

INDEX

- " (null string), 63
- ! (logical NOT operator), 41, 303
- !Ai, 253, 254
- !Gubi, 253, 254
- #! (shebang), 51
- =~ (binding operator), 13, 23
- => (in hashes), 14
- \$ (true at end of string), 39, 306
- \$&, 305
- \$_, 38, 39, 41, 130, 302
- \$_ [0], 61, 312
- \$1, 60, 231, 308
- \$ARGV [0], 169
- & (ampersand), 287
- && (logical AND operator), 40, 70, 303
- *, 39
- *=, 49
- . (concatenation operator), 303
- .=, 39
- ; (semicolon), 13
- @_ array, 61, 62, 312
- @ARGV, 169, 316
- \ (backslash), 187, 191, 231, 307
- \d (digit), 191, 307
- \D (non-digit), 307
- \n (new line), 13, 307
- \S (non-whitespace), 118, 191, 235, 307
- \s (whitespace), 307
- \t (tab), 23, 67
- \W (non-word character), 307
- \w (word character), 307
- ^ (true at beginning of string), 39, 306
- ' ' (backticks), 318
- | (pipe), 51, 285
- || (logical OR operator), 38, 70, 303
- = (assignment operator), 70, 302
- =~ (binding operator), 306
- == (numerical equality), 70
- 1,4-beta-glycosidic bond, 141
- 1000 Genomes Project, 241, 245
- 3' end, 5
- 5' end, 4

- ABL, 77, 81, 83, 89
- ABL1, 77, 83, 86, 88, 89
- ABL2, 88
- acceptor site, 209
- acute lymphoblastic leukaemia (ALL), 76
- adenine, 5, 6
- Africa, 252, 253
- AIDS, 106
- Alexei, son of Tsar Nicholas II, 150, 151
- algae, 228
- alignment, 77, 78, 80, 81, 89, 90
- allele, 244, 247, 252, 263, 270, 272
- alpha tectorin, 93
- Altman, Sidney, 223
- Alu elements, 202
- AluI, 22
- ambiguity symbols, 245, 256
- amino acid, 8, 9, 10, 14, 47, 48, 49, 99
 - amino acid composition, 164, 170, 171, 208
 - amino acid sequence, 10, 47, 48, 154
 - amino acid substitution matrix, 80, 89, 174
- ancient DNA, 46
- AND operator, 40
- angular brackets, 230
- annealing, 45
- annotation, 177, 179, 181, 192
- anticodon, 10, 224
- antiparallel, 6, 12
- antisense, 36
- APC (adenomatous polyposis coli), 74
- apoptosis, 74, 75
- Arber, Werner, 19, 20
- archaea, 2
- Argonaute, 32
- arguments, 316
- array, 14, 301
 - two-dimensional, 215, 247, 263
- array of hashes, 273
- arrays of arrays, 313
- arrays of hashes, 314
- assembly, 166
- assignment, 70
- ATP (adenosine triphosphate), 223, 227, 228
- AUG, 10
- Australia, 123, 124

- backslash, 187, 191, 231, 307
- backticks, 238, 318
- bacteria, 139
- bacteriophage, 174
- BamHI, 22
- Bantu, 253
- base-pairing, 37, 69, 70, 73, 224, 226, 230
- bases, 2, 4, 6
- BCR, 76, 77, 83, 88, 89, 90
- BCR-ABL, 76, 77, 83, 86, 87, 89
- Bertani, Guisepppe, 20
- beta-1,4 glycosidic bond, 140
- Bevasiranib, 34
- bg, 287
- binding operator, 13, 23, 306
- BioPerl, 318
- biotechnology, 34
- bitter taste, 260, 261, 262, 263, 264
 - bitter-taste receptor, 261
- BLAST, 36, 37, 74, 81, 82, 83, 86, 87, 88, 89, 90, 138, 142, 143, 144, 145, 146, 147, 148, 154, 155, 168, 179, 224, 290, 291, 320
- BLAST+, 83
- blastn, 142
- blastp, 86, 142
- blastx, 142
- Blat, 292
- blood coagulation
 - factor V, 243
 - factor VII, 162
 - factor VIII, 151, 160
 - factor IX, 151, 153, 162
 - factor X, 151, 162
 - factor XI, 153, 154, 156
- blood coagulation (clotting), 150, 151, 152, 153, 156, 157, 160, 161, 243
- BLOSUM62, 80
- blue-green algae, 228
- bootstrapping, 112, 113, 117, 118, 119
- branch site, 209, 210, 211
- BRCA1, 37, 137
- breast cancer, 37, 74
- bunzip2, 288
- Bushmen, 245, 249, 253, 256, 257, 261, 263

