

スギの基盤遺伝情報の高精度化と 有用遺伝子の機能解明

スギの全染色体を網羅する塩基配列を解読し、約 5 万個の遺伝子とその位置を特定して、種を代表する標準配列「参照ゲノム配列」を構築しました。



背景と目的

スギは青森から鹿児島まで天然に分布するヒノキ科の常緑針葉樹です。加工のしやすさから古くから幅広く活用され、積極的に植林されてきました。しかし、花粉を大量に飛散するようになり花粉症が大きな社会問題になっています。無花粉品種の開発に向け、スギのゲノムの塩基配列を解読し、無花粉をもたらす遺伝子の解明が求められています。

スギはイネの20倍以上という巨大で、複雑なゲノムを持つため、その配列解読は困難でした。また、個体も寿命が長く、巨大なため、人工交配による遺伝的な解析にも時間と労力がかかっていました。しかし、近年のゲノム解析技術の進歩によって、巨大なゲノムであってもその解読が可能になってきました。

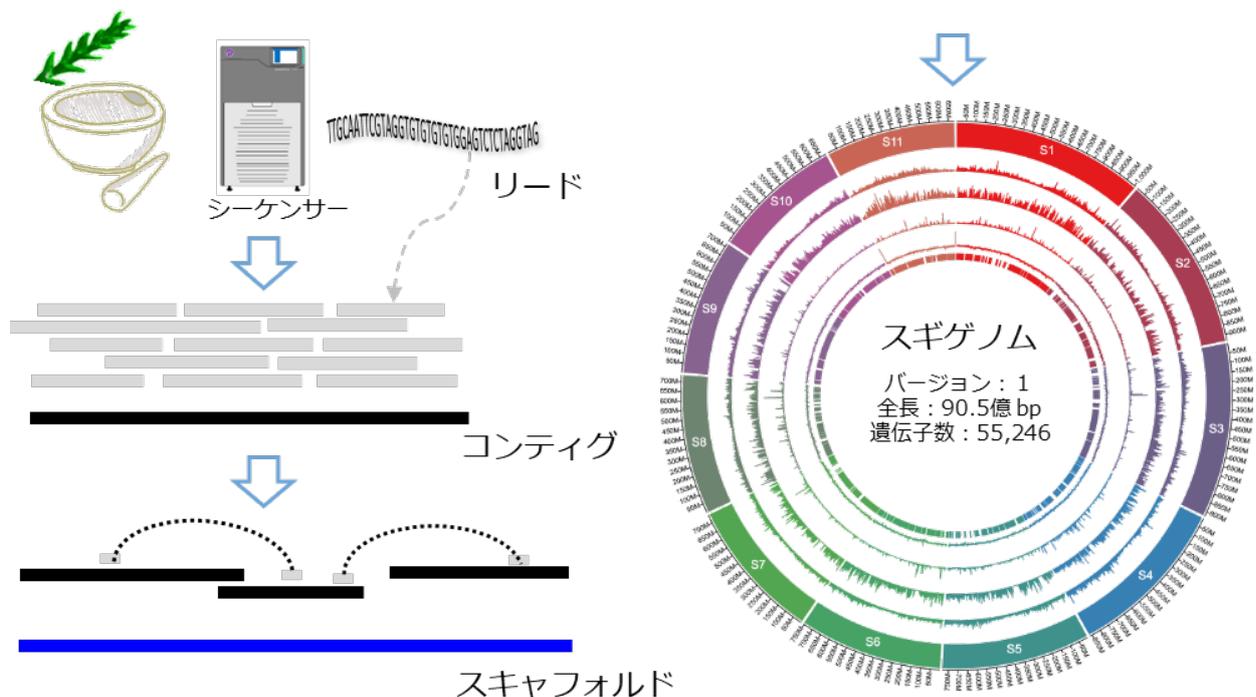
ゲノム解読には自家受精した個体が使われた

スギは種子親（母親）と花粉親（父親）から受け取った2セットのゲノムを持っています。自家受精を繰り返すことによって両セットのゲノムをほぼ同一の塩基配列に近づけることができれば、解読に必要なデータ量を約半分に減らすことができます。そこで自家受精を3回繰り返したところ、2セットのゲノムのほとんど（96%）の塩基配列が同一と推定される個体が得られました。このスギを使ってゲノムを解読しました。

