

スギゲノムの解読と解析

Genome sequencing and analysis of Japanese cedar

東京大学新領域創成科学研究科メディカル情報生命専攻 笠原 雅弘

研究成果概要

本研究では、京都大学化学研究所スーパーコンピュータシステムを利用し、スギゲノム配列の解読と解析を行った。著者の本システム利用の大部分は本システム上で解析を行っている共同研究者とのデータ共有であり、独自の解析はあまり行っていない。データサイズが非常に大きく、パソコンのローカルストレージにダウンロードできないサイズのデータ共有や解析データの統計生成に主に用いていた。

以下、独自解析部分について述べる。Oxford Nanopore Technologies 社が発売している PromethION DNA シークエンサーの試薬が高性能化した。新しい試薬は R10 と呼ばれており、塩基精度が R9 世代のものと比べて大きく上がっていることが謳われている。このため、R10 世代の出力配列がゲノムアセンブリ等の問題にどのように適用できるか(スギではない DNA 配列を用いてではあるが)予備解析を行った。

R10 世代の PromethION シークエンサーの出力を minimap2 を用いて参照ゲノムにマッピングし、samtools と whatshap を用いて一塩基多型の発見とフェージング解析を行った。塩基精度が PacBio HiFi リードに近いこと、PacBio HiFi リード向けのオプションをそのまま用いて解析を行った。この結果、R10 世代の PromethION リードは PacBio HiFi リードと同等の解析オプションとパイプラインでも基本的には解析が可能であることを確認した。ただし、PromethION リードの特徴に合わせたチューニングを行うことで将来的には更なる解析精度向上ができるであろうことが推察された。また、R10 世代の試薬は頻繁にアップデートが行われていたため、今後の Nanopore リードの活用とそのパフォーマンス予測のためには kit14 などのより新しい試薬によりパフォーマンス解析を行っていく必要があると考えられる。

発表論文(謝辞あり)

Fujino, Takeshi, Yamaguchi Katsushi, Toshiyuki T. Yokoyama, Toshiya Hamanaka, Yoritaka Harazono, Hiroaki Kamada, Wataru Kobayashi, et al. 2023. “A Chromosome-Level Genome Assembly of a Model Conifer Plant, the Japanese Cedar, *Cryptomeria Japonica* D. Don.” *bioRxiv*. <https://doi.org/10.1101/2023.02.24.529822>.