία δημιουργία διαφορετικών προσανατολισμών και θέσεων του μορίου σε σχέση με την πρωτεΐνη.
 - ii. Βαθμολόγηση: Αυτό το βήμα περιλαμβάνει την αξιολόγηση της σταθερότητας και της ενεργειακής συμπεριφοράς κάθε συμπλόκου που δημιουργήθηκε στο προηγούμενο βήμα. Χρησιμοποιούνται διάφορες συναρτήσεις βαθμολόγησης για την αξιολόγηση της σταθερότητας και αξιοπιστίας των συμπλόκων, λαμβάνοντας υπόψη παράγοντες όπως: ηλεκτροστατικές αλληλεπιδράσεις, δεσμούς υδρογόνου και αλληλεπιδράσεις Van der Waals μεταξύ πρωτεΐνης και μορίου.
- Το σύμπλοκο με την υψηλότερη βαθμολογία θεωρείται το πιο πιθανό να εμφανιστεί στη φύση. Το αποτέλεσμα της διαδικασίας πρόσδεσης είναι ένα τρισδιάστατο μοντέλο του συμπλόκου που μπορεί να χρησιμοποιηθεί για την κατανόηση των αλληλεπιδράσεων μεταξύ της πρωτεΐνης και του μορίου.



<https://www.futuremedicine.com/doi/abs/10.4155/ebo.13.181>

Τομείς Εφαρμογής της Βιοπληροφορικής

Γλώσσες Προγραμματισμού και Συστήματα (I)

- Η R, η Perl και η Python είναι δημοφιλείς γλώσσες προγραμματισμού που χρησιμοποιούνται στη βιοπληροφορική.
 - i. R: Η R είναι μια «γλώσσα προγραμματισμού» και ένα περιβάλλον λογισμικού για στατιστικούς υπολογισμούς. Χρησιμοποιείται ευρέως στη βιοπληροφορική για την ανάλυση δεδομένων, την οπτικοποίηση και τη στατιστική μοντελοποίηση. Η R έχει μια μεγάλη και ενεργή κοινότητα χρηστών και υπάρχουν πολλά πακέτα βιοπληροφορικής διαθέσιμα για την R, όπως το Bioconductor, το οποίο περιέχει ένα ευρύ φάσμα εργαλείων για την ανάλυση δεδομένων βιοπληροφορικής.
 - ii. Perl: Η Perl είναι μια γλώσσα προγραμματισμού γενικού σκοπού που χρησιμοποιείται ευρέως στη βιοπληροφορική για τον χειρισμό κειμένου. Η Perl διαθέτει μεγάλο αριθμό διαθέσιμων πακέτων, όπως η BioPerl, η οποία παρέχει ένα ευρύ φάσμα εργαλείων βιοπληροφορικής για την ανάλυση ακολουθιών, τον σχολιασμό γονιδίων και τη φυλογενετική.
 - iii. Python: Η Python είναι μια γλώσσα προγραμματισμού γενικού σκοπού, υψηλού επιπέδου, η οποία χρησιμοποιείται στη βιοπληροφορική για την ανάλυση δεδομένων, την οπτικοποίηση και την ανάπτυξη ροών εργασίας. Η Python έχει μια μεγάλη και ενεργή κοινότητα χρηστών και υπάρχουν πολλά πακέτα βιοπληροφορικής διαθέσιμα για την Python, όπως το Biopython.
- Και οι τρεις γλώσσες έχουν δυνατά και αδύνατα σημεία και η επιλογή της γλώσσας εξαρτάται από το συγκεκριμένο ερευνητικό ερώτημα και τον τύπο των δεδομένων που αναλύονται. Η R είναι καλή για στατιστική ανάλυση και οπτικοποίηση, η Perl είναι καλή για χειρισμό κειμένου και η Python είναι καλή για AI και μηχανική μάθηση.



