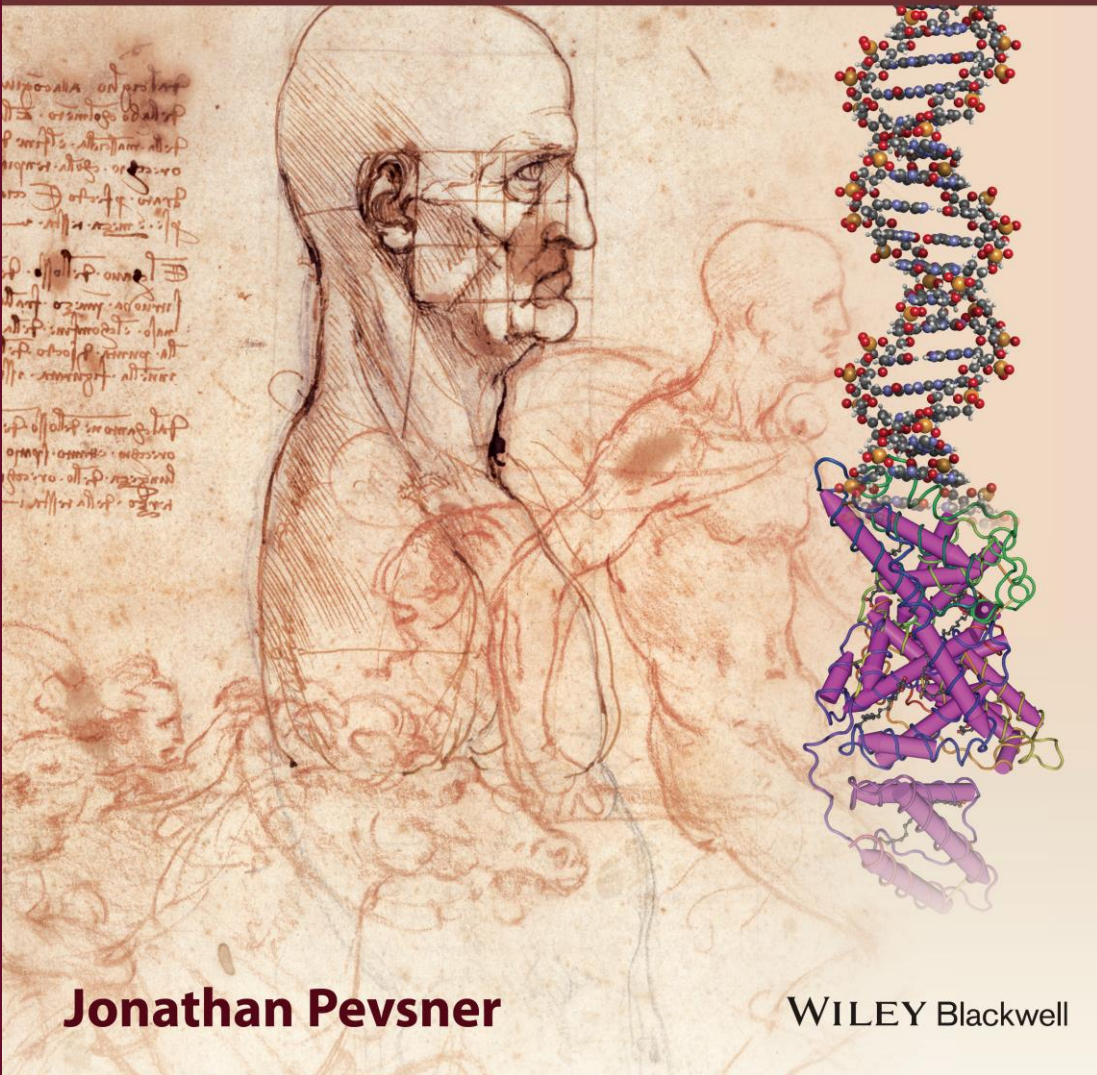


BIOINFORMATICS AND FUNCTIONAL GENOMICS

third edition



Jonathan Pevsner

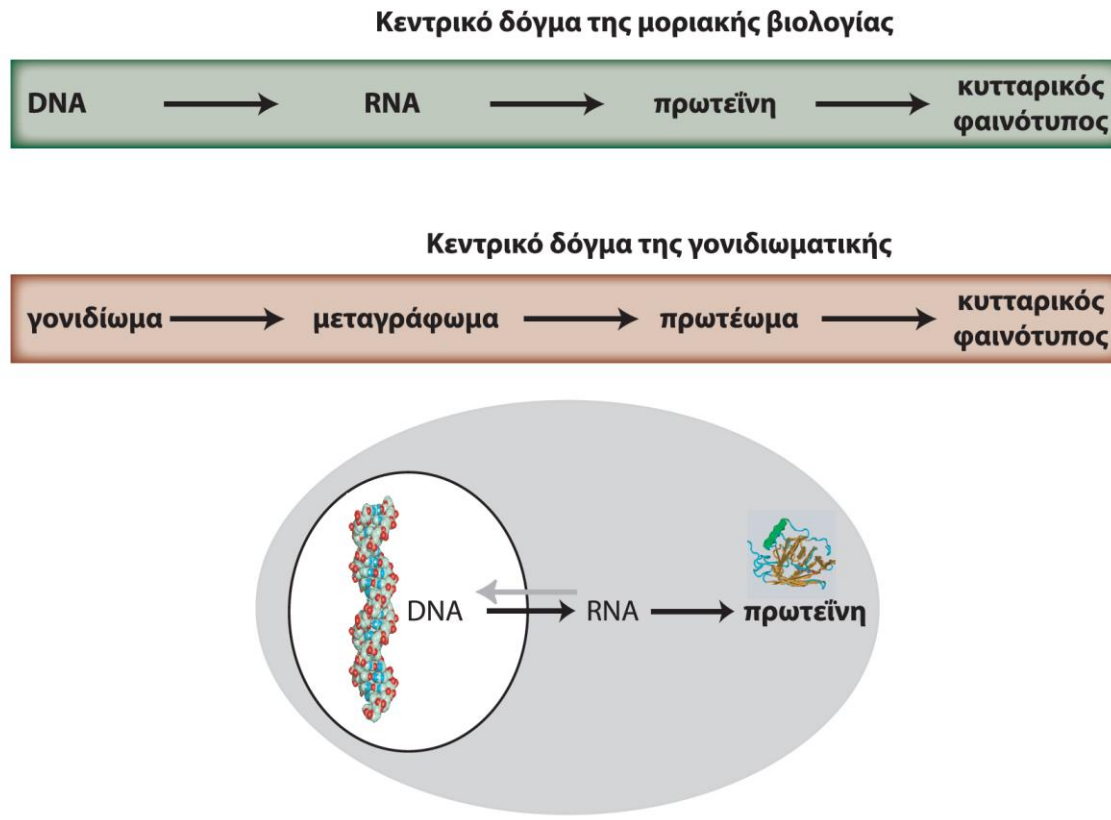
WILEY Blackwell

Κεφάλαιο 1

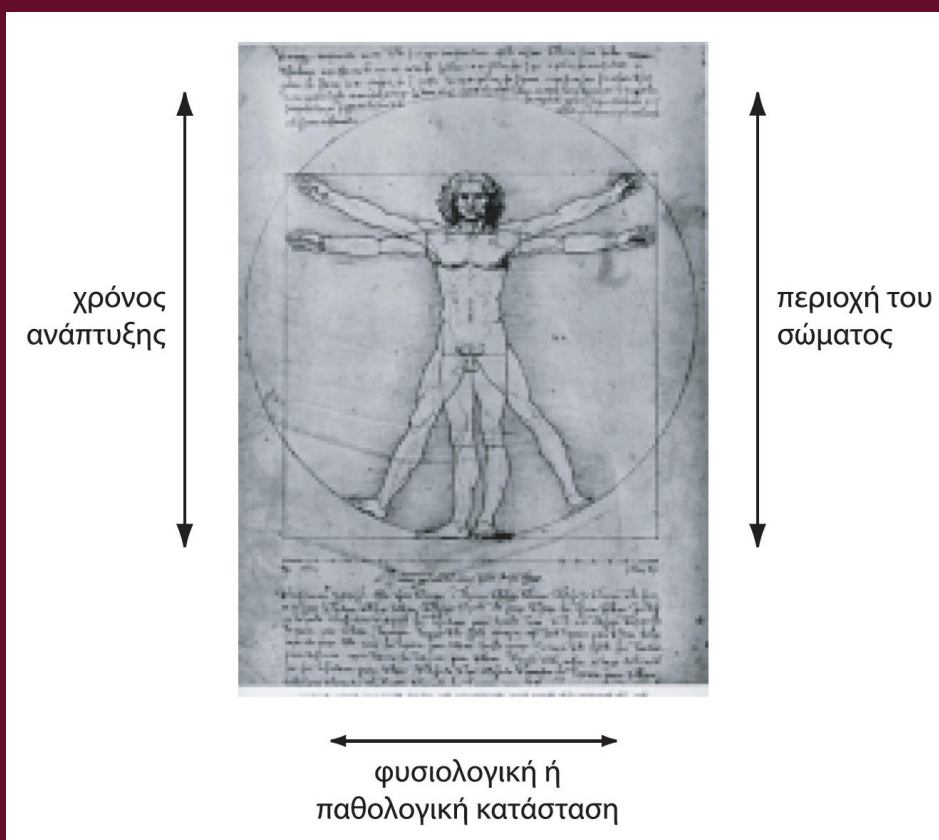
Εισαγωγή

Ακαδημαϊκές
Εκδόσεις

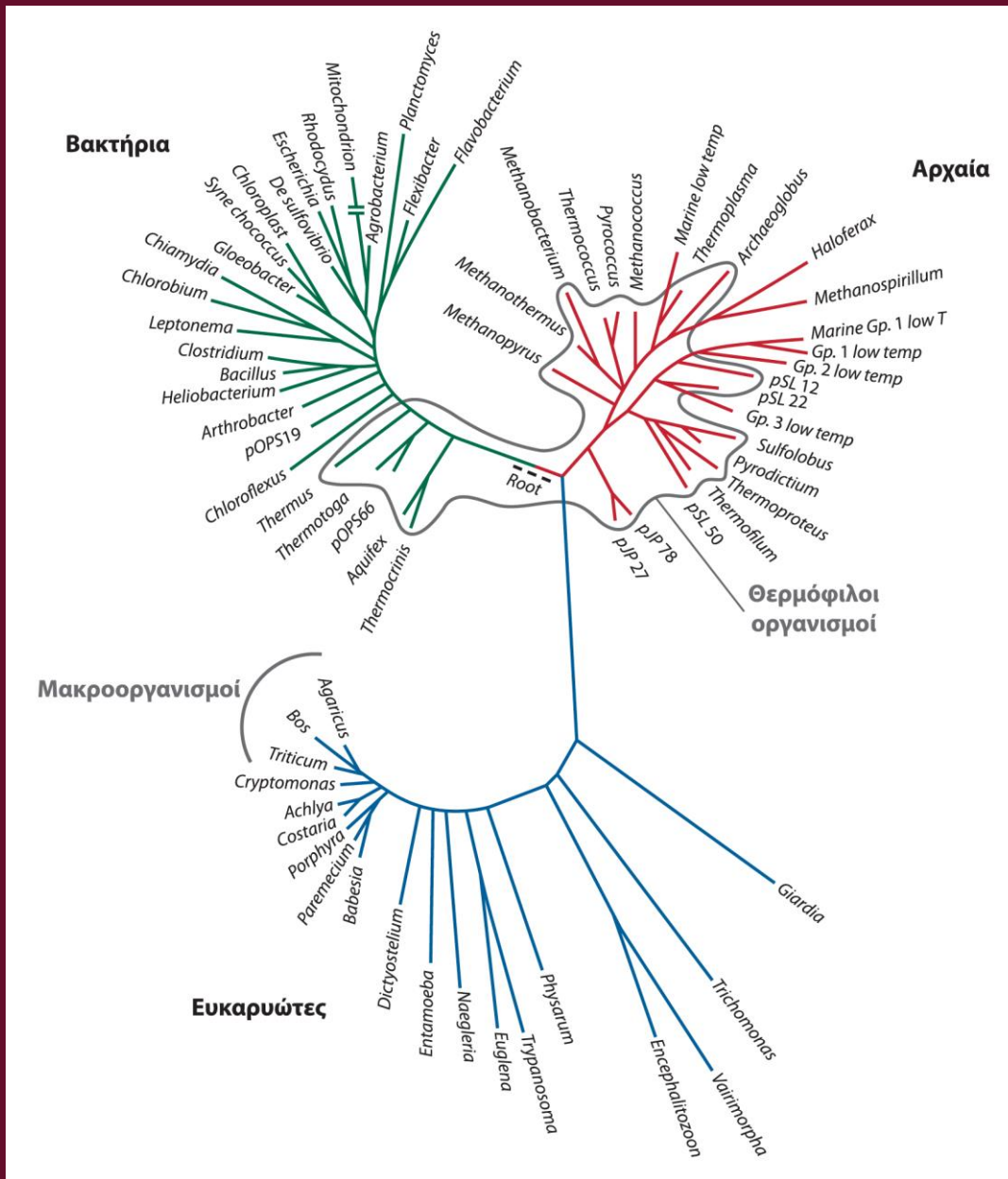




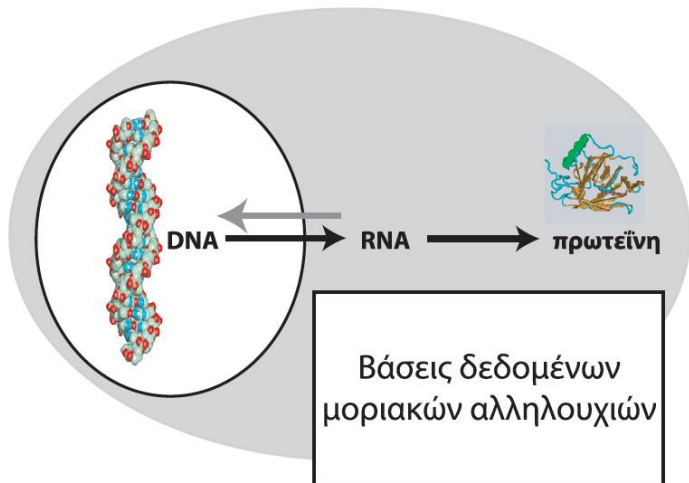
Εικόνα 1.1 Μια πρώτη οπτική γωνία της βιοπληροφορικής αφορά τη διερεύνηση βιολογικών προβλημάτων στο επίπεδο του κυττάρου. Η βιοπληροφορική έχει