- Caenorhabditis elegans*, 32
- CAG repeat, 57, 58, 59, 64, 65
- cancer, 74, 75, 76, 77
 - colorectal, 74
 - therapy, 46
- Candidatus Desulforudis audaxviator*, 139

- capturing matched patterns, 308
- cat, 193, 195, 282
- cd, 280
- Cech, Thomas, 223
- Celera, 1
- cell cycle, 75, 137
- cellulase, 140, 141, 142, 143, 144, 146
- cellulose, 140, 141
- cellulose-binding domain, 146
- central nervous system (CNS), 93
- Charcot-Marie-Tooth disease, 242
- chimpanzee, 92, 111, 245
- chlorarachniophytes, 228
- chlorophyta, 230, 235, 236
- chloroplasts, 228, 229, 230, 235, 236
- chmod, 52, 316
- chomp function, 39
- chromatin, 199
- chromosomal translocation, 56, 76
- chromosome, 7, 11, 76, 205, 266, 267, 269, 270, 272, 273, 276, 277
 - homologous, 266
 - non-homologous, 76
- chronic myelogenous leukaemia (CML), 76, 77
- cloning, 19
- close function, 37, 38, 310
- clot, 151, 152
- ClustalW, 96, 97, 98, 102, 118, 119, 127, 292, 293
- cmbuild, 297
- cmcalibrate, 297
- cmsearch, 233, 234, 235, 236, 297
- CNTNAP2 (contactin-associated protein-like 2), 94
- cockroaches, 140
- codon, 8, 10, 14, 15, 47, 48, 49, 143
- colon cancer, 164
- comm, 285
- command line, 193
- commenting code, 315
- comparative genomics, 177, 178
- comparison operators, 40, 303
- complementarity, 6, 7, 12, 45
- Consense, 294
- constitutive heterochromatin, 166
- convergent evolution, 123, 129
- copy-number variation, 243
- cotton, 34
- counting, 40, 170, 309
- covariance models, 226, 227, 229, 233, 235
- cp, 281
- CPAN, 319, 321

