

博士論文

バクテリア 16S rRNA 遺伝子の進化

佐藤 允治

目次	ページ
第一章 研究の背景と目的	1
リボソームの立体構造と 16S リボソーマル RNA(16S rRNA)	
リボソーマル RNA オペロンの遺伝子構造とコピー数	
多重遺伝子と適応進化	
リボソーマル RNA オペロンごとの役割	
同一ゲノム上の 16S rRNA 遺伝子同士の均一化	
バクテリアにおける遺伝子の水平伝播	
系統解析の指標として用いられてきた 16S rRNA 遺伝子	
自然界における 16S rRNA 遺伝子の水平伝播	
他の翻訳系遺伝子の水平伝播	
16S rRNA 遺伝子の水平伝播実験	
系統樹による解析の原理的限界	
複数通りの進化史を同時に表現できる系統ネットワーク法	
系統ネットワーク法と系統樹	
Bandelt の方法を用いた系統ネットワークの作成方法	
系統ネットワーク法による水平伝播・遺伝的組換えの検出	
Bandelt の方法で作成した系統ネットワークの特徴	
研究の目的	
引用文献	
第二章 <i>Escherichia coli-Shigella</i> 集団の 16S rRNA 遺伝子の解析	37
序論	37
<i>Escherichia coli</i> と <i>Shigella</i> 属	
<i>Escherichia coli</i> と <i>Shigella</i> 属の分類と系統関係	
材料と方法	37
データの取得と整理	
系統ネットワーク解析	
結果と考察	41
<i>Escherichia coli, Shigella</i> 集団の 5 つの 16S rRNA 遺伝子グループ	
<i>Escherichia coli, Shigella</i> 集団のゲノム内多型	
<i>Escherichia coli, Shigella</i> 集団の系統ネットワーク解析	
各 16S rRNA 遺伝子グループの進化史	
遺伝的組換えか平行進化か	
16S rRNA 遺伝子グループ C の進化	
遺伝的組換えが生じた場	
引用文献	
第三章 Enterobacteriales 目の 16S rRNA 遺伝子の解析	51

序論	51
Enterobacteriales 目とは	
材料と方法	51
データの取得と整理	
系統ネットワーク解析	
属ごとの代表配列の選択	
結果と考察	59
属内で 16S rRNA 遺伝子の水平伝播が見られた分類	
a) <i>Escherichia</i> 属	
b) <i>Salmonella</i> 属	
c) <i>Enterobacter</i> 属	
d) <i>Klebsiella</i> 属	
e) <i>Erwinia</i> 属	
f) <i>Dickeya</i> 属	
g) <i>Pectobacterium</i> 属	
属を跨いで 16S rRNA 遺伝子の水平伝播が見られた分類	
a) <i>Escherichia</i> 属と <i>Salmonella</i> 属、 <i>Citrobacter</i> 属	
b) <i>Enterobacter</i> 属と <i>Klebsiella</i> 属	
c) <i>Erwinia</i> 属と <i>Pantoea</i> 属	
d) <i>Pectobacterium</i> 属と <i>Dickeya</i> 属	
e) <i>Yersinia</i> 属と <i>Escherichia</i> 属	
Enterobacteriales 目において普遍的にみられる 16S rRNA 遺伝子の水平伝播	
水平伝播によって大きく変化する 16S rRNA 遺伝子	
16S rRNA 遺伝子の水平伝播が与える影響	
引用文献	
第四章 16S rRNA 遺伝子の水平伝播の網羅的な検出	85
序論	85
遺伝的組換えの検出に有益な系統ネットワーク法	
節約ネットワーク法に基づいて遺伝的組換えを検出する PNarec	
PNarec 法をバクテリアに用いることの問題点	
材料と方法	87
遺伝的組換え体の網羅的な検出方法	
遺伝的組換えを判断する検定方法	
2 sample runs test の計算方法	
Ezawa ら (2006) の手法の計算例	
データの取得と整理	
遺伝的組換えが生じている領域の求め方	

