

Appendix A: SQL Script for Table Creation

SQL Script which has been designed and developed for storing DNA sequenced data and DNA analysis data, that can be executed with Oracle 10g RDBMS, to create various tables.

```
create table aau_species (  
  species_id number(10),  
  species_name varchar2(100),  
  constraint pk_as primary key(species_id)  
);
```

```
create table aau_reference (  
  reference_id number(10),  
  reference_name varchar2(100),  
  constraint pk_aref primary key(reference_id)  
);
```

```
create table aau_reference_details (  
  reference_id number(10),  
  genbank_accession_no varchar2(100)  
);
```

```
create table aau_run  
(  
  species_id number(10),  
  run_id number(10),  
  dt_of_run date,  
  constraint pk_arun primary key(species_id, run_id),  
  constraint fk_arun1 foreign key(species_id) REFERENCES  
  aau_species(species_id)  
);
```

```
create table aau_read (  
  id number(10),  
  species_id number(10),  
  run_id number(2),  
  file_name varchar2(100),  
  read_id number(10),
```

```

read_name varchar2(100),
  read_rank number(10),
read_x number(10,3),
read_y number(10,3),
read_length number(6),
read_sequence clob,
constraint pk_aread primary key(species_id, run_id,
file_name, read_id),
constraint fk_aread1 foreign key (species_id, run_id)
references aa_run(species_id, run_id)
);

```

```

create table aa_contig (
id number(10),
species_id number(10),
run_id number(2),
contig_id number(10),
locus          varchar2(100),
contig_name varchar2(100),
no_of_basepair number(6),
type_of_sequence varchar2(100),
linear          varchar2(100),
organism_type  varchar2(100),
date_of_run    date,
definition     varchar2(100),
strain varchar2(100),
substrain varchar2(100),
accession_no   varchar2(100),
version        varchar2(100),
gi             varchar2(100),
dblink         varchar2(100),
keywords  varchar2(100),
source        clob,
organism clob,
comments  clob,
assembly_method clob,
genaau_coverage varchar2(10),
sequencing_technology varchar2(100),
feature_source_start number(10),
feature_source_end   number(10),
feature_organism varchar2(100),
feature_mol_type varchar2(100),
feature_strain varchar2(100),
feature_sub_strain varchar2(100),
feature_db_xref varchar2(100),
contig_sequence      clob,
constraint pk_ac primary key(species_id, run_id,
contig_id),
constraint fk_acl foreign key (species_id, run_id)
references aa_run(species_id, run_id)

```

```
);
```

```
create table aau_contig_reference (  
  id number(10),  
  species_id number(10),  
  run_id number(2),  
  contig_id number(10),  
  locus varchar2(100),  
  reference number(10),  
  title clob,  
  journal clob,  
  pubmed number(10),  
  constraint pk_acref primary key(species_id, run_id,  
  contig_id, reference),  
  constraint fk_acref1 foreign key (species_id, run_id)  
  references aau_run(species_id, run_id));
```

```
create table aau_contig_reference_author (  
  id number(10),  
  species_id number(10),  
  run_id number(2),  
  contig_id number(10),  
  locus varchar2(100),  
  reference number(10),  
  author varchar(100),  
  
  constraint pk_acrauth primary key(species_id, run_id,  
  contig_id, reference, author),  
  constraint fk_acrauth1 foreign key (species_id, run_id,  
  contig_id, reference) references  
  aau_contig_reference(species_id, run_id, contig_id,  
  reference)  
);
```

```
create table aau_contig_promoter (  
  id number(10),  
  species_id number(10),  
  run_id number(2),  
  contig_id number(10),  
  promoter_id number(10),  
  transcription_start_site number(10),  
  ldf_score number(6,3),  
  tatabox_start_site number(10),  
  tatabox_score number(6,3),  
  tatabox_region varchar2(20),  
  enhancer_start_site number(10),  
  enhancer_score number(6,3),  
  constraint pk_acp primary key(species_id, run_id,  
  contig_id, promoter_id),
```

```

constraint fk_acpl foreign key (species_id, run_id,
contig_id) references aau_contig(species_id, run_id,
contig_id)
);

```

```

create table aau_contig_repeat (
id number(10),
species_id number(10),
run_id number(2),
contig_id number(10),
repeat_id number(10),
constraint pk_acr primary key(species_id, run_id,
contig_id, repeat_id),
constraint fk_acr1 foreign key (species_id, run_id,
contig_id) references aau_contig(species_id, run_id,
contig_id)
);