INDEX

- CpG island, 198, 200, 201, 202, 203, 204, 205, 206, 207
- CpG site, 200, 201, 202, 203, 204, 206
- Crick, Francis, 6
- crossing-over, 76. *See* recombination (crossing-over)
- cut, 283
- cyanobacteria, 228, 229
- Cygwin, 278, 301
- cytosine, 5, 6
- D#kgao, 253, 254
- Darwin termite, 140
- Darwin, Charles, 3
- Dawkins, Richard, 141, 222
- Dayhoff, Margaret, 174
- DDBJ, 175
- deafness, 93
- decimal places, 215, 216
- deoxyribose, 8
- diatom, 228
- Dicer, 32
- die function, 38
- diff, 285
- dingoes, 123
- dinoflagellates, 228
- disease, 7, 243
- distance matrix, 107, 245, 247, 248, 249
- DNA, 2, 4, 6, 8, 11, 12, 14, 198, 199, 223, 228
- DNA ligase, 22
- DNA methylation, 201, 202
- DNA methyltransferases, 201
- DNA mismatch repair, 75
- DNA recombination, 56
- DNA repair, 75
- DNA replication, 4, 8, 11, 12
slippage, 58, 166
- DNA sequencing, 138, 241
- Dnapars, 127, 128, 294
- Dobzhansky, Theodosius, 2
- dog, 122
- domain accretion, 161
- donor site, 209
- dotplot, 78, 165, 171
- double quotes, 12, 13
- dsRNA, double-stranded RNA, 32, 35
- dynamic programming, 78, 79, 81, 90
- EBI (European Bioinformatics Institute), 175, 179
- EcoRI, 22, 23
- EcoRV, 21, 22, 29
- Eddy, Sean, 155, 226
- EFetch, 185, 187, 188, 195
- egg cell, 266, 267
- eiRNA, 33
- ELink, 185, 195
- Emacs, 283
- EMBL database, 174, 175, 177, 178, 180
- EMBOSS, 167, 289
- end-of-line character, 39
- endo-1,4-beta-glucanase, 143
- endoplasmic reticulum (ER) membrane, 227
- endosymbiosis, 228, 229
- Ensembl, 177, 179
genome browser, 180
- Entrez (NCBI), 96, 114, 126, 181, 183, 185, 192, 196, 230
- Entrez Programming Utilities, eUtils, 181, 182, 185, 186, 188, 192
- environmental sequencing, 138
- enzymatic cascade, 152
- enzyme, 4
- epigenetic inheritance, 201
- eq, string equality operator, 70
- error message, 38
- Escherichia coli*, 19, 22, 198, 227
- ESearch, 185, 188, 189
- EST (expressed sequence tag), 177
- eubacteria, 2
- eukaryotes, 2
- E-value, 82, 88, 145, 148
- evolution, 2, 3, 4, 93, 95, 160, 161, 178, 222, 223, 224, 228, 229, 252
- exon, 5, 11, 208, 209, 211, 217, 219, 257, 258, 259, 260
- extrinsic pathway, 151
- false, 305
- family quartet, 269, 271
- FASTA format, 38, 62, 86, 98, 114, 117, 127, 157, 188, 204, 216
- feature table, 84, 177, 187, 188
- ferritin, 66, 67, 68, 69
- fg, 287
- fibrin, 152
- fibrinogen, 152
- fieldholder, 132
- file, 37
- file system, 280
- filehandle, 37, 312
- find, 286
- Fire, Andrew, 32

- FMN1, 58
- folding, 10
- for loop, 16, 39, 49, 53, 54, 204, 215, 247, 259, 304
- foreach loop, 25, 26, 99, 100, 304
- format declaration, 130, 131
- forward strand, 7
- Fox, Arthur L., 261
- FOXP2, 93, 94, 95, 96, 98, 101, 102
- fragile X syndrome, 58, 59
- function, 60, 311
- fusion protein, 76, 77, 81, 83, 89

- G/aq'o, 253, 254
- Galaxy, 253, 255
- gamete, 266
- gap, 78
- gap penalty, 78
- GC content, 36, 47, 202, 203
- gedit, 283
- GenBank database, 174, 175, 176, 177, 178, 180, 185, 192
 - database entry, 177
- gene amplification, 56
- gene expression*, 11, 31, 199, 201
- gene finding
 - ab initio*, 208
 - de novo*, 208
 - homology based, 208
- gene knock-down, 31
- gene knock-in, 31
- gene knock-out, 31
- gene prediction, 208
- gene technology, 19, 22, 151, 173
- gene(s), 7, 8, 198
- genetic code, 8, 10, 14, 49, 143
 - degenerate, 9, 47
- genetic information, 2, 4
- genome, 7, 92, 230
- genomic imprinting, 202
- genomic localization, 235
- genomic sequences, 177
- genomics, 7
- get function, 188
- giant northern termite, 140
- global alignment, 78, 79, 96
- global modifier, 24, 64, 308
- glutamine, 58
- glycosylation, 164
- gossypol, 35, 43
- greedy matching, 60, 158

- green algae, 230
- grep, 193, 195, 261, 285
- guanine, 5, 6
- guide tree, 96, 97, 117
- gunzip, 288, 319
- gut, 139