εξελιχθεί ως επιστημονικός κλάδος, καθώς η συσσώρευση δεδομένων που αφορούν μοριακές αλληλουχίες οδηγεί στον μετασχηματισμό της βιολογίας. Βάσεις δεδομένων όπως αυτές του EMBL (European Molecular Biology Laboratory), η GenBank, η SRA (Sequence Read Archive) και η DDBJ (DNA Database of Japan) χρησιμεύουν ως αποθετήρια για αλληλουχίες συνολικού μεγέθους της τάξης των τετράκις εκατομμυρίων (10^{15}) νουκλεοτιδίων DNA (βλ. Κεφάλαιο 2). Υπάρχουν αντίστοιχες βάσεις δεδομένων για μετάγραφα RNA και για πρωτεΐνες. Μια κύρια κατεύθυνση του πεδίου της βιοπληροφορικής αφορά την άντληση πληροφοριών από τις αλληλουχίες αυτές, με σκοπό την απόκτηση γνώσεων που θα συμβάλουν στην επίλυση ενός ευρέος φάσματος βιολογικών προβλημάτων.



Εικόνα 1.2 Μια δεύτερη οπτική γωνία της βιοπληροφορικής αφορά τη διερεύνηση βιολογικών προβλημάτων στο επίπεδο του οργανισμού. Διευρύνοντας τη θεώρησή μας από το επίπεδο του κυττάρου στο επίπεδο του οργανισμού, μπορούμε να εξετάσουμε τη δυναμική του γονιδιώματος (το οποίο περιλαμβάνει το σύνολο των γονιδίων του οργανισμού που μεταγράφονται σε RNA) και των πρωτεϊνών ενός οργανισμού. Για έναν μεμονωμένο οργανισμό, μπορούμε λοιπόν, χρησιμοποιώντας εργαλεία της βιοπληροφορικής, να περιγράψουμε τις αλλαγές που συμβαίνουν κατά την ανάπτυξή του, τις διαφοροποιήσεις που χαρακτηρίζουν τις διάφορες περιοχές του σώματός του και τέλος τις μεταβολές στη φυσιολογία του λόγω μιας ποικιλίας εγγενών και περιβαλλοντικών ερεθισμάτων ή λόγω παθολογικών καταστάσεων.