```

```

create table aau_contig_repeat_reference (
id number(10),
species_id number(10),
run_id number(2),
contig_id number(10),
repeat_id number(10),
reference_id number(10),
reference_repeat_id number(10),
repeat_score number(5),
substitution_perc number(6,3),
deletion_perc number(6,3),
insertion_perc number(6,3),
repeat_start_site number(10),
repeat_stop_site number(10),
repeat_length_left number(10),
orientation char(1),
mtchnng_repeat_name varchar2(100),
mtchnng_repeat_class_family varchar2(100),
mtchnng_repeat_subclass_family varchar2(100),
reference_start_site number(10),
reference_stop_site number(10),
reference_length_left number(10),
constraint pk_acrr primary key(species_id, run_id,
contig_id, repeat_id, reference_id, reference_repeat_id),
constraint fk_acrr1 foreign key (species_id, run_id,
contig_id, repeat_id)
references aau_contig_repeat(species_id, run_id,
contig_id, repeat_id),
constraint fk_acrr2 foreign key (reference_id) references
aau_reference(reference_id)
);

```

```

create table aau_contig_codingregion (
id number(10),
species_id number(10),
run_id number(2),
contig_id number(10),
codingregion_id number(10),
sequence clob,
constraint pk_acc primary key(species_id, run_id,
contig_id, codingregion_id),
constraint fk_acc1 foreign key (species_id, run_id,
contig_id) references aau_contig(species_id, run_id,
contig_id)
);

```

```

create table aau_contig_codingregion_gene (
id number(10),
species_id number(10),
run_id number(2),
contig_id number(10),
codingregion_id number(10),
gene_id number(10),
gene_name varchar2(100),
kegg clob,
protein_id number(10),
protein_name varchar2(100),
constraint pk_accg primary key(species_id, run_id,
contig_id, codingregion_id, gene_id),
constraint fk_accg1 foreign key (species_id, run_id,
contig_id, codingregion_id)
references aau_contig_codingregion(species_id, run_id,
contig_id, codingregion_id));

```

```

create table aau_contig_cds_gene_chromo_ref (
id number(10),
species_id number(10),
run_id number(2),
contig_id number(10),
codingregion_id number(10),
gene_id number(10),
reference_id number(10),
species_chromosome_no varchar2(15),
reference_chromosome_no varchar2(15),
gene_start_site_in_cds number(10),
gene_stop_site_in_cds number(10),
gene_orientation_in_cds char(1),
reference_start_site number(10),
reference_stop_site number(10),

```

```

reference_orientation char(1),
total_alignment_score number(5),
error_value varchar2(15),
alignment_identity number(6,3),
alignment_length number(10),
constraint pk_accgcr primary key(species_id, run_id,
contig_id, codingregion_id, gene_id, reference_id,
species_chromosome_no, reference_chromosome_no),
constraint fk_accgcr1 foreign key (species_id, run_id,
contig_id, codingregion_id, gene_id) references
aau_contig_codingregion_gene(species_id, run_id,
contig_id, codingregion_id, gene_id),
constraint fk_accgcr2 foreign key (reference_id)
references aau_reference (reference_id) );

```

```

create table aau_contig_cds_protein (
id number(10),
species_id number(10),
run_id number(2),
contig_id number(10),
codingregion_id number(10),
protein_id number(10),
sequence clob,
constraint pk_accp primary key(species_id, run_id,
contig_id, codingregion_id, protein_id),
constraint fk_accp1 foreign key(species_id, run_id,
contig_id, codingregion_id) references
aau_contig_codingregion(species_id, run_id, contig_id,
codingregion_id));

```

```

create table aau_cellular_type (
cellular_type_id number(2),
cellular_type_name varchar2(100),
constraint pk_acty primary key(cellular_type_id)
);

```

```

create table aau_contig_protein_cellular_type (
id number(10),
species_id number(10),
run_id number(2),
contig_id number(10),
codingregion_id number(10),
protein_id number(10),
cellular_type_id number(2),
reliability_perc number(6,3),
constraint pk_accpcty primary key(species_id, run_id,
contig_id, codingregion_id, protein_id,
cellular_type_id),

```

```

constraint fk_accpcty1 foreign key (species_id, run_id,
contig_id, codingregion_id, protein_id) references
aau_contig_cds_protein(species_id, run_id, contig_id,
codingregion_id, protein_id),
constraint fk_accpcty2 foreign key (cellular_type_id)
references aau_cellular_type(cellular_type_id)
);

```

```

create table aau_contig_codingregion_protein_reference (
id number(10),
species_id number(10),
run_id number(2),
contig_id number(10),
codingregion_id number(10),
protein_id number(10),
reference_id number(10),
match_by_blast varchar2(100),
identity_score number(6,3),
positive_score number(6,3),
residue_match number(3),
no_of_mismatch number(3),
gap number(3),
species_start_site number(10),
species_stop_site number(10),
reference_start_site number(10),
reference_stop_site number(10),
error_value varchar2(10),
score number(10),
protein varchar2(100),
description varchar2(100),
constraint pk_accpr primary key(species_id, run_id,
contig_id, codingregion_id, protein_id, reference_id),
constraint fk_accpr1 foreign key (species_id, run_id,
contig_id, codingregion_id, protein_id) references
aau_contig_cds_protein(species_id, run_id, contig_id,
codingregion_id, protein_id),
constraint fk_accpr2 foreign key (reference_id)
references aau_reference(reference_id)
);

```