**ゲノム情報解析基礎（2019年5月13日）課題（提出期限2019年5月27日）**

受講ID（5桁）：

学生証番号（ない人は空欄で可）：

名前：

**課題1：DFASTへの入力として与えたout\_gapClosed.faの配列数は117個だったが、 実行結果として配列数が61個となった。この理由について考えを述べよ。DFAST Help画面内の記載事項をもとに述べてもよいし、R上で検証した結果に基づいて述べてもよい。**

**課題2：以下の中から興味ある事柄を1つ選び、簡単に説明せよ。**

* Sequence logosの計算手順について（情報屋さん向け。内部で何を計算しているのかなど）。

http://www.iu.a.u-tokyo.ac.jp/~kadota/20140430\_kadota.pdf

のスライド30-35あたりにも記載あり。

* GC skewについて（バクテリア解析屋さん向け）。

乳酸菌NGS連載第9回のp3-4あたりにも記載あり。

* 配列のドットプロット（比較ゲノム屋さん向け）。

乳酸菌NGS連載第7回のp106や、第13回のp39-41あたりにも記載あり。

* GFF3ファイルannotation.gffからのCDS情報抽出手段について（チャレンジャー向け）

様々な戦略が考えられます。例えば、makeTxDbFromGFF関数が読み込める形式にすべく、annotation.gffの中身をいじる方向性。他には、annotation.gffの中身は変更せずに最終的にうまく2,311個のCDS領域情報を読み込んでTxDbオブジェクトを作成する方向性。GFF3ファイルを読み込める別の関数を利用する手もあるかもしれません。

**感想やコメントなど（optional）：**