

Sommario

Presentazione dell'opera	XI
Ringraziamenti	XII
Capitolo 1 Introduzione alla bioinformatica	1
1.1 Cenni introduttivi	1
1.2 Pietre miliari della bioinformatica	2
1.3 Infrastrutture bioinformatiche	4
1.4 Indirizzi web delle più rilevanti infrastrutture bioinformatiche	5
1.5 Letture di approfondimento consigliate	6
Capitolo 2 Banche dati biologiche	7
2.1 Introduzione	7
2.2 Sistemi di interrogazione delle banche dati biologiche	8
2.2.1 SRS	9
2.2.2 Entrez	11
2.2.3 ACNUC	12
2.2.4 ACEDB	13
2.3 Banche dati primarie e banche dati specializzate	13
2.4 Banche dati di sequenze nucleotidiche	16
2.5 Banche dati di sequenze proteiche	18
2.6 Banche dati di motivi e domini proteici	19
2.7 Banche dati di strutture proteiche	25
2.8 Banche dati biologiche per il sistema immunitario	26
2.8.1 IMGT	26
2.8.2 Banche dati di peptidi correlati al complesso MHC	26
2.9 Risorse genomiche	27
2.10 Banche dati di geni	28
2.11 Banche dati del trascrittoma	30
2.12 Banche dati di profili di espressione	30
2.13 Banche dati di polimorfismi e mutazioni	32
2.14 Banche dati di <i>pathways</i> metabolici	34
2.15 Banche dati mitocondriali	36
2.16 Indirizzi web per banche dati e risorse biologiche	39



Capitolo 3 Allineamento di sequenze di acidi nucleici e proteine	41
3.1 Introduzione al problema degli allineamenti di sequenze	41
3.1.1 Similarità di sequenze e algoritmi di allineamento	41
3.1.2 Allineamenti di sequenze biologiche con gap	43
3.1.3 Visualizzazione di dot matrix	44
3.2 Le matrici di sostituzione	45
3.2.1 Le matrici PAM	46
3.2.2 Le matrici BLOSUM	48
3.3 Metodi di allineamento esatto	49
3.3.1 Allineamenti globali e locali	49
3.3.2 Algoritmi dinamici di allineamento	50
3.3.3 L'algoritmo di Smith e Waterman per la ricerca di similarità locali	53
3.4 Metodi euristici di allineamento	54
3.4.1 Ricerca di similarità in banche dati: FASTA	54
3.4.2 BLAST	58
3.4.3 Altri programmi di allineamento	62
3.5 Principali risorse disponibili in rete	63
3.6 Letture di approfondimento consigliate	63
Capitolo 4 Allineamento multiplo di sequenze	65
4.1 Introduzione	65
4.2 Algoritmi per l'allineamento multiplo	66
4.3 Misura della qualità di un allineamento multiplo	70
4.4 Strumenti per la visualizzazione e manipolazione dei multiallineamenti	71
4.5 Generazione e applicazione dei profili dei multiallineamenti	73
4.6 Principali risorse disponibili in rete	78
4.7 Letture di approfondimento consigliate	79
Capitolo 5 Ricerca di pattern e di motivi funzionali	81
5.1 Introduzione	81
5.1.1 Definizione di motivo	81
5.1.2 Misure della sensibilità e della specificità di un motivo	82
5.2 Ricerca di pattern e di motivi funzionali in sequenze nucleotidiche	84
5.2.1 Generalità	84
5.2.2 Filtri per l'esclusione delle sequenze ripetitive	85
5.2.3 Ricerca e interpretazione dei promotori eucariotici	85
5.2.4 Siti di giunzione tra introni ed esoni	87
5.2.5 Siti di inizio della traduzione	88
5.2.6 Segnali di terminazione	89
5.2.7 Analisi della strategia dell'uso del codice genetico	89
5.2.8 Conclusioni	92
5.3 Ricerca di pattern e di motivi funzionali in sequenze proteiche	92
5.3.1 Generalità	92
5.3.2 PROSITE	93
5.3.3 Sequenze segnale	94
5.4 Metodi di apprendimento automatico: reti neurali, <i>Hidden Markov Models</i> e algoritmi genetici	95

