Genome Properties Documentation

Release 0.1b

Lorna Richardson, Neil Rawlings, Alex Mitchell, Gustavo Salazar C

Sep 19, 2018

Contents

1 About Genome Properties **3**
- 1.1 How to access Genome Properties data . . . 3
- 1.2 Background . . . 4

2 PATHWAYS **5**

3 METAPATH **13**

4 SYSTEM **15**

5 GUILD **19**

6 CATEGORY **21**

7 Genome Property Types **23**

8 Flatfile Format **25**
- 8.1 DESC file . . . 25
- 8.2 FASTA file . . . 28
- 8.3 Status file . . . 29

9 Calculating Genome Properties **31**

10 Contributing to Genome Properties **33**

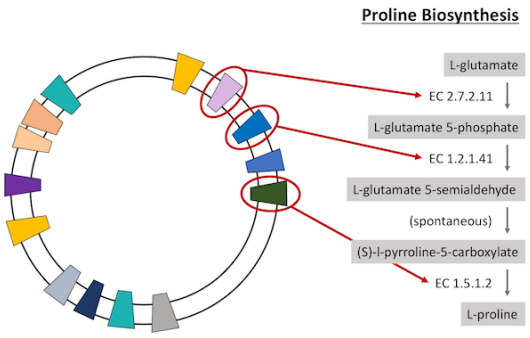
11 Funding **35**

Contents:

CHAPTER 1

About Genome Properties

Genome properties (GP) is an annotation system whereby functional attributes can be assigned to a genome, based on the presence of a defined set of protein family markers within that genome. For example, a species can be proposed to synthesise proline if it can be shown that the genome for that species encodes all the necessary proteins required to carry out the various biochemical steps in the proline biosynthesis pathway.



While it is possible to infer this kind of information through analysis of genomic sequence, using protein family models, like those utilised by InterPro, reduces the number of calculations while at the same time increasing the sensitivity of the query. Integrating the GP annotations into InterPro makes the process of calculating a genome property for any given genome/proteome faster and more accurate than by sequence comparison. The process is further streamlined by the fact that all UniProt sequences, already come with InterPro matches calculated. Each GP is defined as a numbered series of biochemical steps, which in turn have some form of evidence determining the presence of the protein required for that step (usually a profile hidden Markov model, or HMM). The property may include steps which are not strictly necessary, but often take place within the process being described. In the calculation of a GP, the evidence for each *required* step is tested against the genome/proteome under query. Where all steps can be encoded by the query genome, the output is a YES. Where some (above a defined threshold level) steps can be shown to be encoded, the output is a PARTIAL. Where no steps can be shown to be encoded (or fewer than the threshold level), the output is a NO.

For more information about calculating Genome Properties, see here [link to Calculation]

How to access Genome Properties data

Browse by Genome Property: You can browse to your GP of interest using the hierarchy browser. GPs are arranged into categories and subcategories for easy naviagtion. Each GP contains a descriptive abstract describing the property,

as well as the individual steps comprising it. For more infomration on the content of the GP files, see here (link to flatfile description).

You can also navigate to your GP of interest using the lists of various property types (PATHWAY, SYTEM, GUILD, etc) provided under the Browse tab.

Browse by Proteome/Genome: You can browse GPs by species/genome/proteome using our customisable viewer. All GPs are calculated against a reference set of proteomes as provided by UniProt. The output of each GP calculation (yes, no or partial) is reported for each species, in the form of a colour coded matrix. This viewer is easily customisable by the user to allow only specified species/proteoms of interest, as well as GPs of interest, to be included and compared. This viewer allows you to quickly and easily compare the overall "fingerprint" of GP content for a set of species, or indeed the species distribution of a set of GPs.

Upload your data: Further to the features of our interactive viewer already described, you are also able to upload your own proteome data and compare this against the reference set of proteomes available. By analysing your proteome of interest using InterProScan, an output file of InterPro matches is produced. This file (tsv version) can be uploaded to the viewer page, allowing you to view the GP results for your proteome of interest in the colour-coded matrix viewer alongside your chosen comparison set of proteomes/species.

Background

Genome Properties were developed at J. Craig Venter Institute within the TIGRFAM group [PMID:15347579, PMID:23197656] as a way to improve the functional annotation of proteins, as well as providing a resource to assist in comparative genomics. They were based predominantly on TIGRFAM HMM models for step evidence, supplemented with some Pfam models.

One of the main benefits of integrating GP into InterPro is that through the InterPro system of pulling together protein signatures from a range of member databases, there is now a very large potential pool of protein models available to use as evidence for steps, including multiple levels of specificity. In general, we use specific family models in the calculation of GP steps.

Below is a breakdown of the contents of the current release.

PATHWAY 664 COMPLEX 132 METAPATH 60 SYSTEM 327 GUILD 16 CATEGORY 87

CHAPTER 2

PATHWAYS

GenProp0001 Chorismate biosynthesis via shikimate GenProp0023 Glyoxylate shunt GenProp0030 Glutathione biosynthesis GenProp0036 Biotin biosynthesis GenProp0037 Tryptophan biosynthesis from ribose-5-phosphate GenProp0038 Tetrahydrofolate biosynthesis from GTP and PABA GenProp0047 IPP biosynthesis via mevalonate GenProp0048 IPP biosynthesis via deoxyxylulose GenProp0057 NAD(P) biosynthesis from L-aspartate and DHAP GenProp0058 Menaquinone biosynthesis via SEPHCHC GenProp0109 Histidine biosynthesis from ribose-5-phosphate GenProp0110 Purine (inosine-5'-phosphate) biosynthesis from ribose-5-phosphate GenProp0111 Proline biosynthesis from glutamate GenProp0117 Arginine biosynthesis from ornithine, carbamoyl-p and aspartate GenProp0118 Ornithine biosynthesis from glutamate, acetylated branch GenProp0120 Pentose phosphate cycle GenProp0124 Pantothenate biosynthesis from aspartate and 2-oxoisovalerate GenProp0136 Ubiquinone biosynthesis from chorismate, aerobic GenProp0139 Degradation of tyrosine via homogentisate GenProp0141 Galactose-6-phosphate degradation via tagatose-6-phosphate GenProp0143 Leloir pathway (galactose/glucose interconversion) GenProp0146 Bacteriochlorophyll biosynthesis from chlorophyllide a GenProp0147 Glycine betaine biosynthesis from choline GenProp0150 Chlorophyll biosynthesis from chlorophyllide a GenProp0155 Dissimilatory sulfate reduction GenProp0159 Threonine biosynthesis from aspartate semialdehyde GenProp0160 Aspartate semialdehyde biosynthesis from aspartate GenProp0162 Isoleucine biosynthesis from threonine and pyruvate GenProp0163 Valine biosynthesis from pyruvate GenProp0164 Leucine biosynthesis from pyruvate and acetyl-CoA GenProp0171 Coenzyme A biosynthesis from pantothenate GenProp0187 Pyrimidine (uridine-5'-phosphate) de novo biosynthesis GenProp0188 GatABC aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase complex GenProp0189 tRNA-Gln direct aminoacylation GenProp0193 Lysine biosynthesis via alpha-aminoadipate (AAA pathway) GenProp0203 LPS inner core biosynthesis, ketodeoxyoctonate and heptose type GenProp0217 5-aminolevulinate biosynthesis (glutamate pathway) GenProp0218 Cysteine biosynthesis from serine GenProp0220 Uroporhyrinogen III biosynthesis from 5-aminolevulinate GenProp0221 Protoporphyrin IX biosynthesis from uroporphyrinogen III GenProp0222 Protoheme from protoporphyrin IX GenProp0223 5-aminolevulinate biosynthesis from succinyl-CoA and glycine GenProp0231 4-hydroxyphenylacetate degradation GenProp0233 GABA utilization GenProp0238 2-aminoethylphosphonate catabolism to acetaldehyde GenProp0240 Propionyl-CoA catabolism via methylcitric acid GenProp0250 Hydroxyethylthiazole (HET) biosynthesis from 1-deoxy-D-xylulose-5-phosphate, Cysteine, and either Tyr or Gly GenProp0253 Hydroxymethylpyrimidine pyrophosphate (HMP-P) biosynthesis from aminoimidazole ribotide GenProp0254 Thiamine pyrophosphate (TPP) biosynthesis from HMP-PP and HET-P GenProp0259 tRNA-Asn direct aminoacylation GenProp0264 Glucosylglycerol biosynthesis GenProp0265 Trehalose biosynthesis, OtsAB pathway GenProp0266 Trehalose biosynthesis, TreYZ pathway GenProp0268 Ectoine biosynthesis GenProp0269 Coenzyme B12 biosynthesis from cob(II)yrinate diamide GenProp0272 Pyruvate fermentation to acetoin GenProp0273 Protocatechuate degradation to beta-ketoadipate GenProp0275 Cobyrinic acid diamide biosynthesis, anaerobic pathway GenProp0281 Mannosylglycerate biosynthesis GenProp0283 Beta-ketoadipate degradation to succinyl-CoA and acetyl-CoA GenProp0304 Cysteine biosynthesis, tRNA-dependent GenProp0309 Arginine catabolism to glutamate via arginine succinyltransferase (AST) GenProp0466 Molybdopterin biosynthesis GenProp0468 Entner-Doudoroff pathway GenProp0473 Arsenite oxidase GenProp0478 Acetyl-CoA – acetate interconversion via acetyl phosphate GenProp0480 Acetyl-CoA biosynthesis from acetate, direct GenProp0481 Urea carboxylase/allophanate hydrolase pathway GenProp0562 Formaldehyde detoxification, glutathione-dependent GenProp0624 Sulfite reductase complex, Salmonella/Clostridium type GenProp0639 Arginine degradation via citrulline,