Τομείς Εφαρμογής της Βιοπληροφορικής

Γλώσσες Προγραμματισμού και Συστήματα (II)

- Η C++, η Java και η Rust είναι γλώσσες προγραμματισμού, που χρησιμοποιούνται επίσης στη βιοπληροφορική.
 - i. C++: Η C++ είναι μια γλώσσα προγραμματισμού γενικού σκοπού που χρησιμοποιείται ευρέως στη βιοπληροφορική για την ανάπτυξη υπολογιστικών εργαλείων και λογισμικού υψηλής απόδοσης. Η C++ είναι μια compiled γλώσσα, γεγονός που την καθιστά κατάλληλη για υπολογιστικά εντατικές εργασίες, όπως η ευθυγράμμιση ακολουθιών και η πρόβλεψη πρωτεϊνικών δομών. Η C++ χρησιμοποιείται για την υλοποίηση βιβλιοθηκών, όπως η βιβλιοθήκη Eigen για τη γραμμική άλγεβρα και οι βιβλιοθήκες Boost C++ για τη βιοπληροφορική.
 - ii. Java: Η Java είναι μια γλώσσα προγραμματισμού γενικού σκοπού που χρησιμοποιείται στη βιοπληροφορική για την ανάπτυξη εφαρμογών. Η Java είναι platform-independent, αντικειμενοστραφής γλώσσα, γεγονός που την καθιστά κατάλληλη για την ανάπτυξη λογισμικού βιοπληροφορικής που μπορεί να τρέξει σε διαφορετικά λειτουργικά συστήματα. Υπάρχουν πολλές βιβλιοθήκες βιοπληροφορικής διαθέσιμες σε Java, όπως η Java Bioinformatics Library (JBL), η οποία παρέχει ένα ευρύ φάσμα εργαλείων βιοπληροφορικής για την ανάλυση ακολουθιών, τον σχολιασμό γονιδίων και τη φυλογενετική.
 - iii. Rust: Η Rust είναι μια γλώσσα προγραμματισμού συστημάτων που είναι σχετικά νέα για τη βιοπληροφορική, αλλά κερδίζει ολοένα και μεγαλύτερη δημοτικότητα. Η Rust είναι γνωστή για την εστίασή της στην ασφάλεια, την ταχύτητα και την ταυτόχρονη χρήση, γεγονός που την καθιστά κατάλληλη για λογισμικό βιοπληροφορικής που πρέπει να επεξεργάζεται γρήγορα και αποτελεσματικά μεγάλες ποσότητες δεδομένων.
 - iv. SQL: Η SQL (Structured Query Language) είναι μια γλώσσα προγραμματισμού που χρησιμοποιείται για τη διαχείριση και την ανάλυση μεγάλου όγκου βιολογικών δεδομένων. Η SQL χρησιμοποιείται συνήθως για την αποθήκευση, την ανάκτηση και τον χειρισμό δεδομένων σε σχεσιακές βάσεις δεδομένων.



Τομείς Εφαρμογής της Βιοπληροφορικής

Γλώσσες Προγραμματισμού και Συστήματα (III)

- Υπάρχουν διάφορες πλατφόρμες που χρησιμοποιούνται στη βιοπληροφορική για την ανάλυση και την οπτικοποίηση δεδομένων. Μερικά παραδείγματα είναι τα εξής:
 - i. Galaxy: Το Galaxy είναι μια διαδικτυακή πλατφόρμα αφιερωμένη στη βιοϊατρική έρευνα. Παρέχει μια φιλική προς το χρήστη διεπαφή, επιτρέποντας στους ερευνητές να εκτελούν διάφορα εργαλεία και pipelines, όπως ευθυγράμμιση ακολουθιών και φυλογενετική, χωρίς να χρειάζονται δεξιότητες στη γραμμή εντολών.
 - ii. Biopython: Η Biopython είναι ένα σύνολο ενοτήτων Python για τη βιοπληροφορική, παρέχει ένα ευρύ φάσμα εργαλείων για την ανάλυση ακολουθιών, τον σχολιασμό γονιδίων και τη φυλογενετική. Η Biopython είναι ανοικτού κώδικα και διαθέτει μια μεγάλη και ενεργή κοινότητα χρηστών.
 - iii. Bioconda: Η Bioconda είναι μια διανομή λογισμικού βιοπληροφορικού ενδιαφέροντος για τη διαχείριση πακέτων, έχει σχεδιαστεί για να διευκολύνει την εγκατάσταση λογισμικού και τις λοιπές εξαρτήσεις των αλγορίθμων. Η Bioconda παρέχει μεγάλο αριθμό εργαλείων και χρησιμοποιείται ευρέως στην έρευνα και την εκπαίδευση στον τομέα της βιοπληροφορικής.
 - iv. Bioconductor: Το Bioconductor είναι ένα έργο λογισμικού ανοικτού κώδικα και ανοικτής ανάπτυξης για την ανάλυση και κατανόηση γονιδιωματικών high-throughput δεδομένων.
- Αυτά είναι μερικά μόνο παραδείγματα από τις πολλές πλατφόρμες που είναι διαθέσιμες στην έρευνα. Η επιλογή της πλατφόρμας εξαρτάται από το συγκεκριμένο ερευνητικό ερώτημα, τον τύπο των δεδομένων που αναλύονται και τις δεξιότητες και την εμπειρία του ερευνητή.

