

論文提出者	工学府博士後期課程 生命工学 専攻 平成 23 年度入学 学籍番号 11831102 氏名 黒谷 篤之 印
主指導教員氏名	黒田 裕
論文題目	植物プロテオームの網羅的アノテーションとそのデータベース化、及び、被子植物におけるディスオーダー領域と翻訳後修飾の生物情報学的解析
<p>論文要旨 (2000 字程度)</p> <p>本論文は、植物種のタンパク質の構造的・機能的等の大規模な計算をし、その結果をプロテオームデータとしてデータベース化し、更に、被子植物におけるタンパク質のディスオーダー領域と翻訳後修飾について生物情報学的な解析について記載したものである。</p> <p>近年のゲノムの解読技術の発展により、大量なゲノムの配列情報が蓄積しており、植物種においても多数種のゲノム解読がされている。しかしながら、ポストゲノムの解析による植物種の情報量は、動物や微生物と比較して劣っており、特に植物種のタンパク質の構造・機能情報は、著しく劣っているのが現状である。植物の研究はその規模から排出されるデータにも限りがある為、バイオインフォマティクスの力を利用してこれらの情報欠如に対する穴埋めや効率的に実験を行う為の情報の付加が要求されている。</p> <p>「第 1 章」では、本研究に際しての基本情報や解析の意義、タンパク質の翻訳後修飾の種類やタンパク質のディスオーダー領域の特徴、本研究に利用する植物種の特徴等を記載した。</p> <p>「第 2 章」では、植物種の大規模な配列解析として、全ゲノム解読がされた代表的なモデル植物種（アラビドプシス、ダイズ、ポプラ、イネ、ヒメツリガネゴケ、シアニディオシゾン）のタンパク質配列からタンパク質の構造的・機能的・配列的計算、および、新規機能領域等の様々な計算をし、これらのデータを公開する為のデータベースサイトの構築について記載をした。このデータベースサイトは、上記で得た種々の計算結果を複合的に検索できる仕様になっている為、実験者が研究材料の抽出として大量配列から目的の性質を有する配列セットを絞り込むケース等に有効に利用ができるものとした。</p> <p>「第 3 章」では、第 2 章における各植物種のタンパク質の配列的傾向の比較から派生して、単子葉植物、および、双子葉植物についてのタンパク質のディスオーダー領域、及び、翻訳後修飾であるリン酸化、並びに O,N-結合型糖鎖付加について生物情報学的な解析をし、その結果を記載した。第 2 章でのデータベースサイトの構築後、そのデータとしての各植物種のタンパク質の配列的傾向を比較したところ、タンパク質のディスオーダー領域の割合、および、翻訳後修飾である O-結合型糖鎖付加部位数において、双子葉植物（シロイヌナズナ、ダイズ、ポプラ）よりも単子葉植物（イネ）の方が高い値を示しており、一方、N-結合型糖鎖付加部位数は単子葉植物よりも双子葉植物の方が高い値を示していた。そこで、全ゲノム解読がされている 3 種の単子葉植物（イネ、ブラキポディウム、ソルガ</p>	

ム)、及び、5種の双子葉植物（アラビドプシス、ダイズ、ポプラ、ブドウ、トマト）の第2章のデータセットから植物種を増やした形のデータセットとした8種のモデル植物種を解析対象とし、それらの全タンパク質配列について以下の解析を行った。

第3章第1節では、上記の8植物種の全タンパク質の配列セットにおけるディスオーダー領域の予測による割合を比較した。結果としては、ディスオーダー領域の割合は、双子葉植物よりも単子葉植物の方がディスオーダー領域を多く含む傾向があることを確認した。これまでに原核生物と真核生物での間でのディスオーダー領域の割合に差があることは知られていたが、近縁の種間での相違の報告はなく、この傾向は単子葉植物および双子葉植物間の違いとしての新たな知見となった。更に、植物種間での配列類似度とディスオーダー領域の関係を確認したところ、双子葉植物および単子葉植物の分類に関係なく、種間での類似度が高い配列同士よりも類似度が低い配列同士の方がディスオーダー領域の割合は高いという傾向を得た。種間の差をディスオーダー領域の多いタンパク質に絞り込んで観察することへの利用が考えられる。

第3章第2節では、上記の全8植物種について、主なタンパク質の翻訳後修飾であるS,T,Y-リン酸化、およびO,N-結合型糖鎖付加の計算をし、これと計算によるタンパク質のディスオーダー領域の割合との間の関係を観察し、その傾向の解析について記載した。結果としては、S,T,Yリン酸化部位数は、種に関係なく同程度のヒットがあり、特に植物種間における特徴は確認できなかった。また、O-結合型糖鎖付加の部位数は、単子葉植物の方が双子葉植物よりも多く、逆にN-結合型糖鎖付加の部位数は、双子葉植物の方が単子葉植物よりも多かったことを確認した。更に、リン酸化、O-結合型糖鎖付加は、ディスオーダー領域で生じやすいという傾向を確認した。植物種において、O-結合型糖鎖付加とディスオーダー領域のこのような傾向確認の報告はなく、新規な情報となる。

植物において全ゲノム解読が終了している種は、現在のところ実験的に扱いやすいモデル植物が中心となっている。植物は、食料、飼料、材料資源、環境改善等多様な利用価値がある一方、未解明な部分が多く、応用研究が進む余地の広い分野でもある。今後は、現在のモデル植物で蓄積されたデータを基として、実用植物の研究へとシフトしていくと考えられる為、今回の解析データや配列属性等の情報は公開により広範に利用され、研究の支援・促進材料となることを願いたい。

TITLE	Comprehensive annotation of plant proteomes and bioinformatic analysis of disorder regions and post-translational modifications in angiosperms
NAME	Atushi Kurotani
<p style="text-align: center;">ABSTRACT</p> <p>This paper is comprehensive sequence analysis of plant proteins and correlations of protein disorder with post-translational modifications (PTMs) in angiosperms</p> <p>The development of DNA sequencing techniques achieved a lot of genome sequencing projects in recent times, such as Arabidopsis, Rice, and so on, in plant genomes. After that, those sequence information leads the post-genomic works. Although, protein research in plants lags behind that in animal and bacterial species.</p> <p>Therefore, I analyzed the protein secondary structural features and novel functional regions with prediction tools using whole amino acid sequences of 6 main model plant species (Arabidopsis, Soybean, Poplar, Rice, Moss, Shyzon). Those results were published on the web as database site which is available for multiple search.</p> <p>So far, there is no research systematically focusing on large-scale analysis of degree of protein disorder and number of PTMs in angiosperm. Therefore, I investigated the global degree of protein disorder and the multiple sites of phosphorylation, N, O-linked glycosylation in prediction in the complete proteomes for various plant species including 5 typical dicots (Arabidopsis, Soybean, Poplar, Grape, Tomato) and 3 typical monocots (Rice, Brachypodium, Sorghum). As a result, the disorder degree in monocots was found to be significantly higher than that of in dicots. Phosphorylation sites did not have a universal preference in monocots and dicots. O-linked glycosylation sites in monocots was found to be higher than those in dicots, whereas, N-linked glycosylation sites in dicots was found to be higher than those in monocots. In addition, phosphorylation and O-glycosylation displayed a clear preference for occurrence in disordered regions of plant proteins. In contrast, N-glycosylation did not show a universal preference in disordered and ordered regions.</p> <p>It is expected that this website together with these data will be effectively utilized to the next step of plant science research.</p>	