ATP-generating GenProp0641 Spermidine biosynthesis from putrescine and S-adenosylmethionine GenProp0642 Putrescine biosynthesis from arginine utilizing agmatinase GenProp0653 Ectoine catabolism GenProp0659 Tryptophan degradation to anthranilate GenProp0671 H4MPT-linked C1 transfer pathway GenProp0673 Ribulose monophosphate pathway GenProp0677 Queuosine (Q-tRNA) biosynthesis from preQ0 GenProp0681 Fatty acid biosynthesis from acetyl-CoA GenProp0687 Allantoin catabolism to oxamate and carbamoyl-phosphate GenProp0688 Urate catabolism to allantoin GenProp0689 Glyoxalate conversion to phosphoglycerate GenProp0691 Glycolysis GenProp0692 Selenouridine-containing tRNA GenProp0697 Guanine monophosphate (GMP) biosynthesis from inosine monophosphate (IMP) GenProp0704 tRNA U34 carboxymethylaminomethyl modification GenProp0706 Cyclohexa-1,5-diene-1-carbonyl-CoA to 3-hydroxypimelyl-CoA, dch/had/oah Pathway GenProp0708 Catechol meta-cleavage pathway GenProp0709 Benzoate catabolism BoxABC pathway GenProp0711 Catechol ortho-cleavage upper pathway GenProp0713 2-aminoethylphosphonate catabolism via phosphonoacetate GenProp0718 Oxalate degradation GenProp0724 Phosphonoacetaldehyde biosynthesis from phosphoenolpyruvate GenProp0728 AMP metabolism using type III Rubisco GenProp0730 Methylthioadenosine to methylthioribose-1-phosphate via MTR GenProp0732 Methionine salvage enolase-phosphate pair MtnW/MtnX GenProp0736 Proposed phosphonate catabolism pathway HpnWXZ GenProp0738 Agmatine to putrescine via N-carbamoylputrescine GenProp0744 Glyoxalase pathway, glutathione dependent GenProp0747 Adenosine monophosphate (AMP) biosynthesis from inosine monophosphate (IMP) GenProp0750 UDP-N-acetylglucosamine biosynthesis from fructose-6-phosphate GenProp0753 Mycothiol biosynthesis GenProp0754 Acetate production from acetylphosphate GenProp0759 Para-aminobenzoic acid (PABA) biosynthesis from chorismate GenProp0766 Anhydro-N-acetylmuramic acid to N-acetylglucosamine-phosphate GenProp0786 Lysine biosynthesis via diaminopimelate (DAP), succinylated branch GenProp0787 Lysine biosynthesis via diaminopimelate (DAP), acetylated branch GenProp0792 7,8-didemethyl-8-hydroxy-5-deazariboflavin (FO) biosynthesis GenProp0797 Selenocysteinyl-tRNA, PSTK/SepSecS pathway GenProp0813 Pyrimidine utilization GenProp0822 Poly(gamma-glutamic acid) biosynthesis GenProp0829 Menaquinone biosynthesis via futalosine GenProp0837 1-acylglycerol-3-phosphate biosynthesis, PlsX/Y pathway GenProp0862 Pyridoxal phosphate biosynthesis, PdxA/PdxJ pathway GenProp0878 N(epsilon)-acetyl-beta-lysine biosynthesis GenProp0893 Sulfopyruvate biosynthesis via L-cysteate GenProp0908 2,3-diaminopropionic acid biosynthesis GenProp0909 Capreomycidine biosynthesis GenProp0927 Bacillithiol biosynthesis GenProp0942 Alkane biosynthesis, cyanobacterial GenProp0950 dTDP-4-dehydro-6-deoxy-alpha-D-glucose biosynthesis from glucose-1-phosphate GenProp0991 ATP-grasp maturase system, microviridin/marinostatin class GenProp1017 GDP-D-mannose biosynthesis from b-D-fructose-6-phosphate GenProp1058 2,3-dihydroxybenzoate biosynthesis GenProp1064 Ovothiol biosynthesis GenProp1070 Beta-glutamate biosynthesis GenProp1084 Enduracididine biosynthesis GenProp1098 3-methylarginine biosynthesis GenProp1215 Siroheme biosynthesis GenProp1216 Tetrahydrofolate salvage from 5,10-methenyltetrahydrofolate GenProp1217 Superpathway of steroid hormone biosynthesis GenProp1218 Hydroxymethylpyrimidine salvage GenProp1219 Thiamine salvage I GenProp1220 Fatty acid biosynthesis initiation II GenProp1221 Pyrimidine deoxyribonucleosides salvage GenProp1222 UDP-N-acetyl-D-glucosamine biosynthesis II GenProp1223 L-threonine degradation V GenProp1224 Lipoate salvage I GenProp1225 D-galactarate degradation I GenProp1226 Pyruvate fermentation to ethanol II GenProp1227 Guanosine deoxyribonucleotides de novo biosynthesis I GenProp1228 L-idonate degradation GenProp1229 D-myo-inositol (1,3,4)-trisphosphate biosynthesis GenProp1230 Aerobic respiration III (alternative oxidase pathway) GenProp1231 2-Hydroxypenta-2,4-dienoate degradation GenProp1233 L-glutamate degradation II GenProp1234 Superpathway of L-tyrosine biosynthesis GenProp1235 Adenine and adenosine salvage III GenProp1236 Retinoate biosynthesis II GenProp1237 Lipoate biosynthesis and incorporation I GenProp1238 L-cysteine biosynthesis IV (fungi) GenProp1239 ADP-L-glycero-beta-D-manno-heptose biosynthesis GenProp1240 Ketolysis GenProp1241 Wybutosine biosynthesis GenProp1242 Muropeptide degradation GenProp1243 Salicylate biosynthesis I GenProp1244 Myo-inositol biosynthesis GenProp1245 L-tyrosine degradation III GenProp1246 Zymosterol biosynthesis GenProp1247 Glycogen biosynthesis I (from ADP-D-Glucose) GenProp1248 Fructose 2,6-bisphosphate biosynthesis GenProp1249 1D-myo-inositol hexakisphosphate biosynthesis I (from Ins(1,4,5)P3) GenProp1250 Aspartate superpathway GenProp1251 L-tyrosine biosynthesis I GenProp1252 Cardiolipin biosynthesis III GenProp1253 Superpathway of arginine and polyamine biosynthesis GenProp1254 NADH to hydrogen peroxide electron transfer GenProp1255 Adenosine nucleotides degradation II GenProp1256 Pyruvate to cytochrome bo oxidase electron transfer GenProp1257 Allantoin degradation to glyoxylate I GenProp1258 Superpathway of S-adenosyl-L-methionine biosynthesis GenProp1259 Glycogen degradation II GenProp1260 UDP-alpha-D-xylose biosynthesis GenProp1261 Cytokinins 7-N-glucoside biosynthesis GenProp1262 UTP and CTP de novo biosynthesis GenProp1263 Methylglyoxal degradation III GenProp1264 Superpathway of pyrimidine deoxyribonucleoside salvage GenProp1265 Glyoxylate cycle GenProp1266 Superpathway of thiamine diphosphate biosynthesis II GenProp1267 Mixed acid fer-

mentation GenProp1268 Adenine salvage GenProp1269 Pyruvate to cytochrome bd oxidase electron transfer GenProp1270 Enterobacterial common antigen biosynthesis GenProp1271 Myo-, chiro- and scyllo-inositol degradation GenProp1272 Spermine and spermidine degradation III GenProp1273 Thymine degradation GenProp1274 D-serine metabolism GenProp1275 Glycine cleavage GenProp1276 Putrescine degradation V GenProp1277 Sphingosine and sphingosine-1-phosphate metabolism GenProp1278 Purine ribonucleosides degradation GenProp1279 Superpathway of L-arginine, putrescine, and 4-aminobutanoate degradation GenProp1280 L-arginine degradation II (AST pathway) GenProp1281 Hydrogen sulfide biosynthesis I GenProp1282 L-arginine degradation III (arginine decarboxylase/agmatinase pathway) GenProp1283 Sulfate reduction I (assimilatory) GenProp1284 5-Aminoimidazole ribonucleotide biosynthesis I GenProp1286 Myo-inositol degradation I GenProp1287 3-Phenylpropanoate and 3-(3-hydroxyphenyl)propanoate degradation to 2-hydroxypentadienoate GenProp1288 GABA shunt GenProp1289 Thiamine salvage II GenProp1290 Lipid IVA biosynthesis GenProp1291 Superpathway of chorismate metabolism GenProp1292 L-tryptophan degradation VIII (to tryptophol) GenProp1293 Chitobiose degradation GenProp1294 Pentose phosphate pathway (non-oxidative branch) GenProp1295 Methylerythritol phosphate pathway I GenProp1296 NAD biosynthesis III GenProp1297 Superpathway of glycol metabolism and degradation GenProp1298 Lipoate biosynthesis and incorporation II GenProp1299 2-O-alpha-mannosyl-D-glycerate degradation GenProp1300 Urea cycle GenProp1301 Superpathway of sulfate assimilation and cysteine biosynthesis GenProp1302 Galactolipid biosynthesis I GenProp1303 N-acetylglucosamine degradation I GenProp1304 Globo-series glycosphingolipids biosynthesis GenProp1305 NAD phosphorylation and transhydrogenation GenProp1306 Glycolysis I (from glucose 6-phosphate) GenProp1307 Superpathway of pyrimidine nucleobases salvage GenProp1308 Fatty acid beta-oxidation VI (peroxisome) GenProp1309 Superpathway of L-phenylalanine biosynthesis GenProp1310 D-galactose degradation I (Leloir pathway) GenProp1311 Protein NEDDylation GenProp1312 Aminopropanol phosphate biosynthesis I GenProp1313 tRNA-uridine 2-thiolation (mammalian mitochondria) GenProp1315 Fatty acid biosynthesis initiation III GenProp1316 Phosphatidate biosynthesis (yeast) GenProp1317 Protein Pupylation and dePupylation GenProp1318 Superpathway of pyrimidine ribonucleotides de novo biosynthesis GenProp1320 Nitric oxide biosynthesis II (mammals) GenProp1321 Retinoate biosynthesis I GenProp1322 Superpathway of geranylgeranyl diphosphate biosynthesis II (via MEP) GenProp1323 Purine deoxyribonucleosides degradation I GenProp1324 Glycerol degradation V GenProp1325 Superpathway of (Kdo)2-lipid A biosynthesis GenProp1326 Allantoin degradation to ureidoglycolate II (ammonia producing) GenProp1327 Lipoprotein posttranslational modification GenProp1328 L-isoleucine degradation II GenProp1329 Nitrate reduction III (dissimilatory) GenProp1330 UDP-beta-L-arabinose biosynthesis II (from beta-L-arabinose) GenProp1331 Pyocyanin biosynthesis GenProp1332 Superpathway of tetrahydrofolate biosynthesis GenProp1333 Superpathway of L-alanine biosynthesis GenProp1334 L-leucine degradation III GenProp1335 Ceramide degradation GenProp1336 2,3-Dihydroxybenzoate biosynthesis GenProp1337 Vitamin E biosynthesis (tocopherols) GenProp1338 Estradiol biosynthesis I (via estrone) GenProp1339 Alpha-linolenate biosynthesis I (plants and red algae) GenProp1340 5-(Carboxymethoxy)uridine biosynthesis GenProp1341 NADH to trimethylamine N-oxide electron transfer GenProp1342 Pyruvate fermentation to isobutanol (engineered) GenProp1343 Guanosine deoxyribonucleotides de novo biosynthesis II GenProp1344 Gluconeogenesis I GenProp1345 Acetate formation from acetyl-CoA I GenProp1346 Glutathione degradation (DUG pathway - yeast) GenProp1347 UDP-alpha-D-glucuronate biosynthesis (from myo-inositol) GenProp1348 2-Oxoglutarate decarboxylation to succinyl-CoA GenProp1349 Superpathway of adenosylcobalamin salvage from cobinamide I GenProp1351 Phosphatidylcholine biosynthesis I GenProp1352 Phosphatidylcholine resynthesis via glycerophosphocholine GenProp1353 Superpathway of photosynthetic hydrogen production GenProp1354 mRNA capping I GenProp1355 Chlorophyll a biosynthesis II GenProp1356 N10-formyl-tetrahydrofolate biosynthesis GenProp1357 L-citrulline degradation GenProp1358 Superpathway of L-threonine biosynthesis GenProp1359 Glutathione biosynthesis GenProp1360 tRNA processing GenProp1361 Pyrimidine nucleobases salvage II GenProp1362 Putrescine degradation I GenProp1363 Sphingolipid recycling and degradation (yeast) GenProp1364 Hentriaconta-3,6,9,12,15,19,22,25,28-nonaene biosynthesis GenProp1365 L-asparagine degradation III (mammalian) GenProp1366 2-Methyladeninyl adenosylcobamide biosynthesis from adenosylcobinamide-GDP GenProp1367 D-lactate to cytochrome bo oxidase electron transfer GenProp1368 L-arabinose degradation I GenProp1369 Superpathway of pyrimidine ribonucleosides salvage GenProp1370 Ethylene glycol degradation GenProp1371 Uracil degradation I (reductive) GenProp1372 Adenosine ribonucleotides de novo biosynthesis GenProp1373 NADH to cytochrome bd oxidase electron transfer II GenProp1374 Glycolate and glyoxylate degradation I GenProp1375 L-tyrosine degradation I GenProp1376 Very long chain fatty acid biosynthesis II GenProp1377 Biotin biosynthesis from 8-amino-7-oxononanoate I GenProp1378 Adeninyl adenosylcobamide biosynthesis from adenosylcobinamide-GDP GenProp1379 Ethanol degradation IV GenProp1380 Superpathway of pyrimidine ribonucleosides degradation GenProp1381 Methylphosphonate degradation I GenProp1382

