

第 1 章 序論

1.1 背景

グラム陰性細菌の一部の細菌は、III 型分泌装置と呼ばれるタンパク質分泌装置を持つ。III 型分泌装置は鞭毛の基部に類似した構造を持つ注射器状の形状をしており、細菌が宿主細胞にタンパク質を注入するために使われる。裸のタンパク質であれば宿主細胞の免疫系の標的にされるが、このような構造をしていることでタンパク質を露出することなく標的の細胞に注入することができる。このとき注入されるタンパク質はエフェクタータンパク質と呼ばれる。

このエフェクタータンパク質は宿主細胞の機能に様々な影響を与える。例えば、あるエフェクタータンパク質は宿主細胞の表面にある分子と結合し、細菌を細胞表面に繋ぎとめる役割を果たす。また別のエフェクタータンパク質は、宿主細胞の骨格に影響を与えて細菌が取り付くための台座を作る役割を果たす。このように、分泌装置を持つ細菌はエフェクタータンパク質を使って宿主細胞を自分自身に適した環境に変化させていくのである。

したがって、III 型分泌装置とエフェクタータンパク質は病原細菌の病原性を表す重要な機構であると考えられている。ところが近年、共生細菌にも III 型分泌装置を持つものがいることが明らかになってきた。病原細菌の病原性を表す III 型分泌装置をなぜ共生細菌が持つのか。III 型分泌装置とエフェクタータンパク質の役割と機能に注目が集まっている。

1.2 先行研究

矢原耕史, 姜英, 柳川堯による先行研究 [1] では、病原細菌と共生細菌のエフェクタータンパク質の物理化学的指標を収集し、非線形多変量解析のカーネルロジスティック回帰を用いて解析することで、両者のエフェクタータンパク質を判別することに成功した。また、その判別性能をサポートベクターマシン (2 クラスの教師あり分類問題を解くための統計的学習アルゴリズム) によって評価し、さらにその相互依存関係及び最も特徴的な領域を、グラフィカルモデリング (複数の確率変数間の関連を分析する多変量解析の 1 つ) と sliding window 解析 (高速乗算法の 1 つ) によって明らかにした。この先行研究における病原細菌と共生細菌のエフェクタータンパク質の正判別率 (正しく判別されたエフェクタータンパク質の個数の割合) は 84.2 % であり、両者を区別する特徴的な指標を 7 つ同定した。

1.3 本研究の目的

本研究では、より簡単な主成分分析と線形判別分析を用いて、線形解析のみで病原細菌と共生細菌のエフェクタータンパク質を判別することを目的とした。その結果、線形解析のみによる判別でも先行研究と同等以上の判別精度を得られたことを報告する。

第 2 章 データセット

共生細菌のエフェクタータンパク質 30 種、病原細菌のエフェクタータンパク質 30 種、計 60 種を解析対象

とした。解析対象のタンパク質 ID は先行研究 [1] と同様のものである。この ID を Genbank データベースで検索し、タンパク質のアミノ酸配列を得た。アミノ酸配列から、タンパク質の物理化学的指標 44 成分を計算する。求めた物理化学的指標は [表1]の通りであり、先行研究 [1] と同様の指標である。計算には、配列解析のソフトウェア EMBOSS とタンパク質のシグナルペプチド予測ツール SignalP を用いた。

表 1: 物理化学的指標の一覧

1	抗原性領域の個数	23	アミノ酸の割合 : Leu
2	切断部位の個数	24	アミノ酸の割合 : Met
3	二次構造 (ヘリックス) の個数	25	アミノ酸の割合 : Asn
4	二次構造 (シート) の個数	26	アミノ酸の割合 : Pro
5	二次構造 (ターン) の個数	27	アミノ酸の割合 : Gln
6	二次構造 (コイル) の個数	28	アミノ酸の割合 : Arg
7	疎水性モーメント	29	アミノ酸の割合 : Ser
8	残基重量の平均	30	アミノ酸の割合 : Thr
9	電荷	31	アミノ酸の割合 : Val
10	等電点	32	アミノ酸の割合 : Trp
11	モル吸光係数	33	アミノ酸の割合 : Tyr
12	吸光係数	34	アミノ酸の割合 : 短いアミノ酸
13	封入体形成の確率	35	アミノ酸の割合 : 小さいアミノ酸
14	アミノ酸の割合 : Ala	36	アミノ酸の割合 : 脂肪族アミノ酸
15	アミノ酸の割合 : Cys	37	アミノ酸の割合 : 芳香族アミノ酸
16	アミノ酸の割合 : Asp	38	アミノ酸の割合 : 非極性アミノ酸
17	アミノ酸の割合 : Glu	39	アミノ酸の割合 : 極性アミノ酸
18	アミノ酸の割合 : Phe	40	アミノ酸の割合 : 荷電性アミノ酸
19	アミノ酸の割合 : Gly	41	アミノ酸の割合 : 塩基性アミノ酸
20	アミノ酸の割合 : His	42	アミノ酸の割合 : 酸性アミノ酸
21	アミノ酸の割合 : Ile	43	シグナルペプチドから成熟タンパク質までの切断部位の個数
22	アミノ酸の割合 : Lys	44	シグナルペプチドの割合