5.4.1	Generalità	95
5.4.2	Le reti neurali	95
5.4.3	Catene di Markov e Hidden Markov Models (HMM)	98
5.4.4	Algoritmi genetici	101
5.5	Principali risorse disponibili in rete	102
5.6	Letture di approfondimento consigliate	103
Capitolo 6 Evoluzione molecolare		105
6.1	Introduzione	105
6.2	Meccanismi molecolari alla base dei processi evolutivi	106
6.3	Geni ortologi e paraloghi	110
6.4	Determinazione delle distanze genetiche tra sequenze nucleotidiche e aminoacidiche	111
6.5	L'orologio molecolare	119
6.6	Filogenesi molecolare	121
6.6.1	Metodi per la costruzione degli alberi filogenetici	123
6.6.2	Determinazione dei tempi di divergenza	129
6.7	Test statistici per valutare l'accuratezza delle ricostruzioni filogenetiche	131
6.8	Letture di approfondimento consigliate	133
Capitolo 7 Analisi strutturale delle proteine		135
7.1	Introduzione	135
7.1.1	Struttura delle proteine	136
7.2	Metodi classici per la determinazione della struttura proteica	137
7.2.1	Diffrazione ai raggi X	137
7.2.2	Spettroscopia a risonanza magnetica nucleare (NMR)	138
7.2.3	Conclusioni	138
7.3	Banche dati di strutture proteiche	140
7.3.1	PDB	140
7.3.2	MMDB	142
7.3.3	DSSP	142
7.3.4	HSSP	143
7.3.5	FSSP	144
7.3.6	SCOP	144
7.3.7	CATH	145
7.4	Programmi di visualizzazione	145
7.4.1	Metodi di rappresentazione delle strutture proteiche	145
7.4.2	Programmi di grafica molecolare	147
7.5	Ricerche di similarità strutturale	147
7.5.1	DALI	148
7.5.2	SSAP	148
7.5.3	CE	149
7.6	Metodi per la predizione della struttura secondaria	149
7.6.1	PHD	151
7.6.2	PSIPRED	152
7.6.3	PREDATOR	152
7.6.4	JPRED	153

7.7	Metodi per la predizione della struttura terziaria	153
7.7.1	Homology modelling	154
7.7.2	Threading	156
7.8	Principali risorse disponibili in rete	157
7.9	Lecture di approfondimento consigliate	158
Capitolo 8 Analisi della struttura secondaria di molecole di RNA		159
8.1	Introduzione	159
8.2	Componenti della struttura secondaria	159
8.3	Metodi per la predizione della struttura secondaria	160
8.4	Metodi per la ricerca di strutture secondarie	167
8.5	Principali risorse disponibili in rete	169
8.6	Lecture di approfondimento consigliate	169
Capitolo 9 Sequenziamento e analisi di genomi		171
9.1	Introduzione	171
9.2	Il sequenziamento di genomi	172
9.2.1	L'approccio shotgun	172
9.2.2	Strategie di sequenziamento	176
9.2.3	Determinazione della qualità delle sequenze: Phred	179
9.2.4	Assemblaggio di sequenze: Phrap	180
9.2.5	Assemblaggio di sequenze: Arachne	181
9.2.6	Finishing: Consed	183
9.3	Metodi predittivi per l'analisi di genomi	186
9.3.1	Analisi delle Open Reading Frames (ORF)	187
9.3.2	Ulteriori analisi per l'identificazione delle sequenze codificanti: approcci intrinseci	188
9.3.3	Approcci estrinseci per l'analisi di sequenza codificanti	189
9.3.4	Orpheus	190
9.3.5	Metodi predittivi per l'analisi di genomi eucariotici	191
9.4	Analisi di sintenie	195
9.5	Principali risorse disponibili in rete	196
9.6	Lecture di approfondimento consigliate	198
Capitolo 10 Analisi del trascrittoma e del proteoma		199
10.1	Il trascrittoma	199
10.1.1	Sequenziamento di librerie di cDNA	200
10.1.2	Librerie normalizzate	201
10.1.3	Expressed sequence tags (EST) e sequenziamento full length	202
10.1.4	Clustering e assemblaggio di sequenze di trascritti	203
10.1.5	Allineamento di sequenze di trascritti su DNA genomico	205
10.1.6	Analisi e predizione di splicing alternativi	207
10.1.7	Silenziamento del RNA e ruolo dei trascritti non codificanti proteine	208
10.2	I microarray di cDNA	210
10.2.1	Acquisizione e normalizzazione dei dati	211
10.2.2	Matrice dei dati	212
10.2.3	Identificazione di geni differenzialmente espressi	212