Εικόνα 1.3 Μια τρίτη οπτική γωνία της βιοπληροφορικής αφο-ρά τη διερεύνηση βιολογικών προβλημάτων στο επίπεδο του δέντρου της ζωής. Το πεδίο της βιοπληροφορικής μελετά όλες τις μορφές ζωής στη Γη, δηλαδή και τους τρεις κύριους κλάδους του δέντρου της ζωής: τα βακτήρια, τα αρχαία και τους ευκαρυώτες. Οι ιοί, που βρίσκονται στο όριο του ορισμού της ζωής, δεν απεικονίζονται εδώ. Για όλα τα είδη, η συλλογή και η ανάλυση των μοριακών αλληλουχιών τους μας επιτρέπει να περιγράψουμε το γονιδίωμά τους (το σύνολο του DNA τους). Μπορούμε να μελετήσουμε περαιτέρω τις παραλλαγές που παρατηρούνται μεταξύ των ειδών ή μεταξύ των ατόμων ενός είδους και έτσι να συναγάγουμε την εξελικτική ιστορία της ζωής στη Γη.



Μέρος I: Βιοπληροφορική: αναλύοντας το DNA, το RNA και τις πρωτεΐνες

Κεφάλαιο 1: Εισαγωγή

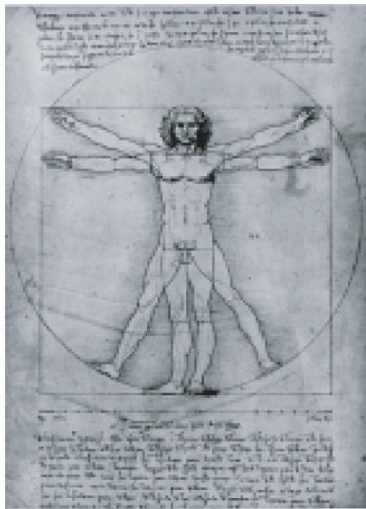
Κεφάλαιο 2: Πώς να αποκτήσετε πρόσβαση σε αλληλουχίες

Κεφάλαιο 3: Πώς να συγκρίνετε δύο αλληλουχίες

Κεφάλαια 4 και 5: Πώς να συγκρίνετε μία αλληλουχία με τις αλληλουχίες των βάσεων δεδομένων

Κεφάλαιο 6: Πώς να πραγματοποιήσετε πολλαπλές στοιχίσεις αλληλουχιών

Κεφάλαιο 7: Πώς οι πολλαπλές στοιχίσεις αλληλουχιών παρουσιάζονται ως φυλογενετικά δέντρα



Μέρος II: Λειτουργική γονιδιωματική: από το DNA στο RNA και στην πρωτεΐνη

Κεφάλαιο 8: DNA: Το ευκαρυωτικό χρωμόσωμα

Κεφάλαιο 9: Ανάλυση DNA: αλληλούχιση επόμενης γενιάς

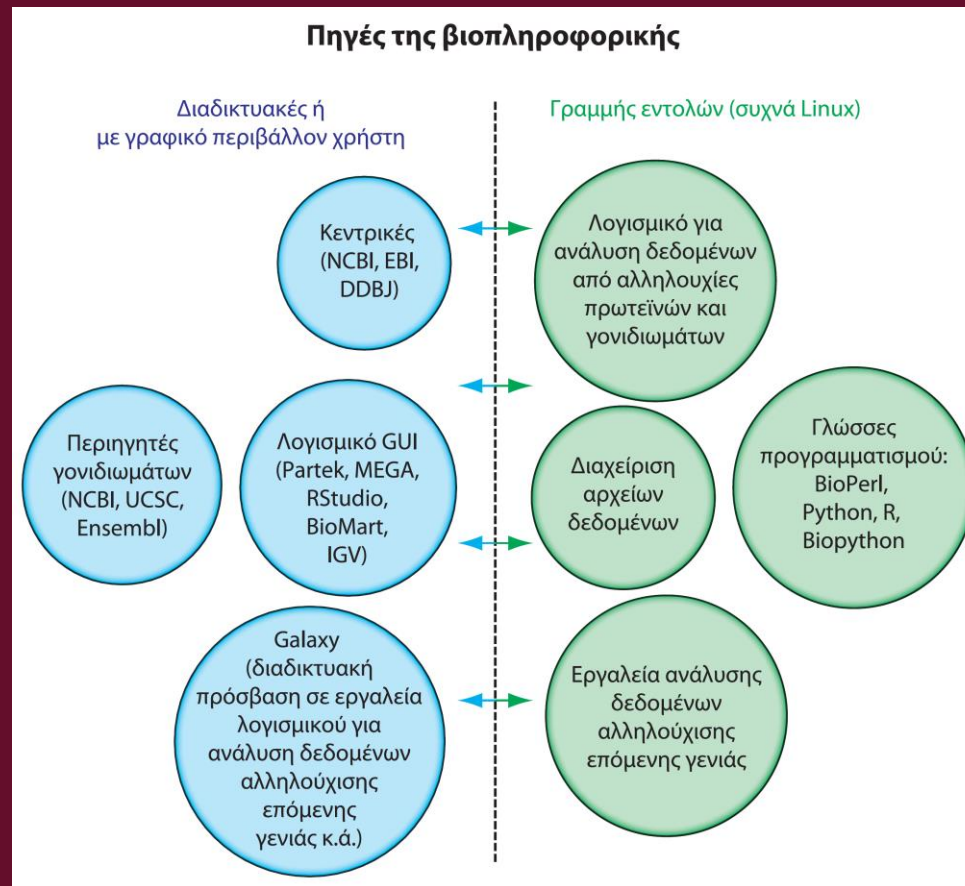
Κεφάλαιο 10: Βιοπληροφορικές προσεγγίσεις στο RNA