Phytyl diphosphate biosynthesis GenProp1383 Glutaminyl-tRNAgln biosynthesis via transamidation GenProp1384 Pyrimidine ribonucleosides salvage I GenProp1385 Progesterone biosynthesis GenProp1386 Superpathway of L-serine and glycine biosynthesis I GenProp1387 Superpathway of beta-D-glucuronosides degradation GenProp1388 Pterostilbene biosynthesis GenProp1389 Superpathway of guanosine nucleotides de novo biosynthesis I GenProp1390 Choline degradation I GenProp1391 Glycerol-3-phosphate to fumarate electron transfer GenProp1392 Sulfate reduction II (assimilatory) GenProp1393 Allantoin degradation to glyoxylate II GenProp1394 Trehalose degradation I (low osmolarity) GenProp1395 Superpathway of D-myo-inositol (1,4,5)-trisphosphate metabolism GenProp1396 UTP and CTP dephosphorylation II GenProp1397 (Kdo)2-lipid A biosynthesis I GenProp1398 Poly(glycerol phosphate) wall teichoic acid biosynthesis GenProp1399 Superpathway of L-cysteine biosynthesis (mammalian) GenProp1400 Queuosine biosynthesis GenProp1401 Proline to cytochrome bo oxidase electron transfer GenProp1402 Phenylethylamine degradation I GenProp1403 dTDP-N-acetylthomosamine biosynthesis GenProp1404 Superpathway of L-aspartate and L-asparagine biosynthesis GenProp1405 Superpathway of branched chain amino acid biosynthesis GenProp1406 Superpathway of purine nucleotides de novo biosynthesis I GenProp1407 Glycolysis II (from fructose 6-phosphate) GenProp1408 2-Oxoisovalerate decarboxylation to isobutanoyl-CoA GenProp1409 Bis(guanylyl molybdenum cofactor) biosynthesis GenProp1410 Pyrimidine deoxyribonucleosides degradation GenProp1411 All-trans-farnesol biosynthesis GenProp1412 Glycogen degradation I GenProp1413 Mitochondrial L-carnitine shuttle GenProp1414 Glycine betaine degradation II (mammalian) GenProp1415 Superpathway of fatty acid biosynthesis initiation (E. coli) GenProp1416 Allantoin degradation to glyoxylate III GenProp1417 Estradiol biosynthesis II GenProp1418 UMP biosynthesis III GenProp1419 Superpathway of L-methionine biosynthesis (transsulfuration) GenProp1420 Gibberellin inactivation II (methylation) GenProp1421 Heme b biosynthesis I (aerobic) GenProp1423 Sphingolipid biosynthesis (yeast) GenProp1424 Two-component alkanesulfonate monooxygenase GenProp1425 Pyruvate fermentation to ethanol I GenProp1426 Aerobic respiration II (cytochrome c) (yeast) GenProp1427 UMP biosynthesis I GenProp1428 L-arginine degradation I (arginase pathway) GenProp1429 GDP-L-fucose biosynthesis I (from GDP-D-mannose) GenProp1430 Adenine and adenosine salvage II GenProp1431 L-arginine degradation IV (arginine decarboxylase/agmatine deiminase pathway) GenProp1432 Mevalonate pathway I GenProp1433 Aminopropylcadaverine biosynthesis GenProp1434 Putrescine degradation II GenProp1436 Superpathway of mycolate biosynthesis GenProp1437 Ethanol degradation III GenProp1438 Pentose phosphate pathway GenProp1439 2-Oxobutanoate degradation I GenProp1441 Acetoacetate degradation (to acetyl CoA) GenProp1442 Protein O-[N-acetyl]-glucosylation GenProp1443 UDP-N-acetyl-D-glucosamine biosynthesis I GenProp1444 Complex N-linked glycan biosynthesis (plants) GenProp1445 L-threonine degradation IV GenProp1446 Adenosine deoxyribonucleotides de novo biosynthesis II GenProp1447 Methanogenesis from methylthiopropanoate GenProp1448 UDP-N-acetylmuramoyl-pentapeptide biosynthesis I (meso-diaminopimelate containing) GenProp1449 D-fructuronate degradation GenProp1450 L-tryptophan biosynthesis GenProp1451 Beta-carotene biosynthesis GenProp1452 Gamma-linolenate biosynthesis II (animals) GenProp1453 Cardiolipin and phosphatidylethanolamine biosynthesis (Xanthomonas) GenProp1454 Adenosylcobalamin biosynthesis from adenosylcobinamide-GDP I GenProp1455 Protein O-mannosylation III (mammals, core M3) GenProp1457 Thiamine diphosphate biosynthesis I (E. coli) GenProp1458 Cinnamate and 3-hydroxycinnamate degradation to 2-hydroxypentadienoate GenProp1459 D-galactose detoxification GenProp1460 D-myo-inositol (3,4,5,6)-tetrakisphosphate biosynthesis GenProp1461 tRNA-uridine 2-thiolation (cytoplasmic) GenProp1462 Yersiniabactin biosynthesis GenProp1463 D-allose degradation GenProp1464 Indole-3-acetate biosynthesis I GenProp1465 Pyrimidine ribonucleosides salvage II GenProp1466 L-ornithine biosynthesis I GenProp1467 L-valine degradation II GenProp1468 UDP-2,3-diacetamido-2,3-dideoxy-alpha-D-mannuronate biosynthesis GenProp1469 Guanosine nucleotides degradation III GenProp1470 Superpathway of heme b biosynthesis from uroporphyrinogen-III GenProp1471 Beta-caryophyllene biosynthesis GenProp1472 L-ornithine biosynthesis II GenProp1473 Cis-alkene biosynthesis GenProp1474 Nitrate reduction VIIIb (dissimilatory) GenProp1475 Superpathway of L-lysine, L-threonine and L-methionine biosynthesis I GenProp1476 Autoinducer AI-2 degradation GenProp1477 Coenzyme B/coenzyme M regeneration III (coenzyme F420-dependent) GenProp1478 Chorismate biosynthesis from 3-dehydroquinate GenProp1479 Polyisoprenoid biosynthesis (E. coli) GenProp1480 Peptidoglycan biosynthesis I (meso-diaminopimelate containing) GenProp1481 Superpathway of L-citrulline metabolism GenProp1482 S-methyl-5'-thioadenosine degradation I GenProp1483 Glycogen biosynthesis II (from UDP-D-Glucose) GenProp1484 Superpathway of guanosine nucleotides de novo biosynthesis II GenProp1485 Pyruvate fermentation to acetoin III GenProp1486 Oleate beta-oxidation GenProp1487 L-lysine degradation XI (mammalian) GenProp1488 Oleate beta-oxidation (thioesterase-dependent, yeast) GenProp1489 Ammonia assimilation cycle I GenProp1490 Indole-3-acetate inactivation VIII GenProp1491 Ac/N-end rule pathway GenProp1492 Sucrose degradation III (sucrose invertase) GenProp1493 Succinate to cytochrome bo oxidase electron transfer GenProp1494 Pyrimidine deoxyri-