10.2.4	Clustering: similarità/dissimilarità	213
10.2.5	Algoritmi di clustering	214
10.2.6	Gestione e condivisione dei dati	216
10.3	Il proteoma	217
10.3.1	Gel elettroforesi bidimensionale	217
10.3.2	Spettrometria di massa	218
10.4	Principali risorse disponibili in rete	221
10.5	Letture di approfondimento consigliate	221
 Capitolo 11 Genomica funzionale		223
11.1	Metodi sperimentali per l'analisi dell'interazione proteina-proteina	223
11.1.1	Tecnica del doppio ibrido in lievito.	224
11.1.2	Spettrometria di massa per l'identificazione di complessi proteici	225
11.1.3	Interazioni genetiche	226
11.1.4	FRET	227
11.1.5	Conclusioni	227
11.2	Predizioni bioinformatiche di interazioni proteina-proteina	228
11.2.1	Mutazioni correlate	228
11.2.2	Docking	229
11.2.3	SPOT: uso di frequenze di residui in posizioni di contatto tra proteine	230
11.3	Associazioni funzionali indirette	232
11.3.1	Predizione dell'interazione tra proteine omologhe a proteine che interagiscono	232
11.3.2	Profili filogenetici	232
11.3.3	Stele di Rosetta	233
11.3.4	Similarità di alberi filogenetici	234
11.3.5	Prossimità genomica	235
11.4	Banche dati di interazioni proteiche	235
11.5	Principali risorse disponibili in rete	237
11.6	Letture di approfondimento consigliate	237
 Appendice A Esercitazioni di bioinformatica		239
Esercizio 1	Ricerca nelle banche dati di sequenze genomiche umane codificanti la subunità 4 della citocromo c ossidasi.	239
Esercizio 2	Ricerca di sequenze di citocromo c ossidasi umane mediante ricerca di similarità in banche dati (Database Similarity Searching)	248
Esercizio 3	Costruzione di un multiallineamento o di un suo «profilo»	251
Esercizio 4	Localizzazione del gene cox4 sul genoma umano mediante Ensembl e il Genome Browser UCSC	257
Esercizio 5	Caratterizzazione funzionale di un frammento genomico umano mediante Ensembl	258
Esercizio 6	Predizione di geni codificanti proteine in sequenze genomiche	262
Esercizio 7	Caratterizzazione di ipotetiche proteine predette da programmi di gene finding	262
Esercizio 8	Determinazione della struttura di un gene mediante il confronto tra la sequenza genomica e l'mRNA maturo	266
Esercizio 9	Verifica della espressione dei geni predetti mediante confronto con le banche dati EMBL ed EST	267

Esercizio 10	Analisi funzionale delle ipotetiche proteine predette nell'analisi di sequenze genomiche	270
Esercizio 11	Consultazione di banche dati metaboliche	273
Esercizio 12	Predizione della struttura terziaria delle proteine	273
Appendice B Elementi di informatica		281
B.1	Struttura e funzionamento di un personal computer	281
B.1.1	Calcolatori multiutente	281
B.1.2	Struttura di un calcolatore multiutente	282
B.1.3	Rete di computer	282
B.1.4	Cos'è un sistema operativo	282
B.1.5	Cos'è un «file»	283
B.1.6	Cos'è una «directory»	283
B.1.7	Cos'è un programma per calcolatore	284
B.2	Introduzione ai sistemi operativi <i>linux</i> e <i>unix</i>	285
B.2.1	Unix e linux	285
B.2.2	Alcuni comandi fondamentali di unix e linux	285
B.2.3	I file in linux e unix	288
B.2.4	Segnali	288
Appendice C Cenni sul linguaggio di programmazione Perl		291
C.1	Disponibilità e installazione di Perl	292
C.2	Somma.pl: un semplice programma per sommare numeri	292
C.2.1	Argomenti passati con la linea di comando	293
C.2.2	Operazioni cicliche: l'istruzione «for»	293
C.3	Solodueparole.pl: un programma per cercare due parole vicine in un file di testo	294
C.3.1	Controllo del flusso: die, if, unless	295
C.3.2	L'uguale non è sempre uguale!	296
C.3.3	Creazione di una stringa di spazi	297
C.3.4	Apertura di un file	297
C.3.5	Ciclo principale del programma solodueparole.pl	298
C.4	Un programma per lanciare Blast e analizzare automaticamente i risultati	299
C.4.1	Installazione del programma Blast	299
C.4.2	Reperimento dei dati di sequenza e formattazione con formatdb	300
C.4.3	Uso di Blast con comandi di linea	302
C.4.4	SmistaBlast.pl	304
C.4.5	Le subroutine	304
C.4.6	Gli array	305
C.4.7	Pattern matching	308
C.4.8	Altre particolarità di SmistaBlast.pl	308
C.5	Principali risorse disponibili in rete	310
Glossario		311
Indice analitico		321