

スギゲノムの解読と解析

Genome sequencing and analysis of Japanese cedar

東京大学 大学院新領域創成科学研究科 メディカル情報生命専攻 笠原研究室 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。

スギ花粉症は日本で大きな社会問題となっており、その解決は喫緊の課題である。特に、高度成長期に植えられ、花粉を大量に放出している(伐採にはちょうど適した時期の)スギ人工林を伐採し新たに植林を行う(条件が悪い土地ではあるいは植えない)ことが重要である。しかし、輸入材と比べた場合に日本林業の商業的な採算性は厳しく、特に私有地のスギ人工林の植え替えを促進するには多額の補助金を注ぎ込む必要があり、財政事情が厳しい政府あるいは地方自治体では二の足を踏んでいた。

また、スギ花粉を出さないいわゆる無花粉スギ(雄性不稔スギ)はスギ花粉症への抜本的対策として期待されているが、自然変異体を利用しているため原因遺伝子は不明であり、各地の気候に適し商業的に優れた特性を持つ様々な雄性不稔スギを大量生産することはできない。

これらの課題を解決するための基礎研究として我々はスギゲノム配列を解読している。スギゲノム(約11 Gb)はヒトゲノム(約3 Gb)の3.8倍の大きさを持ち、予備解析の結果、ヒトゲノムより遙かに解読が難しいゲノム構造(極めて頻度の高い反復配列)を持っていることが分かっており、ゲノム解読には予測不能な極めて大きな計算量を必要とする。

昨年度までは主として Oxford Nanopore Technologies のシーケンサーから得られた配列データを用いてゲノムアセンブリを試みていたが、新たに PacBio の HiFi リードが得られたことから、HiFi リードを用いたゲノムアセンブリの研究に取り組み始めた。HiFi リードはリード長に関して Oxford Nanopore Technologies のリードより短い傾向にあるがシーケンシングの精度が高いことからアセンブリの品質向上が期待できる。今後は複数のシーケンシング技術の結果を合わせたアセンブリ手法の研究なども進めていく予定である。

合わせてアセンブリ結果の評価手法についても開発を進めている。マーカーによって得られた連鎖地図とアセンブリとの整合性を評価するプログラムを開発したが、さらに評価手法を洗練させていく予定である。