bonucleotide phosphorylation GenProp1495 N-end rule pathway I (prokaryotic) GenProp1496 Hydrogen production V GenProp1497 Ethanol degradation I GenProp1498 Superpathway of menaquinol-8 biosynthesis I GenProp1499 L-selenocysteine biosynthesis I (bacteria) GenProp1501 Allantoin degradation to ureidoglycolate I (urea producing) GenProp1502 NAD/NADP-NADH/NADPH cytosolic interconversion (yeast) GenProp1503 L-tryptophan degradation to 2-amino-3-carboxymuconate semialdehyde GenProp1504 Nitrate reduction X (dissimilatory, periplasmic) GenProp1505 Bacillithiol biosynthesis GenProp1506 4-Aminobutanoate degradation I GenProp1507 L-homocysteine biosynthesis GenProp1508 Violaxanthin, antheraxanthin and zeaxanthin interconversion GenProp1509 1D-myo-inositol hexakisphosphate biosynthesis II (mammalian) GenProp1510 Fatty acid beta-oxidation II (peroxisome) GenProp1511 D-myo-inositol-5-phosphate metabolism GenProp1512 Fatty acid biosynthesis initiation I GenProp1513 Superpathway of thiamine diphosphate biosynthesis III (eukaryotes) GenProp1514 D-glucarate degradation I GenProp1515 Succinate to cytochrome bd oxidase electron transfer GenProp1516 Phosphatidylcholine biosynthesis V GenProp1517 Ganglio-series glycosphingolipids biosynthesis GenProp1518 Lacto-series glycosphingolipids biosynthesis GenProp1520 Pentose phosphate pathway (oxidative branch) I GenProp1521 tRNA methylation (yeast) GenProp1522 Chitin degradation I (archaea) GenProp1523 4-Aminobutanoate degradation III GenProp1524 Complex N-linked glycan biosynthesis (vertebrates) GenProp1525 Tetrapyrrole biosynthesis II (from glycine) GenProp1526 Chondroitin sulfate biosynthesis GenProp1527 Myo-inositol degradation II GenProp1528 Guanine and guanosine salvage GenProp1529 Superpathway of fucose and rhamnose degradation GenProp1530 Superpathway of cholesterol biosynthesis GenProp1533 L-isoleucine degradation I GenProp1534 Kanosamine biosynthesis II GenProp1535 Formate to trimethylamine N-oxide electron transfer GenProp1536 Pyruvate decarboxylation to acetyl CoA GenProp1537 NADH to fumarate electron transfer GenProp1538 Superpathway of aromatic amino acid biosynthesis GenProp1539 Neolacto-series glycosphingolipids biosynthesis GenProp1540 Pyrimidine ribonucleosides salvage III GenProp1541 Superpathway of geranylgeranyldiphosphate biosynthesis I (via mevalonate) GenProp1542 Phosphatidylethanolamine biosynthesis I GenProp1543 Pyruvate fermentation to acetate IV GenProp1544 Fatty acid beta-oxidation (peroxisome, yeast) GenProp1545 i antigen and I antigen biosynthesis GenProp1546 Sulfoquinovosyl diacylglycerol biosynthesis GenProp1547 Anthocyanidin modification (Arabidopsis) GenProp1548 D-myo-inositol (1,4,5)-trisphosphate biosynthesis GenProp1549 UDP-N-acetyl-alpha-D-mannosaminouronate biosynthesis GenProp1550 Superpathway of L-tryptophan biosynthesis GenProp1551 L-tryptophan degradation X (mammalian, via tryptamine) GenProp1552 Allopregnanolone biosynthesis GenProp1553 L-homoserine biosynthesis GenProp1554 Nitrate reduction V (assimilatory) GenProp1555 tRNA-uridine 2-thiolation (bacteria) GenProp1556 Glycolate and glyoxylate degradation II GenProp1557 Entner-Doudoroff pathway I GenProp1558 Sterol:steryl ester interconversion (yeast) GenProp1559 2'-Deoxy-alpha-D-ribose 1-phosphate degradation GenProp1560 2-Carboxy-1,4-naphthoquinol biosynthesis GenProp1561 Superpathway of coenzyme A biosynthesis III (mammals) GenProp1562 Fatty acid salvage GenProp1563 Glycerol-3-phosphate to cytochrome bo oxidase electron transfer GenProp1564 Trehalose degradation II (trehalase) GenProp1565 N-acetylglucosamine degradation II GenProp1566 D-galactonate degradation GenProp1567 Glycerol degradation I GenProp1568 Purine deoxyribonucleosides degradation II GenProp1569 Very long chain fatty acid biosynthesis I GenProp1570 N6-L-threonylcarbamoyladenosine37-modified tRNA biosynthesis GenProp1571 Spermine biosynthesis GenProp1572 L-carnitine degradation I GenProp1573 Sulfate activation for sulfonation GenProp1574 Inositol diphosphates biosynthesis GenProp1575 Phospholipid remodeling (phosphatidylethanolamine, yeast) GenProp1576 Pyrimidine ribonucleosides degradation GenProp1577 Biotin biosynthesis I GenProp1578 Methylglyoxal degradation I GenProp1579 4-Aminobutanoate degradation II GenProp1580 Homolactic fermentation GenProp1581 L-homoserine and L-methionine biosynthesis GenProp1582 Hydrogen to trimethylamine N-oxide electron transfer GenProp1583 Nitrate reduction VIII (dissimilatory) GenProp1584 L-aspartate degradation II GenProp1585 Superpathway of ubiquinol-8 biosynthesis (prokaryotic) GenProp1586 Adenosine deoxyribonucleotides de novo biosynthesis GenProp1587 Oleate biosynthesis II (animals and fungi) GenProp1588 Lipoxin biosynthesis GenProp1589 D-arabinose degradation I GenProp1590 4-Amino-2-methyl-5-diphosphomethylpyrimidine biosynthesis GenProp1591 Heme degradation I GenProp1592 Superpathway of purine nucleotides de novo biosynthesis II GenProp1593 Staphyloferrin A biosynthesis GenProp1594 Superpathway of ergosterol biosynthesis I GenProp1595 Heparan sulfate biosynthesis GenProp1596 Superpathway of polyamine biosynthesis I GenProp1597 Pinolenate and coniferonate biosynthesis GenProp1598 Superoxide radicals degradation GenProp1599 Glycolysis III (from glucose) GenProp1600 Abscisic acid degradation by glucosylation GenProp1601 Superpathway of coenzyme A biosynthesis I (bacteria) GenProp1602 PreQ0 biosynthesis GenProp1603 C20 prostanoid biosynthesis GenProp1604 Pyridoxal 5'-phosphate salvage I GenProp1605 D-myo-inositol (1,4,5)-trisphosphate degradation GenProp1606 Octopamine biosynthesis GenProp1607 Purine deoxyribonucleosides salvage GenProp1608 NADH to cytochrome bd oxidase electron transfer I GenProp1609 Ergosterol biosynthesis I GenProp1610 Spermine and spermidine degradation I Gen-

Prop1611 Xanthine and xanthosine salvage GenProp1612 Gluconeogenesis III GenProp1613 Formate assimilation into 5,10-methylenetetrahydrofolate GenProp1614 UMP biosynthesis II GenProp1615 Ammonia assimilation cycle III GenProp1616 Tetrahydrofolate biosynthesis GenProp1617 Pyrimidine deoxyribonucleotides de novo biosynthesis I GenProp1619 L-histidine degradation I GenProp1620 Pyruvate fermentation to acetate VIII GenProp1621 Pyrimidine deoxyribonucleotides de novo biosynthesis II GenProp1622 Mycocyclosin biosynthesis GenProp1623 Anhydromuropeptides recycling I GenProp1624 Fructoselysine and psicoselysine degradation GenProp1625 Lipoate biosynthesis and incorporation III (Bacillus) GenProp1626 L-alanine degradation I GenProp1627 Cardiolipin biosynthesis I GenProp1628 5-Aminoimidazole ribonucleotide biosynthesis II GenProp1629 Propanoyl CoA degradation I GenProp1630 Aminomethylphosphonate degradation GenProp1631 Superpathway of pyrimidine deoxyribonucleosides degradation GenProp1632 Erythro-tetrahydrobiopterin biosynthesis I GenProp1633 Pyridoxal 5'-phosphate biosynthesis I GenProp1634 CMP phosphorylation GenProp1635 Superpathway of pyrimidine deoxyribonucleotides de novo biosynthesis GenProp1636 D-galacturonate degradation I GenProp1637 Aerobic respiration I (cytochrome c) GenProp1638 (E,E)-4,8,12-trimethyltrideca-1,3,7,11-tetraene biosynthesis GenProp1639 Sucrose degradation II (sucrose synthase) GenProp1640 L-serine biosynthesis GenProp1641 NADH to cytochrome bo oxidase electron transfer II GenProp1643 Chorismate biosynthesis I GenProp1644 Mineralocorticoid biosynthesis GenProp1645 Zeaxanthin biosynthesis GenProp1646 L-threonine degradation II GenProp1647 Superpathway of lipopolysaccharide biosynthesis GenProp1648 Colanic acid building blocks biosynthesis GenProp1649 Dehydro-D-arabinono-1,4-lactone biosynthesis GenProp1650 Guanosine ribonucleotides de novo biosynthesis GenProp1651 Lipid A-core biosynthesis GenProp1652 Urate conversion to allantoin I GenProp1653 Leukotriene biosynthesis GenProp1654 Homocysteine and cysteine interconversion GenProp1655 L-methionine degradation III GenProp1656 Xylose degradation I GenProp1657 Ethylene biosynthesis I (plants) GenProp1658 NAD salvage pathway III GenProp1660 L-asparagine biosynthesis III (tRNA-dependent) GenProp1661 D-galactose degradation V (Leloir pathway) GenProp1662 L-lactaldehyde degradation (aerobic) GenProp1663 Heme b biosynthesis IV (Gram-positive bacteria) GenProp1664 Gamma-glutamyl cycle GenProp1665 Sphingolipid biosynthesis (mammals) GenProp1666 Glucocorticoid biosynthesis GenProp1667 L-alanine biosynthesis I GenProp1668 Gadusol biosynthesis GenProp1669 CO2 fixation into oxaloacetate (anaplerotic) GenProp1670 Di-trans,poly-cis-undecaprenyl phosphate biosynthesis GenProp1671 Ceramide de novo biosynthesis GenProp1672 Hydrogen to fumarate electron transfer GenProp1673 L-leucine degradation I GenProp1674 Thiamine formation from pyrithiamine and oxythiamine (yeast) GenProp1675 Superpathway of D-glucarate and D-galactarate degradation GenProp1676 Nitrate reduction IX (dissimilatory) GenProp1677 NAD/NADP-NADH/NADPH mitochondrial interconversion (yeast) GenProp1678 Beta-alanine biosynthesis IV GenProp1679 L-phenylalanine degradation I (aerobic) GenProp1680 Fucose degradation GenProp1681 L-cysteine degradation III GenProp1682 Factor 420 polyglutamylation GenProp1683 Epoxysqualene biosynthesis GenProp1684 Kdo transfer to lipid IVA I GenProp1685 Hydrogen sulfide biosynthesis II (mammalian) GenProp1686 Heme b biosynthesis II (anaerobic) GenProp1687 2-Methylcitrate cycle I GenProp1688 Heme degradation VII GenProp1689 Superpathway of 5-aminoimidazole ribonucleotide biosynthesis GenProp1690 Glycolipid desaturation GenProp1691 Superpathway of glycolysis and the Entner-Doudoroff pathway GenProp1692 D-myo-inositol (1,4,5,6)-tetrakisphosphate biosynthesis GenProp1693 TCA cycle III (animals) GenProp1694 Superpathway of demethylmenaquinol-8 biosynthesis I GenProp1695 Serotonin and melatonin biosynthesis GenProp1696 Curcumin degradation GenProp1697 Thiazole biosynthesis III (eukaryotes) GenProp1698 L-phenylalanine degradation III GenProp1699 Glycosaminoglycan-protein linkage region biosynthesis GenProp1701 Tetrapyrrole biosynthesis I (from glutamate) GenProp1702 Thiamine diphosphate biosynthesis IV (eukaryotes) GenProp1703 Retinol biosynthesis GenProp1704 Trans, trans-farnesyl diphosphate biosynthesis GenProp1705 Sedoheptulose bisphosphate bypass GenProp1706 Protein N-glycosylation processing phase (yeast) GenProp1707 tRNA splicing II GenProp1708 L-phenylalanine biosynthesis I GenProp1709 Trichome monoterpenes biosynthesis GenProp1710 Superpathway of glycolysis, pyruvate dehydrogenase, TCA, and glyoxylate bypass GenProp1711 Molybdenum cofactor biosynthesis GenProp1712 Mucin core 1 and core 2 O-glycosylation GenProp1713 UDP-beta-L-rhamnose biosynthesis GenProp1715 Ethanol degradation II GenProp1716 Superpathay of heme b biosynthesis from glutamate GenProp1717 Fatty acid beta-oxidation I GenProp1718 Phospholipid remodeling (phosphatidylcholine, yeast) GenProp1719 L-methionine salvage from L-homocysteine GenProp1720 Superpathway of b heme biosynthesis from glycine GenProp1721 Glutathione-peroxide redox reactions GenProp1722 Chitin degradation to ethanol GenProp1723 Conversion of succinate to propanoate GenProp1724 Chlorophyll a biosynthesis I GenProp1725 Thiamine diphosphate biosynthesis II (Bacillus) GenProp1726 Starch degradation II GenProp1727 4-Aminobenzoate biosynthesis GenProp1728 Thyroid hormone biosynthesis GenProp1729 Glycerol-3-phosphate to hydrogen peroxide electron transport GenProp1730 Inosine-5'-phosphate biosynthesis I GenProp1731 3-Dehydroquinate biosynthesis I GenProp1732 Pentose phosphate pathway (oxidative branch) II GenProp1733 Pu-

