5. Phylogenetic trees

- The tree of life

- The tree, an abstract object

- Building an array of distances

- The UPGMA algorithm

- Differences are not always what they look like

- **The diversity of bioinformatics algorithms**

- The application domains in microbiology
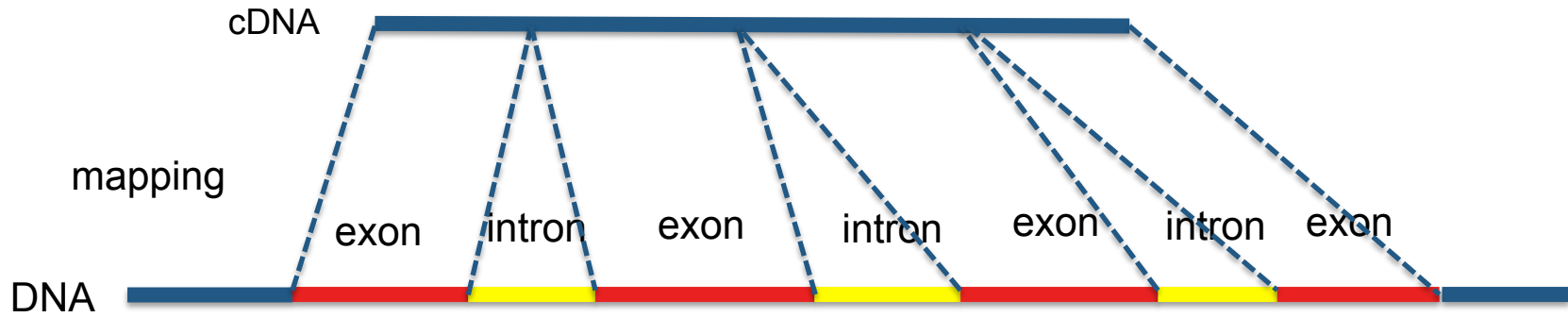
Multiple classes of problems

- **Read assembly**
- Sequence mapping and comparison
- Detection of repeated sequences
- Gene prediction
- Simple and multiple sequence alignments
- Phylogenetic tree reconstruction
- Structure prediction of
  - RNA
  - Proteins
- Network analysis and simulation
- …

```
CGCTGGAACGG
     AACGGCTAGTTA
          TAGTTAGAA
CGCTGGAACGGCTAGTTAGAA
```

Multiple classes of problems

- Read assembly
- **Sequence mapping and comparison**
- Detection of repeated sequences
- Gene prediction
- Simple and multiple sequence alignments
- Phylogenetic tree reconstruction
- Structure prediction of
  - RNA
  - Proteins
- Network analysis and simulation
- …

cDNA

mapping

DNA

exon   intron   exon   intron   exon   intron   exon

Multiple classes of problems

- Read assembly
- Sequence mapping and comparison
- **Detection of repeated sequences**
- Gene prediction
- Simple and multiple sequence alignments
- Phylogenetic tree reconstruction
- Structure prediction of
  - RNA
  - Proteins
- Network analysis and simulation
- …

Multiple classes of problems

- Read assembly
- Sequence mapping and comparison
- Detection of repeated sequences
- **Gene prediction**
- Simple and multiple sequence alignments
- Phylogenetic tree reconstruction
- Structure prediction of
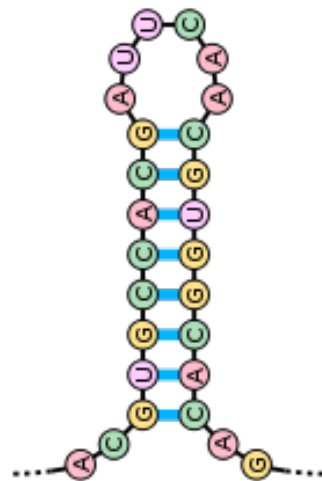 - RNA
  - Proteins
- Network analysis and simulation
- …

Multiple classes of problems

- Read assembly
- Sequence mapping and comparison
- Detection of repeated sequences
- Gene prediction
- **Simple and multiple sequence alignments**
- Phylogenetic tree reconstruction
- Structure prediction of
  - RNA
  - Proteins
- Network analysis and simulation
- …

```
HglK        78  117   CDLRFANLEGADLTDANLIAASLHKSNLRRANLCRATLNR
HglK       118  157   CNLSEADLTESDANEALFCQAVFTEVEAHGLRLYRAKVSQ
HglK       158  197   AQLMGAHLHQAYAPEADFSAVAAIAVDLRWANLRKTNFRG
HglK       203  242   GNFRGANLTQADFTGANLKGANLRGANLVGTNLQRADLSD
spkB       454  493   QNLVGIVLAKAFVPGINCYQANLTNANFEQAELTRADFGK
spkB       504  543   ANLSDAYFGYADLRGADLRGANLNGVNFKYANLQGANFSG
slr1819     20   59   LKLPGINLEAADLIGIVLNEADLRGANLLFCYLNRANLGQ
slr1819     60   99   ANLVAANLSGASLNQADLAGADLRSANFHGAMLQGAILRD
slr1819    100  139   SDMTLATLQDTNLIGADLRGADLSGATLTGACLRGANMRQ
slr1819    151  190   AILGRADLQGANMKGVDLSRADLSYANLKEANLRDVDLRK
slr1819    191  230   ADLSYANLKGALLTDANLSGAKLNGADLQNANLMRAKISE
slr1819    231  270   AEMTAVNCQGAIMTHVNLNRTNLTGSNLSFTRMNSADLSR
slr1819    271  310   ANLTKANLQEAELIEAFFARANLTEANFINANLVRADLMS

consensus             A L   A L   A L   A L   A L   A L   A L   A L
```
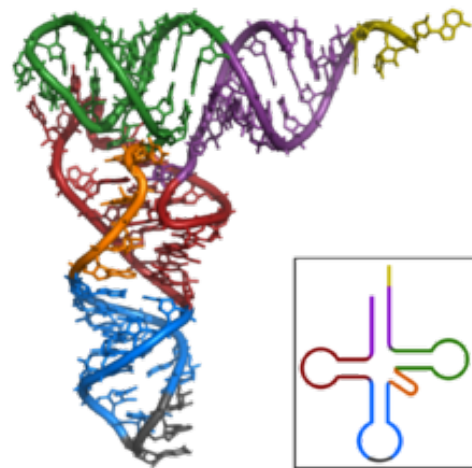
8