	どの分類群とどの分類群で遺伝的組換えが生じたのかを求める	
	結果と考察	91
	分類群ごとの組換え体の頻度	
	遺伝的組換えと 16S rRNA 遺伝子のコピー数	
	遺伝的組換えが生じている領域	
	16S rRNA 遺伝子に種固有の領域は存在しない	
	水平伝播で飛び越える分類群	
	網を超えて生ずる 16S rRNA 遺伝子の水平伝播	
	23S rRNA 遺伝子における遺伝的組換え	
	平行進化サイト数の期待値と実測値の比較	
	平行進化サイト数の期待値の計算方法	
	実際の r^* の計算	
	引用文献	
第五章	結論と展望	105
	結論	105
	過去の 16S rRNA 遺伝子の水平伝播の報告と本研究の成果の総括	
	a)16S rRNA 遺伝子の水平伝播は普遍的な現象である	
	b)ゲノム内組換えと水平伝播による進化モデル	
	c)16S rRNA 遺伝子の水平伝播と他の遺伝子の水平伝播の違い	
	d)水平伝播によって進化しやすい 16S rRNA 遺伝子	
	Complexity Hypothesis を再解釈する	
	16S rRNA 遺伝子の水平伝播について新たに分かったこと	
	a)16S rRNA 遺伝子の水平伝播は領域を問わずに生じている	
	b)分類群ごとに水平伝播が高頻度に生じる領域が異なる	
	16S rRNA 遺伝子の水平伝播は何かをもたらしているのか？	
	16S rRNA 遺伝子を用いた種の同定について	
	展望	109
	16S rRNA 遺伝子の領域的な水平伝播をもたらす機能的変化の解明	
	遺伝的組換えを検出する基準の精査	
	引用文献	
	謝辞	112

第一章 研究の背景と目的

1. リボソームの立体構造と 16S リボソーマル RNA (16S rRNA)

リボソームは全ての生物に存在し、DNA から mRNA に写し取られた遺伝情報をポリペプチドへと翻訳する役割を担っている。*Escherichia coli* における 70S リボソームの 3 次元構造を図 1.1 に示す。バクテリアにおいては、リボソームは 30S と 50S の 2 つのサブユニットから成り立っている。30S サブユニットは 16S rRNA と 21 のリボソーマルタンパク質から構成され、50S サブユニットは 5S rRNA と 23S rRNA、36 のリボソーマルタンパク質から構成されており、それぞれが、rRNA とタンパク質の巨大な複合体である。16S rRNA と 23S rRNA はそれぞれ 30S サブユニットと 50S サブユニットの中心骨格を成す分子である。

mRNA の SD 配列 (Shine-Dalgarno 配列) と 30S サブユニットが結合するところから翻訳は開始される。開始コドンである f-メチオニル tRNA が 30S の P 部位に結合し、f-メチオニル tRNA のアンチコドンが mRNA の開始コドンに対合する。続いて、30S サブユニットの A 部位に、mRNA 上の次のコドンとそれに対応するアミノアシル tRNA が対合すると、16S rRNA に構造変化が生じ、そのシグナルが 50S へと伝わり、EF-Tu の GTP が加水分解される。16S rRNA の暗号解読中心 (Decoding Center) は、A 部位のアミノアシル tRNA と mRNA のアンチコドン領域の対合を厳密に監視する。その後、23S rRNA のペプチド転位反応活性中心の触媒作用により、P 部位の tRNA からアミノ酸が切り離され、A 部位の tRNA が運んできたアミノ酸と結合する。A 部位に結合していた mRNA とアミノアシル tRNA はそのまま P 部位へとシフトする。この一連の動作を終始コドンを認識するまで繰り返すことで、リボソームは mRNA の情報をポリペプチドへと翻訳する。

2. リボソーマル RNA オペロンの遺伝子構造とコピー数

5S rRNA、16S rRNA、23S rRNA の 3 つの rRNA は、リボソーマル RNA オペロン上にコードされており、ほとんどのバクテリアにおいて 5' 末端側から 16S rRNA 遺伝子、23S rRNA 遺伝子、5S rRNA 遺伝子という並び順が保存されている(図 1.2)。それぞれの rRNA 遺伝子の長さは、*Escherichia coli* において、5S rRNA 遺伝子は 120 塩基、16S rRNA 遺伝子は 1542 塩基、23S rRNA 遺伝子は 2906 塩基である。多くのバクテリアは、リボソーマル RNA オペロンを複数コピー持っており、例えば、*Escherichia coli* MG1655 はゲノム中に 7 コピーのリボソーマル RNA オペロンを持つ。バクテリアの間でリボソーマル RNA オペロンのコピー数はまちまちで、例えば、*Bacillus subtilis* ATCC 23857 は 10 コピーであるが、*Helicobacter pylori* 26695 や *Aquifex aeolicus* VF5 などは 2 コピーである(Klappenbach et al., 2001)。