trescine biosynthesis II GenProp1734 Flavin biosynthesis I (bacteria and plants) GenProp1735 Archaeosine biosynthesis I GenProp1736 NADH repair GenProp1737 CMP-N-acetylneuraminate biosynthesis I (eukaryotes) GenProp1739 Ent-kaurene biosynthesis I GenProp1740 Pregnenolone biosynthesis GenProp1741 Adenine and adenosine salvage V GenProp1742 Androgen biosynthesis GenProp1743 UDP-alpha-D-glucose biosynthesis I GenProp1744 Ubiquinol-8 biosynthesis (prokaryotic) GenProp1745 GA12 biosynthesis GenProp1746 Alkylnitronates degradation GenProp1747 Photorespiration GenProp1748 Phosphopantothenate biosynthesis I GenProp1749 Superpathway of acetate utilization and formation GenProp1750 Glycerol-3-phosphate shuttle GenProp1751 NADH to cytochrome bo oxidase electron transfer I GenProp1752 Superpathway of purine deoxyribonucleosides degradation GenProp1753 Inosine 5'-phosphate degradation GenProp1754 Protein ubiquitination GenProp1755 PpGpp biosynthesis GenProp1756 Poly(3-O-beta-D-glucopyranosyl-N-acetylgalactosamine 1-phosphate) wall teichoic acid biosynthesis GenProp1757 Inosine-5'-phosphate biosynthesis II GenProp1758 3-Phosphoinositide biosynthesis GenProp1759 Spermidine hydroxycinnamic acid conjugates biosynthesis GenProp1760 Superpathway of gibberellin GA12 biosynthesis GenProp1761 Linoleate biosynthesis I (plants) GenProp1762 Ethanolamine utilization GenProp1763 Superpathway of carotenoid biosynthesis in plants GenProp1764 Flavin biosynthesis IV (mammalian) GenProp2007 Complement activation, common pathway 1 GenProp2088 Insulin signaling pathway - glycogenesis GenProp2089 Glucagon signaling pathway for glycogenolysis GenProp2090 Insulin signaling pathway - glucose uptake GenProp2097 Insulin release from secretory granules

CHAPTER 3

METAPATH

GenProp0029 Nitrogen fixation GenProp0033 TCA cycle GenProp0046 IPP biosynthesis GenProp0125 Lysine biosynthesis via diaminopimelate (DAP) GenProp0144 Chlorophyllide a biosynthesis from protoporphyrin IX GenProp0183 dTDP-L-rhamnose biosynthesis from dTDP-4-dehydro-L-rhamnose GenProp0199 Lysine biosynthesis GenProp0204 KDO(2)-lipid A (Re LPS) biosynthesis and delivery GenProp0219 Spermidine biosynthesis from arginine (via agmatine) GenProp0241 Phosphonate catabolism GenProp0255 5-aminolevulinate biosynthesis GenProp0261 Thiamine pyrophosphate (TPP) de novo biosynthesis GenProp0287 Ribonucleotide reduction GenProp0305 Cysteine biosynthesis GenProp0479 Acetate – acetyl-CoA interconversions GenProp0489 F-type conjugation system GenProp0611 Aerobic respiration GenProp0612 Aerobic electron transfer (transport) chain, ETC GenProp0615 Cytochrome c based oxygen reduction and quinone re-oxidation GenProp0616 Single complex oxygen reduction and quinone re-oxidation GenProp0643 Putrescine biosynthesis from arginine via ornithine GenProp0644 Putrescine biosynthesis from glutamate via ornithine GenProp0645 Putrescine biosynthesis GenProp0686 Allantoin catabolism to glyoxalate and urea GenProp0698 Xanthine catabolism to urate GenProp0700 Purine catabolism via urate, xanthine and allantoin GenProp0710 Generic phosphonates utilization GenProp0720 2-aminoethylphosphonate utilization GenProp0725 Outer membrane protein assembly complex GenProp0729 Methionine salvage from methylthioadenosine GenProp0758 Lycopene biosynthesis from IPP GenProp0788 Lysine biosynthesis via diaminopimelate (DAP) utilizing ammonia and NADPH GenProp0789 Homocysteine regeneration from S-adenosylhomocysteine GenProp0791 Coenzyme F420 biosynthesis GenProp0793 CMP-N-acetylneuraminate biosynthesis from UDP-N-acetylglucosamine GenProp0796 CMP-pseudaminic acid biosynthesis from UDP-N-acetylglucosamine GenProp0812 Rho-dependent termination GenProp0814 Urea utilization GenProp0836 Menaquinone biosynthesis GenProp0841 2-oxoacid:ferredoxin oxidoreductase GenProp0930 GDP-4-keto-6-deoxymannose biosynthesis from beta-D-fructose-6-phosphate GenProp0969 Biosynthesis of UDP-N-acetyl-D-mannosaminuronic acid (UDP-ManNAcA) GenProp0970 Biosynthesis of Und-PP-GlcNAc GenProp0971 Biosynthesis of undecaprenyl phosphate (Und-P) GenProp0972 dTDP-4-acetamido-4,6-dideoxy-D-galactose (dTDP-Fuc4NAc) biosynthesis GenProp0975 Biosynthesis of UDP-N-acetyl-D-mannosamine (UDP-ManNAc) GenProp1032 dTDP-4-dehydro-L-rhamnose biosynthesis from dTDP-4-dehydro-6-deoxy-alpha-D-glucose GenProp1034 Decarboxylation/antiport proton-motive cycle: oxalate-formate GenProp2010 Fibrinolysis GenProp2013 26S proteasome GenProp2033 mTORC1 activation GenProp2035 mTORC1 pathway GenProp2050 TFIIH basal transcription factor GenProp2052 SAGA-like (SLIK) complex GenProp2086 Pro-insulin activation GenProp2087 Proglucagon activation GenProp2095 Insulin secretion via PKA GenProp2096 Insulin secretion via PKC GenProp2098 Insulin secretion via IP3 GenProp2099 Insulin secretion via glucose

CHAPTER 4

SYSTEM

GenProp0010 Inteins GenProp0051 Urease GenProp0052 Type III secretion GenProp0053 Type II secretion GenProp0055 Polyhydroxyalkanoic acids GenProp0059 Type I secretion GenProp0061 Lipoprotein system lgt/lsp/lnt GenProp0114 Nucleotide excision repair GenProp0119 PTS transport system GenProp0127 Tat (Sec-independent) protein export GenProp0128 F1/F0 ATPase GenProp0129 Na+-translocating NADH-quinone reductase GenProp0130 Rnf-type electron transport complex GenProp0132 Transcription termination/antitermination Nus factors GenProp0135 NADH dehydrogenase complex I GenProp0137 Iron-sulfur cluster assembly SUF system GenProp0138 Iron-sulfur cluster assembly iscSUA-hscBA-fdx system GenProp0151 Resistance to mercury GenProp0156 Cyanophycin-like storage polymers GenProp0165 MinCDE system GenProp0166 MreBCD system GenProp0168 Glycogen system GenProp0170 Coenzyme PQQ biosynthesis GenProp0172 Potassium-transporting ATPase KdpFABC GenProp0176 TRAP-T (tripartite ATP-independent periplasmic transporters) family transporters GenProp0178 Glycine cleavage system GenProp0190 Phosphate ABC transporter (pstSCAB-phoU) GenProp0191 Sulfate/thiosulfate ABC transporter GenProp0192 Molybdate ABC transporter GenProp0198 RuvABC Holliday junction complex GenProp0201 SMC-ScpA-ScpB complex GenProp0207 Lipoprotein localization system lolABCDE GenProp0209 Sec system preprotein translocase GenProp0214 Acetyl-CoA carboxylase complex GenProp0215 SOS response GenProp0216 RecBCD pathway GenProp0225 Mismatch repair GenProp0226 Integron GenProp0232 Phosphonates C-P lyase system GenProp0236 Phosphonates ABC transport GenProp0244 Chaperone system: DnaK-DnaJ-GrpE GenProp0245 Chaperone system: GroEL/GroES GenProp0246 Chaperone system: thermosome/prefoldin GenProp0251 ClpXP degradation machine GenProp0252 Thiamine/TPP ABC transporter ThiBPQ GenProp0262 RNA polymerase, bacterial GenProp0263 DNA polymerase III, bacterial GenProp0271 Trehalose utilization GenProp0277 Cobalt import system (ABC transporter: 3.A.1.18.1) GenProp0288 Tetrahydromethanopterin S-methyltransferase complex GenProp0289 Class I (aerobic) ribonucleotide reductase GenProp0290 Class II (B12-dependent) ribonucleotide reductase GenProp0291 Class III (anaerobic) ribonucleotide reductase GenProp0292 Ethanolamine degradation organelle GenProp0294 Ethanolamine degradation proteinaceous organelle GenProp0295 Prokaryotic N-terminal cleavage/methylation, type IV pilin-like GenProp0310 CRISPR system, I-F/Ypest subtype GenProp0311 Competence-related genes GenProp0313 CRISPR system, I-C/Dvulg subtype GenProp0314 CRISPR system, II/Nmeni subtype GenProp0315 CRISPR system, I-E/Ecoli subtype GenProp0316 CRISPR system, III-B/RAMP module subtype GenProp0317 CRISPR system, I-B subtype, Tneap branch GenProp0318 CRISPR system, III-A/Mtube subtype GenProp0319 CRISPR system, I-A/Apern subtype GenProp0320 CRISPR system, I-B subtype, Hmari branch GenProp0322 Addiction module, Hig class GenProp0323 Addiction module, ParE class GenProp0324 Addiction module, RelE-RelB class GenProp0326 Protein sorting system, PEP-CTERM/exosortase (generic) GenProp0455 Restriction enzyme system, type I GenProp0456 Attenuation system: tryptophanase GenProp0457 Rhamnose catabolism GenProp0458 Fucose catabolism GenProp0460 Gas vesicle GenProp0465 Circadian clock system KaiABC GenProp0469 CRISPR system, Dpsyc subtype GenProp0470 Sporadically distributed four-gene operon GenProp0471 Uncharacterized gene pair TIGR02683/TIGR02684 GenProp0472 Uncharacterized gene pair TIGR02687/TIGR02688 GenProp0474 Arsenical resistance system GenProp0483 Cytochrome c oxidase, cbb3-type GenProp0484 F-type conjugation system (type IV secretion) specific components GenProp0485 F and P-type conjugation systems (type IV secretion), common components GenProp0487 RepABC-type alpha-proteobacterial replicon(s) GenProp0488 Superoxide dismutase, nickel-type GenProp0490 Ti-type conjugation system GenProp0491 RecFOR recombination pathway GenProp0492 Nonhomologous end-joining, bacterial type GenProp0493 AddAB pathway GenProp0494 Nickel