染色体の全体像が塩基配列として初めて明らかになった

複雑なゲノムを解読するには、ひと続きで長い（高分子量の）DNAを抽出する必要があります。そこでスギの若葉から高分子量DNAを抽出する方法を開発し、平均で50,000塩基対を上回る長さのDNAを抽出してゲノム解読に使用しました。その結果、平均長16,960塩基対の約1,888万本の読取配列（リード）からなる総計3,166億塩基対のデータを収集できました（図）。これはスギのゲノムサイズの28.8倍の量に相当し、ゲノム全体が網羅されていると考えられました。これらのリードを2,650個の部分配列（コンティグ）にまとめました（図）。

生物のゲノムは、染色体というひとつながりのDNA配列でできています。その配列を得るため、近接したコンティグをつなげる方法を検討しました。ゲノムDNAは細胞核の中で折りたたまれて存在するため、その状態で隣り合っているDNAは近接したコンティグである可能性が高くなります。この情報を用いて、コンティグを連結することに成功し、長大な11本の連結配列（スキュフォールド）が得られました（図）。これが、スギの11本の染色体に相当する「参照ゲノム配列」となります。



図：（左）高分子量DNAを抽出し、リードと呼ばれるひと続きの塩基配列（灰色の線分）を読み取りました。それらのリードをつなぎあわせてコンティグと呼ばれる部分配列（黒い線分）を得ました。近接したコンティグを連結して、染色体全体を網羅するスキュフォールドと呼ばれる連結配列（青い線分）を構築しました。（右）実際に解読されたスギゲノムを環状に表現した図です。内側の環から遺伝マーカーの密度、塩基配列の組成（グアニンとシトシンの含量%）、配列の不確かさの割合（%）、繰り返し配列の密度、遺伝子の密度、染色体番号（S1~S11）、および物理距離（100万塩基対）を表します（Fujino et al. 2023を改変）。

スギゲノムの大半は「繰り返し配列」だった

完成したスギの参照ゲノム配列は90.5億塩基対で、ヒトゲノムの約3倍の大きさに相当します。タンパク質を合成するための情報を持っているDNA領域（遺伝子）をこの参照ゲノム配列から探索したところ、植物に存在するはずの遺伝子の91.4%を含んだ55,246個を特定しました。これは既往の針葉樹研究の最高値89.4%を上回る値です。

また、タンパク質を合成する情報を持っていない部分（ゲノム全体の99.4%）の配列を解析した結果、ゲノム全体の83.6%が繰り返し配列であることが分かりました。繰り返し配列は、それ自身の配列をコピーして増えてきたと考えられ、未知の機能をもっていると言われていています。スギの遺伝子の数は他の植物とあまり変わりませんが、このような繰り返し配列が増えることによりゲノムが巨大になってきたと考えられます。

成果の利活用

今回構築した参照ゲノム配列により、雄性不稔など特定の性質に関する遺伝子を迅速に単離できるようになるため、有用な品種の開発・育成が加速されます。さらに、ゲノムや遺伝子には進化の道筋が記録されていることから、地球規模の気候変動への適応をより正確に予測できるようになります。

要旨

スギの全染色体（11本）を網羅する塩基配列の解読に成功しました。新型の塩基配列解読装置（シーケンサー）で配列データを収集し、染色体ごとにつながった約91億塩基対を解読しました。さらに55,246個の遺伝子を同定し、針葉樹では最も完成度が高いゲノム解読を達成しました。また、ゲノム配列の83.6%を「繰り返し配列」として特定しました。これらのゲノム情報をForestGENで公開したことにより、スギの進化の歴史の推定や有用な遺伝子の同定が加速します。

Fujino et al. (2023) A chromosome-level genome assembly of a model conifer plant, the Japanese cedar, *Cryptomeria japonica* D. Don. bioRxiv, doi: <https://doi.org/10.1101/2023.02.24.529822>

ForestGEN, https://forestgen.ffpri.go.jp/jp/info_sugi1.html

研究代表者

樹木分子遺伝研究領域 永光 輝義



プロフィール

森林植物の遺伝的変異を調べ、環境への適応進化や生物間相互作用を研究しています。

担当研究機関

森林総合研究所（樹木分子遺伝研究領域）
東京大学・基礎生物学研究所・国立遺伝学研究所・筑波大学・新潟大学

問い合わせ先 TEL 029-829-8377（相談窓口）

表紙写真：全染色体の塩基配列が解読されたスギ



ISSN 1349-0605
森林総合研究所交付金プロジェクト研究 成果 No. 97
「スギの基盤遺伝情報の高精度化と有用遺伝子の機能解明」
発行日 令和5（2023）年5月8日
発行者 国立研究開発法人 森林研究・整備機構 森林総合研究所
〒305-8687 茨城県つくば市松の里1番地
電話 029-873-3211（代表）

※本誌掲載記事及び写真の無断転載を禁じます。