- H19, 202
- H1N1, 188, 190, 192, 195
- H2N2, 190
- H3N2, 190
- Haemophilus influenzae*, 20
- hairpin, 69, 72, 224, 227, 229, 235
- hash, 14, 15, 16, 24, 25, 99, 130, 235, 258, 259, 260, 302
- hashes of arrays, 313
- hashes of hashes, 314
- head, 282
- helix, 70
- haemagglutinin, 190
- haemoglobin, 56
- haemophilia, 150, 151
 - haemophilia A, 151
 - haemophilia B, 151
- heterochromatin, 199
- heterozygous, 244, 245, 256, 270
- histone, 199, 202
- histone acetylation, 199
- HIV, 106, 113, 114, 117, 119, 223
 - env*, 114, 115, 117
 - genes, 115
 - genome, 113, 114
 - reverse transcriptase (RT), 114, 115, 117, 118
- HMMER, 155, 156, 157, 236, 296
- hmmscan, 155, 156, 157, 159
- homologous recombination, 76
- homologue, 82, 227, 236
- homology, 138, 154, 224, 236
- homozygous, 244, 245, 256, 270
- HSP (high scoring pair), 88, 89, 90
- humans, 92, 101, 111
- human genetic diversity, 249, 250
- human genome, 7, 11, 139, 198, 202, 218, 222, 241, 243
- Human Genome Project, 1, 241
- Hungerford, David, 76
- huntingtin, 58
- Huntington, George, 57
- Huntington's disease, 55, 57, 58
- hydrogen bonds, 6

INDEX

- hyperferritinaemia cataract syndrome, 66, 68
 hypoxanthine-guanine phosphoribosyltransferase (HGPRT), 57
- if statement, 40, 303
 indices, 247
 Infernal, 227, 229, 230, 233, 237, 297
 influenza virus, 173, 188, 190, 191
 influenza virus A, 177, 190
 info, 287
 Ingram, Vernon, 56
 insulin, 174
 int function, 273, 310
 interleaved format, 98
 intestine, 139
 intrinsic pathway, 151
 intron, 11, 209, 211
 inverted repeat, 21
 iron, 66, 67, 68, 69
 iron responsive element (IRE), 66, 67, 68, 69, 70, 72, 226
 IRP, 67, 68
 irradiation, 75
- Jalview, 293
 join function, 310
- Kent, Jim, 1
 keys, in hashes, 14, 15
 Khoisan, 245, 249, 253
 kill, 287
 Kim, Seong-Jin, 242, 245
 kringle domain, 161
- lariat structure, 209
 last function, 100, 304, 305
 lc function, 218, 309
 length function, 17, 26, 39, 99, 204, 309
 Lesch–Nyhan syndrome, 57
 less, 282
 leukaemia, 242
 Li–Fraumeni syndrome, 75
 lineage, 129
 local alignment, 78, 79, 80, 88, 236
 log function, 216
 logical operators AND, OR and NOT (Entrez), 182
 log-odds matrix, 213
 loop statement, 15
 low complexity sequence, 146
 lower triangular matrix, 250
 ls, 280
- Lupski, James, 242
 Luria, Salvador, 20
 Lynch syndrome, 75
- malaria, 57
 mammals, 93
 man, 287
 marsupials, 125, 126, 127, 128, 134, 136
Mastotermes darwiniensis, 140
 match operator `m//`, 23, 24, 308
 meiosis, 76, 266, 267, 268
 Mello, Craig, 32
 membrane, 2
 metagenomics, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 148
 Metatheria, 135
 methionine, 9
 methylating enzyme, 21
 methylation, 201
 MFOLD, 235
 Miller syndrome, 269
 minimal evolution, 107
 minus strand, 7
 miRNA, 32, 33, 34
 mispriming, 48
 mitochondria, 227, 228
 mitochondrial DNA, 252
 mitochondrial genome, 126
Mixotricha paradoxa, 141
 mkdir, 282
 molecular phylogeny, 105, 106, 107, 192
 MrBayes, 127, 128, 295
 mRNA, 5, 8, 10, 35, 36, 59, 62, 66, 67, 68, 69, 177, 198, 208, 209, 226
 MS2, 174
 MUC5AC, 166
 MUC6, 168, 171
 mucins, 164, 165, 166, 172
 Mullis, Kary, 44
 multiple sequence alignment, 92, 95, 96, 107, 117, 127, 155, 211, 227, 237, 249
 Muscle, 96, 293
 mutation, 3, 68
 non-synonymous, 243
 synonymous, 243
 my declaration, 52, 54
Mycobacterium tuberculosis, 46
 Myers, Gene, 1
- Nasutitermes*, 142
 Nathans, Daniel, 20