Multiple classes of problems

- Read assembly
- Sequence mapping and comparison
- Detection of repeated sequences
- Gene prediction
- Simple and multiple sequence alignments
- **Phylogenetic tree reconstruction**
- Structure prediction of
  - RNA
  - Proteins
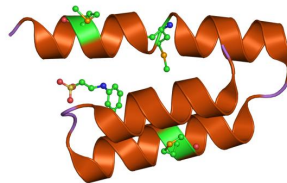- Network analysis and simulation
- …

Multiple classes of problems

- Read assembly
- Sequence mapping and comparison
- Detection of repeated sequences
- Gene prediction
- Simple and multiple sequence alignments
- Phylogenetic tree reconstruction
- **Structure prediction of**
  - **RNA**
  - Proteins
- Network analysis and simulation
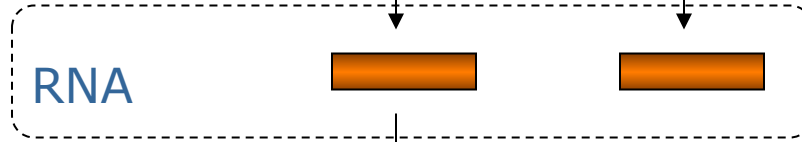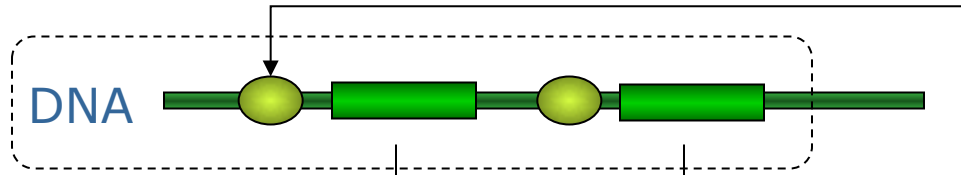- …

Multiple classes of problems

- Read assembly
- Sequence mapping and comparison
- Detection of repeated sequences
- Gene prediction
- Simple and multiple sequence alignments
- Phylogenetic tree reconstruction
- **Structure prediction of**
  - **RNA**
  - Proteins
- Network analysis and simulation
- …

Multiple classes of problems

- Read assembly
- Sequence mapping and comparison
- Detection of repeated sequences
- Gene prediction
- Simple and multiple sequence alignments
- Phylogenetic tree reconstruction
- **Structure prediction of**
  - RNA
  - **Proteins**
- Network analysis and simulation
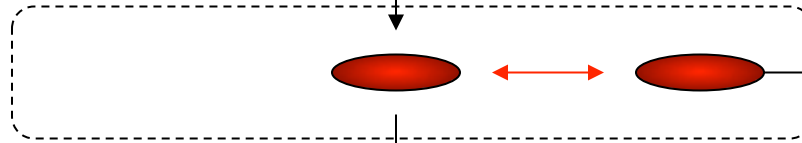- …

Multiple classes of problems

- Read assembly
- Sequence mapping and comparison
- Detection of repeated sequences
- Gene prediction
- Simple and multiple sequence alignments
- Phylogenetic tree reconstruction
- Structure prediction of
  - RNA
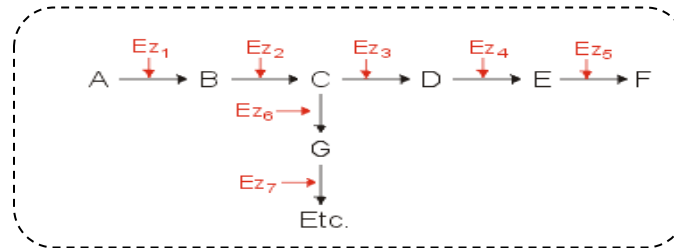 - Proteins
- **Network analysis and simulation**
- …

Genes

DNA

RNA

Regulation

Proteins

Enzymes

Metabolic reactions
and pathways

Ez_1 Ez_2 Ez_3 Ez_4 Ez_5

A $\rightarrow$ B $\rightarrow$ C $\rightarrow$ D $\rightarrow$ E $\rightarrow$ F

$Ez_6 \rightarrow$

G

$Ez_7 \rightarrow$

Etc.

Multiple classes of problems

- Read assembly
- Sequence mapping and comparison
- Detection of repeated sequences
- Gene prediction
- Simple and multiple sequence alignments
- Phylogenetic tree reconstruction
- Structure prediction of
  - RNA
  - Proteins
- Network analysis and simulation
- …

Material licensing