import ABC transporter GenProp0542 Tol-Pal system GenProp0543 TonB-dependent transport GenProp0610 Endospore formation marker gene set GenProp0613 Cytochrome c reductase GenProp0614 Cytochrome c oxidase, mitochondrial, caa3-type GenProp0617 Cytochrome (quinone) oxidase, bd type GenProp0618 Cytochrome (quinone) oxidase, bo type GenProp0620 Cytochrome (quinone) oxidase, aa3 type, QoxABCD GenProp0626 Protein sorting system, sortase type, LPXTG/SrtA class GenProp0629 A1/A0 ATPase GenProp0630 2-oxoglutarate dehydrogenase system GenProp0631 Nitrogenase, Fe-only GenProp0632 Nitrogenase, V-containing GenProp0633 Nitrogenase, Mo-containing GenProp0634 Sigma-54 (RpoN) systems GenProp0636 Respiratory nitrate reductase GenProp0637 Anaerobic dimethyl sulfoxide reductase GenProp0640 Xanthine dehydrogenase GenProp0648 Phage shock protein regulon GenProp0649 Heterocyst-related ABC exporter DevBCA GenProp0651 Ectoine ABC transporter EhuABCD GenProp0652 Exopolysaccharide biosynthesis, exosortase A-associated GenProp0654 Intracellular sulfur oxidation GenProp0655 Exopolysaccharide biosynthesis, exosortase B-associated GenProp0658 Cellulose biosynthesis GenProp0660 Photosystem I GenProp0661 Photosystem II GenProp0662 Photosynthetic reaction center, alphaproteobacterial type GenProp0663 Verrucomicrobium/Chthoniobacter four-plus-PEP_CTERM cassette GenProp0664 Protein sorting system, sortase type, SrtB class GenProp0665 Peptide chain release operon, RctB-like/PrfH GenProp0666 Membrane-associated monooxygenase (methane/ammonia) GenProp0668 Acetyltransferase/synthase/peptidase cassette GenProp0669 Protein sorting system, putative, exosortase G class GenProp0670 CRISPR system, Aferr subtype GenProp0672 Citrate lyase system GenProp0674 Malonate decarboxylase GenProp0676 Alkylhydroperoxide reductase AhpCF (peroxiredoxin) GenProp0678 C-type cytochrome biogenesis, system I GenProp0680 C-type cytochrome biogenesis, system II GenProp0682 Formate-dependent nitrite reductase, seven-gene type GenProp0683 Cytochrome c nitrite reductase NrfHA GenProp0684 FtsH/HflKC protease complex GenProp0685 CRISPR system, I-D/Cyano subtype GenProp0693 Fructose utilization as fructose-1,6-bisphosphate GenProp0694 Glucose utilization as fructose-1,6-bisphosphate GenProp0695 Cytochrome b558/566 GenProp0696 Xanthine utilization as a source of guanine-monophosphate (GMP) GenProp0699 DNA gyrase GenProp0701 DNA sulfur modification system dnd GenProp0702 Benzoyl-CoA reductase GenProp0703 4-hydroxybenzoyl-CoA reductase GenProp0705 Xanthine dehydrogenase, bacillus type (pucABCDE) GenProp0707 Botulinum toxin system GenProp0712 2-aminoethylphosphonate (AEP) ABC transporter, type I, PhnSTUV GenProp0714 Galactarate utilization via tartronate semi-aldehyde GenProp0715 Galactarate utilization via 2-oxoglutarate semialdehyde GenProp0716 Glucarate utilization via tartronate semi-aldehyde GenProp0717 Glucarate utilization via 2-oxoglutarate semialdehyde GenProp0719 Methyl-coenzyme M reductase GenProp0721 2-aminoethylphosphonate (AEP) ABC transporter, type II GenProp0722 Methanogenesis marker set GenProp0723 Hydrogenase, nickel and F420-dependent GenProp0726 Selenium-dependent molybdenum hydroxylase system GenProp0735 Type VI secretion GenProp0737 Anaerobic glycerol-3-phosphate dehydrogenase complex GenProp0739 Iron-sulfur cluster assembly CsdA-CsdE system GenProp0740 Translation initiation, bacterial GenProp0741 Translation elongation, bacterial GenProp0742 Iron-sulfur cluster assembly NIF system GenProp0743 Urea ABC transporter UrtABCDE GenProp0745 Lipoyl-protein attachment GenProp0746 Translation termination, bacterial GenProp0748 Choline ABC transporter, ChoXWV family GenProp0749 Choline sulfate utilization GenProp0751 Acidobacterial ADOP/PadR gene pairs GenProp0752 Ergothioneine biosynthesis GenProp0756 Fumarate reductase complex GenProp0757 Quorum-sensing, autoinducer-2 system GenProp0764 DMSO reductase family type II enzyme GenProp0767 Very short patch repair GenProp0768 CRISPR system, Pging subtype GenProp0778 Protein sorting system, GlyGly-CTERM/rhombosortase GenProp0781 Gliding motility, Bacteroidetes type GenProp0790 Sporadic pair TIGR03545/TIGR03546 GenProp0794 Proposed N-acetyl sugar amidation module WbuXYZ GenProp0798 Enterococcus/Aeromonas extended locus GenProp0802 Ribosome biogenesis proteins, bacteria GenProp0806 Replication initiation, bacterial GenProp0809 Bacteriocin system, lactococcin 972 group GenProp0810 Encapsulin proteinaceous organelle GenProp0811 S-layer homology domain-mediated cell wall binding GenProp0821 Putative Na:solute symporter two-gene cassette GenProp0828 Heme uptake system, NEAT-domain mediated GenProp0830 Integral membrane mystery pair GenProp0833 Proteasome-targeting modification by pupylation GenProp0834 Proteasome, bacterial GenProp0835 ATP-dependent protease HslVU GenProp0839 2-oxoacid:ferredoxin oxidoreductase, multisubunit form GenProp0840 Polyphosphate kinase/exopolyphosphatase system GenProp0842 2-oxoacid:acceptor oxidoreductase, 2 subunit form GenProp0845 Protein sorting system, KxYKxGKxW class GenProp0847 Exosporium GenProp0853 Lantibiotic system, gallidermin/epidermin family GenProp0854 ParB-rel/ThiF-rel cassette PRTRC GenProp0855 Integrating conjugative element, PFGI-1 class GenProp0857 Anchored repeat-class ABC transporter, P. acnes type GenProp0858 Conjugative transposon, Bacteroides type GenProp0859 Protein sorting system, proteobacterial dedicated sortase type GenProp0860 Tryptophan tryptophylquinone modification of methylamine dehydrogenase GenProp0861 Bacteriocin system, NHLP (nif11/nitrile hydratase leader peptide) transport group GenProp0867 Decarboxylation/antiport proton-motive cycle: aspartate-alanine GenProp0871 Decarboxy-

lation/antiport proton-motive cycle: ornithine-putrescine GenProp0872 Decarboxylation/antiport proton-motive cycle: histidine-histamine GenProp0877 Flagellar motor stator complex GenProp0879 Flagellar export apparatus GenProp0880 Flagellar basal body complex GenProp0881 Flagellar assembly apparatus GenProp0882 Flagellar filament and hook complex GenProp0883 Flagellar motor switch (rotor) complex GenProp0885 Flagellar post-translational modification components GenProp0891 Coenzyme M biosynthesis GenProp0894 Actinobacterial uncharacterized trio GenProp0895 Alcohol ABC transporter, PedABC-type GenProp0896 Proposed F420-0 transporter GenProp0898 RNA ligase/unknown protein pair GenProp0899 Nitrile hydratase GenProp0901 Post-ribosomal natural product synthesis system, Burkholderia TOMM-type GenProp0902 Quinohemoprotein amine dehydrogenase GenProp0903 Radical SAM/uracil DNA glycosylase system GenProp0904 Type VII secretion, Actinobacteria form GenProp0905 Type VII secretion, Firmicutes form GenProp0907 Poly-beta-1,6 N-acetyl-D-glucosamine system, PgaABCD type GenProp0910 Butyryl-CoA to butyrate, phosphotransbutyrylase pathway GenProp0911 Butyryl-CoA to butyrate, acetate CoA-transferase pathway GenProp0914 Hydrogenase, [FeFe]-dependent GenProp0915 Hydrogenase, mono-iron type GenProp0916 Radical SAM/uncharacterized protein TIGR03936 system GenProp0917 Mycofactocin system GenProp0918 Anaerobic sulfatase/maturase system GenProp0919 SCIFF/radical SAM Clostridial gene pair GenProp0920 Radical SAM Y_X(10)_GDL system GenProp0921 Radical SAM pair and His-Xaa-Ser repeats peptide GenProp0922 CRISPR system, Myxan subtype GenProp0923 Mycothiol system GenProp0928 OMP chaperone system: SurA-Skp-DegP GenProp0934 Dimethyl sulphone utilization (aerobic) GenProp0935 Uptake/catabolism system KPN_01854/KPN_01858 GenProp0936 Bacteriocin system, sporulation delay protein group GenProp0938 Glycine radical enzyme system, YjjI/YjjW pair GenProp0939 RSAM-containing biosynthetic cluster, MSMEG_0568 system GenProp0940 Metallo-mystery pair GenProp0941 TonB-dependent nutrient uptake pairs, RagAB/SusCD type GenProp0943 Pyruvate formate-lyase system GenProp0944 Cohesin/dockerin (cellulosome-like) system GenProp0945 Radical SAM maturase bacteriocin system, CLI_3235 type GenProp0948 RNA repair, Hen1/Pnkp system GenProp0954 Radical SAM-cylized peptide, Pep1357C family GenProp0955 Modified peptide/radical SAM maturase system, YydFG family GenProp0956 Radical SAM maturase/selenobacteriocin system GenProp0957 Protein sorting system, SipW class GenProp0958 D-alanyl-lipoteichoic acid biosynthesis GenProp0961 Heme metabolism pair HutWX GenProp0962 Methanobactin biosynthesis, Mb-OB3b family GenProp0967 Radical SAM maturase system, CXXX repeats type GenProp0977 DNA ligase/helicase system GenProp0978 Protein sorting system, PGF-CTERM/archaeosortase A GenProp0979 Protein sorting system, PEF-CTERM/archaeosortase C GenProp0980 Protein sorting system, putative, exosortase F class GenProp0981 Radical SAM/lipoprotein system GenProp0982 Radical SAM maturase system, FibroRumin system GenProp0983 Protein sorting system, VPXXXP-CTERM/archaeosortase B GenProp0984 Radical SAM maturase system, GG-Bacteroidales group GenProp0985 Protein sorting system, VPDSG-CTERM/exosortase C GenProp0987 Protein sorting system, cyano-PEP-CTERM class GenProp0988 Elongation factor P beta-lysylation GenProp0989 Protein sorting system, VPEID-CTERM/exosortase E GenProp0992 Radical SAM maturase system, methanogen GenProp0993 Protein sorting system, PIP-CTERM/archaeosortase D GenProp0996 Protein sorting system, IPTLxxWG-CTERM/exosortase H GenProp0999 Por secretion system GenProp1000 ATP-grasp maturase system, uncharacterized GenProp1002 Grasp-with-spasm peptide maturase system GenProp1003 Cysteine S-glycopeptide biosynthesis, sublancin family GenProp1004 Light-independent protochlorophyllide reductase GenProp1005 Protein sorting system, putative, exosortase J class GenProp1006 Orbivirus protein set GenProp1007 Orthobunyavirus protein set GenProp1009 Coronavirus protein set GenProp1010 Accessory Sec system, Actinobacterial type GenProp1011 Quorum-sensing system, cyclic peptide-mediated GenProp1012 Influenzae A virus protein set GenProp1013 Influenzae B virus protein set GenProp1016 Seadornavirus protein set GenProp1026 Nodulation factor production GenProp1037 Radical SAM/SPASM system GRRM GenProp1052 Geopeptide radical SAM/SPASM maturase system GenProp1053 RSAM/selenodomain system GenProp1054 Protein sorting system, MSEP-CTERM/exosortase K GenProp1055 Protein sorting system, PEFG-CTERM/thaumarchaeosortase GenProp1057 Radical SAM/SPASM system GeoRSP GenProp1060 tRNA N6-threonylcarbamoyladenosine modification GenProp1061 CRISPR system, PreFran subtype GenProp1062 Radical SAM/SPASM system Clo7bot GenProp1063 Memo/AMMECR1/rSAM family trio system GenProp1065 Radical SAM/SPASM TIGR04347/TIGR04031 system GenProp1069 FxLD lantipeptide system GenProp1071 Myo-inositol catabolism GenProp1072 Cyanase system GenProp1074 Type V secretion GenProp1075 Ferrous iron transport Feo system GenProp1076 Choline TMA-lyase system GenProp1077 tRNA(Pro) cmo5UGG modification GenProp1078 Sporulation killing factor system GenProp1079 Lpt lipopolysaccharide export transenvelope protein complex GenProp1080 Protein O-heptosyltransferase with autotransporter target GenProp1081 Replication restart, PriA/PriB primosome pathway GenProp1082 16S rRNA C1402 m(4)Cm modification GenProp1083 Cyanobactin-like ribosomal natural product biosynthesis GenProp1085 Pyranoside core peptidyl nucleoside antibiotic biosynthesis GenProp1087 Protein sorting system, vault protein/exosortase

