

針葉樹では世界で最も完成度の高い スギ全染色体の塩基配列解読に成功

樹木分子遺伝研究領域:上野 真義、伊原 徳子、内山 憲太郎、伊津野 彩子 企画部:松本 麻子
 東京大学:藤野 健、横山 稔之、濱中 俊哉、原 蘭 陸正、鎌田 寛彬、小林 航、笠原 雅弘
 基礎生物学研究所:山口 勝司、重信 秀治 国立遺伝学研究所:豊田 敦
 筑波大学:津村 義彦 新潟大学:森口 喜成

ゲノムはあらゆる生物の設計図です。スギが属する針葉樹はゲノムが大きく(ポプラやユーカリの15~50倍)構造が複雑なため解読が困難でしたが、最新の技術でスギの全染色体(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て染色体ごとにつながった約91億塩基対を解読し、55,246個の遺伝子を同定しました。これは針葉樹では世界で最も完成度の高いゲノム解読です。また、ゲノム全体の83.6%を「繰り返し配列」として特定しました。解読した配列を標準配列「参照ゲノム配列」として公開したことで、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。

■ 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平均長50,000塩基対(bp)以上の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 bp、約1,888万本の読取配列(リード)からなる総計3,166億bpのデータを収集できました。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1)。

ゲノムは、染色体というひとつつながりのDNA配列でできています。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に存在する可能性が高いと考えられます。この情報を用いてコンティグの連結に成功し、長大な11本の連結配列(スキファールド)を得ました。これは、スギの11本の染色体に相当しました。この連結配列を、スギという種を代表する標準配列「参照ゲノム配列」としてとりまとめ、森林総合研究所のウェブサイト(ForestGEN)で公開しました。

■ スギゲノムの大半は「繰り返し配列」だった

完成した参照ゲノム配列は90.5億bp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で最高値だった89.4%を上回り、

さらに染色体上でそれらの位置が特定されたことから、針葉樹では世界で最も完成度の高いゲノム解読となります。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、また、未知の機能をもっていると言われていています。スギは、このような繰り返し配列が増えることでゲノムが巨大になってきたと考えられます。

研究資金

- ・本研究所の交付金プロジェクト1「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」、「有用遺伝子の特定に向けたスギ全ゲノム走査」、および交付金プロジェクトFS「有用遺伝子の特定に向けたスギ全ゲノム塩基配列の概要解読」
- ・科研費(JP16H06279)「先進ゲノム解析研究推進プラットフォーム」、同(JP20H03239)「植物の巨大なゲノムを解読・解析する手法」
- ・農林水産省 農林水産業・食品産業科学技術研究推進事業および生研支援センター イノベーション創出強化研究推進事業(JPJ007097)「無花粉スギの普及拡大に向けた DNA マーカー育種技術と効率的な苗木生産技術の開発(Project ID 28013B)」
- ・基礎生物学研究所共同利用研究「スギの全ゲノム配列の解読(15-829、16-403、17-405、18-408、19-420、20-428、21-302、22NIBB402)」

参照文献・サイト

- 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

専門用語

ゲノム:生物がもつ1セットの遺伝情報の総体をゲノムと呼びます。遺伝情報の最小の単位はDNAを構成する4種の塩基(アデニン、チミン、グアニン、シトシン)で、その長さは塩基対(base pair: bp)で表されます。