Τομείς Εφαρμογής της Βιοπληροφορικής

Γλώσσες Προγραμματισμού και Συστήματα (IV)

- Τα pipelines και οι ροές εργασίας είναι σύνολα υπολογιστικών εργαλείων και διαδικασιών που χρησιμοποιούνται για την ανάλυση βιολογικών δεδομένων. Έχουν σχεδιαστεί για την αυτοματοποίηση και τον εξορθολογισμό της ανάλυσης μεγάλου όγκου δεδομένων, όπως αλληλουχίες DNA ή πρωτεϊνών, και για την πραγματοποίηση προβλέψεων σχετικά με βιολογικά συστήματα.
 - i. Pipelines: Pipelines είναι σειρές υπολογιστικών εργαλείων που χρησιμοποιούνται για την εκτέλεση μιας συγκεκριμένης ανάλυσης, όπως η ευθυγράμμιση ακολουθιών ή ο σχολιασμός γονιδίων. Αυτά τα εργαλεία εκτελούνται συνήθως με συγκεκριμένη σειρά, με την έξοδο ενός εργαλείου να χρησιμοποιείται ως είσοδος για το επόμενο. Τα pipelines μπορούν να χρησιμοποιηθούν για την ανάλυση μεγάλων δεδομένων, όπως δεδομένων NGS, καθώς και για την εκτέλεση σύνθετων αναλύσεων, όπως *de novo* assembly και ανάλυση μεταγραφωμάτων.
 - ii. Ροές εργασίας: Η ροή εργασίας είναι μια γενικότερη έννοια που αναφέρεται στη συνολική διαδικασία ανάλυσης βιολογικών δεδομένων. Οι ροές εργασίας μπορεί να περιλαμβάνουν βιοπληροφορικά pipelines, αλλά μπορεί επίσης να περιλαμβάνουν και άλλους τύπους αναλύσεων, όπως στατιστικές αναλύσεις ή οπτικοποίηση δεδομένων. Οι ροές εργασίας μπορούν να χρησιμοποιηθούν για την αυτοματοποίηση και τον εξορθολογισμό ολόκληρης της διαδικασίας ανάλυσης, από την απόκτηση δεδομένων έως τα τελικά αποτελέσματα.
- Αυτές οι ροές εργασίας είναι απαραίτητες για την έρευνα της βιοπληροφορικής, καθώς επιτρέπουν στους ερευνητές να αναλύουν γρήγορα και αποτελεσματικά μεγάλες ποσότητες δεδομένων και να κάνουν προβλέψεις για τα βιολογικά συστήματα. Υπάρχουν διάφορες πλατφόρμες που είναι διαθέσιμες για τη δημιουργία και την κοινή χρήση pipeline και ροών εργασίας βιοπληροφορικής, όπως Galaxy, Nextflow και Snakemake, οι οποίες επιτρέπουν στους ερευνητές να μοιράζονται, να επαναχρησιμοποιούν και να αναπαράγουν εύκολα την ανάλυση (καθώς επίσης να επιτρέπουν τη διαλειτουργικότητα/παραμετροποίηση και τον ευκολότερο διαμοιρασμό).