N GenProp1088 Thiosulfate oxidation to sulfate GenProp1089 B12-dependent trimeric diol/glycerol dehydratase GenProp1090 Radical SAM/SPASM maturase system XYE GenProp1095 Exodeoxyribonuclease VII GenProp1099 8-oxoguanine DNA repair GenProp1100 Mobile element associated three-Cys-motif pair GenProp1101 Quinoprotein cytochrome relay system GenProp1102 Protein sorting system, CGP-CTERM/unknown of Thermococcaceae GenProp1103 Protein sorting system, Synergist-CTERM/unknown of Synergistetes GenProp1104 Acyl carrier protein phosphopantetheinylation GenProp2039 Mitochondrial protein processing GenProp2085 Renin-Angiotensin system GenProp2092 GLP-1 signalled increase in cAMP GenProp2093 GIP signalled increase in cAMP GenProp2094 PACAP signalled increase in cAMP

CHAPTER 5

GUILD

GenProp0002 Coenzyme F420 utilization GenProp0021 CRISPR region GenProp0182 Reduction of oxidized methionine GenProp0208 Phage: major features GenProp0213 Resistance to Reactive Oxygen Species (ROS) GenProp0258 tRNA aminoacylation GenProp0321 Toxin-antitoxin system, type II GenProp0476 Protein-coding palindromic elements GenProp0886 Flagellar components of unknown function GenProp0933 Bacillithiol utilization GenProp1091 Toxin-antitoxin system, type I GenProp1092 Toxin-antitoxin system, type III GenProp1093 Abortive infection proteins GenProp1094 Energy-coupling factor transporters GenProp2018 Initiator caspases of the apoptosis extrinsic pathway GenProp2019 Executor caspases of apoptosis

CHAPTER 6

CATEGORY

GenProp0007 Flagella GenProp0011 Methanogenesis GenProp0017 Metabolism GenProp0027 Photosynthesis GenProp0049 Spore formation GenProp0063 Biosynthesis GenProp0064 Biological niche GenProp0065 Genome properties GenProp0066 Selfish genetic elements GenProp0069 Cell surface component GenProp0070 Energy metabolism GenProp0071 Transport GenProp0072 Central intermediary metabolism GenProp0073 Cell motility GenProp0074 Virulence GenProp0075 Protein modification, prosthetic groups and cofactors GenProp0076 DNA handling GenProp0077 Natural products biosynthesis GenProp0115 DNA repair GenProp0116 DNA metabolism GenProp0126 Amino acid biosynthesis GenProp0131 Electron transport GenProp0133 Nucleic acid metabolism GenProp0140 Iron-sulfur cluster assembly systems GenProp0142 Catabolism GenProp0157 Storage polymer systems GenProp0167 Cellular growth, organization and division GenProp0179 Protein transport GenProp0180 Small molecule transport GenProp0184 Cofactor biosynthesis GenProp0185 Nucleotide biosynthesis GenProp0186 Storage and structural polymer biosynthesis GenProp0196 Sulfur metabolism GenProp0197 Protein biosynthesis GenProp0202 Nitrogen metabolism GenProp0239 Propionyl-CoA catabolism GenProp0242 Phosphorus metabolism GenProp0243 ATP synthases GenProp0247 Protein metabolism GenProp0248 Protein folding GenProp0267 Osmolyte biosynthesis GenProp0297 Lipopolysaccharide biosynthesis GenProp0298 Fermentation GenProp0299 Fatty acid biosynthesis GenProp0300 Photosynthetic cofactor biosynthesis GenProp0301 Protein degradation GenProp0302 Protein repair GenProp0303 Detoxification GenProp0462 Response to stress GenProp0475 Unknown biological process GenProp0551 Gene regulatory systems GenProp0603 Lifestyle GenProp0731 Amino acid metabolism GenProp0733 Amino acid salvage GenProp0734 Amino acid catabolism GenProp0795 Carbohydrate biosynthesis GenProp0816 Nucleoside catabolism/utilization pathways GenProp0817 Nucleotide catabolism/utilization pathways GenProp0819 RNA metabolism GenProp0820 Iron metabolism GenProp0838 Lipid biosynthesis GenProp0864 Response to external environment GenProp0866 Carbon metabolism GenProp0873 Ion-motive antiporter cycles GenProp0874 Response to acidic (low pH) environment GenProp1048 Quorum-sensing, biofilm formation, and developmental systems GenProp1067 Defense systems GenProp1106 CRISPR systems GenProp1107 Viral protein sets GenProp1211 Regulatory systems GenProp1212 Regulation of biological processes GenProp1213 Response to stimulus GenProp1214 Transcription GenProp1765 Vitamin Biosynthesis GenProp1766 Hormone biosynthesis GenProp1767 Secondary metabolites biosynthesis GenProp1768 Alcohols degradation GenProp1769 Amines and polyamines catabolism GenProp1770 Carbohydrates degradation GenProp1771 Carboxylates degradation GenProp1772 Fatty acid degradation GenProp1773 Lipid degradation GenProp1774 Inorganic nutrients metabolism GenProp1775 Secondary metabolites degradation GenProp1776 Nucleosides and nucleotides degradation GenProp1777 Nucleic acids processing GenProp1778 Hormone systems

CHAPTER 7

Genome Property Types

There are five different types of Genome Properties (GP) represented.

PATHWAY These represent groups of proteins that perform biochemical steps in order within a recognised enzymatic pathway.

METAPATH These represent a specific type of PATHWAY where one or more of the steps within the pathway are described by another Genome Property. For the purposes of calculation, METAPATHs are dependent on their respective GP steps.

SYSTEM These represent groups of proteins that work together to fulfil a role, but do not necessarily represent a traditional enzymatic pathway (for example, transport systems).

GUILD These represent groups of proteins sharing overall function, but which do not represent a system.

CATEGORY These are organisational properties that allow the full set of genome properties to be arranged into a hierarchy. These properties are not calculated.

CHAPTER 8

Flatfile Format

Each Genome Property is represented by up to three files

- **DESC file** - a description of the GP and the constituent steps
- **FASTA file** - a concatenation of fasta files that resolve to a yes for each of the constituent steps
- **status file** - a record of binary flags recording whether the GP has been curated and is to be made public

DESC file

The tags used in the DESC file are listed below, along with the description of the field they relate to.

AC	Accession ID
DE	Description/name of Genome Property
TP	Type
AU	Author
TH	Threshold
RN	Reference number
RM	PMID of reference
RT	Reference title
RA	Reference author
RL	Reference citation
DC	Database title
DR	Database link
PN	Parent accession ID
CC	Property description
**	Private notes
–	Separator
SN	Step number
ID	Step ID
DN	Step display name (includes EC number if available)
RQ	Required step
EV	Evidence (includes whether sufficient)
TG	Gene Ontology (GO) ID
//	End

The DESC file is formatted such that a single tag is included on each line, followed by 2 blank spaces, followed by the value of the field.

In the case of the property description (CC) and private notes (**) fields, the information may stretch accross multiple lines. The line length is limited to 80 characters (including the tag) and so any subsequent lines used must also carry the tag. See the PATHWAY example below.

```
AC  GenProp0001
DE  chorismate biosynthesis via shikimate
TP  PATHWAY
AU  Haft D
TH  2
RN  [1]
RM  12636087
RT  The biosynthesis of shikimate metabolites.
RA  Knaggs AR;
RL  Nat Prod Rep. 2003;20:119-136.
DC  Shikimate and Chorismate Biosynthesis
DR  IUBMB; misc; shikim;
DC  Phenylalanine, Tyrosine and Tryptophan Biosynthesis
DR  KEGG; map00400;
DC  Chorismate biosynthesis
DR  MetaCyc; ARO-PWY;
CC  Chorismate is the final common intermediate in the biosynthesis of
CC  phenylalanine, tyrosine and tryptophan (the aromatic amino acids) as
CC  well as menaquinone, ubiquinone, salicylate and phenazine.  Chorismate
CC  D-erythrose 4-phosphate and phosphoenolpyruvate.  This pathway is
CC  widely distributed among microorganisms.  Certain methanogenic
CC  euarchaeota appear to be missing the first two steps of the pathway -
CC  these may be catalyzed by as of yet uncharacterized enzymes.
**  Archaea are hard to assign.  It would appear that the chorismate
**  pathway is at least partially present in archaea, since many synthesize
**  aromatic amino acids and contain some of the enzymes covered by the
**  present HMM-set.  In particular, shikimate kinase appears to be absent
**  in many archaea which contain the enzymes before and after it in the
**  pathway (Archaeoglobus, Methanobacterium, Methanococcus, Pyrococcus)
**  implying that there may be a non-orthologous enzyme.  Similarly, the
**  first two steps of the pathway appear to be missing in certain archaea
**  but are found in others.
--
SN  1
ID  phospho-2-dehydro-3-deoxyheptonate aldolase
DN  phospho-2-dehydro-3-deoxyheptonate aldolase (EC 2.5.1.54)
RQ  1
EV  IPR006219; TIGR00034; sufficient;
TG  GO:0009423;
EV  IPR002480; TIGR01358; sufficient;
TG  GO:0009423;
EV  IPR006268; TIGR01361; sufficient;
TG  GO:0009423;
EV  IPR010210; TIGR01949; sufficient;
TG  GO:0009423;
--
SN  2
ID  3-dehydroquinate synthase
DN  3-dehydroquinate synthase (EC 4.2.3.4)
RQ  1
EV  IPR016037; TIGR01357; sufficient;
TG  GO:0009423;
EV  IPR002812; PF01959; sufficient;
TG  GO:0009423;
```

```
--
SN  3
ID  3-dehydroquinate dehydratase
DN  3-dehydroquinate dehydratase (EC  4.2.1.10)
RQ  1
EV  IPR001874; TIGR01088; sufficient;
TG  GO:0009423;
EV  IPR001381; TIGR01093; sufficient;
TG  GO:0009423;
--
SN  4
ID  shikimate 5-dehydrogenase
DN  shikimate 5-dehydrogenase (EC  1.1.1.25)
RQ  1
EV  IPR011342; TIGR00507; sufficient;
TG  GO:0009423;
EV  IPR010110; TIGR01809; sufficient;
TG  GO:0009423;
--
SN  5
ID  shikimate kinase
DN  shikimate kinase (EC  2.7.1.71)
RQ  1
EV  IPR031322; PF01202; sufficient;
TG  GO:0009423;
EV  IPR010189; TIGR01920; sufficient;
TG  GO:0009423;
--
SN  6
ID  3-phosphoshikimate 1-carboxyvinyltransferase
DN  3-phosphoshikimate 1-carboxyvinyltransferase (EC  2.5.1.19)
RQ  1
EV  IPR006264; TIGR01356; sufficient;
TG  GO:0009423;
--
SN  7
ID  chorismate synthase
DN  chorismate synthase (EC  4.2.3.5)
RQ  1
EV  IPR000453; TIGR00033; sufficient;
TG  GO:0009423;
//
```