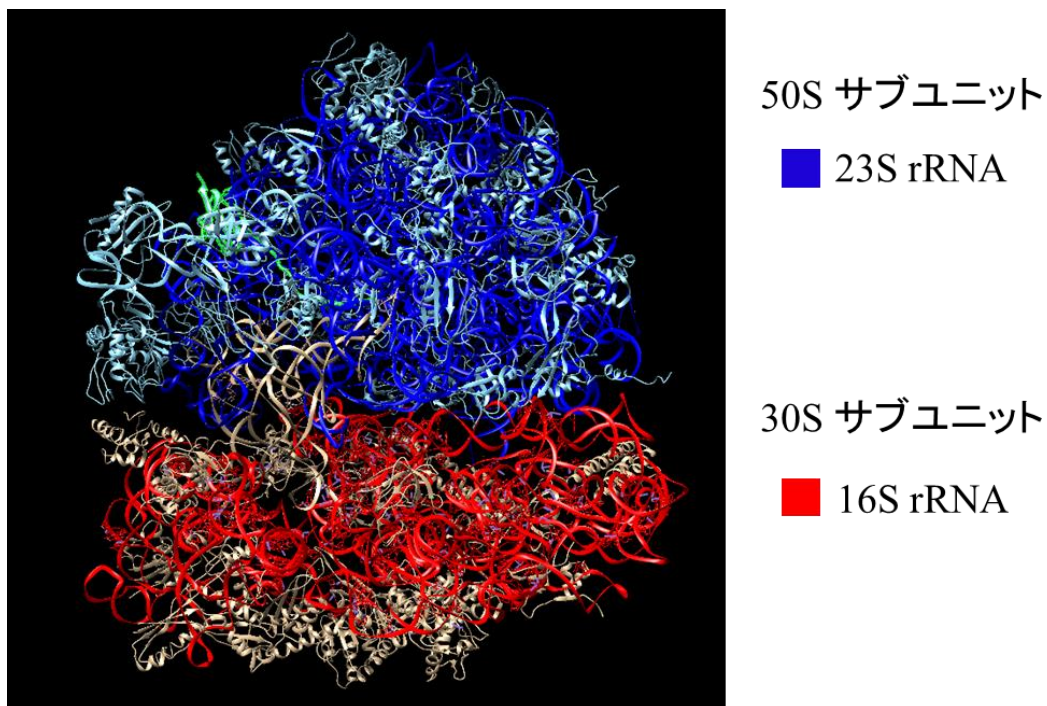


図 1.1 *Escherichia coli* のリボソーム 70S サブユニットの 3 次元構造。構造は、30S (PDB id 2WDK) と 50S (PDB id 2WDL) を合わせて表示した (Voortees et al., 2009)。青色のリボンは、23S rRNA を表し、赤色のリボンは、16S rRNA を表す。

Escherichia coli MG1655 リボソーマル RNA オペロン B の構造

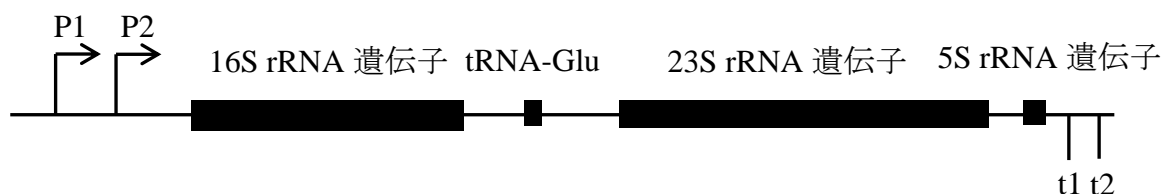


図 1.2 *Escherichia coli* MG1655 株のリボソーマル RNA オペロン B の構造

3. 多重遺伝子と適応進化

リボソーマル RNA オペロンに含まれる 16S rRNA 遺伝子、23S rRNA 遺伝子、5S rRNA 遺伝子のように、ゲノム上に複数コピー存在する遺伝子は多重遺伝子と呼ばれ、適応進化において重要な役割を果たしていることが知られている (Nei and Rooney, 2005)。リボソーマル RNA オペロンのほかには、低温ショックタンパク質 (CSP) が多重遺伝子の例として知られている。CSP 遺伝子は、70 アミノ酸残基程度の大きさの遺伝子で、大腸菌においては 9

