

## 真核生物における SMC タンパク質ファミリーの多様化及び二次的喪失

吉永 真理 (筑波大学 生物学類) 指導教員: 稲垣 祐司 (筑波大学 計算科学研究センター)

## ■背景

ATPase スーパーファミリーに属する SMC タンパク質ファミリーは染色体構造の維持に必須で、原核生物から真核生物まで広く保存されると考えられている。原核生物では 1 種類の SMC がホモ二量体を形成するのにに対し、真核生物では 6 つの SMC パラログ (SMC1- SMC6) が、3 種類のヘテロ二量体を形成する。SMC 1-3 と SMC 2-4 二量体は、それぞれ染色体分離の際に姉妹染色体を束ねるコヒーシンと染色体を凝集させるコンデンシンを構成する。SMC5-6 二量体は DNA 修復に関わる。原核生物では主に DNA 鎖を凝集させる役割を果たす SMC だが、遺伝子重複とその後の機能分化により、染色体の維持という真核生物特有の機構の要となった。

Cobbe & Heck (2004) は SMC1 から 6 の配列に基づく分子系統解析を行ったところ、SMC1 と 4 間、2 と 3 間、5 と 6 間の近縁性が復元された。しかし、当該論文ではオピストコンタに含まれる 5 種とトリパノソーマ原虫からの配列データだけが解析されており、真核生物の多様性を十分に俯瞰しているとはいえない。そこで、本研究では広域な真核生物を対象に SMC1 から 6 を探索し、新たな配列データをもとに系統解析を行うことで、SMC の多様性と進化の再検証を行った。

## ■方法

## i) SMC タンパク質の相同配列探索

NCBI nr (non-redundant protein sequences) データベース中の原核生物 14 種と真核生物 28 種について、ヒトもしくは検索対象の近縁種の SMC タンパク質を問い合わせ配列として BLASTP による同源性検索を行い、SMC 相同配列を得た。NCBI nr データベースに登録されているタンパク質では不十分な真核生物 9 種については、NCBI SRA (Short Read Archive) からトランスクリプトームデータを取得した。sra-toolkit v2.10.0 と fastp v0.20.0 による配列のクオリティコントロール後、Trinity v2.8.5 でアセンブルして mRNA 配列を復元した。復元後の配列に対して、ヒトや検索対象の近縁種の SMC を問い合わせ配列として TBLASTN による同源性検索を行い、SMC 相同配列を得た。最終的に SMC のアミノ酸配列を、原核生物 14 種、系統的に広範な真核生物 37 種から同定し、系統解析に用いた。

## ii) SMC に進化的に近縁なタンパク質の相同配列探索

真核生物の Rad50 や原核生物の SbcC、RecN は SMC とアミノ酸配列レベルで同源性が検出できる。また、細菌類では SMC とアミノ酸配列レベルで同源性をもつ MukB をもつことが知られている。4 種類のタンパク質のアミノ酸配列を探索し、SMC と共に系統解析に用いた。データベース内の探索は、SMC と同様に行った。

## iii) 系統解析

SMC およびその近縁タンパク質のアミノ酸配列を、MAFFT v7.453 を用いてアラインし、配列間で保存性が低い座位を手動で排除した。最終的に配列の保存性がある 220 アミノ酸座位を系統

解析に用いた。SMC アライメントから、LG + Γ4 モデルを用いた最尤 (ML) 法により ML 系統樹を復元した。系統関係の頑健性は、リサンプリングデータ (100 個) をもちいた ML 法によるブートストラップ解析により検討した。上記系統解析には IQ-TREE v1.6.12 を用いた。

## ■結果と考察

真核生物 37 種の SMC 配列から復元した ML 系統樹では、SMC1 から 6 の単系統群 (クレード) が形成された。広範な真核生物系統に SMC1-6 が存在することから、真核生物の初期段階ですでに SMC の 6 つのパラログが確立していたことが再確認できた。サブファミリー間の系統関係は、SMC1 と 4 間および 5 と 6 間の近縁性が復元されたが、先行研究とは異なり 2 と 3 は互いに近縁とはならなかった。ただしアライメント中で欠失座位が多い SRA データから取得した配列を系統解析から削除したところ、SMC2 と 3 間の近縁性も復元された。従って、39 種からの SMC 配列に基づく解析で、SMC2 と 3 間の近縁性が復元できなかったのは、欠失座位の多い配列が原因である可能性がある。

SMC1、2、3 および 4 は、解析したほとんど全ての種で検出できたものの、SMC5 と 6 の相同配列が検出できなかった系統が複数存在した。例えば繊毛虫 *Paramecium tetraurelia* および *Tetrahymena thermophila* には高品質なゲノム・トランスクリプトームデータが整備されているにも関わらず、SMC5 と 6 が検出されなかった。同様にトリパノソーマ原虫 (*Trypanosoma brucei*) でも SMC5 および 6 が検出されなかった。これらの結果は DNA 修復に関わる SMC5 と 6 が二次的に欠失している系統が存在することを示唆する。そこで上記 2 種以外の繊毛虫 11 種について、主にトランスクリプトームデータ中に SMC1-6 配列を探索した。その結果、解析した繊毛虫 11 種中 10 種で SMC 5 と 6 とともに検出できず、繊毛虫の進化における SMC5 と 6 の欠失を支持する結果となった。同様の探索を、トリパノソーマ原虫がふくまれるキネトプラスト綱に含まれる 4 種、キネトプラスト綱の姉妹群であるディプロネマ綱とユーグレナ綱に含まれる 4 種について行った。興味深いことに、キネトプラスト綱とディプロネマ綱に属する 6 種でも SMC5 と 6 は検出されなかった。この結果は、両綱の共通祖先で SMC5 と 6 の欠失が起こった可能性を示唆する。

本研究の結果は、SMC1-6 が真核生物初期進化で確立したが、真核生物多様化の過程で 2 次的欠失が複数回起こっていることを示唆する。今後は SMC の欠失がどのように補完されているのかを実験的に検証する必要がある。

## ■参考文献

Cobbe N, Heck MM (2004) The evolution of SMC proteins: phylogenetic analysis and structural implications. *Mol Biol Evol.* 2004 21:332-47.