While the layout of the DESC file for CATEGORY type properties follows the same format, the steps do not refer to calculable evidence. In the case of CATEGORY, the steps define the properties (including other sub-categories) that exist as children of the CATEGORY. See the CATEGORY example below.

```
AC  GenProp0063
DE  Biosynthesis
TP  CATEGORY
AU  Haft DH
TH  0
CC  The process of creating complex biomolecules from simpler starting
CC  materials.
--
SN  1
ID  Natural products biosynthesis
RQ  0
```

```
EV  GenProp0077;
--
SN  2
ID  Amino acid biosynthesis
RQ  0
EV  GenProp0126;
--
SN  3
ID  Cofactor biosynthesis
RQ  0
EV  GenProp0184;
--
SN  4
ID  Nucleotide biosynthesis
RQ  0
EV  GenProp0185;
--
SN  5
ID  Storage and structural polymer biosynthesis
RQ  0
EV  GenProp0186;
//
```

FASTA file

The FASTA file includes fasta sequences that are a match for each constituent step of the property. GPs of type CATEGORY do not have associated FASTA files as they do not contain any calculable steps. Similarly, a METAPATH which contains only GPs as evidence for its steps, would not have an associated FASTA file. The FASTA file is formatted such that each individual block of fasta sequence includes a descriptive header line, in the format provided by UniProt. The appropriate step number is then added to this header line in parenthesis, as shown below.

```
>sp|P0AB91|AROG_ECOLI (Step num: 1) Phospho-2-dehydro-3-deoxyheptonate aldolase, Phe-
↪sensitive OS=Escherichia coli (strain K12) GN=aroG PE=1 SV=1
```

An example FASTA file is shown here:

```
>sp|P0AB91|AROG_ECOLI (Step num: 1) Phospho-2-dehydro-3-deoxyheptonate aldolase, Phe-
↪sensitive OS=Escherichia coli (strain   K12) GN=aroG PE=1 SV=1
MNYQNDDLRIKEIKELLPPVALLEKFPATENAANTVAHARKAIHKILKGNDDRLLVVIGP
CSIHDPVAAKEYATRLLALREELKDELEIVMRVYFEKPRTTVGWKGLINDPHMDNSFQIN
DGLRIARKLLLDINDSGLPAAGEFLDMITPQYLADLMSWGAIGARTTESQVHRELASGLS
CPVGFKNGTDGTIKVAIDAINAAGAPHCFLSVTKWGHSAIVNTSGNGDCHIILRGGKEPN
YSAKHVAEVKEGLNKAGLPAQVMIDFSHANSSKQFKKQMDVCADVCQQIAGGEKAIIGVM
VESHLVEGNQSLESGEPLAYGKSITDACIGWEDTDALLRQLANAVKARRG

>sp|P07639|AROB_ECOLI (Step num: 2) 3-dehydroquinate synthase OS=Escherichia coli␣
↪(strain K12) GN=aroB PE=1 SV=1
MERIVVTLGERSYPITIASGLFNEPASFLPLKSGEQVMLVTNETLAPLYLDKVRGVLEQA
GVNVDSVILPDGEQYKSLAVLDTVFTALLQKPHGRDTTLVALGGGVVGDLTGFAAASYQR
GVRFIQVPTTLLSQVDSSVGGKTAVNHPLGKNMIGAFYQPASVVVDLDCLKTLPPRELAS
GLAEVIKYGIILDGAFFNWLEENLDALLRLDGPAMAYCIRRCCELKAEVVAADERETGLR
ALLNLGHTFGHAIEAEMGYGNWLHGEAVAAGMVMAARTSERLGQFSSAETQRIITLLKRA
GLPVNGPREMSAQAYLPHMLRDKKVLAGEMRLILPLAIGKSEVRSGVSHELVLNAIADCQ
SA
```

```
>sp|P05194|AROD_ECOLI (Step num: 3) 3-dehydroquinate dehydratase OS=Escherichia coli␣
↪(strain K12) GN=aroD PE=1 SV=2
MKTVTVKDLVIGTGAPKIIVSLMAKDIASVKSEALAYREADFDILEWRVDHYADLSNVES
VMAAAKILRETMPEKPLLFTFRSAKEGGEQAISTEAYIALNRAAIDSGLVDMIDLELFTG
DDQVKETVAYAHAHDVKVVMSNHDFHKTPEAEEIIARLRKMQSFDADIPKIALMPQSTSD
VLTLLAATLEMQEQYADRPIITMSMAKTGVISRLAGEVFGSAATFGAVKKASAPGQISVN
DLRTVLTILHQA

>sp|P15770|AROE_ECOLI (Step num: 4) Shikimate dehydrogenase (NADP(+)) OS=Escherichia␣
↪coli (strain K12) GN=aroE PE=1 SV=1
METYAVFGNPIAHSKSPFIHQQFAQQLNIEHPYGRVLAPINDFINTLNAFFSAGGKGANV
TVPFKEEAFARADELTERAALAGAVNTLMRLEDGRLLGDNTDGVGLLSDLERLSFIRPGL
RILLIGAGGASRGVLLPLLSLDCAVTITNRTVSRAEELAKLFAHTGSIQALSMDELEGHE
FDLIINATSSGISGDIPAIPSSLIHPGIYCYDMFYQKGKTPFLAWCEQRGSKRNADGLGM
LVAQAAHAFLLWHGVLPDVEPVIKQLQEELSA

>sp|P0A6D7|AROK_ECOLI (Step num: 5) Shikimate kinase 1 OS=Escherichia coli (strain␣
↪K12) GN=aroK PE=1 SV=2
MAEKRNIFLVGPMGAGKSTIGRQLAQQLNMEFYDSDQEIEKRTGADVGWVFDLEGEEGFR
DREEKVINELTEKQGIVLATGGGSVKSRETRNRLSARGVVVYLETTIEKQLARTQRDKKR
PLLHVETPPREVLEALANERNPLYEEIADVTIRTDDQSAKVVANQIIHMLESN

>sp|P0A6D3|AROA_ECOLI (Step num: 6) 3-phosphoshikimate 1-carboxyvinyltransferase␣
↪OS=Escherichia coli (strain K12) GN=aroA   PE=1 SV=1
MESLTLQPIARVDGTINLPGSKSVSNRALLLAALAHGKTVLTNLLDSDDVRHMLNALTAL
GVSYTLSADRTRCEIIGNGGPLHAEGALELFLGNAGTAMRPLAAALCLGSNDIVLTGEPR
MKERPIGHLVDALRLGGAKITYLEQENYPPLRLQGGFTGGNVDVDGSVSSQFLTALLMTA
PLAPEDTVIRIKGDLVSKPYIDITLNLMKTFGVEIENQHYQQFVVKGGQSYQSPGTYLVE
GDASSASYFLAAAAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICWGDDYISCTRGE
LNAIDMDMNHIPDAAMTIATAALFAKGTTTLRNIYNWRVKETDRLFAMATELRKVGAEVE
EGHDYIRITPPEKLNFAEIATYNDHRMAMCFSLVALSDTPVTILDPKCTAKTFPDYFEQL
ARISQAA

>sp|P12008|AROC_ECOLI (Step num: 7) Chorismate synthase OS=Escherichia coli (strain␣
↪K12) GN=aroC PE=1 SV=4
MAGNTIGQLFRVTTFGESHGLALGCIVDGVPPGIPLTEADLQHDLDRRRPGTSRYTTQRR
EPDQVKILSGVFEGVTTGTSIGLLIENTDQRSQDYSAIKDVFRPGHADYTYEQKYGLRDY
RGGGRSSARETAMRVAAGAIAKKYLAEKFGIEIRGCLTQMGDIPLDIKDWSQVEQNPFFC
PDPDKIDALDELMRALKKEGDSIGAKVTVVASGVPAGLGEPVFDRLDADIAHALMSINAV
KGVEIGDGFDVVALRGSQNRDEITKDGFQSNHAGGILGGISSGQQIIAHMALKPTSSITV
PGRTINRFGEEVEMITKGRHDPCVGIRAVPIAEAMLAIVLMDHLLRQRAQNADVKTDIPR
W
```

Status file

Each GP has an associated 'status' file which records (using binary flags) whether the property has been curated, and whether it is to be made public. This file is edited by the curator as part of the curation process prior to release. Private curator notes can be included below the double hyphen. An example status file is shown here:

```
checked:  1
public: 0
--
contains a dependent GP that is not yet public
```

CHAPTER 9

Calculating Genome Properties

The presence or absence of genome properties (GPs) within a given genome or proteome, is calculated based on the matches to the relevant InterPro entries and their associated protein signatures. The evidence (HMM) for each constituent step is tested against the genome in question, and each step defined as a hit or a miss. The total number of hits is then compared with the threshold level defined for the GP, to determine if the GP resolves to a YES (all required steps are present), NO (no required steps are present) or PARTIAL (the number of required steps present is greater than the threshold, indicating that some evidence of the presence of the GP can be assumed).

The pipeline for how users can calculate GPs for their own data described below.

Users begin with their own query "genome/proteome" in the form of a list of FASTA format sequences. The InterProScan matches for this genome must then be calcualted. This can either be done using EBI webservices, or by downloading a local install of InterProScan and running the calculation locally. By either method, the InterProScan matches can be output as TSV format. It is then possible to upload this TSV file to the GP viewer available on this website. This allows the user to easily compare the pattern of assertions for the user-defined genome, with the set of reference genomes provided on the webiste. For more information on how to use the viewer, see here #viewer .

CHAPTER 10

Contributing to Genome Properties

The majority of the current set of Genome Properties were produced within the TIGRFAM project, and refer to predominantly prokaryotic genomes. In parallel with the integration of Genome Properties into InterPro, further (predominantly eukaryotic) properties were produced by the InterPro group, and have been added to the total set. We welcome collaboration with any groups with an interest in defining further Genome Properties based on their own area of interest/expertise.

If you would like to contribute to Genome Properties in this way, please contact us at:

GenProp@ebi.ac.uk

CHAPTER 11

Funding

National Science Foundation, grant number 1458808

Genome Properties is supported by the following organisations:

EMBL	EMBL is EMBL-EBI's parent organisation
BBSRC	Biotechnology and Biological Sciences Research Council
NSF	National Science Foundation