個 (CpsA-CpsI) が存在しており、それぞれに機能が異なっている。例えば、CspA と CspB は低温環境下において発現し、mRNA 上の余分な二次構造を一本鎖へと解く分子シャペロンとしての機能を有している。この一方で、CspC と CspE は転写アンチターミネーターとしての機能を有している (Bae et al., 2000)。配列相同性の点においても、大腸菌の 9 個 (CpsA-CpsI) が平均してアミノ酸配列レベルで 67% 程度の相同性を有する中で、CspA と CspB は 80% 程度の、CspC と CspE は 84% 程度の相同性を有している。

霊長類においては、Zhang ら(1998)は、旧世界猿が新世界猿から分岐したのちに、EDN 遺伝子(Eosinophil Derived Neurotoxin gene)が重複して、ECP 遺伝子(Eosinophil Cationic Protein gene)に進化したことを報告している。ECP 遺伝子の系統では、EDN 遺伝子からの重複後に、アルギニンへの置換が多く生じており、EPC は正に荷電したアルギニンが病原体の細胞膜に穴をあけることによってバクテリアや寄生虫を殺傷することが知られているため、重複遺伝子による適応進化の例として考えられている(Young et al., 1986; Rosenberg et al., 1989; Nei and Kumar, 2006)。

また、リボソーマル RNA オペロンの P1、P2 プロモーターは非常に強力で、生体内に存在する分子は、rRNA とリボソーマルタンパク質が圧倒的に多く、対数増殖期には細胞内のリボソームは 71,000 分子とも言われている(Bremer and Dennis, 2008)。このように、リボソーマル RNA オペロンの多重化はリボソーマル RNA を大量に発現するための方策である。

4. リボソーマル RNA オペロンごとの役割

リボソーマル RNA オペロンは、大量に発現していることが一つの特徴であるが、さらに生育状況に応じて転写量が変化することが知られている。例えば、*Escherichia coli* の 7 つのリボソーマル RNA オペロン間では、リボソーマル RNA オペロン E は、他のリボソーマル RNA オペロンに比べて、最少培地においてプロモーター活性が下がることが知られており、貧栄養状態での生育速度の調節に関わっていると考えられている(Condon et al., 1992)。また、リボソーマル RNA オペロン G は、転写調節因子である Fis タンパク質による転写調節を、他のリボソーマル RNA オペロンに比べて非常に受けにくいいため、Fis を介した転写活性の影響を減少させている(Condon et al., 1992)。このように、リボソーマル RNA オペロンは、EDN 遺伝子や ECP 遺伝子のように、種分化に沿ってそれぞれが分岐進化している訳ではないが、オペロン自身はほぼ相同でありつつも、コピー間の転写量を調節することにより環境変化に対し応答する仕組みを備えている例である。

5. 同一ゲノム上の 16S rRNA 遺伝子同士の均一化

重複遺伝子において、染色体の不等乗換えや遺伝子変換、遺伝的組換えによって配列の均一化が生ずることが哺乳類の Alu 配列の解析などによって知られている (Ohta, 1976, 1985; Smith, 1976; Jeffreys, 1979; Slightom et al., 1980)。バクテリアにおいては、調査した限りリボソーマル RNA においてのみ、このような配列の均一化が生じているようだ。同一ゲノム上

に複数コピー存在するリボソーマル RNA オペロン間の 16S rRNA 遺伝子同士は、ほぼ相同な配列を有しており（たとえば、*Escherichia coli* MG1655 では、7 コピーの 16S rRNA 遺伝子間で平均 99.5%の相同性を有している）、同一ゲノム上の 16S rRNA 遺伝子同士で配列の均一化が働いていると考えられる（Hashimoto et al., 2003; Acinas et al., 2004）。Hashimoto ら (2003)によると、*Escherichia coli* において、リボソーマル RNA オペロン B 上の 16S rRNA 遺伝子は世代あたり 5×10^{-9} 程度の頻度で遺伝的組換えにより同一ゲノム上の他の 16S rRNA 遺伝子と均一化されており、Lee ら(2012)によってゲノムワイドに計算された塩基置換の頻度（世代あたりサイトあたり 2.2×10^{-10} ）よりも 20 倍ほど高い。

6. バクテリアにおける遺伝子の水平伝播

個体から個体へと遺伝子が伝達されることを遺伝子の水平伝播という。水平伝播の遺伝学的な仕組みについては、自然形質転換 (Griffith, 1928) や接合伝達 (Lederberg and Tatum, 1946)、トランスポゾンを介した DNA の伝播 (Grinsted, 1990; Tsuda and Iino, 1987) などが知られている。

