Index

absolute value, 275
abstract (I)LP formulation, 13
abstract problem formulation, 13
statement, 13
adjacency matrix, 84
alignment of contact maps, 373
alignment score, 374
allele, 49, 254, 332
Alzheimer’s, 24
amino acid, 122
backbone, 122
chain, 122
contact, 129
hydrophilic, 129
hydrophobic, 129
residues, 122
rotamer, 124
rotations, 123
side chain, 122
stability, 125
AND, 206
ants, 18
arborescence, 288, 308
minimum weight, 288
minimum-weight spanning, 293
spanning, 288
arginine, 123
artifact problem in cancer, 266
assignment ILP, 171, 183
assignment inequalities, 170
association mapping, 253
candidate gene, 253
genome wide, 253
ataxia, 21
ATCAY, 21
bacteria, 153
balance of molecules, 215
bananas, 385
Bayesian analysis, 341
bbminph.pl, 356
bbmisstest.pl, 264
beauty or strength, 100
bi-objective optimization, 273
biclique, 363
balanced, 364
near, 364
bicluster, 365
binary
character, 49
formulation, 11
tree, 357
variable, 11
biological model, xiv
biological networks, 15
bipartite graph, 138, 311
black box, 389
Boolean network, 205
BRCA1, 332
breakpoint
cost, 329
distance, 329
median problem, 329
phylogeny problem, 329
breakpoint lower bound, 326
breast cancer, 66
breversals.pl, 327
bulk sampling, 291
Buneman graph, 101
cabbage, 318

cancer, 28, 53, 66, 309
artifact problem, 62
cervical, 169
endometrial, 53
MCR problem, 60
metastasis, 60
secondary, 60
Cape of South Africa, 4, 370
captive breeding, 342
case, 253
causal site, 255
central strings, 192
cervical cancer, 169
character, 49
binary, 49
clique, 55
compatibility, 51
compatible, 52
convergent, 89
ideal, 50
incompatible, 55
perfect, 50
removal, 54
removal problem, 72
chemical
compound, 206
reaction, 206
chromosome, 332
chromosome reversals, 316
CL2graphfile.py, 42
class of a partition, 236
cleanres.pl, 168
cleanres.py, 117
CLgraphfile.py, 38
CLgraphgen.py, 38
clique, 26
edgeweighted, 126
ILP for second largest, 41
ILP solution, 29
in MCR problem, 55
maximal, 26
maximum, 26, 55, 84, 387
maximum weight, 39, 267
most probable, 40
node weights, 39
second largest, 40, 229
clone, 60
cluster, 358
clustering, 357, 362
clustering coefficient, 26
clustering-layou, 358
CMO, 372
extended scoring scheme, 376
problem, 374
CNV, 290
coalescent, 351
coefficients, 5
coevolution, 142
column, see also character, site, 49
column pattern, 267
column permutation
problem, 164
via TSP, 165
combinatorial drug therapy, 366
common subsequence, 313

communities, 18, 235
community detection, 235, 361, 365
compact TSP formulation, 171
compatibility, 52
complementary nucleotide pair, 105
completion
constraint, 269
of an extended matrix, 269
theorem, 270
compound, 206
concrete LP formulation, 5, 8
conflicted triples, 284
connected subgraph, 303
problem, 309
consecutive ones, 164
conservation, 4
conservation corridors, 309
conservation of atoms, 213
contact map
alignment, 373
overlap (CMO), 372
control, 253
convergent
character, 89
mutation, 89
phylogeny, 89
copy number variation (CNV), 290
costs
fixed, 371
Cplex, xii, xiv, 10
creative commons, xvii
cut, 247
(6,1), 247
capacity, 247
dgetic, 25, 247
in protein graphs, 248
maximum, 253, 255
multi, 247
node, 25
DAG, 331
Dantzig, George, 10, 181, 385
DDI, 138
cover, 138
de novo sequencing, 159
degree of a node, 23
in and out, 25
dense near clique, 76
density of a graph, 25, 77
diploid, 331
directed acyclic graph (DAG), 331
DNA sequence assembly, 160
DNA sequencing, 159
reads, 160
Dollo parsimony, 268
domain-domain interaction, 137
driver gene, 309
problem, 310
drug
bad side effects, 369
drug cover, 367
drug therapy
combinatorial, 366
drug/protein model, 367
ductal carcinoma, 290
dysbindin, 51
Index