- NCBI, 175, 177, 178, 179, 180, 181, 182, 185,
 192, 196
 documentation resources, 196
 SNP database, 243
 Taxonomy database, 129, 135, 181
- ne (not equal), 63, 70
- Neanderthal, 46
- nedit, 283
- Neighbor, 248
- neighbour-joining, 107, 108, 112, 113, 119, 127,
 128, 245
- neocortex, 93
- neuraminidase, 190
- new line character, 13
- Newick format, 97
- next function, 305
- Nexus format, 295
- Nicholas II, 150, 151
- NMR, 180
- Nobel Prize, 6, 20, 32, 223
- NOT, logical operator, 41
- Nowell, Peter, 76
- nucleomorph, 228
- nucleosome, 199
- nucleotide excision repair (NER), 75
- nucleotide(s), 6, 8
- nucleus, 199
- number, 14
- numerical equality, 70
- oestrogen response element
- oligonucleotide, 48
- OMIM (Online Mendelian Inheritance in Man),
 181
- oncogenes, 34, 75
- open function, 37, 38, 310
- open reading frame (ORF), 10
- operating system, 278
- OR operator, 38
- organism, 2
- orthologue, 88, 98, 224, 226
- orthology, 83, 224
- Out-of-Africa hypothesis, 252
- p53, 74, 75
- palindrome, 21, 29
- PAN domain, 153, 156
- paralogue, 88, 224
- paralogy, 83, 93, 224
- parentheses, 59, 60, 62, 231, 308
- parsimony, 127, 128
- paternity testing, 46
- pattern matching, 23, 24, 26, 49, 229, 230,
 305
- PCR (polymerase chain reaction), 44, 45, 46, 47,
 48, 50, 54, 114, 125, 151
- Perl modules, 188, 263, 264, 318, 319, 321
- Perl one-liner, 193, 317
- personal genomes, 241, 242, 252, 266
- Pfam database*, 155, 156, 157, 158, 159, 160, 161,
 162, 179
- phenylthiocarbamide (PTC), 261, 262, 263
- Philadelphia chromosome, 76, 77
- phosphate, 4
- photosynthesis, 228
- PHYLIP, 112, 127, 248, 294
- phylogenetic tree, 106
- pipe, 193, 231
- piRNA*, 33, 34
- plant, 228
- plant toxins, 261
- plasmid, 23, 44
- Plasmodium falciparum*, 57
- plus strand, 7
- POD, 316
- point mutation, 56
- polarity, 4, 5, 7, 8
- polyA tail, 5, 209
- polyadenylation, 209
- polymerase chain reaction (PCR), 45
- polyQ (poly-glutamine), 58
- position specific scoring matrix (PSSM), 154, 211,
 212, 213, 215, 216, 218, 219
- Prader–Willi disease, 225
- predator, 123
- primary ciliary dyskinesia, 269
- primary structure, 179
- primate, 92, 102
- primer, 45, 47, 48
- print function, 12, 13, 38, 310
- profile, 154, 155, 211, 237
- profile hidden markov model (profile HMM), 155,
 227
- progressive alignment, 96
- protease domain, 153, 156, 160, 161
- protein, 2, 4, 5, 8, 11, 14, 198
- Protein Data Bank (PDB), 6, 22, 153, 179, 180,
 181, 226
- protein domain architecture, 146, 153, 154, 160
- Protein Information Resource (PIR), 174
- protein sequence, 178
- protein synthesis, 8, 222

