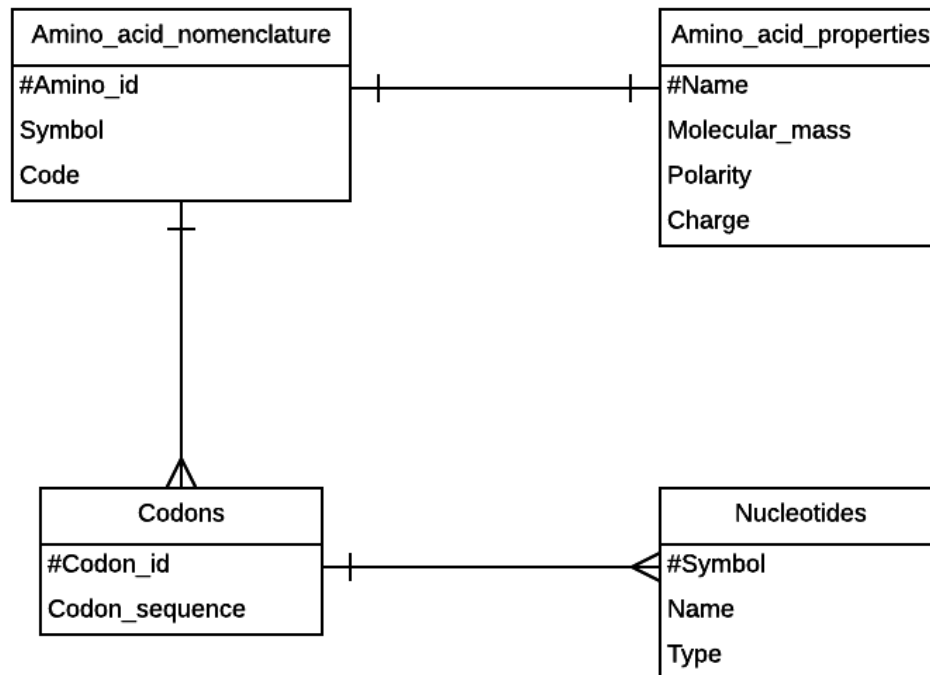INF115 – Assignment Two

1. One Nomenclature can have one set of properties, one Nomenclature can have many Codons and one Codon can have many Nucleotides.
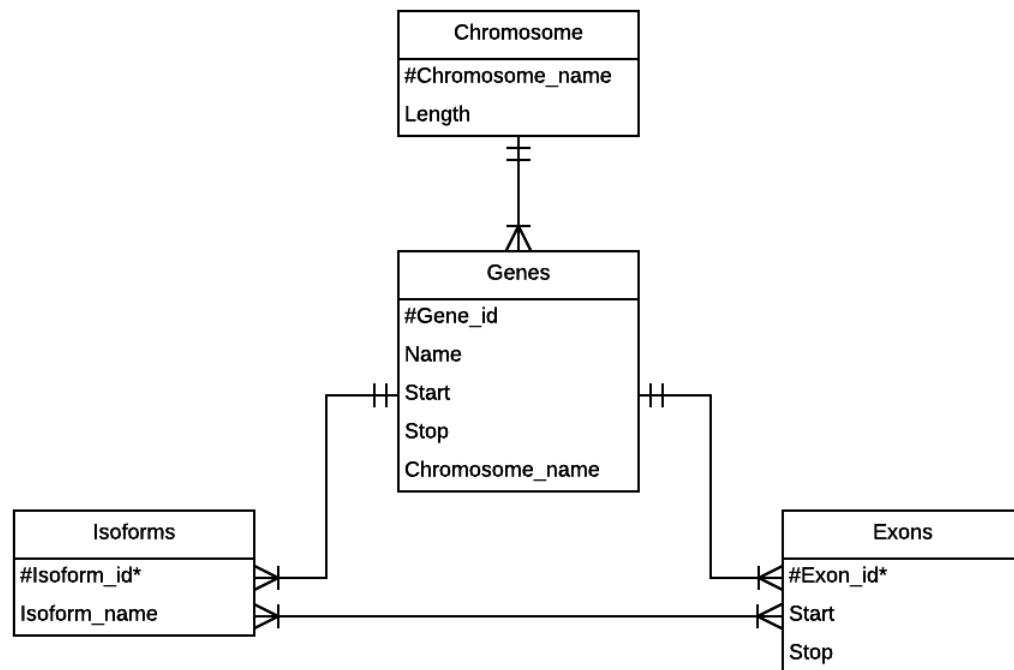
Amino_acid_nomenclature
#Amino_id
Symbol
Code