ecology, 3
edge cut, 25, 247
elephant testes, 90
enzyme, 206
epilepsy, 296
equilibrium behavior, 219
evolutionary tree
drawing, 145
expectation, 239
expected number, 238
extended matrix, 269
false negative, 267
false positive, 267
family, 378
feasible solution, 6, 7, 11
FGG TSP formulation, 179
first-rna.py, 111
fixed costs, 371
flipping
in one column, 266
values, 265
flow, 212
flow of integers, 319
flux, 212
flux balance analysis, 213
fMRI, 23
food webs, 16
forward convergent phylogeny, 91
founder, 336
cell, 60
selection, 341
fourth-rna.py, 117
free energy, 125
function assignment, 250
problem, 250
gender, 380, 381
gene
expression, 20
data, 362
influence, 311
interaction graph, 21
generic
model for haplotyping, 351
genome-wide association study (GWAS), 253
genotype, 332, 349
data, 349
definition, 350
permitted, 333
GG formulation, 169
GG TSP formulation, 171, 174
efficiency, 176
global min-cut, 253
globin family, 191
GLPK, 10
graph
bipartite, 138, 363
connected, 247
density, 25, 78
disconnected, 247
gene interaction, 21
product, 83
sparse, 257
undirected, 15
graphgen.py, 38

