

Objects in DNA analysis

Nasser M. Abbasi

May 16, 2000

This is a description of objects used in DNA/Genomics software development. I need to first come up with such a list, and the relationship between them to base any further software development of a sw library. I think I will write the library in Java.

For each object, it is described, and what other objects it contains, and what attributes the objects has.

The idea is that each such object, will be represented as a class. The object attributes will be used to define the access methods of the object (messages on the object).

I might even one day use these object to write a molecular life simulation program.

1. Chromosome.
contains: contig objects. one DNA Molecule
Attributes:
 - (a) Enochromatic size, in kb.
 - (b) number of contigs it contains.
 - (c) number of bases.
 - (d) Name or ID of chromosom (1, 2, ... Y, X).
2. contig.
contains: physical position object, cytogenetic position object.
Attributes:
 - (a) Name or ID. such as "NT_00039"
 - (b) size (in kb in bases).
3. physical position of contig in the chromosom.
Attributes (these are estimates from RH map positions):
 - (a) start position. a number
 - (b) end position. a number.
4. Cytogenetic position of contig in the chromosom.
Attributes (based on annotation from GenBank entry):
 - (a) start id. such as "1p34.3"
 - (b) end id. such as "p36.13"
5. DNA molecule (or DNA sequence)
contains number of Gene objects, contains one sequence of DNA base pairs objects. attributes:

- (a) number of genes.
- (b) number of base pairs.

6. Gene

contains number of exons and number of introns.

Attributes:

- (a) number of exons.
- (b) number of introns.

7. exon. a region in a gene that is a coding region for protein.

Contains number of codons, and a DNA sequence. Attributes:

- (a) size in bases.
- (b) number of codons.
- (c) start position in gene
- (d) end position in gene.

8. intron. a region in a gene that is a not a coding region for protein.

Contains a DNA sequence Attributes:

- (a) size in bases.
- (b) start position in gene
- (c) end position in gene.