Amino_acid_properties
#Name
Molecular_mass
Polarity
Charge

Codons
#Codon_id
Codon_sequence

Nucleotides
#Symbol
Name
Type

2.

1. Entities:
   -Genes
   -Exons
   -Isoforms
   -Chromosome

2. ER-Diagram:

Chromosome
#Chromosome_name
Length

Genes
#Gene_id
Name
Start
Stop
Chromosome_name

Isoforms
#Isoform_id*
Isoform_name

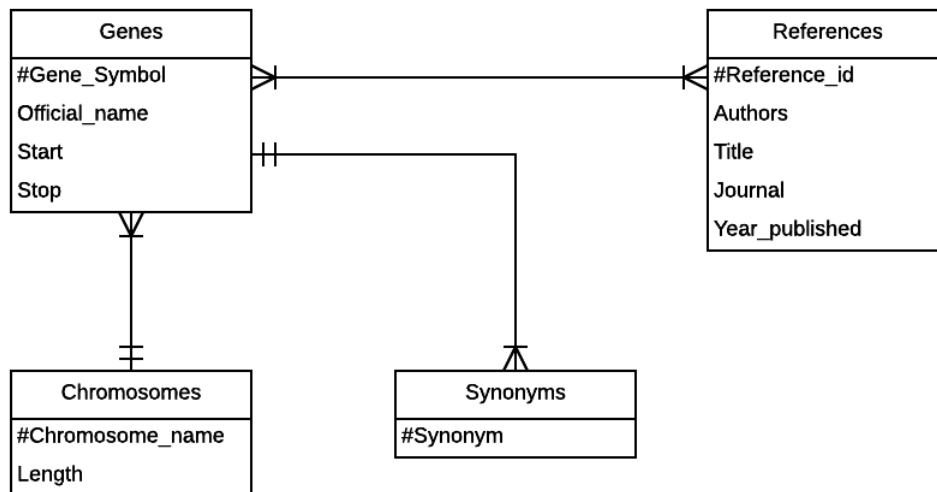**Exons**
#Exon_id*
Start
Stop

3. DB in Third Normal Form:

Chromosome(#Chromosome_name, Length)
Genes(#Gene_id, Name, Start, Stop, Chromosome_name*)
Exons(#Exon_id, Start, Stop, Gene_id*)
Isofom(#Isoform_id, Isoform_name, Gene_id*)
Isoform_exons(#Isoform_id*, #Exon_id*)

3.
1. Entities:
   -Chromosomes
   -Genes
   -Synonyms
   -References

2. ER-Diagram:

3. First Normal Form: Breakes second normal form since Chromosome_Name is not dependent on Gene_symbol.

Synonyms(#Synonym, Gene_symbol*)

Genes(#Gene_symbol, Chromosome_name, Chromosome_length, Official_name, Start_coordinate, Stop_coordinate, Reference_id*, Year_published)

References(#Reference_id, Authors, Title, Journal)