Τομείς Εφαρμογής της Βιοπληροφορικής

Βιολογία Συστημάτων (I)

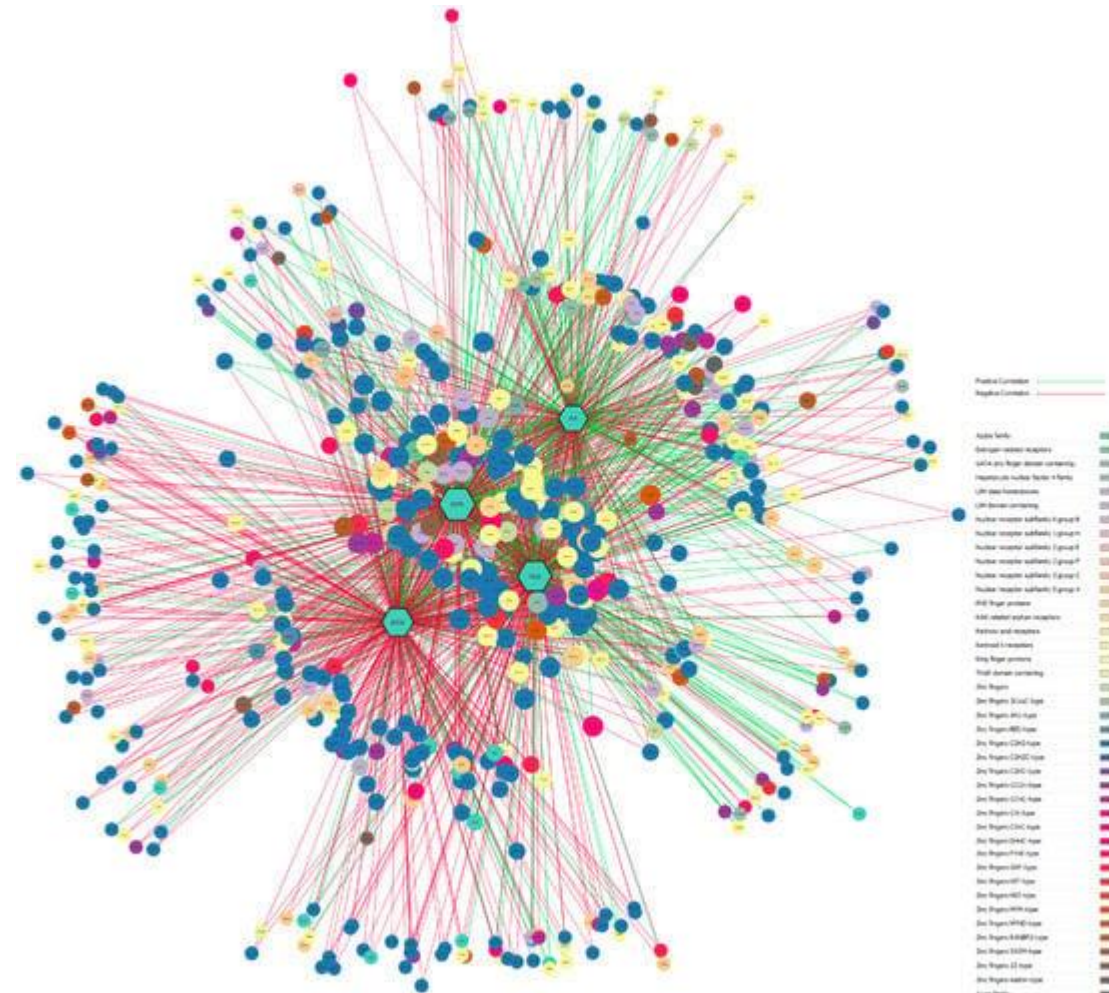
- Η βιολογία συστημάτων είναι μια ολιστική προσέγγιση της μελέτης των βιολογικών συστημάτων που αποσκοπεί στην κατανόηση των πολύπλοκων αλληλεπιδράσεων μεταξύ των διαφόρων στοιχείων του συστήματος, όπως τα γονίδια, οι πρωτεΐνες και τα μονοπάτια αλληλεπιδράσεων. Μια σημαντική πτυχή της βιολογίας συστημάτων είναι η χρήση υπολογιστικών εργαλείων και τεχνικών για την ανάλυση μεγάλου όγκου δεδομένων και τη μοντελοποίηση της συμπεριφοράς των βιολογικών συστημάτων.
 - ❑ Μελέτη των δικτύων αλληλεπίδρασης πρωτεϊνών (protein-protein interaction networks). Οι πρωτεΐνες είναι οι μοριακές «μηχανές» που εκτελούν την πλειονότητα των λειτουργιών στα κύτταρα και συχνά αλληλεπιδρούν με άλλες πρωτεΐνες για την εκτέλεση αυτών των λειτουργιών. Αυτές οι αλληλεπιδράσεις μπορούν να αναπαρασταθούν ως δίκτυα, όπου οι κόμβοι αντιπροσωπεύουν πρωτεΐνες και οι ακμές αντιπροσωπεύουν αλληλεπιδράσεις μεταξύ πρωτεϊνών.