guaranteed bounds, 44
guilt by association, 249, 253
Gurobi, xii, xiv, 10
V*, 44
BestBd, 44
bounds, 44
command line execution, 34
GAP, 46
incumbent solution, 44
lazy constraint, 184
opt, 44
parameters, 46
heuristics, 47
MIPFocus, 47
migapp, 46
migappers, 46
time limit, 48
progress reporting, 44
resultfile, 34
GWAS, 253
half-sibling, 380
Hamming distance, 192
handshake lemma, 238
haplotype, 332, 343, 378
assembly, 343, 345
inference problem
in individuals, 343
in populations, 350
inference problem (HI), 343
maternal, 378
paternal, 378
haplotyping, 343
gene model, 351
maximum parsimony, 353
MP-PPH, 353
parsimony with PPH, 353
perfect phylogeny, 352
population variant, 348
pure parsimony, 351, 353, 355, 390
head spinning, 219
heterozygous site, 350
HI
problem, 350
solution, 350
hierarchical clustering, 357
high density near clique, 76
high-density subgraph, 15, 25, 77, 256, 296
HIV, 366
homolog, 332
homology modeling, 128
homozygous site, 350
horizontal transmission, 281
HP model, 128
embedding, 129
hub, 23
human population expansion, 353
Huntington's disease, 290
hypercube, 90, 293
construction, 90
leaves and paths, 90
idiom, 74
δ-relaxation, 223, 224
absolute value for variables, 278
absolute value in an inequality, 279
idiom, (cont.)
absolute value in the objective, 275
AND for binary variables, 230
at least k inequalities, 229
common form, 229
conditional forced equality, 230
equal number of inequalities, 229
exploiting and extending, 225
forced vs. testing, 227
If-Then for binary variables, 74
If-XOR for binary variables, 154
implied satisfaction for inequalities, 227
linear function with binary variables
greater-than case, 223
less-than case, 221
max of two values, 231
min of two values, 231
minimizing absolute value, 278
NAND for inequalities, 226
NOT-EQUAL, 347
for linear functions, 228
for two integer variables, 227, 228
testing, 227
Only-If for binary variables, 74, 75
only-if linear function with binary variables
greater-than case, 224
less-than case, 225
Only-If XOR for variables, 155
OR for inequalities, 276
the OR idiom for inequalities, 226
nonnegative values, 226
XOR for inequalities, 226
If-Then idiom, 74
IIS in Gurobi, 102
ILP
expressibility, 14
formulation
abstract, 13
congest, 13
solves, 12
IM, 261
IMCR problem, 264
IMM, 261
inconsistent idea, 387
impossible idea, 387
in vitro, 360
in vivo, 360
incumbent value, 44
induced subgraph, 25, 78
inverse clique problem, 72
inheritance
probabilities, 333
inhibition, 212, 219
integer linear function, 11
inequality, 11
interaction network, 18
inverse genetics, 341
inverse near-clique problem, 72
inversesgraph.py, 72
iPoint, 288
irreducible inconsistent subsystems (IIS), 103
K cut, 250, 361
k partition, 256
problem, 253
carma, 385
kcut.py, 251
knockout, 288
largest dense.pl, 71
largest high-density subgraph, 256
lea, 148
LCS for multiple strings, 313
LCM, Multi.py, 316
least common ancestor, 148
leukemia, 24
lice, 153
linear constraints, 5
equality, 5
function, 5, 339
coefficients, 5
inequality, 5
objective function, 5
linear programming
concrete, 7
formulation, 5
congest, 5
model, 5
relaxation, 12
what if?, 9
linear programming (LP), 3
linear vs. nonlinear, 385
linearity of expectation, 239
local-global hybrid, 288
logarithm, 339
Louvain, 236
LP algorithms, 10
LP solvers, 7, 10
lp.py, 102
LSD, 83
M1graphTSPfile.pl, 168
MAF, 282
maps
inconsistent, 313
marker ordering, 162, 177
column-permutation problem, 164
marker pipeline.pl, 168
markerTSP.py, 168
Markov Chain Monte Carlo, 391
mass-flow analysis, 213
matched pair, 105
math, xvii
mathematical model, xiv
matrix M, 49
max likelihood, 331
maximum clique, 55
in RNA folding, 112
maximum cut, 253, 255
maximum parsimony, 90
a smaller ILP formulation, 98
for CNV reconstruction, 292
haplotyping, 353
improving the practicality, 98
practicality of the ILP solution, 96
problem, 287
ILP solution, 95
speeding up the solution, 101
maximum-agreement forest, 282
maximum-clique problem, 30
MCR problem, 55, 260
node weighted, 62
Mendelian genetics, 379
metabolic engineering, 205
metabolic network, 205
metabolite, 19, 213
metastasis, 60
Mighty Mouse, 299
MILP, 11
min-max optimization, 281
minimize a maximum, 194
model
biological, xiv
LP, 5
mathematical, xiv
modeling tricks
see idiom, 74
modular2.py, 244
modularity, 235, 366
aberration, 242
assigning nodes, 245
counting classes, 245
module, 82, 236
most strings with a few bad columns, 189
motif, 81, 82
search, 81, 83
speedup, 87
search problem, 83
mountain lions, 17
MP-PPH problem, 353
MTZ TSP formulation
efficiency, 178
multi-set cover, 311
multiple sclerosis, 24
Mus Musculus, 299
mustard, 327
mutation, 53
and cancer, 53
convergent, 89
forward, 89
parallel, 89
Nclique.py, 71
near character clique, 72
tenear clique, 65, 67, 75, 88
alternative definition, 70, 71, 76
high density, 76
inverse problem, 72
largest, dense, 76
nested pairing, 106
network
Boolean, 205
Boolean metabolic, 206
brain connections, 23
correlation, 23
definition, 16
flow, 509
gene influence, 20
genomic, 19
interaction, 18
mass flow, 213
metabolic, 18, 205
NOTCH PPI, 21
PPI, 21
protein-protein, 21
quantitative flow model, 213
network analysis
fantasy, 216
network flow inequalities, 309
Newick format, 153
specialized for tanglegrams, 153
NO-OP, 319
node cut, 25
node degree, 23
non-crossing
alignment, 374
NOTCH
gene, 22
pathway, 21
PPI network, 21
NP hard, 14
NP hardness, 12, 36, 386
oak tree, 381
objective function, 7
objective value, 7
oncogene, 290
Only-If idiom, 74
optimal
solution, 7
OR, 206
oscillation with time, 220
overlap scores, 160
pair
complementary, 105
pairing, 105
nested, 106
non-crossing, 106
paleo-genetics, 164
pancreatic cancer, 206
metastases, 62
parasite, 153
parasite Ω, 143
parent pair, 379
parsimony, 379
partition, 236, 250, 361
passenger gene, 309
pedigree, 331
designing, 341
reconstruction, 336
perfect character, 50
perfect phylogeny, 62, 260, 267, 268
and cancer, 53
dysbindin, 51
haplotype (PPH) problem, 351
in cancer, 54
model, 51
theorem, 52
tree, 51
Perl, 80
Perl program
bbminhap.pl, 356
bbmisstest.pl, 264
breversals.pl, 327
cleanres.pl, 168
denseILP.pl, 80
densesub.pl, 80
HPb.pl, 133
HPb1.pl, 136
Perl program (cont.)
HPb1mid.pl, 135
largest-dense.pl, 71
M1graphTSPfile.pl, 168
marker-pipeline.pl, 168
M1graphTSPfile.pl, 175
persistent.pl, 271
randomrna.pl, 111
sperfreduce.pl, 271
sperse-persistent.pl, 271
tangILP.pl, 153
persistent phylogeny, 268
persistent.pl, 271
Peter Pan, 378
Phase, 390
phase, 350
phasing, see haplotyping, 350
phylogenetic trees, 50
phylogeny
artifact problem, 62
convergent, 89
forward convergent, 91
perfect, 51
planar drawing, 358
plurality sequence, 191
pmarkersolcheck.pl, 168
population structure, 254
PPH
model, 351
problem, 351
PPH network, 21, 22, 249, 289, 295
yeast, 87
product graph, 83
MG, 83
proline, 123
protein, 122
complex, 27, 249, 296, 299
contact map, 572
design, 126
domains, 137
folding, 128
globular, 129
HP model, 128
module, 27
pathways, 369
stability, 125
structure, 122, 372
threading, 376
protein-protein
interaction
network, 22
protein-protein interaction, 137
protein problem, 130
empirical results, 134
speedups, 135
pure parsimony, 390
pure-parsimony
model, 351
Python, 38
Python program
4part-random.py, 253
fourth-rna.py, 117
CLgraphfile.py, 42
cleanrex.py, 117
CLgraphfile.py, 38
CLgraphgen.py, 38
first-rna.py, 111
graphgen.py, 38
HPb-3D.py, 136
inversecliques.py, 72
kcutf.py, 251
LCS_Mult.py, 316
lp.py, 102
modular2.py, 244
N clique.py, 71
randomgraph.py, 244
tsp.py, 184
tspjulia.py, 184
Q(C), 236
Q(G), 236
randomgraph.py, 244
randomrna.pl, 111
rate limits, 215
reaction, 206
recombination, 332
reference genome, 344
relaxation/separation method, 184
representative sequence, 191
reversal-phylogeny problem, 328
reversals, 316
RNA
base stacking, 114
binding strength, 113
cloverleaf, 115
crossing matched pairs, 120
dynamic programming, 107
fold stability, 106
folding, 105, 375
helix, 115
loop, 113
matched pair, 105
non-complementary matching, 114
pairing, 105
pseudo-knot, 120
secondary structure, 105
stability, 115
stacked quartets, 114
threading, 377
tRNA, 115
weighting stacked quartets, 117
rooted subtree-prune-and-regraft (rSPR), 281
rotamer, 124
prediction, 125
row, see also taxon, 49
rSPR, 281
ILP formulation, 285
s.t, cut, 247
sanity check, 240
SAT-solver, 388
satisfiability, 388
satisfied inequalities, 67
Schizophrenia, 24
semantics, 389
separation, 184
sequence
analysis, 186
two string, 187
Index