4.
Chromosome(#Chromosome_name, Chromosome_length)
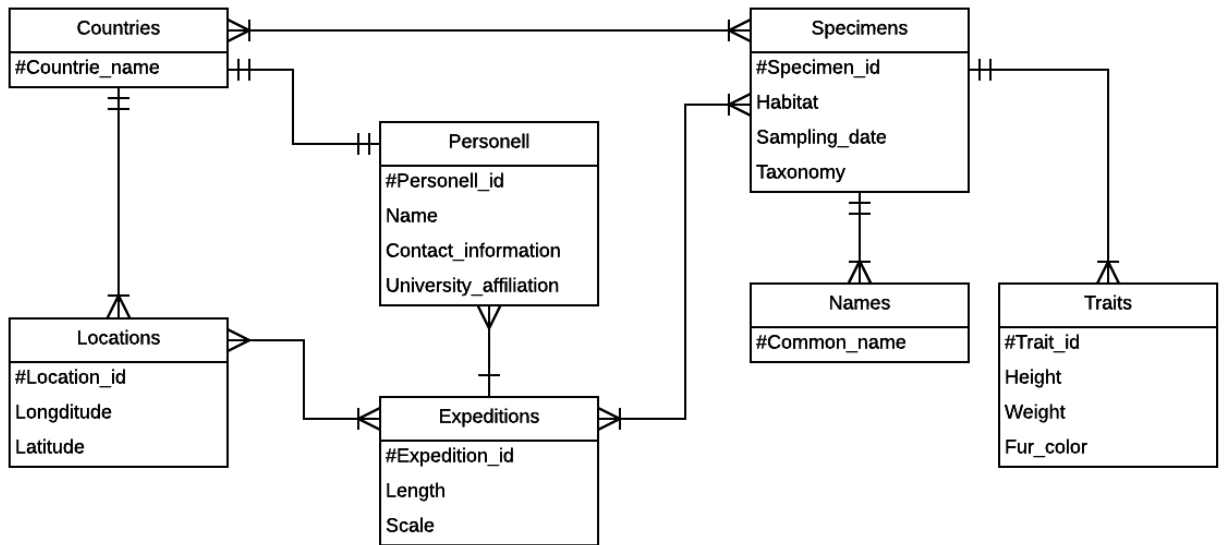
Synonyms(#Synonym, #Gene_symbol*)

Gene(#Gene_symbol, Chromosome_name*, Official_name, Start_coordinate, Stop_coordinate)

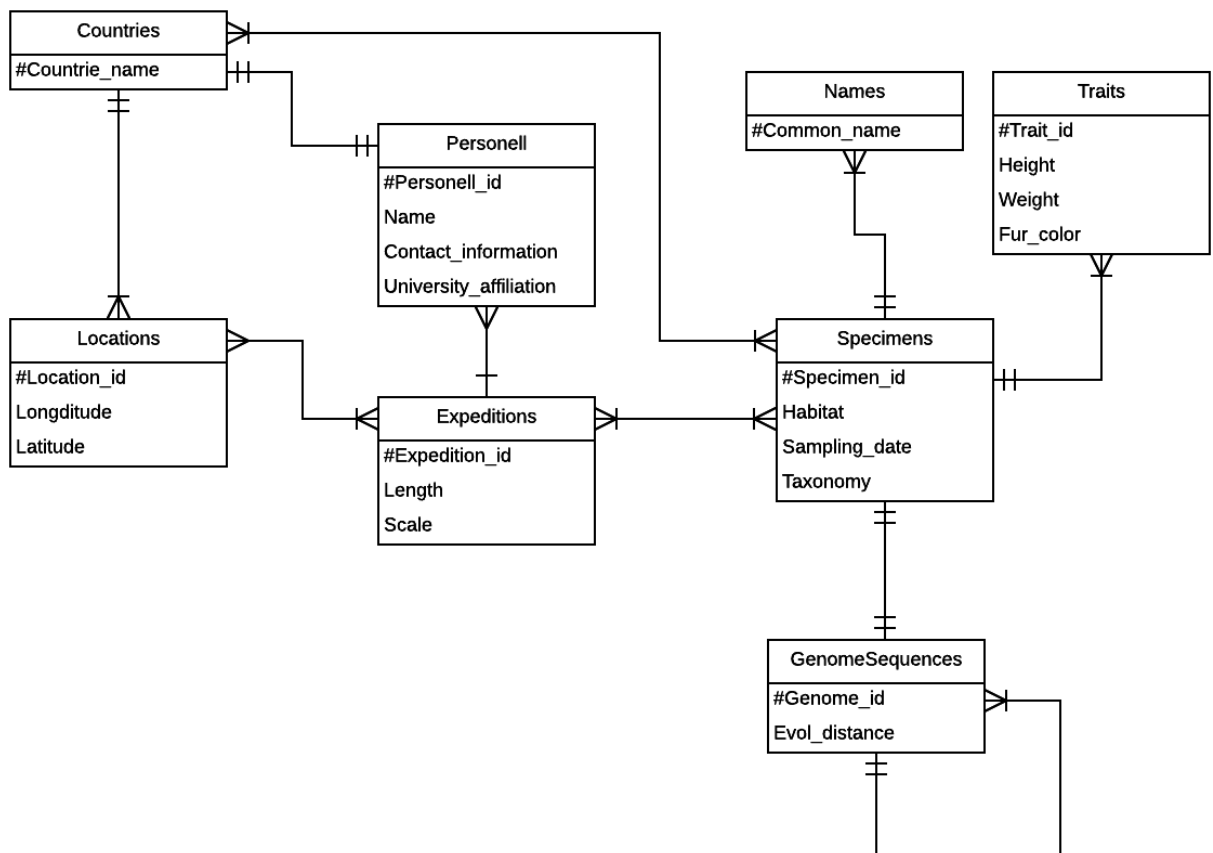GeneReference(#Gene_symbol*, #Reference_id*)