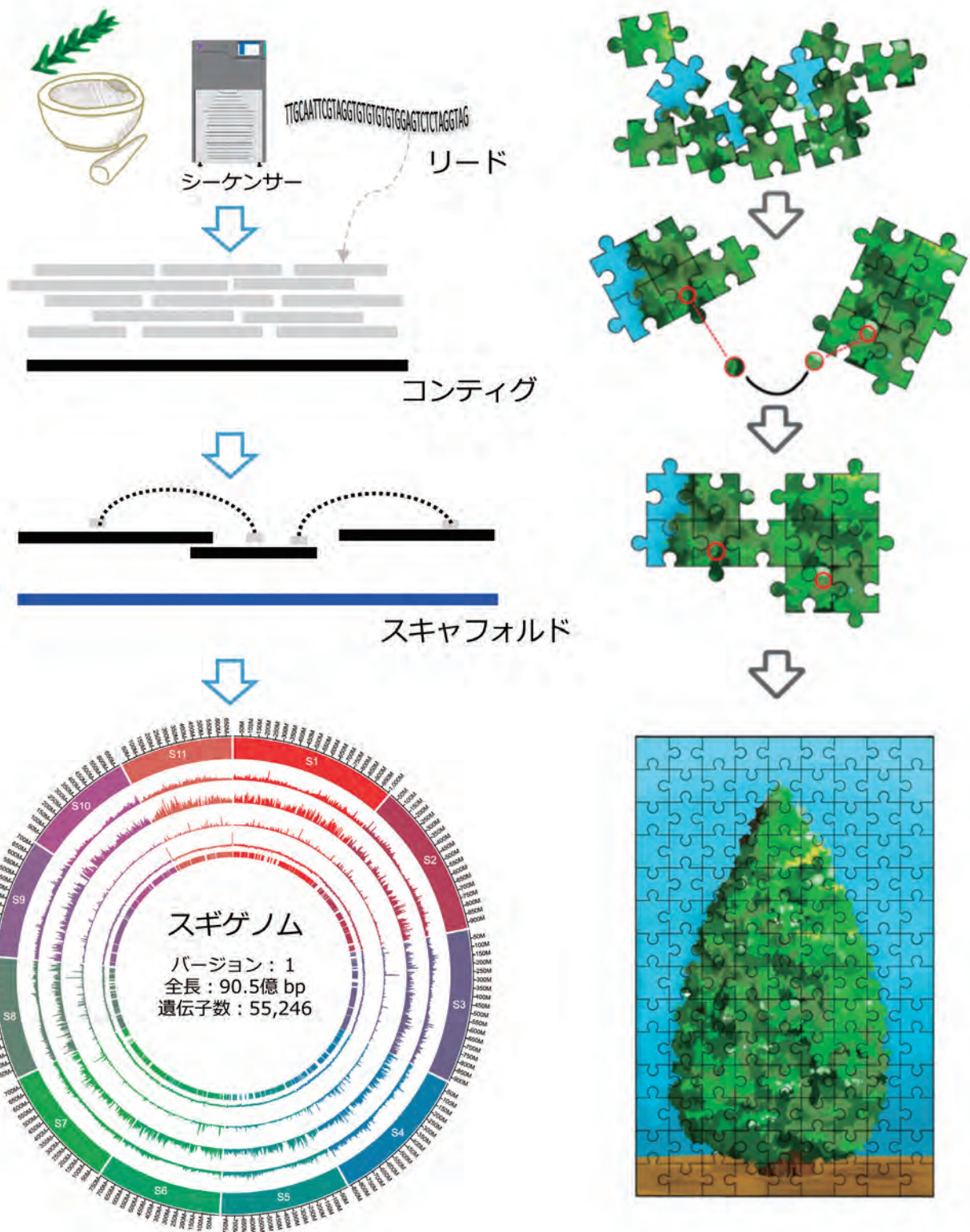


図1 ジグゾーパーズル(右)になぞらえたスギゲノム解読手順のイメージ図(左)

スギのゲノムが解読されていく様子をジグゾーパーズルが完成していく様子に例えて表現しました。高分子量DNAを抽出することで、ひと続きの長い塩基配列(リードと呼ばれる読み取り配列:灰色の線分)を読み取ることができます。リードの長さはジグゾーパーズルのピースの大きさに例えられ、ピースサイズが大きいほど、そしてユニークなピースが多いほど組み立てが容易になります。そういったリードをつなぎあわせてもう少し大きなピースを組み立てました(コンティグと

呼ばれる部分配列:黒い線分)。スギのゲノムは83.6%が繰り返し配列でユニークな配列が少なくゲノム解読を難しくしていましたが、近接したコンティグ同士の連結によって染色体全体を網羅した配列(スカフォールドと呼ばれる最終的な連結配列:青い線分)を構築することが出来ました。左下の図が、実際に解読されたスギゲノムの全貌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(1~11)、および物理距離(100万bp)を表します。(Fujino et al. 2023を改変)