 - ❑ Τα δίκτυα αλληλεπίδρασης πρωτεϊνών μπορούν να χρησιμοποιηθούν για την κατανόηση της οργάνωσης και της λειτουργίας των βιολογικών συστημάτων σε μοριακό επίπεδο. Για παράδειγμα, αναλύοντας δίκτυα αλληλεπίδρασης πρωτεϊνών, είναι δυνατόν να εντοπιστούν πρωτεΐνες-κλειδιά που διαδραματίζουν κεντρικό ρόλο στο δίκτυο (γνωστές ως protein-hubs) και να κατανοηθεί πώς αυτές οι πρωτεΐνες συμμετέχουν στη ρύθμιση βιολογικών διεργασιών.
 - ❑ Μια άλλη σημαντική πτυχή της βιολογίας συστημάτων είναι η ανάλυση λειτουργικού εμπλουτισμού (functional enrichment analysis). Πρόκειται για μια στατιστική μέθοδο που χρησιμοποιείται για τον εντοπισμό υπερεκπροσωπούμενων λειτουργικών κατηγοριών σε ένα σύνολο γονιδίων ή πρωτεϊνών. Για παράδειγμα, εάν διαπιστωθεί ότι ένα σύνολο γονιδίων εκφράζεται διαφορετικά σε μια συγκεκριμένη ασθένεια ή πάθηση, η ανάλυση λειτουργικού εμπλουτισμού μπορεί να χρησιμοποιηθεί για τον εντοπισμό των βιολογικών διεργασιών και μονοπατιών που είναι πιθανότερο να εμπλέκονται στην ασθένεια ή την πάθηση.