バクテリアにおいては、遺伝子の水平伝播は普遍的に存在することが知られている。例えば *Escherichia coli* のゲノムのおよそ 18% は *Salmonella* 属と分岐して以降の一億年の間に水平伝播によって蓄積したものと見積もられている (Lawrence and Ochman, 1997)。また、Nelson ら(1999) は、*Thermotoga maritima* においては、ゲノムの 24% がアーキア由来であると見積もっている。このように、バクテリアにおいて普遍的に観察される水平伝播であるが、Rivera ら(1997) は、転写、翻訳、複製に関連する遺伝子に加え、GTPases、ATPase、tRNA 合成酵素などを含む情報系遺伝子と、アミノ酸合成、補因子の生合成、エンベロップ、エネルギー代謝などを担う操作系遺伝子 (Operational Gene) の間で水平伝播の頻度は異なると指摘している。

Rivera ら(1997) の結果を踏まえ、Jain ら(1999) は、複雑な相互作用の中におかれている情報系遺伝子は水平伝播を受け入れないという Complexity Hypothesis を提唱している。一例として、リボソームを構成する因子間の複雑な相互作用を示した (図 1.3)。この仮説を支持する結果は、Wellner と Gophna(2008)、Lercher と Pal(2008)、Cohen ら(2011) などによって得られており、例えば Cohen ら(2011) は、STRING (version 8) データベースに登録されているバクテリア全種に渡るタンパク質-タンパク質相互作用と、遺伝子ファミリーごとの水平伝播頻度を解析することによって、タンパク質-タンパク質相互作用が大きい遺伝子ほど水平伝播の頻度が少ないことを報告している。

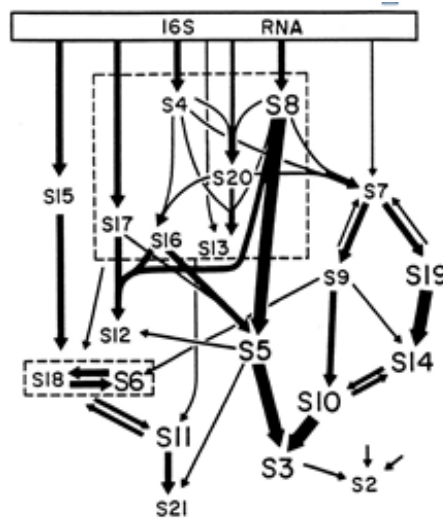


図 1.3 リボソームの構成成分間の複雑な相互作用。Jain ら(1999)より転載（出版元：米国科学アカデミー）。矢印の太さは相互作用の強さを表す。Jain ら(1999)によって提唱された Complexity Hypothesis では、情報系遺伝子はこのように複雑な相互作用の中におかれているために、水平伝播を受け入れないとされている。

7. 系統解析の指標として用いられてきた 16S rRNA 遺伝子

16S rRNA 遺伝子は、系統関係の推定や環境中の微生物の多様度の測定、種の同定などに最も良く用いられてきた遺伝子である (Woese, 1987; Woese et al., 1990; Head, 1998; Hugenholtz et al., 1998)。古くは、バクテリアの系統解析を行う上でシトクローム c が優れた分子時計として考えられ用いられてきた (Errede et al., 1978)。しかし、シトクローム c は、一部のバクテリアの種において遺伝子サイズが大きく変化しており、バクテリアを通して遺伝子サイズが似通っている 16S rRNA の方がより分子時計としてふさわしいと考えられるようになった (Ambler et al., 1979; Dickerson, 1980; Woese, 1987)。16S rRNA 遺伝子が最もよく用いられてきた理由の前提として DNA の配列決定方法が発展したことが挙げられるが、他の遺伝子に対する優位性はその他にもいくつかあり、主に以下の 3 つである。

- 1) 全生物に共通して保存されている遺伝子である。
- 2) 相同性の高い領域と低い領域が存在するため、ユニバーサルプライマーの設計が容易であるうえに、様々な幅の系統群を比較することが可能。
- 3) 翻訳に関わる遺伝子(情報系遺伝子)なので、水平伝播の影響を受けない(Complexity Hypothesis)。

特に、3)に記したような 16S rRNA 遺伝子の種特異性は幅広く信じられており、過去 30 年にわたって信じられてきた原核生物の系統分類は 16S rRNA 遺伝子を用いて行われている (Woese, 1987; Woese et al., 1990)。また、16S rRNA 遺伝子の相同性から種の判定をする