assembly

- added constraints, 185
- via TSP, 161
- plurality, 191
- representative, 191
- similarity, 186
- sequencing read, 344
- set cover, 140, 367, 379
- shared contact, 374
- shotgun sequencing, 160
- sibling reconstruction problem, 378
- siblings, 377
- full, 377
- side effect, 75, 222
- signaling pathway, 20, 169
- via TSP, 169
- signed sorting-by-reversals, 327
- simple RNA-folding problem, 107
- simplifications, 111
- simplex algorithm, 10
- single nucleotide polymorphism, 254, 345
- single-cell
- sampling, 291
- sequencing, 260
- site
- heterozygous, 350
- homozygous, 350
- site removal
- string, 188
- small is beautiful, 100
- snake oil, 174
- SNP, 254, 345, 349
- causal, 255
- social interaction network, 18
- solution
- feasible, 6, 7, 11
- optimal, 7
- unbounded, 8
- sorting-by-reversals, 317
- abstract ILP formulation, 324
- signed, 327
- sparse graph, 257
- species-protection problem, 3
- sperfreduce.pl, 271
- spersistent.pl, 271
- sprepare-persistent.pl, 271
- stability, 106
- state, 49
- steady-state, 219
- Steiner node, 287
- Steiner tree, 287, 308
- strength, 184
- string
- site removal problem, 187
- strings
- central, 192
- structural genome variants, 29
- stub, 240
- sub-clone, 60
- subgraph, 15
- common induced, 86
- high-density, 249, 256
- induced, 25, 78, 86
- largest high density, 78
- subsequence, 313
- subset
- strict, 25
- symbol, 25
- subtour, 171
- subtree elimination, 181
- subtree exchange
- minimizing, 273
- suffix-prefix overlap, 160
- super families, 121
- tangILP.pl, 153
- tanglegram, 142, 359
- backstory, 143
- bi-objective, 273
- crosses, 146
- formal definition, 146
- interleaf distance, 274
- min-max distance, 281
- ordering Lemma, 147
- problem, 147
- ILP formulation, 148
- logic, 147
- subtree exchange, 146
- tasks A through D, xiii
- taxon, 49
- threatened species, 3, 389
- time
- synchronized, 220
- Topo-free, 299
- abstract ILP formulation, 306
- query problem, 300
- toxicity, 360
- TP53, 290
- trait, 49
- translocations, 316
- traveling salesman
- path problem, 329
- traveling salesman problem, 156
- compact formulation, 169
- converting path to tour, 157
- GG TSP formulation, 169
- MTZ formulation, 177
- path, 157
- strong vs. weak formulations, 179
- subtree elimination, 181
- tour, 157
- TSPLIB, 176
- tree
- binary, 357
- perfect phylogeny, 51
- phylogenetic, 50
- tsp.py, 184
- tspjulia.py, 184
- tumor
- colon, 366
- oncogene, 290
- suppressor gene, 290
- turnips, 318
- twin cells, 269
- unbounded solution, 8
- union
- symbol, 41
- variable, 5
- binary, 11
- bounded, 225
variable (cont.)
  indicator, 221, 236
  integer, 11
  variants, 49

wasting disease, 378
wildlife corridor, 309

worst case, 10
Xenopus, 220
yeast, 143
yeast PPI network, 38, 70, 81, 87
Zachary Karate Club, 241