Τομείς Εφαρμογής της Βιοπληροφορικής

Βιολογία Συστημάτων (II)



- Ένα δίκτυο μπορεί να γίνει αρκετά πολύπλοκο.

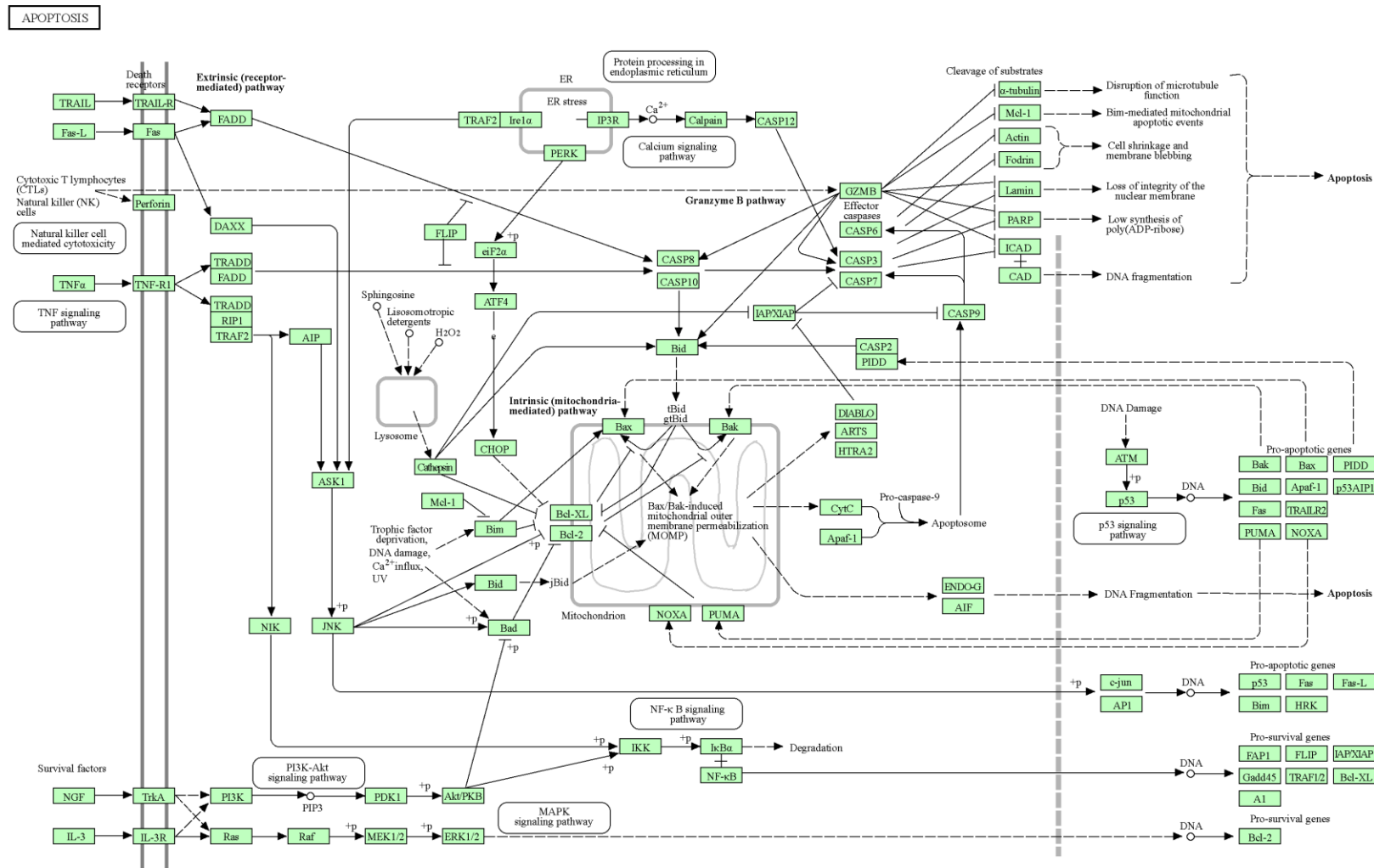


Τομείς Εφαρμογής της Βιοπληροφορικής

Βιολογία Συστημάτων (III)



- Το μονοπάτι του μηχανισμού απόπτωσης απεικονισμένο σε μορφή KEGG.



Τομείς Εφαρμογής της Βιοπληροφορικής

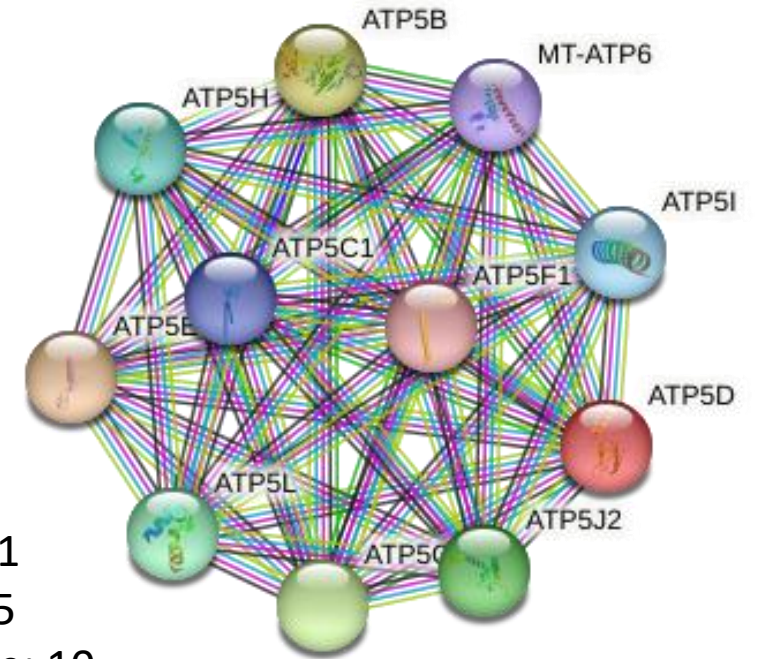
Βιολογία Συστημάτων (IV)



- Το αποτέλεσμα της αναζήτησης της ανθρώπινης μιτοχονδριακής ATP synthase subunit delta στο εργαλείο STRING.
- Στην εικόνα απεικονίζονται οι functional partners της ATP5D.