第 3 章 判 別 結 果

本研究では統計解析ソフト R を用いて解析を行った。解析の始めに、第 2 章で作成したデータセットを平均 0、分散 1 に標準化した。

3.1 判 別 精 度

主成分分析で得た合成変数に対して交差検定法で判別分析を行った。この判別に使用した主成分数は第 9 主成分までである。判別結果は [表 2] の通りである。共生細菌のエフェクタータンパク質 30 種のうち、27 種が正しく共生と判別され 3 種が誤って病原と判別された。また、病原細菌のエフェクタータンパク質 30 種のうち、2 種が誤って共生と判別され 28 種が正しく病原と判定された。

表 2: 判別結果

		判 別	
		共 生	病 原
真 実	共 生	(A) 27	(B) 3
	病 原	(C) 2	(D) 28

3.2 共 生 性 ・ 病 原 性 に 特 徴 的 な 指 標

判別に用いた主成分を構成する 44 成分の物理化学的指標のうち、共生性・病原性に大きく寄与している指標を調べて [表 3][表 4] にまとめた。

表 3: 共生性に寄与

抗原性領域の個数	切断部位の個数
二次構造（ヘリックス）の個数	二次構造（シート）の個数
疎水性モーメント	残基重量の平均
アミノ酸の割合：Ile	アミノ酸の割合：Lys
アミノ酸の割合：Leu	アミノ酸の割合：Asn
アミノ酸の割合：脂肪族アミノ酸	シグナルペプチドの割合

表 4: 病原性に寄与

アミノ酸の割合：Ala	アミノ酸の割合：Asp
アミノ酸の割合：Gly	アミノ酸の割合：Pro
アミノ酸の割合：Arg	アミノ酸の割合：短いアミノ酸

アミノ酸の割合：小さいアミノ酸	アミノ酸の割合：芳香族アミノ酸
アミノ酸の割合：荷電性アミノ酸	アミノ酸の割合：酸性アミノ酸

3.3 主成分数の決定方法

主成分分析を判別に用いる際には、主成分をいくつまでとるのが問題となる。これは、使用する主成分数が少なすぎれば元のデータを再現しきれず、逆に使用する主成分数を増やしすぎるとノイズが増えてデータの再現性が落ちてしまうからである。

3.1 の判別に使用した主成分数は9であった。これは使用する主成分の数を変えて判別を行い、最も正判別率の良い主成分数を採択したものである。したがって、未知の細菌のエフェクタータンパク質を判別するには適切な主成分数を決定する方法を考える必要がある。

本研究では、CV（交差検定）ありの判別分析の結果と CV なしの判別分析の結果を比較することで、適切な主成分数を決定する方法を考えた。

第4章 結論

4.1 判別精度

先行研究 [1] の正判別率 84.2 %に対して、本研究の正判別率は 91.7 %である。したがって、共生細菌と病原細菌のエフェクタータンパク質を判別する判別精度と解析手法を改善することができた。

4.2 共生性・病原性に特徴的な指標

タンパク質の 44 成分のうち共生性・病原性に特徴的な指標を示した [表 3][表4]。この結果には先行研究 [1] で明らかにされた全7指標も含まれているため、適切な判別ができていると考えられる。また、今回の結果から細菌の病原性が解明されることを期待したい。

4.3 適切な主成分数の決定

正解がわからない判別、つまり未知の細菌のエフェクタータンパク質の判別の際にも、交差検定ありの判別と交差検定なしの判別との一致率の標準誤差のプロットから適切な主成分数を決定することが可能である。

参考文献

[1] Yahara, K., Jiang, Y., and Yanagawa, T.: Computational identification of discriminating features of pathogenic and symbiotic type III secreted effector proteins, IPSJ SIG Technical Report, 2010-BIO-21, No.16, pp.1-8, (2010).

[2] Yuichi Nakano, Y-h. Taguchi: Feature extraction for discriminance of symbiotic/parasitic bacterial type III effector protein using principal component analysis, IPSJ SIG Technical Report, no.37 (2011).

[3] Yuichi Nakano, Y-h. Taguchi: Feature extraction for discriminance of symbiotic/parasitic bacterial type III effector protein using principal component analysis, 2011 IEEE International Conference on Bioinformatics & Biomedicine, vol.2, pp.964-965 (2011).