Κεφάλαιο 11: Ανάλυση μικροσυστοιχιών και RNA-seq

Κεφάλαιο 12: Ανάλυση πρωτεϊνών και πρωτεϊνικών οικογενειών

Κεφάλαιο 13: Πρωτεϊνική δομή

Κεφάλαιο 14: Λειτουργική γονιδιωματική



Εικόνα 1.5 Πηγές της βιοπληροφορικής. Αριστερά εμφανίζονται πηγές πληροφορίας που βασίζονται στο διαδίκτυο (συχνά αναφέρονται ως point-and-click). Σε αυτές περιλαμβάνονται οι βασικότερες διαδικτυακές πύλες εισόδου σε δεδομένα βιοπληροφορικής [το NCBI (National Center for Biotechnology Information) και το EBI (European Bioinformatics Institute)], οι σημαντικότεροι περιηγητές γονιδιωμάτων [αυτός της Ensembl και αυτός του UCSC (University of California at Santa Cruz)], βάσεις δεδομένων και εξειδικευμένες ιστοσελίδες. Δεξιά εμφανίζονται πηγές πληροφορίας που βασίζονται στη γραμμή εντολών. Σε αυτές περιλαμβάνονται οι γλώσσες προγραμματισμού (όπως η Biopython, η BioPerl και η R) και το λογισμικό γραμμής εντολών (η πρόσβαση στο οποίο κατά κανόνα γίνεται μέσω του λειτουργικού συστήματος Linux). (GUI: Graphical User Interface, Γραφικό Περιβάλλον Χρήστη.)

Πίνακας 1.1 Συνοπτική παρουσίαση μερικών παραδειγμάτων διαδικτυακού λογισμικού (ή λογισμικού που διαθέτει GUI) και λογισμικού γραμμής εντολών που χρησιμοποιούνται σε διάφορα κεφάλαια αυτού του βιβλίου.

Τμήμα: Κεφάλαιο	Αντικείμενο	Διαδικτυακό λογισμικό ή λογισμικό που διαθέτει GUI	Λογισμικό γραμμής εντολών
I: 2	Πρόσβαση σε πληροφορία	BioMart Genome Workbench	EDirect
I: 3	Στοίχιση κατά ζεύγη	BLAST	BLAST+ Biopython needle (EMBOSS) water (EMBOSS)
I: 4	BLAST	BLAST	BLAST+
I: 5	Αναζήτηση σε βάση δεδομένων	DELTA-BLAST Megablast	HMMER
I: 6	Πολλαπλή στοίχιση αλληλουχιών	Pfam, MUSCLE	MAFFT
I: 7	Φυλογένεση	MEGA	MrBayes
II: 8	Χρωμοσώματα	Galaxy	geecee (EMBOSS) isochore (EMBOSS)
II: 9	Αλληλούχιση επόμενης γενιάς	Galaxy, SIFT, PolyPhen2	SAMTools, tabix, VCFtools
II: 10	RNA	RNAfam, tRNAscan	
II: 11	RNA-seq	Galaxy	affy (R package), RSEM
II: 12	Πρωτεωμική	ExPASy	pepstats (EMBOSS)
II: 13	Δομή πρωτεϊνών	Cn3D, Pymol	psiphi (EMBOSS)
II: 14	Λειτουργική γονιδιωματική	FLink, Cytoscape	