[GO:0042776](#)*

[GO:0042407](#)**



number of nodes: 11

number of edges: 55

average node degree: 10

avg. local clustering coefficient: 1

expected number of edges: 12

PPI enrichment p-value: $< 1.0e-16$

Πρακτική Εφαρμογή της Βιοπληροφορικής στον Τομέα της Μυοσκελετικής Ογκολογίας

Εισαγωγή (I)

- Υπάρχουν διάφοροι τρόποι με τους οποίους η βιοπληροφορική χρησιμοποιείται στον τομέα της μυοσκελετικής ογκολογίας για τη βελτίωση της διάγνωσης, της θεραπείας και των αποτελεσμάτων των ασθενών:
 - i. Ανάλυση γονιδιακής έκφρασης: Η βιοπληροφορική μπορεί να χρησιμοποιηθεί για την ανάλυση δεδομένων γονιδιακής έκφρασης από μυοσκελετικούς όγκους για τον εντοπισμό γονιδίων που εκφράζονται διαφορετικά σε καλοήθεις και κακοήθεις όγκους.
 - ii. Ανάλυση γονιδιωματικών δεδομένων: Η βιοπληροφορική μπορεί να χρησιμοποιηθεί για την ανάλυση γονιδιωματικών δεδομένων από μυοσκελετικούς όγκους, όπως δεδομένα αλληλούχησης DNA, για τον εντοπισμό γενετικών μεταλλάξεων και παραλλαγών που σχετίζονται με την ανάπτυξη και την εξέλιξη αυτών των όγκων.
 - iii. Προβλεπτική μοντελοποίηση: Η βιοπληροφορική μπορεί να χρησιμοποιηθεί για την ανάπτυξη προγνωστικών μοντέλων για την πρόβλεψη των αποτελεσμάτων των ασθενών, όπως η επιβίωση ή η υποτροπή. Αυτά τα μοντέλα μπορούν να βασίζονται σε διάφορους τύπους δεδομένων, όπως απεικονιστικά δεδομένα, κλινικά δεδομένα και γονιδιωματικά δεδομένα, και μπορούν να χρησιμοποιηθούν για την ενημέρωση των θεραπευτικών αποφάσεων και για τον εντοπισμό ασθενών που διατρέχουν υψηλό κίνδυνο υποτροπής.
 - iv. Ανάλυση κλινικών δεδομένων: Η βιοπληροφορική μπορεί να χρησιμοποιηθεί για την ανάλυση μεγάλου όγκου κλινικών δεδομένων (όπως ηλεκτρονικά αρχεία υγείας) και για τον εντοπισμό μοτίβων και τάσεων που μπορούν να χρησιμοποιηθούν για τη βελτίωση της διάγνωσης και της θεραπείας των μυοσκελετικών όγκων.
 - v. Ανάπτυξη φαρμάκων.

Πρακτική Εφαρμογή της Βιοπληροφορικής στον Τομέα της Μυοσκελετικής Ογκολογίας

Παράδειγμα Ροής Εργασίας Μεταγραφομικών Δεδομένων (II)

FastQC

Εξέταση της ποιότητας των πρωτογενών ακατέργαστων δεδομένων που ελήφθησαν από τη βάση δεδομένων

Trim Galore!

Φιλτράρισμα και βελτιστοποίηση πρωτογενών δεδομένων

adDEr

προ-επεξεργασία για την ανάλυση των δικτύων βιολογικών αλληλεπιδράσεων και την ευκολότερη εισαγωγή τους στα επόμενα εργαλεία

Salmon

Επιλεκτική ευθυγράμμιση στο μεταγράμμα και το γονιδίωμα αναφοράς

Meta-Analysis

Ποσοτική ανασκόπηση και σύνθεση αποτελεσμάτων

MultiQC

Οπτικοποίηση και περαιτέρω ανάλυση των πρωτογενών ακατέργαστων δεδομένων

DESeq2

Εκτίμηση της διαφορικής έκφρασης των δεδομένων

tximeta

Ημι-αυτόματη εισαγωγή ποσοτικοποιημένων μεταγραφομικών δεδομένων στην R

STRING & WebGestalt

Λειτουργική ανάλυση εμπλουτισμού

□ Ως ροή εργασίας στην Πληροφορική, καλείται ένα σύνολο διαδικασιών του οποίου οι λειτουργίες φιλτράρουν και μετασχηματίζουν δεδομένα.