Reference(#Reference_id, Authors, Title, Journal, Year_published)

4.

1.

2. Two

5.
1. Errors
   1. Dependency error in Patient: Where postcode is dependent on Region, not PatientID
   2. Patient can only be in one location at one time so PatientToLocation is not needed.
   3. The highest normalization level these tables upholds are 2NF

2. There will still be dependencies on dependencies. So BCNF is not fulfilled.
   Here featureId will be dependent on OutbreakId and OutbreakID is dependent on
   GenomeID. There exists transitive dependencies. And the same value for OutbreakIDand
   FeatureID will appear many times duplicates.
   Not all values are directly dependent on the primary key

3. Functional Dependencies: x is dependent on y
   1. OutbreakId > GenomeId
   2. FeatureType > FeatureID

4. Candidate Keys: GenomeID, PatientID, OutbreakID, FeatureID

5. Normalization:
   1. First Normal Form:
      System (#PatientID, FirstName, SurName, Postcode, Region #SampleID, SampleDate,
      #TestID, TestName, ResitantToAntibiotic, #LocationID, LocationName, #GenomeID,
      OutbreakID, FeatureID, FeatureType, AntibioticName)

   2. Second Normal Form
      Patient(#PatientID, FirstName, SurName, Postcode, Region)

      Sample(#SampleID, SampleDate, PatientID*)

      Labtest(#TestID, TestName, SampleID*, ResitantToAntibiotic)

      HospitalLocation(#LocationID, LocationName)

      PatientToLocation(#PatientID*, #LocationID*)

      GenomeSequence(#GenomeID, PatientID*, OutbreakID, FeatureID, FeatureType,
      AntibioticName)

3. Third Normal Form

    Patient(#PatientID, FirstName, SurName, Postcode*)

    Region(#Postcode, Region)

    Sample(#SampleID, SampleDate, PatientID*)

    Labtest(#TestID, TestName, SampleID*, ResitantToAntibiotic, OutbreakID)

    HospitalLocation(#LocationID, LocationName)

    PatientToLocation(#PatientID*, #LocationID*)

    GenomeSequence(#GenomeID, PatientID*, OutbreakID, FeatureID*)

    Feature(#FeatureID, FeatureType, AntibioticName)

4. BCNF

    Same as Third Normal Form.