



第2回 2013年10月18日(金) 14:50~15:40

## fasta,fastq data processing with Perl

Perlによるfastq,fasta形式データの処理

Paul Horton / ポール・ホートン

Computational Biology Research Center (CBRC)

National Institute of Advanced Industrial Science and Technology (AIST)

産業技術総合研究所 生命情報工学研究センター

In general, it is not possible to simply forward data output from sequencers to pre-existing software analysis packages. Usually simple but ad hoc tasks, such as extraction of reads with high quality or trimming of Poly-A tails from transcripts is necessary. In this lecture I will illustrate some usefule Perl techniques while showing how to use the tool fastapl/fastqpl to perform typical sequence processing tasks with 1-3 lines of Perl code. In addition I will use fastapl as and example to describe how to prepare a debian software package and increase computation speed of Perl program via XS.

Keywords: fasta, fastq, sequence processing, ad hoc processing, program generation

次世代シーケンサーから出力されるデータが、そのまま既存の解析ソフトウェアに渡せることが少ない。信頼度の高いリードだけを抽出したり、転写配列からPoly-A末端を除いたりするなど、比較的単純ではあるが、応用ごとに異なる処理を行うことが多い。本講義では、Perl言語の基礎知識を交えながら、fastaとfastq形式のデータ処理を1~3行程度のPerlコードで技術できるツール、fastapl/fastqplを紹介する。また、fastaplを例にして、ソフトウェアツールのdebianパッケージ化と、XSによるPerlプログラムの計算速度の最適化について解説する。

キーワード : fasta, fastq, 配列処理、単発的処理、プログラム自動生成