

課題 0 : 準備課題

マイコプラズマ菌の全 DNA 配列ファイルから、各自に割り当てられた部分配列を切り出して別のファイルにする。

0-1 : ファイル

以下のディレクトリに様々なバクテリアのゲノム配列が格納されている :

```
/pub/sfc/dnadb/genomes/bacteria/
```

その中の

```
Mgen/mgen.gb
```

というファイルにマイコプラズマ (*M.genitalium*) の遺伝子の塩基配列 (約 0.6 メガ) が格納されている。

0-2 : 特定遺伝子配列の切り出し

例えば割り当ての列が以下の場合、

```
127178 - 128811 , 237246 - 239251
```

127178 文字目から 128811 文字目までと、237246 文字目から 239251 文字目までを切り出して 2 つのファイルを作る。

切り出しの例 : 100 文字目から 200 文字目までを切り出す場合

```
1 taagttatta ttagttaat actttaaca atattattaa ggtatttaa aaatactatt
61 atagtattta acatagttaa ataccttcct taatactggt aaattatatt caatcaatac
121 atatataata ttattaaaat acttgataag tattatttag atattagaca aatactaatt
181 ttatattgct ttaacttta ataaacta cttatgtatt aagtaaatat tactgtaata
241 ctaataacaa tattattaca atagctaga ataattttgc tagtatcaat aattactaat
```

emacs 等のテキストエディタを駆使して、100 文字目から 200 文字目まで、

```
tcaatcaatacatatataatattattaaaacttgataagattatttagatat tagacaaactaattttatatgtcttaacttta
```

これを 1 つのファイルに入れる。

今回の皆さんの割当は以下の通り。

```
1 つ目の配列      2 つ目の配列。
11152 - 12140 , 136079 - 137367
```

課題 1 : 塩基使用頻度解析

基本課題

1-1: 担当遺伝子配列をファイルから読み込み、その A,T,G,C 塩基の数をカウントするプログラムを書く。

1-2: マイコプラズマ菌全ゲノムファイルから配列を読み込めるようにプログラムを変更し、担当遺伝子配列の塩基使用頻度に偏りがなかったか考察する。

ヒント 2: 文字列どうしを比較するには `==` ではなく `eq` を用いる。

応用課題

1-3: 二連続塩基 (dinucleotide、16 種類) の頻度解析を行うようにプログラムを変更する。マイコプラズマ全ゲノム DNA 配列の二連続塩基の頻度パターンを調べ、1-2 の結果からの期待値と比較、考察する。