

Index

- >, 82, 83
- ase, 4
- <=>, 101
- =~, 10
- \$_, 52, 117, 164, 202, 239
- @ARGV, 100, 101
- @_, 36, 50, 82, 190
- 3' UTR (GenBank), 74
- 5' UTR (GenBank), 74

- abstract class, 88
- accession number, 72, 174
- Ackermann's function, 277
- active site, 93, 127, 173
- adenine, 1, 2
- adjacency list, 163–5, 198, 199, 201, 228, 229
- adjacent (in the graph), 156
- AIDS, 2
- alignment
 - compact, 309
 - gap-free, 96, 97, 109, 114, 125
 - multiple, 65, 127, 128, 131, 132, 134, 136–41, 147, 149, 151, 154–6, 161
 - pairwise, 128, 248
- allele, 245
- amino acid, 4–7, 10, 16, 55, 56, 58–61, 70, 74, 75, 97, 103, 114, 145, 155, 173, 193, 231, 232, 234
- aminoacyl-tRNA synthetase, 16
- ancestor
 - species, 57, 65, 155–157
 - in tree, 156, 157, 167, 184, 315
- and**, 266
- annealing, DNA, 31
- annotation, 72–5, 81, 89, 91, 174, 204, 244

- anticodon, 16
- April, 97
- argument list, 50, 82, 84, 120
- ASCII, 72
- Ashbya gossypii*, 237
- asparagine, 56
- atan2**, 302
- AUG, 7
- autosome, 3, 4

- backbone, 2
- background frequencies, 58, 70, 117
- backtracking, 258, 262, 263, 268
- BACs, 203
- bacteria, 2, 4, 31, 57, 72, 203, 257
- bacteriophage, 4
- bifurcating trees, 156, 157
- bit
 - unit of information, 46–8, 53, 64, 70, 123–5, 233
 - unit of memory, 145, 178
- bitwise binary operations, 145
- BLAST, 73, 93, 96, 97, 100, 102, 107–9, 114, 116, 123–5, 139
- bless**, 83, 88
- block
 - motif, 173
 - in Perl, 52, 101
 - in wish list, 101, 275, 279–85, 288, 289, 291–5, 297–9
- BLOSUM, 65, 71, 98, 100, 109, 123, 124, 149
- bond, 24, 27, 29, 93, 174, 300, 304
 - covalent, 6, 19
 - hydrogen, 2, 6, 17–22, 24, 26–9, 301, 303, 305
 - phosphodiester, 2