基準が報告されている (Stackebrandt と Goebel(1994)によって 97.0%、後に Stackebrandt と Ebers(2006)によって 98.7-99.0%)。これらの基準は、DNA-DNA 分子交雑法によって交雑形成率が 70%を越す場合同種と考えるという古典的な分類学の基準を満たす、16S rRNA 遺伝子の塩基配列相同性である(Wayne et al., 1987)。

8. 16S rRNA 遺伝子の自然界における水平伝播の報告

この一方で、16S rRNA 遺伝子の自然界における水平伝播の報告も存在する(Mylvaganam and Dennis, 1992; Eardly et al., 1996; Wang et al., 1997; Lan and Reeves, 1998; Ueda et al., 1999; Yap et al., 1999; Wang and Zhang, 2000; Parker., 2001; Schouls et al., 2003)。これらの報告を表 1.1 にまとめた。これらの報告は主に、(1)16S rRNA 遺伝子の部分的な水平伝播。(2)リボソーマル RNA オペロンが丸ごと水平伝播している例。(3)同一ゲノム上の 16S rRNA 遺伝子同士の間での配列の大規模な不一致に大別される。

部分的な 16S rRNA 遺伝子の水平伝播の例として、例えば、Wang と Zhang (2000)は、目視による配列比較と系統樹解析によって、Actinobacteria 門、Streptosporangiaceae 科における 16S rRNA 遺伝子の部分的な水平伝播を報告している (*Nonomuraea* 属と *Microtetrastroma* 属間、*Nonomuraea* 属と *Streptosporangium* 属間、*Microtetrastroma* 属と *Streptosporangium* 属間)。Wang と Zhang (2000)は、Jain ら(1999)が提唱した Complexity Hypothesis によって水平伝播しない遺伝子であると考えられてきた 16S rRNA 遺伝子の水平伝播について、部分的な構造の置換ならば許容されうるとする Simplified Complexity Hypothesis という考え方を提唱した。また、Schouls ら(2003)は、16S rRNA 遺伝子の 1014 残基-1055 残基の領域が *Streptococcus anginosus* 内の系統間で水平伝播していることを、配列の目視による比較と系統樹解析、サザンハイブリダイゼーション法によって報告している。Schouls ら(2003)は、比較した *Streptococcus anginosus* の系統の持つ全ての 16S rRNA 遺伝子が組換え体であったことを踏まえ、部分的な水平伝播により異種由来の断片を持つ 16S rRNA 遺伝子が他の 16S rRNA 遺伝子を乗っ取り、組換え 16S rRNA 遺伝子のコピー数が増えることを報告した。

リボソーマル RNA オペロンが丸ごと水平伝播している例として、Yap ら(1999)は *Thermomonospora chromogena* のゲノム上に存在する異なる 2 タイプのリボソーマル RNA オペロンのうち片方は、*Thermomonospora chromogena* と同目異亜目に属する *Thermobispora bispora* に対して高い相同性を示し、系統樹解析においても *Thermomonospora chromogena* の他のリボソーマル RNA オペロンではなく、*Thermobispora bispora* のリボソーマル RNA オペロンと近縁であったことなどから、リボソーマル RNA オペロンが丸ごと水平伝播した例であるとして報告している。

同一ゲノム上の 16S rRNA 遺伝子同士の間での配列の大規模な不一致の例としては、*Thermomonospora chromogena*, *Thermobispora bispora*, *Haloarcula marismortui* において同一ゲノム上の 16S rRNA 遺伝子や 23S rRNA 遺伝子のコピー間での 10%程度に及ぶ配列の不一致が報告されており、16S rRNA 遺伝子の水平伝播の存在を示唆する結果として報告されてい