- Proteobacteria, 228
 proteolysis, 152
 proteolytic activity, 152
 prothrombin, 151, 162
 pseudocount, 215, 216
 PSI-BLAST, 81, 155, 156, 162, 292
 PTS domain, 164, 165, 166, 168, 171
 PubMed, 181
 purines, 5
 pwd, 280
 pyrimidine dimer, 75
 pyrimidines, 6

 quantifier, 39, 41, 59
 Queen Victoria of England, 150

 R, 147, 160, 171, 205, 206, 219, 276, 323, 329
 RAD9, 137
 reading frame, 142
 rearrangements, 56
 reassortment, 191
 REBASE, 24
 recombination, 166
 recombination (crossing-over), 266, 267, 268, 269, 270, 272, 275, 276, 277
 RefSeq, 179
 regular expression, 23, 27, 59, 230, 231, 235, 305
 repeat operator, 248
 repeats, 58, 165, 166, 167
 restriction enzyme, 19, 20, 21, 22, 24, 25, 26, 28, 29, 30
 reverse complement, 61, 230
 reverse function, 12, 309
 reverse strand, 7
 reverse transcriptase (RT), 114
 reverse translation, 48
 Rfam database, 179, 227, 233, 238
 ribose, 8
 ribosomal RNA, 222
 ribosome, 227
 riboswitches, 68, 226
 rm, 281
 rmdir, 282
 RNA, 2, 4, 8, 11, 14, 37, 190, 223
 catalytic RNA, 223
 de novo gene finding, 224
 interference, 31, 33
 non-coding RNA (ncRNA), 202, 222, 223, 224, 225, 226, 227
 primary structure, 224, 226, 227
 RNase MRP, 222
 RNase P, 222, 224
 secondary structure, 224, 226, 227, 229, 235
 secondary structure prediction, 235
 silencing, 31
 tertiary structure, 224
 RNA polymerase, 8, 198
 RNA polymerase II, 198
 RNA world hypothesis, 222, 223
 RNAi, 31, 34, 35, 36, 39, 42, 47
 Rutherford, Ernest, 173

Saccharomyces cerevisiae, 31
 Sanger, Frederick, 174
 scalar, 14, 301
 SEA domain, 165, 166
 segregation, 266, 267
 selection, 3
 selenocysteine insertion sequence (SECIS), 68, 226
 sense, 36
 seqboot, 294
 sequence logo, 211, 220
 sequence similarity, 137
 serine-protease domain, 153
 serum amyloid P component, 216
 sexual reproduction, 267
 shebang, 51
 sickle-cell anaemia, 56, 57
 signal peptide, 227
 signal recognition particle (SRP), 227
 single quotes, 12, 13
 siRNA, 32, 33, 34, 36, 37, 42
 Smith, Hamilton, 20
 SNORD116, 225
 SNP (single nucleotide polymorphism), 242, 243, 244, 245, 248, 249, 253, 254, 256, 257, 258, 259, 260, 261, 263, 269, 270, 272
 SNP database, 181
 snRNP, 208, 209, 210, 211
 somatic cells, 74
 sort (Unix), 51, 193, 195, 284
 sort function, 310, 311
 Spanish flu, 173, 188, 190
 species, 2
 speech, 93, 94, 95, 102
 sperm cell, 266, 267
 splice site, 209, 210, 211, 212, 213, 215, 218, 219
 spliceosome, 209, 210
 splicing, 4, 11, 93, 151, 208, 209, 210, 222
 split function, 130, 145, 158, 309
 sreformat, 297