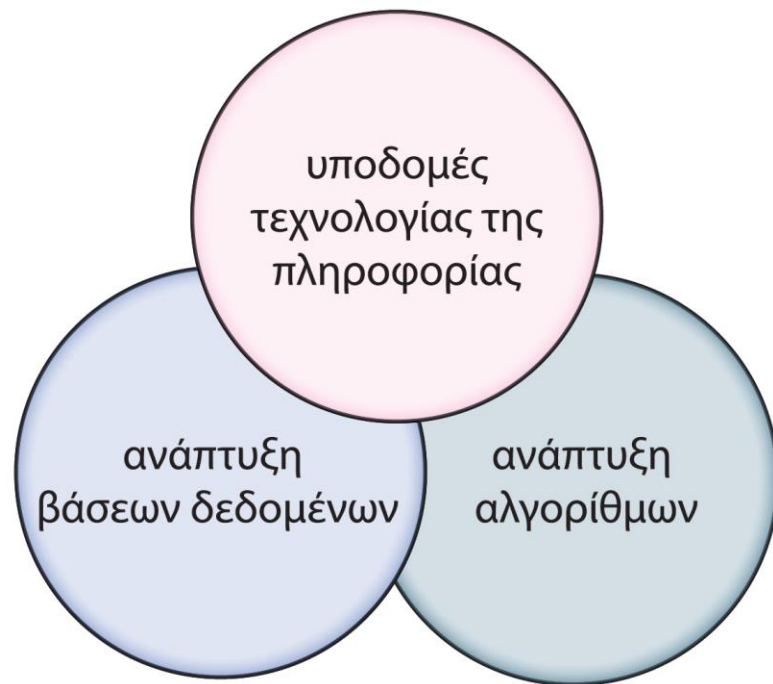
Πίνακας 1.2 Διαγωνισμοί επίλυσης επιστημονικών προβλημάτων του πεδίου της βιοπληροφορικής.

Όνομα/ Ακρωνύμιο	Διαγωνισμός	Κεφάλαιο
Alignathon	Σύγκριση μεθόδων στοίχισης γονιδιωμάτων (Compare whole-genome alignment methods)	6
EGASP	Αξιολόγηση μεθόδων γονιδιωματικού υπομνηματισμού στο ENCODE (ENCODE Genome Annotation Assessment Project)	8
Assemblathon	Σύγκριση της επίδοσης των μεθόδων συναρμολόγησης γονιδιωμάτων (Compare the performance of genome assemblers)	9
GAGE	Αξιολόγηση εδραιωμένων μεθόδων συναρμολόγησης γονιδιωμάτων (Genome Assembly Gold-standard Evaluations)	9
ABRF	Ένωση Μονάδων Βιομοριακών Πηγών (Association of Biomolecular Resource Facilities) – Αξιολόγηση της φωσφορυλίωσης (Assessment of phosphorylation)	12
CASP	Κριτική αξιολόγηση δομικών προβλέψεων (Critical Assessment of Structure Prediction)	13
CAFA	Κριτική αξιολόγηση της πρωτεϊνικής λειτουργίας (Critical Assessment of Protein Function)	14
CAGI	Κριτική αξιολόγηση μεθόδων ερμηνείας της ποικιλότητας του γονιδιώματος (Critical Assessment of Genome Interpretation)	14

Χρήστες εργαλείων



Κατασκευαστές εργαλείων



Εικόνα 1.6 Χρήστες εργαλείων και κατασκευαστές εργαλείων. Η πληροφορική βρίσκει εφαρμογές σε όλο και περισσότερα επιστημονικά πεδία τα τελευταία χρόνια, με αποτέλεσμα τη δημιουργία νέων κλάδων της, όπως η βιοπληροφορική, η πληροφορική δημόσιας υγείας, η ιατρική πληροφορική και η πληροφορική βιβλιοθηκών. Καθένας από αυτούς τους τομείς ασχολείται με τη συστηματοποίηση και την ανάλυση όλο και μεγαλύτερων σετ δεδομένων. Η βιοπληροφορική και η γονιδιωματική επικεντρώνονται κυρίως στις πρωτεΐνες, στα γονίδια και στα γονιδιώματα.