- branch, 156
- branch node, 156–8, 159, 160, 162, 164–6, 168
- branch-and-bound, 149, 152, 161, 162, 170, 198, 200, 202, 228
- Brassica* species, 275
- buffer, 83–5
- C-terminus, 4
- cabbage, 275
- camel book, 15
- carbon, 1, 2
- caret, 175, 177
- catalysts and catalysis, 4, 6, 16, 17, 93, 231, 257
- CDS (GenBank), 74, 75
- Central Dogma of Molecular Biology, 5
- centrifugation, 17, 245
- character
 - in phylogeny, 155, 156, 158, 162, 171
 - in string, 8, 9, 11, 12, 21, 22, 26, 32, 33, 37, 78, 120, 134, 136, 162, 177–9, 181, 183, 248, 249, 260
- characteristic polynomial equation, 112, 115–17, 119–21
- chimera, 203
- chimpanzee, 1
- chomp**, 9
- chromosome, 2–5, 15, 30, 31, 55, 203, 246, 275–7, 299
- class, 79–82, 88, 89, 118, 120, 162, 179, 181, 199, 205
- clone (GenBank), 75
- clone_library (GenBank), 75
- close, 81, 84, 86
- cloverleaf, 18
- CLUSTAL, 139
- cmp**, 101
- coding sequence, 6, 7, 17, 44, 74, 75, 231–5, 237, 238, 244
- codon, 4, 6–8, 12, 16, 55, 56, 61, 75, 231–5, 241, 243, 244
- codon bias, 231, 232, 235, 244
- codon_start (GenBank), 75
- cole vegetables, 275
- collagen, 4
- color blindness, 1
- column of alignment, 32–4, 40, 42, 55, 58, 59, 95, 128, 129, 132–4, 136, 139, 146, 155, 158, 162, 174, 210, 214–16, 311
- complement, 2, 6, 11, 30, 31, 77, 238, 240, 275
- complexity class, 198
- concatenation, string, 12, 22, 28, 75, 78, 174, 178, 193, 194, 196, 252, 310
- connected (in graph), 156, 158
- consensus, 127, 204, 222, 227
- constructor, 82, 88
- contig, 204, 214, 217, 219–26, 228
- corn, 30, 44, 46–8
- coupon collector's problem, 67
- CpG islands, 44
- crystallography, 18, 19, 56
- cystic fibrosis, 56
- cytoplasm, 6, 16, 56
- cytosine, 1, 2
- DATA, 8
- database, 72, 73, 75, 92, 93, 96–8, 100, 102, 105, 107–9, 113, 114, 123–5, 139, 173–5, 177, 193, 195, 231, 318
- db_xref (GenBank), 75
- DDBJ, 72
- deep copy, 164
- deletion
 - from DNA, 55, 56, 210, 246
 - from Perl string, 10, 12
- denaturation (of DNA), 3, 31
- dentistry, 155
- deoxyribonucleic acid (DNA), 1
- deoxyribose, 1, 2
- dereferencing operator, 61, 82, 83, 86, 104, 105, 145
- descendent
 - species, 156
 - in tree, 156, 167, 181, 315
- deterministic computation, 131
- diagonal, scoring matrix, 107, 208, 209, 250, 255
- dicodon, 231, 235, 237, 238
- die**, 69
- digestion experiments, 257, 258, 260, 262, 263, 274
- digits (in strings), 46, 78, 91, 120, 178, 227, 239, 240
- diploidy, 4
- disjoint sets, 222, 289, 313, 317
- divergence, 65
- DNA, 1–7, 10–12, 15–17, 30–2, 41, 44, 47–9, 52, 55–7, 70, 72, 74, 75, 109, 174, 196,

