

Appendices

Appendix IV: Feature Keys and Qualifiers – a brief explanation of what they are and a sample of the ones we use.

1 – Feature Keys: They describe features with DNA coordinates and once marked, they all appear in the Artemis main window. The ones we use are:

CDS: Marks the extent of the coding sequence.

RBS: Ribosomal binding site

misc_feature: Miscellaneous feature in the DNA

rRNA: Ribosomal RNA

repeat_region

repeat_unit

stem_loop

tRNA: Transfer RNA

2 – Qualifiers: They describe features in relation to their coordinates. Once marked they appear in the lower part of the Artemis window. They describe the feature whose coordinates appear in the ‘location’ part of the editing window. The ones we commonly use for annotation at the Sanger Institute are:

/class: Classification scheme we use “in-house” developed from Monica Riley’s MultiFun assignments (see Appendix VI).

/colour: Also used in-house in order to differentiate between different types of genes and other features.

/gene: Descriptive gene name, eg. *ilvE*, *argA* etc.

/label: Allows you to label a gene/feature in the main view panel.

/note: This qualifier allows for the inclusion of free text. This could be a description of the evidence supporting the functional prediction or other notable features/information which cannot be described using other qualifiers.

/product: The assigned possible function for the protein goes here.

/pseudo: Matches in different frames to consecutive segments of the same protein in the databases can be linked or joined as one and edited in one window. They are marked as pseudogenes. They are normally not functional and are considered to have been mutated.

/locus_tag : Systematic gene number, eg *SAS1670*, *Sty2412* etc.

The list of keys and qualifiers accepted by EMBL in sequence/annotation submission files are list at the following web page:

<http://www3.ebi.ac.uk/Services/WebFeat/>

Appendix VI: Prokaryotic Protein Classification Scheme used within the PSU

This scheme was adapted for in-house use from the Monica Riley's protein classification (<http://genprotec.mbl.edu/files/Multifun.html>).

More classes can be added depending on the microorganism that is being annotated (e.g secondary metabolites, sigma factors (ECF or non-ECF), etc).

0.0.0 Unknown function, no known homologs

0.0.1 Conserved in Escherichia coli

0.0.2 Conserved in organism other than Escherichia coli

1.0.0 Cell processes

1.1.1 Chemotaxis and mobility

1.2.1 Chromosome replication

1.3.1 Chaperones

1.4.0 Protection responses

1.4.1 Cell killing

1.4.2 Detoxification

1.4.3 Drug/analog sensitivity

1.4.4 Radiation sensitivity

1.5.0 Transport/binding proteins

1.5.1 Amino acids and amines

1.5.2 Cations

1.5.3 Carbohydrates, organic acids and alcohols

1.5.4 Anions

1.5.5 Other

1.6.0 Adaptation

1.6.1 Adaptations, atypical conditions

1.6.2 Osmotic adaptation

1.6.3 Fe storage

1.7.1 Cell division

2.0.0 Macromolecule metabolism

2.1.0 Macromolecule degradation

2.1.1 Degradation of DNA

2.1.3 Degradation of polysaccharides

2.1.2 Degradation of RNA

2.1.4 Degradation of proteins, peptides, glycoproteins

2.2.0 Macromolecule synthesis, modification

2.2.01 Amino acyl tRNA synthesis; tRNA modification

2.2.07 Phospholipids

2.2.02 Basic proteins - synthesis, modification

2.2.08 Polysaccharides - (cytoplasmic)

2.2.03 DNA - replication, repair, restriction./modification

2.2.09 Protein modification

2.2.04 Glycoprotein

2.2.10 Proteins - translation and modification

2.2.05 Lipopolysaccharide

2.2.11 RNA synthesis, modif., DNA transcrip.

2.2.06 Lipoprotein

2.2.12 tRNA

3.0.0 Metabolism of small molecules

3.1.0 Amino acid biosynthesis

3.1.01 Alanine

3.1.08 Glutamine

3.1.15 Phenylalanine

3.1.02 Arginine

3.1.09 Glycine

3.1.16 Proline

3.1.03 Asparagine

3.1.10 Histidine

3.1.17 Serine

3.1.04 Aspartate

3.1.11 Isoleucine 3.1.18 Threonine

3.1.05 Chorismate

3.1.12 Leucine

3.1.19 Tryptophan

3.1.06 Cysteine

3.1.13 Lysine

3.1.20 Tyrosine

3.1.07 Glutamate

3.1.14 Methionine

3.1.21 Valine

Appendix VI (cont):