る(Mylvaganam and Dennis, 1992; Wang et al., 1997; Yap et al., 1999)。

他方で、水平伝播によってこのような同一ゲノム内での 16S rRNA 遺伝子配列の大規模な不一致が生ずる現象は、稀であるとの報告も存在する (Tian et al., 2015)。Tian ら(2015)は、2,143 個のバクテリアゲノムのうち、16S rRNA 遺伝子におけるゲノム内多型が 2%を超えるゲノム 28 個について個別に調査し、このうち 15 個のゲノムにおいて同一ゲノム内での 16S rRNA 遺伝子配列の大規模な不一致が水平伝播によるものであるとし、水平伝播が生ずる系統的な範囲は種間、ないし属間のみであるとしている。しかしながら、Tian ら(2015)の解析には、(1)ゲノム内多型を持つゲノムのみに着目している。(2)水平伝播か否かの判定に系統樹を用いている。(3)系統樹作成の際の配列選択を、BLAST を通じて行っている。と言った問題点が存在する。(1)については、ゲノム内多型が見られる 16S rRNA 遺伝子に限定して解析を行っているため、外来の 16S rRNA 遺伝子によってゲノム内の 16S rRNA 遺伝子すべてが均一化されたような例を検出することはできない。その上、2%以上のゲノム内多型に着目しているため、それ以下のゲノム内多型を有するゲノムについては解析の手が及んでいない。特に、非常に遠い系統からの水平伝播であったとしても、それが部分的であれば配列全体の多型は 2%以下に納まることは十分に考えられる。(2)については、系統樹解析を通じて水平伝播の検出を行う場合、系統樹の系統関係に支配的に作用するほどの変異量が水平伝播によってもたらされている必要があり、部分的な水平伝播の検出においては適した解析手法とは言えない。(3)については、組換え体の親候補を選ぶ際に BLAST を用いているため、水平伝播が部分的であった場合、必ずしも外来領域と相同性の高い配列が選ばれてくるとは限らない。以上のように、Tian ら(2015)の解析では、十分に 16S rRNA 遺伝子の水平伝播の可能性が網羅的に検証されたとは言えない。

9. 他の翻訳系遺伝子における水平伝播

16S rRNA 遺伝子のほかにも翻訳に関わる遺伝子の水平伝播が報告されている。Yuri ら(1999)は 20 アミノ酸に対応するアミノアシル tRNA 合成酵素の系統解析を 3 界に渡って行うことにより、アミノアシル tRNA 合成酵素が界を超えて水平伝播していることを報告している。リボソームを構成するタンパク質をコードしている遺伝子の水平伝播についても報告が存在する。Garcia ら(2002)は、Actinobacteria 門に属する *Arthrobacter sp.* のリボソーマルタンパク質 L27 が、異なる門である Firmicutes 門に属する *Bacillus* と近縁であることを系統解析により報告しており、リボソーマルタンパク質 L27 が水平伝播していることを報告している。以上のように、水平伝播しないと考えられてきた翻訳系を担う遺伝子であっても水平伝播が皆無であるということではない。

表1.1 これまでに報告された、16S rRNA遺伝子の水平伝播とコピー間での不一致。

年	著者	報告内容	報告分類群
1992	Mylvaganam and Dennis	16S rRNA遺伝子のコピー間での大きな多型	<i>Haloarcula marismortui</i>
1996	Eardly et al	16S rRNA遺伝子の部分的な水平伝播	<i>Rhizobium, Aeromonas</i>
1997	Wang et al	16S rRNA遺伝子のコピー間での大きな多型	<i>Thermobispora bispora</i>
1998	Lan and Reeves	リボソーマルオペロンが丸ごと水平伝播	<i>Vibrio cholerae</i>
1999	Ueda et al	16S rRNA遺伝子の部分的な水平伝播	<i>Streptomyces</i>
1999	Yap et al	リボソーマルオペロンが丸ごと水平伝播	<i>Thermomonaspora chromogena</i>
1999	Yap et al	16S rRNA遺伝子のコピー間での大きな多型	<i>Thermobispora chromogena</i>
2000	Wang and Zhang	16S rRNA遺伝子の部分的な水平伝播	<i>Actinomycetes</i>
2001	Parker	16S rRNA遺伝子の部分的な水平伝播	<i>Bradyrhizobium</i>
2003	Schouls et al	16S rRNA遺伝子の部分的な水平伝播	<i>Streptococcus anginosus</i>