Index

327

- 202, 203, 212, 231–3, 236, 238, 245,
246, 252, 255, 257, 275, 276, 307
- DNA, 228
- DNAse, 31
- Drosophila melanogaster*, 44
- duplication, gene, 55, 57
- dynamic programming, 26, 29, 32, 139, 141,
143, 146, 147, 149, 223, 225, 234, 240,
242, 248, 250, 255
- E-value, 113, 123–5
- EcoRI, 257
- edge, graph, 129, 156–8, 163, 164, 168, 170,
198, 199, 201, 229, 282, 299
- efficiency, algorithmic, 20, 28, 38, 41, 43,
52, 55, 94, 96, 103, 107, 114, 131, 132,
139, 149, 178, 194, 200, 228, 230, 232,
246, 249, 251, 272, 273, 279, 289, 304,
313, 315
- elastin, 4
- electrophoresis, 31, 257, 273
- EMBL, 72
- encapsulation, data, 80, 81, 222
- endonuclease, restriction, 257
- endoplasmic reticulum, rough, 17
- enhancer, 6
- entropy, 47, 48, 53, 57, 125
- enzyme, 4, 6, 16, 31, 257, 258
- Escherichia coli*, 2, 4
- eukaryotes, 4, 6, 75, 231, 237, 244
- eval**, 77, 252
- evolution, 42, 55–7, 65–7, 140, 155, 156,
275
- evolutionary tree, 65, 155, 156
- exhaustive search, 161, 162, 198
- exon, 17, 74, 75, 232, 235, 237
- exon (GenBank), 74
- exponential running time, 45, 130–2, 315
- exporting subroutines and variables, 49, 51,
211, 214, 217, 222
- external nodes, 156
- extreme value distribution, 113
- FASTA, 73, 85, 89, 96, 108, 204
- feature table, GenBank, 74, 75
- filehandle, 8, 83, 84, 303
- fingerprinting, DNA, 245
- folding, RNA and protein, 5, 6, 16, 20, 21,
24–9
- for**, 26, 28, 63, 117
- foreach**, 9, 52, 63
- fragment assembly, 196, 230
- fruit fly, 44, 47, 48
- gamete, 245
- gap, 32, 33, 39, 40, 42, 55, 58, 59, 96, 97,
100, 109, 111, 114, 125, 126, 128, 129,
133, 134, 145, 155, 210, 214–16, 230,
247, 248, 312
- gap penalty, 42, 134
- GCG, 89, 90
- gel, 31, 202, 203, 212, 213, 257, 273
- GenBank, 72, 73, 75, 77, 78, 85, 89, 92, 108,
173, 231, 244
- gene, 1, 4, 5, 44, 55, 57, 74, 231, 232, 234–8,
257, 275–7, 299
- gene** (GenBank), 74, 75
- gene prediction, 6, 231, 235, 238
- genes, 75
- genome, 4, 15, 30, 43, 44, 48, 244, 245, 275,
276
- GenPept, 75
- geysers, 31
- ghostview, 300, 306
- global variable, 117, 199, 252, 260
- globin, 56, 57, 100
- greedy methods, 137, 138, 161, 162, 170, 197,
204, 237
- grep**, 202
- guanine, 1, 2, 18
- guano, 1
- Hamiltonian path problem, 198
- haploidy, 4
- happy clique, 284–8, 295–8
- hash table, 7, 8, 10, 22, 60–4, 68, 81, 83, 84,
86, 88, 98, 103–5, 107, 120–2, 142, 143,
162–4, 174, 175, 179–81, 185, 198, 199,
201, 205–11, 222, 224, 226, 228, 229,
238, 239, 254, 260, 266, 267, 273, 313,
314
- helicase, DNA, 6
- helix, double, 2, 5, 6, 15, 44
- heme group, 57, 93
- hemoglobin, 57, 93, 108
- heuristic, 93, 96, 102, 109, 128, 132, 141
- hexagram, 235, 236, 238–40, 244
- HindIV, 257
- hit, 99, 100, 105–7
- HIV, 2, 155

328

homology, 56–8, 60, 70, 93, 96, 97, 100, 127, 155, 171, 173, 231, 275
 host, 4, 203
 hurdle, 275, 280, 283, 284, 287, 288, 292, 293, 295, 297, 298
 hydrophobicity, 56

 incarcerated, wrongfully, 245
 indel, 55, 58
 inheritance (genetic), 1, 245, 246
 inheritance (object-oriented programming), 79, 80, 83, 85, 86, 88, 190
 insertion, 55, 145, 186, 210, 246
 instantiation of class, 79, 88
 interaction of wishes, 278, 279, 285, 288–90
 intergenic DNA, 245
 internal node, suffix tree, 156, 160, 178, 181, 184, 185
 interpolation, string, 22, 28
 interpreter, Perl, 7, 69, 78, 266
 interval
 arithmetic with, 262–73, 305
 between tandem repeats, 247, 250–2, 254
 in wish list, 278, 279, 283, 290
 intron, 6, 17, 74, 231, 235, 237
 intron (GenBank), 74
 invariant
 data, 81, 308
 loop, 308
 inverted tree, 313
in vitro, 5
in vivo, 5
 iron, 57, 93
 IUPAC, 193