- 3.2.0 Biosynthesis of cofactors, carriers
 - 3.2.01 Acyl carrier protein (ACP)
 - 3.2.02 Biotin
 - 3.2.03 Cobalamin
 - 3.2.04 Enterochelin
 - 3.2.05 Folic acid
 - 3.2.06 Heme, porphyrin
 - 3.2.07 Lipoate
 - 3.2.08 Menaquinone, ubiquinone
 - 3.2.09 Molybdopterin
 - 3.2.10 Pantothenate
 - 3.2.11 Pyridine nucleotide
 - 3.2.12 Pyridoxine
 - 3.2.13 Riboflavin
 - 3.2.14 Thiamin
 - 3.2.15 Thioredoxin, glutaredoxin, glutathione
 - 3.2.16 biotin carboxyl carrier protein (BCCP)
- 3.3.0 Central intermediary metabolism
 - 3.3.01 2'-Deoxyribonucleotide metabolism
 - 3.3.02 Amino sugars
 - 3.3.03 Entner-Doudoroff
 - 3.3.04 Gluconeogenesis
 - 3.3.05 Glyoxylate bypass
 - 3.3.06 Incorporation metal ions
 - 3.3.07 Misc. glucose metabolism
 - 3.3.08 Misc. glycerol metabolism
 - 3.3.09 Non-oxidative branch, pentose pathway
 - 3.3.10 Nucleotide hydrolysis
 - 3.3.11 Nucleotide interconversions
 - 3.3.12 Oligosaccharides
 - 3.3.13 Phosphorus compounds
 - 3.3.14 Polyamine biosynthesis
 - 3.3.15 Pool, multipurpose conversions of intermed. metab.
 - 3.3.16 S-adenosyl methionine
 - 3.3.17 Salvage of nucleosides and nucleotides
 - 3.3.18 Sugar-nucleotide biosynthesis, conversions
 - 3.3.19 Sulfur metabolism
 - 3.3.20 Amino acids
 - 3.3.21 other
- 3.4.0 Degradation of small molecules
 - 3.4.1 Amines
 - 3.4.2 Amino acids
 - 3.4.3 Carbon compounds
 - 3.4.4 Fatty acids
 - 3.4.5 Other
 - 3.4.0 ATP-proton motive force
- 3.5.0 Energy metabolism, carbon
 - 3.5.1 Aerobic respiration
 - 3.5.2 Anaerobic respiration
 - 3.5.3 Electron transport
 - 3.5.4 Fermentation
 - 3.5.5 Glycolysis
 - 3.5.6 Oxidative branch, pentose pathway
 - 3.5.7 Pyruvate dehydrogenase
 - 3.5.8 TCA cycle
- 3.6.0 Fatty acid biosynthesis
 - 3.6.1 Fatty acid and phosphatidic acid biosynthesis
- 3.7.0 Nucleotide biosynthesis
 - 3.7.1 Purine ribonucleotide biosynthesis
 - 3.7.2 Pyrimidine ribonucleotide biosynthesis
- 4.0.0 Cell envelop
 - 4.1.0 Periplasmic/exported/lipoproteins
 - 4.1.1 Inner membrane
 - 4.1.2 Murein sacculus, peptidoglycan
 - 4.1.3 Outer membrane constituents
 - 4.1.4 Surface polysaccharides & antigens
 - 4.1.5 Surface structures
- 4.2.0 Ribosome constituents
 - 4.2.1 Ribosomal and stable RNAs
 - 4.2.2 Ribosomal proteins - synthesis, modification
 - 4.2.3 Ribosomes - maturation and modification
- 5.0.0 Extrachromosomal
 - 5.1.0 Laterally acquired elements
 - 5.1.1 Colicin-related functions
 - 5.1.2 Phage-related functions and prophages
 - 5.1.3 Plasmid-related functions
 - 5.1.4 Transposon-related functions
 - 5.1.5 Pathogenicity island-related function
- 6.0.0 Global functions
 - 6.1.1 Global regulatory functions
- 7.0.0 Not classified (included putative assignments)

Appendix VII: List of colour codes

- 0** (white) - Pathogenicity/Adaptation/Chaperones
- 1** (dark grey) - energy metabolism (glycolysis, electron transport etc.)
- 2** (red) - Information transfer (transcription/translation + DNA/RNA modification)
- 3** (dark green) - Surface (IM, OM, secreted, surface structures)
- 4** (dark blue) - Stable RNA
- 5** (Sky blue) - Degradation of large molecules
- 6** (dark pink) - Degradation of small molecules
- 7** (yellow) - Central/intermediary/miscellaneous metabolism
- 8** (light green) - Unknown
- 9** (light blue) - Regulators
- 10** (orange) - Conserved hypo
- 11** (brown) - Pseudogenes and partial genes (remnants)
- 12** (light pink) - Phage/IS elements
- 13** (light grey) - Some misc. information e.g. Prosite, but no function

Appendix VIII: List of degenerate nucleotide value/IUB Base Codes.

R = A or G

S = G or C

B = C, G or T

Y = C or T

W = A or T

D = A, G or T

K = G or T

N = A, C, G or T

H = A, C or T

M = A or C

V = A, C or G