

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

藤野 健

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、スギゲノムの解読と解析に取り組んだ。

スギのゲノムサイズはヒトゲノム(約 3 Gb)に比べてはるかに巨大であり、フローサイトメトリーを用いた既存の解析で約 11 Gb という推定が得られていた。また、スギゲノムは極めて頻度の高い反復配列を持つことが予備解析から判明している。このため、スギゲノムの解析には大きな計算資源を必要とする。

我々はスギから高分子 DNA を抽出し、PacBio HiFi リードを得ることでゲノムアセンブリを行った。この結果、N50 で約 12Mb, 合計で約 9 Gb のコンティグを得ることができた。さらに Omni-C 技術によって構築した Illumina ライブラリをシーケンシングした結果と合わせることで、コンティグの 97 % を 11 本の染色体上に配置することができた。

本年度は、このアセンブリ結果を査読論文として発表するにあたり、京都大学化学研究所スーパーコンピュータシステムを用いて解析を行った。具体的には HiFi リードに対して Meryl による k-mer 数集計を行い、これをもとに Merqury によるゲノムのホモ接合度推定およびアセンブリの完全性の評価を行った。また、遺伝子アノテーションのフィルタリング基準について説明する図表を論文に掲載するため、計算結果の集計および作図を行った。

発表論文(謝辞あり)

Fujino, Takeshi, et al. "A chromosome-level genome assembly of a model conifer plant, the Japanese cedar, *Cryptomeria japonica* D. Don." *BMC genomics* 25.1 (2024): 1039.