 Java, 80–2, 206
join, 41, 143
 joined (in graph), 156, 163, 290
 junk DNA, 6, 245

 keratin, 4
keys, 61

 laser, 31, 202, 212
last, 62, 63
 layer, in alignment matrix, 141–3, 149, 152
lc, 37
 leaf, of graph or tree, 65, 155–60, 164, 166, 178–80

leghemoglobin, 57
 lenses, scoring matrices as, 67
 leucine, 56, 231
 levels of structure, 17, 18
 lexicographic order, 61, 101
 lg, 46
 ligament, 4
 likelihood, 44–8, 57, 58, 171, 214, 273
 linguistics, 155
 Linux, 300
 local variable, 11, 23, 153, 214
 lod, 46, 52, 59, 60, 64, 66, 68, 233, 234, 236–40, 244
log, 50
 lower bound, 149–151, 200
 lysine, 232

 mammals, 6, 17, 275
map, 52, 53, 70, 164
 methionine, 6, 7, 231
 methylase, DNA, 257
 microsatellite, 245
 Microsoft, 298
 minisatellite, 246
misc_feature (GenBank), 74
 mitochondrial DNA, 75, 275
 morphological features, 155
 mosaic, 30
 Mother Goose, 197
 motif, 173–6, 193, 194
 mRNA (messenger RNA), 6, 7, 16, 17
 MS/DOS, 12
 multidimensional array, 37, 142, 210
 multiset, 258, 263
 muscle, 57
 mustard, 275
 mutation, 30, 55–7, 60, 65–9, 93, 108, 156, 158, 160–2, 164–6, 170, 171, 245, 246, 275
my, 8, 83
 myoglobin, 57, 65

 N-terminus, 4
 nat, 46, 47, 53, 70
 NCBI, 72
 Needleman–Wunsch algorithm, 38, 43, 60, 94, 128, 132, 139, 152, 255, 307
 neighbor, 19, 26, 156, 158, 163–6, 190, 214, 301, 303
 nested parentheses, 22, 77, 301, 302

Index

Index

329

- Nesting Constraint, 19, 20
- Neurospora crassa*, 4
- newline character, 9
- next**, 63, 218
- nickname, 61, 63
- nitrogen, 57
- nodule, 57
- nondeterministic computation, 131
- NP-completeness, 131, 132, 157, 198, 277
- nucleus, 3, 4, 6
- object-oriented programming, 73, 79–81
- observation (statistical), 44, 48
- odds, 45–7, 59, 60, 64, 66, 70, 103, 232, 235, 236
- oligonucleotide, 30
- opsin, 97
- optional argument, 63, 267
- or**, 266
- ORF (open reading frame), 231
- orthology, 57, 97
- overflow, 302
- overlapping sequences, 31, 32, 41, 106, 196–9, 201, 203, 228, 263, 266, 272, 273
- overlapping wishes, 278–80, 282, 284–91, 296–8
- oxygen, 2, 57, 93
- P-value, 113
- package**, 70, 206, 207, 211
- package-level variable, 205, 206
- palindrome, 257
- PAM, 57, 65–8, 71, 100, 109, 123
- paralogy, 57, 97
- parameter
 - statistical, 44, 48, 57, 58
 - subroutine, 11, 176, 181, 252
- parentheses, 9, 12, 20–2, 27–9, 77, 78, 177, 239, 301–4
- parity, 278, 279, 282, 283
- parsimony, 65, 155, 156, 168, 171
- parsing, 77, 91
- Pascal, 80
- paternity lawsuits, 245
- path
 - in alignment matrix, 136, 147, 150, 154, 314, 315
 - in a graph, 156, 178–80, 182, 185, 186, 188, 190, 191, 198, 200, 201, 223, 229
- path compression, 289, 315
- pattern matching, 9, 11, 12, 21, 49, 69, 176, 177, 239
- PAUP, 171
- PCR, 30, 31, 43, 257
- pedigree, 155, 245, 275
- pelagra, 15
- penalties, alignment scoring, 42, 55, 56, 134, 174, 210, 230, 312
- peptide, 4
- permutations, 107, 277–81, 284, 288, 292, 297
- phenylalanine, 56
- phosphate group, 173
- phosphodiester bond, 2
- PHRAP, 70, 196, 202–4, 208, 210, 213, 214, 220, 230
- PHRED, 202, 204, 213, 230
- PHYLP, 171
- phylogenies, 65, 155–7, 160–5, 168, 170–2, 245
- planet Koolman, 70
- planet Thremino, 114, 117, 173
- plasmid, 203
- polymerase, DNA and RNA, 6, 17, 30, 31, 43, 212
- polymerase chain reaction, 17, 30, 43
- polynomial running time, 131, 132, 157, 200, 230, 277, 298
- pop**, 28, 308
- PostScript, 300, 303, 306
- precomputation, 153, 197
- prefix, 31, 32, 35, 38, 39, 41, 95, 98, 179, 184, 189, 197, 198, 228
- pregnancy, 35%, 57
- preprocessing, 96, 98, 99, 102, 103, 105, 178, 191, 240
- Primary Proximity Constraint, 19
- primary structure, 17, 93
- primary transcript, 6
- primate, 57
- primer, oligonucleotide, 30, 31, 203
- print**, 117, 120
- printf**, 120
- product (GenBank), 75
- profile (for protein family), 127, 139, 173, 174
- projection (of alignment), 127, 128, 137
- prokaryotes, 4, 6, 237, 244
- proline, 55, 232
- promoter, 6, 74
- PROSITE, 173–8, 189, 190, 193–5

