

[別紙2]

審査の結果の要旨

申請者氏名 張 翔

様々な生物の研究において、対象とする生物の高精度のゲノムシーケンス情報や連鎖地図が有用であることは論を待たない。2005年の次世代シーケンサ(NGS)の出現以降、全ゲノム解析にとって十分な大量の塩基配列データを得ることは比較的容易になった。それにもかかわらず、主要なNGSである短鎖高密度並列型NGSによって解読されるデータは大量であるものの、1リード長が短いため、ゲノム中に存在する各種反復配列の壁を乗り越えるのは難しいため、多くの生物のゲノム解読において解読データを編集して最終的に得られるデータは多量の短いコンティグ、スキマホールドの集合であるケースがほとんどである。また最近になって脚光を浴びている長鎖解読型のNGSを併用した場合であっても、最終的に構築されるゲノムシーケンスがそれぞれの染色体に対応するところまで至ることは難しい。これらのシーケンシングとアセンブルによって構築されたシーケンスを各染色体に沿って整列化させるためには連鎖地図の活用は極めて有効である。しかし連鎖地図の作成においても、従来法であるマイクロサテライトによるタイピングで構築される連鎖地図にはマーカー密度やタイピングの効率性に限界があった。

一方、魚類では、雌性発生などの生殖技術が活用できることや多卵で、一对の親から多数の子を得ることができるなど、ゲノム解析が先行したヒトやマウスなどの哺乳類などにはない特徴がある。雌性発生では第一卵割阻止を行うことで雌親由来の1本の染色体が倍加したダブルハプロイド個体が作出できる。ダブルハプロイド個体はハプロイドと同等でDNA多型が存在しないため、個々の多型部位に関してヘテロ接合は存在しない。このことは一塩基多型のタイピングなどにおいて極めて大きな利点であり、個々の多型部位をカバーする1リードの塩基配列データがあればタイピングは完了する。さらにSNPタイピングは短鎖型NGSを用いて効率的に行うことができる。本論文はこれらの手法を実際に行い、その有効性を検証したものである。第一章では以上に述べた内容を含め、連鎖地図作成に関する手法を概説するとともに、脊椎動物の中で最小のゲノムを持つトラフグのゲノム解析に関する研究の歴史が概観されている。そして本論文研究では、完全に新規にゲノム解析を行うことを想定して、すでに存在するトラフグゲノムシーケンス(FUGU5)を用いずに、自ら新たにアセンブルを行ったゲノム配列をリファレンスとして用いて、染色体ごとに整列化されたゲノムシーケンスを構築し、それをすでに既存ゲノム配列と比較・評価するため、トラフグを対象としてテストを行ったことが目的として述べられている。

第二章はダブルハプロイド個体の作出からシーケンスデータの取得のプロセスが記述されている。UV照射による不活化精子をトラフグ卵に受精後、コールドショック法によって第一卵割を阻止後、胚発生した卵を選抜し、それらからゲノムDNAを抽出してIllumina社のNGSを使っ

てゲノムの解読を行った。192 検体をシーケンシングしたところ、23 検体では十分なデータが得られなかったため除外し、残りの 169 検体について、さらに解析を進めた。これら 169 検体から得られたデータ量はのべ 71.32Gb であった。

第三章では新規ゲノムアセンブルとそれをリファレンスとした各個体の SNP タイピングについて記述されている。解析過程で見つかった部分的ヘテロの 4 検体を除いた 165 検体を最終的な解析に用いた。これらの検体から得られた 69.85Gb のシーケンスデータをもとにゲノムアセンブルソフトである SOAPdenovo2 を用いてアセンブルを行ったところ、54,127 個のスキマホール、N50 が 22,235 bp、合計 356.59 Mb からなるリファレンスシーケンスが構築された。これらのリファレンスに対して、それぞれの検体のリードをマップした結果、1,070,601 個の SNP サイトが見いだされ、それぞれの検体の各 SNP サイトに関するタイピングデータが得られた。これらの塩基配列データのゲノムに対する平均カバー深度が約1倍と大変低いカバー率であった。

第四章ではSNPのタイピングデータの欠落を補うための手法として、新規に考案された Short Segment Genotype (SSG) の活用について論述されている。各サンプルの各 SNP サイトにおけるデータの深度は約1倍であるため、数十%の SNP サイトではデータが欠落している。しかし、減数分裂において各染色体における交叉頻度は平均1回程度であるため、その交叉部位以外の領域においては、雌親のそれぞれの相同染色体の相がそのまま保たれている。したがって染色体全体に対して相対的に小さなゲノム領域の中に SNP の相は交叉部位含む領域を除いて一定である。そこで SSG の Short Segment として8kbを設定し、その中に存在する SNP のタイピングデータにより SSG を一つのマッピング単位として連鎖地図を作製に用いた。

第五章では得られた連鎖地図の結果と、既存のトラフグゲノムデータであるFUGU5との比較、について述べられている。構築された連鎖地図では 37,343 個の SSG が 3,090 の部位にマップされ、22 個の染色体(連鎖群)にまとめられた。その数は既報の連鎖群数と同一であった。それらに利用された SNP は 802,277 個であった。ゲノム全体の遺伝距離は 2319.65cM であり、平均の 3,090 の部位間(マーカー間)の距離は 0.75cM であった。またこれらの連鎖地図を FUGU5 と照合することで、FUGU5 ではマップされていなかったスキマホールのうち 65.97Mb 相当の配列をマップし、そのうち 28.2Mb については方向も決定できた。FUGU5 との照合においては、ほぼ完全な対応関係が確認できたが、11 個の染色体については、FUGU5 との微小な相違が見いだされた。このことはトラフグ個体間の大きなセグメントレベルの多型、あるいはいずれかのエラーであると考えられた。

以上の論文の内容はダブルハプロイドを用いたタイピングでは非常に低いカバー深度で極めて高分解能の連鎖地図を作成できることを示したものであり、多くの魚類や魚類と同様のダブルハプロイド・ハプロイドを作製できる他の生物におけるゲノム解析にも資する結果である。これらの研究成果は、学術上応用上寄与するところが少なくない。よって、審査委員一同は本論文が博士(農学)の学位論文として価値あるものと認めた。