- SRP RNA, 227, 228, 229, 230, 231, 233, 235, 236, 237
- SSH (secure shell), 278, 280
- statement, 13
- Stein, Lincoln, 300
- Steitz, Joan, 208
- stem-loop, 224
- stochastic context-free grammar models (SCFGs), 226
- strict module, 52, 54
- string, 12, 14
- string equality operator, 70
- sub, 311
- substitution operator *s///*, 24, 26, 65, 204, 308
- substr function, 15, 16, 26, 39, 99, 309
- sugar, 4, 8
- swine flu, 188, 191, 194, 195
- Swiss-Prot, 83, 84, 85, 86, 90, 95, 143, 144, 156, 157, 179
- synaptic plasticity, 95, 102
- system function, 238, 317, 318

- tail, 282
- tandem repeats, 166
- tar, 288, 320
- TAS2R38, 261, 262
- Tasmania, 124
- Tasmanian devil, 125, 129
- Tasmanian tiger, 121, 122
- TATA binding protein, 200
- TATA box, 200
- taxonomy, 126, 129, 130, 134
- Taxonomy database (NCBI), 129, 135
- tblastn, 142, 144
- tblastx, 142
- T-Coffee, 96, 293
- termites, 138, 140, 141, 142
- therapeutics, 34
- three-dimensional structure, 5, 153, 180
- thrombin, 152
- thylacine, 121, 122, 123, 124, 125, 127, 135, 136
- Thylacinus cynocephalus*, 122
- thymine, 5, 6, 8
- tissue factor, 151
- transcription, 4, 5, 8, 11, 14, 93, 198, 199, 200, 202
 - transcription factor, 93, 198, 199, 200
 - transcription start site, 198, 200, 202, 203, 205
- trans-esterification reaction, 209
- transferrin, 67
- transferrin receptor (Tfr), 67, 68, 69
- translation, 4, 5, 8, 10, 11, 14, 67, 68, 142, 222, 226
- transliteration operator *tr///*, 12, 13, 170, 204, 218, 308
- transposon, 34
- tree construction
 - character-based, 107
 - distance-based, 107
- tRNA, 10, 174, 222, 224
- true, 305
- tuberculosis, 46
- tumour, 34
- tumour suppressor, 34
- tumour suppressor genes, 75, 201
- Tutu, Archbishop Desmond, 242, 245, 252, 253, 254

- U1 RNA, 210, 211
- U2 RNA, 210, 211
- U2 snRNP, 210, 211
- uc function, 218, 309
- UCSC Genome Bioinformatics, 177, 179, 213, 244, 257
- UCSC genome browser, 180, 261
- UCSC table browser, 258
- UCSF Chimera, 6, 22, 153, 226
- ulcerative colitis, 164
- UniProt, 83, 84, 86, 95, 156, 179
- uniq, 193, 195, 284
- Unix, xvii, 12, 192, 193, 278
- Unix shell script, 238
- unless statement, 40, 41, 42, 304
- until loop, 304
- untranslated region (UTR), 5, 66, 67, 68
- unwinding, 36
- uracil, 8
- UTR (untranslated region), 226

- values, in hashes, 14, 15
- variable, 12
- variable interpolation, 13
- vascular endothelial growth factor, 34
- Venter, Craig, 138, 139, 241, 242, 245
- vi, 282
- von Willebrand disease, 151
- von Willebrand factor, 151
- VWD domain, 165, 166

- Wall, Larry, 300
- warning messages, 52

INDEX

- Waters, Charles Oscar, 57
Watson, James, 6, 241, 242,
245
wc, 64, 285
wget, 187, 287, 288
while loop, 37, 38, 64, 65, 304
Wnt signalling pathway, 74, 75
wood digestion, 142
woolly mammoth, 46
word matching, 81
word neighbours, 81
write function, 130, 131

x (repeat operator), 303
Xeroderma pigmentosum, 75
XML, 191
X-ray crystallography, 153, 180

YanHuang, 242, 245

zebra finch, 95
zero-based index, 255, 259
Zuker, Michael, 235
 β -globin, 56
 δ -cadinene synthase, 35