330

Index

- pruning search space, 160, 161, 198
- pseudocount, 59, 63
- purine, 1
- push**, 28, 104, 165
- pyrimidines, 1
- quadratic space (memory), 96, 112, 180, 307, 312
- qualifier (GenBank annotation), 74, 75, 92
- quality of sequence data, 196, 203, 204, 212–14, 216, 217, 223, 225, 227, 228
- quaternary structure, 18
- query, 93, 96–100, 102–9, 113, 114, 124, 125, 178, 179, 181
- queue, 181
- qw**, 49
- reading frame, 10–12, 75, 231, 232, 237, 242, 243
- rearrangement, genomic, 275, 276
- recurrence relation, 40, 111, 112, 116–22, 234, 311
- recursion, 20–3, 27–9, 39, 40, 42, 122, 159, 166, 168, 170, 191, 192, 200, 225, 241, 248, 261, 262, 268, 269, 300, 301, 303, 304, 310
- ref**, 87, 88
- reference (Perl), 60, 61, 82–4, 103–5
 - hash, 61, 83, 103, 105
 - list, 103–6, 143
 - object, 83
 - scalar, 61
 - subroutine, 144, 145
- regular expressions, 176
- relevance tests for alignments, 143–5, 148, 153
- repeat_region** (GenBank), 74
- replication, 5, 6, 55, 212, 245, 275
- replication fork, 6, 245
- residue, 4, 5, 7, 8, 56, 58–60, 62, 64–6, 70, 96, 97, 104, 114, 118, 124, 125, 132–4, 152, 155, 171, 173–6, 190–2, 235, 275
- return**, 28, 266
- reverse**, 11
- ribonucleic acid (RNA), 2
- ribose, 2
- ribosome, 6, 7, 16–18, 55, 74, 231, 232
- rice, 30
- RNA, 2, 4–7, 10–12, 16–21, 24, 26, 29, 74, 212, 231, 232, 300, 301, 304, 305
- rooted tree, 156, 158
- rRNA, 6, 7, 17, 232
- rRNA (GenBank), 74
- s**, 10
- satellite, 245–52, 254, 256
- scalar, 8, 60–2, 103, 105, 164, 273
- secondary structure, 16, 18–21, 24, 29, 212, 300
- sedimentation, 17
- sequence alignment, 56, 127, 141, 155, 171, 172, 230, 312
- sequence assembly, 30, 32, 70, 196, 202, 204
- sequencer, automatic, 31, 32
- serial number of object, 205–8, 217, 223
- shift**, 9, 50
- Shine–Dalgarno sequence, 2, 7, 235
- shotgun sequencing projects, 31, 43
- signed permutation, 277, 284
- similarity, 32, 33, 35, 36, 38, 41–3, 56, 57, 93–6, 107, 138, 141, 152, 307, 309, 310
- simple block, 279–82, 284, 285, 288, 295, 298
- simple wish, 278–80, 284–7, 295–8
- Smith–Waterman algorithm, 94, 96, 125, 203, 208, 210
- snippet, 223
- snRNA, 17
- sort**, 61, 62, 101
- soybean, 57
- spiral, 2
- splice**, 41, 165
- spliceosome, 17, 18
- splicing, 6, 17
- split**, 9, 21, 41, 100, 117, 143
- SRP, 17, 305
- stack, 252, 290, 291
- STADEN, 230
- static variables and subroutines, 206
- statistics, 44, 109, 116, 124, 125, 173
- STDERR, 303
- STDIN, 10, 60, 62, 68, 117
- STDOUT, 303
- Sterling’s approximation, 34
- sticky end, 257
- straddle, 283, 284, 293, 298
- strain** (GenBank), 75
- string matching, 32
- subclass, 79–81, 86, 88