Πρακτική Εφαρμογή της Βιοπληροφορικής στον Τομέα της Μυοσκελετικής Ογκολογίας

Πρακτικό Παράδειγμα Αναζήτησης Στοιχείων Πρωτεΐνης Ενδιαφέροντος (III)

Επιλογή γονιδίου που διαδραματίζει σημαντικό ρόλο σε νεοπλασίες, π.χ., IDH1, NRAS, KRAS, TP53.

A. Εύρεση Αλληλουχίας: Πλοήγηση με τον περιηγητή στον ιστότοπο της UniProt: <https://www.uniprot.org/>

1. Αναζήτηση για IDH1.
2. Αντιγραφή της ακολουθίας.

B. Ανίχνευση Ομολογίας: Πλοήγηση με τον περιηγητή στον ιστότοπο BLAST: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

1. Επιλογή Protein BLAST.
2. Επικόλληση της ακολουθίας της IDH1 στο παράθυρο που ονομάζεται *“Enter Query Sequence”* και αφήνουμε τις βασικές παραμέτρους:
MSKKISGGSVNEMQGDDEMTRIIEWELIKEKLIFPYVELDLHSYDLGIENRDATNDQVTKDAAEAIAIKKHNVGVKCATITPDEKRVVEEFKQMWKSPNGTIRNIIILGGTVFRE
AIIICKNIPRLVSGWVVKPIIIIGRHAYGDQYRATDFVVPGPVKVEITYTPSDGTQKVYTYLVHNFEEGGGVAMGMYNQDKSIEDFAHSSFQMAISKGWPLYLSTKNTILKKYD
GRFKDIFQEIYDKQYKSQFEAQKIWYEHRLIDDMVAQAMKSEGGFIWACKNYDGDVQSDSVAQGYGSLGMMTSVLVCPDGKTVEAEAAHGTVTRHYRMYQKQGE
TSTNPIASIFAWTRGLAHRKLDNNKELAFFANALEEVSITETIEAGFMTKDLAACIKGLPNVQRSDYLNTFEFMDKLGLENLKIQLAQAKL

C. Ανίχνευση Αλληλεπιδράσεων Μεταξύ Πρωτεϊνών: Πλοήγηση με τον περιηγητή στον ιστότοπο STRING: <https://string-db.org/>

1. Επιλογή αναζήτησης βάσει ακολουθίας.

D. Πρόβλεψη Δομής και Λειτουργίας της Πρωτεΐνης. Πλοήγηση με τον περιηγητή στον ιστότοπο I-TASSER: <https://zhanggroup.org/I-TASSER/>

1. Εισαγωγή στοιχείων και ακολουθίας.

Πρακτική Εφαρμογή της Βιοπληροφορικής στον Τομέα της Μυοσκελετικής Ογκολογίας

Πρακτικό Παράδειγμα Αναζήτησης Στοιχείων Πρωτεΐνης Ενδιαφέροντος (IV)

- A. Πρόβλεψη Λειτουργίας της Πρωτεΐνης. Πλοήγηση με τον περιηγητή στον ιστότοπο Pfam: <http://pfam.xfam.org/>
1. Εισαγωγή ακολουθίας στο παράθυρο που ονομάζεται “*Search by sequence*” και αφήνουμε τις βασικές παραμέτρους.
- B. Μοριακός Ελλιμενισμός (molecular docking): Πλοήγηση με τον περιηγητή στον ιστότοπο SwissDock: <http://www.swissdock.ch/>
1. Εισαγωγή PDB αρχείου.
 2. Εισαγωγή ligand.
- C. Αναζήτηση Βιολογικών Μονοπατιών: Πλοήγηση με τον περιηγητή στον ιστότοπο KEGG: <https://www.genome.jp/kegg/pathway.html>
1. Αναζήτηση IDH1.
- D. Varsome: <https://varsome.com/>
- E. iTOL: <https://itol.embl.de/>

Βιβλιογραφία και Χρήσιμοι Σύνδεσμοι

- CHAFFEY N. (2003). Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. Molecular biology of the cell. 4th edn. Annals of Botany, 91(3), 401. <https://doi.org/10.1093/aob/mcg023>
- Μπάγκος, Π., 2015. Βιοπληροφορική. [ηλεκτρ. βιβλ.] Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών <https://repository.kallipos.gr/handle/11419/5016>

Σας Ευχαριστούμε
για την Προσοχή
σας!



Τέλος Παρουσίασης

Mail: kanakogloud@biol.uoa.gr
apampalou@gmail.com