- substitution matrix, 55–7, 60, 65, 67, 71, 109, 124, 312
- substr**, 12, 37, 41, 78, 183, 239, 240
- sucrose, 17
- suffix, 32, 41, 95, 151, 152, 177–90, 193–9, 223, 226, 228, 246–53, 255, 256
- suffix tree, 177–82, 184, 185, 189, 190, 193–9, 256
- sugar, 1, 2
- superclass, 79, 80, 86, 87
- superhurdle, 283, 284, 288, 292, 293, 298
- superstring, 196–8, 200–2, 228–30
- Svedberg unit, 17
- SWISS-PROT, 75, 92, 173
- synteny, 275
- tandem repeat, 245, 246, 254, 256
- target sequence, 32, 99, 100, 105–8, 174, 176–83, 185, 187–9, 191, 193, 195, 196, 203, 204, 220, 228, 252
- taxon, 65, 72, 155
- taxonomy, 65, 72, 74, 155
- telomere, 246
- template, 30, 31, 212
- tendon, 4
- terminal, 10, 303
- tertiary structure, 18–20
- Thermus aquaticus*, 31
- thymine, 1, 2
- tissue_type** (GenBank), 75
- topological sort, 226
- topology of phylogeny, 156, 157, 160, 161, 168, 226
- tr**, 11
- training set, 48, 57–60, 65, 66, 232, 235, 237, 238, 244
- transcription, 5–7, 11, 17
- transl_table** (GenBank), 75
- translation (RNA to protein), 2, 5–7, 10–12, 16, 17, 61, 75, 231, 232, 239
- transliteration (strings), 11
- translocation, 55, 275, 277
- traveling salesperson problem, 198
- trichotomous comparison, 178
- trigram, 232, 233, 235, 240, 244
- tRNA, 6, 16–18, 232
- tryptophan, 6
- Turing machine, 131
- turnip, 275, 276
- twist (even permutation), 277–88, 291–4, 297–9
- twisted block, 279–81, 283, 284, 292, 297, 298
- twisted wish, 278–80, 284–6, 288, 297, 298
- tyrosine, 17, 56
- undef**, 9, 84, 266, 271
- underflow, 45, 302
- Unix, 89
- unrooted tree, 156
- unshift**, 308
- untranslated region, 74
- untwisting
 - DNA, 6
 - a wish, 278, 279, 286
- upper bound, 113, 150, 151, 170, 171, 201, 202, 229
- upstream, 235
- uracil, 2, 18
- use**, 49, 51, 190, 206
- value, 209
- vertebrate, 57
- vertices of graph, 198, 200, 201, 229
- virus, 2, 4, 155
- visit, 198, 201, 229
- VNTR, 245, 246
- weather, 48
- well-founded recursion, 23
- while**, 9, 12, 63, 117
- wildcard, 176, 192, 194
- Wisconsin package, 89
- wish list, 278, 280–2, 286, 288–90, 295
- wobble, 18
- x**, 27, 53
- YACs, 203
- yeast, 17, 18, 203
- Yellowstone National Park, 31
- Yersinia pestis*, 91
- Zea mays*, 44
- zebra, 65