10. 16S rRNA 遺伝子の水平伝播実験

自然界において 16S rRNA の水平伝播が報告されている一方で、当研究室では、16S rRNA 遺伝子の水平伝播を模した実験にも成功している (Kitahara et al., 2012)。Kitahara ら(2012)は *Escherichia coli* との相同性が 80%ほどの、綱レベルで異なる分類群の 16S rRNA 遺伝子でも *Escherichia coli* の生育を相補できることを見出し、16S rRNA の機能にとっては配列そのものというよりも二次構造の保守性が重要であることを見出した。この結果は、Jain ら(1999)が提唱した Complexity Hypothesis とは相反する結果で、16S rRNA を異種のもものと置換しても二次構造に大きな変化がなければ、他の成分との相互作用を失わないということを意味している。一般的に、リボソーマルタンパク質と rRNA 分子は、正の電荷を持ったアミノ酸残基と RNA 分子のリン酸酸素の間の塩橋を介して作用していることが知られており、核酸塩基そのものが認識されるわけではない (Conn et al., 1999; Wimberly et al., 1999; Allers et al., 2001; Brodersen et al., 2002)。Kitahara ら(2012)の報告においても、図 1.4 に示すように二次構造が大きく変化しない 16S rRNA 遺伝子の水平伝播ならば十分に許容される余地があることが示された。

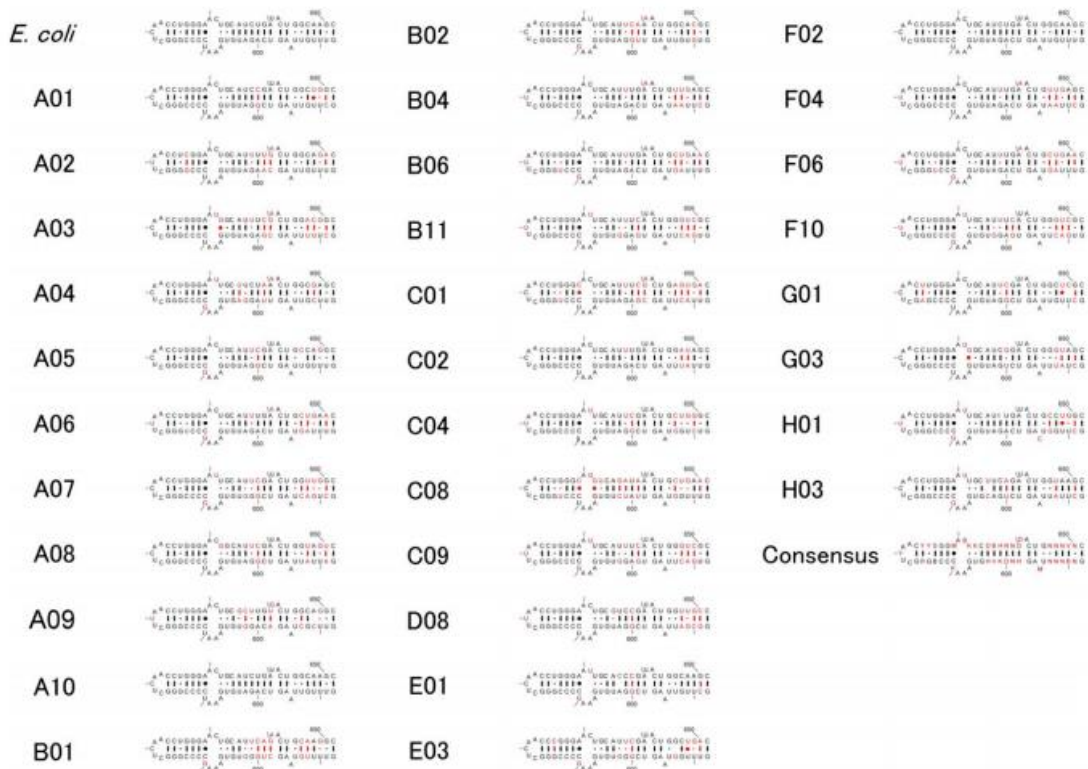


図 1.4 *Escherichia coli* の生育を相補した異種由来の 16S rRNA の、h21(helix 21)領域の二次構造。Kitahara ら(2012)より出版元（米国科学アカデミー）転載。*Escherichia coli* の 16S rRNA と同じ二次構造をとる異種由来の 16S rRNA ならば、*Escherichia coli* の生育を相補した。二次構造は、Cannone ら(2002)のウェブサイトを参考にした。

11. 系統樹による解析の原理的解析限界

遺伝子の進化史を明らかにするために一般的用いられる系統樹作成法は、何かしらの生物学的な原理やモデル、哲学的な思想などを元に系統樹を作成する。例えば、Saitou と Nei (1987)によって開発された近隣結合法は、ある配列数から作成されるすべての樹形において、枝長の総和が最小の樹形が最も優れているという最小進化原理に基づいている (Nei and Kumar, 2006)。他方で、最大節約法は、ある過程を説明するのに最良の方法は最小限の仮定を必要するという William of Ockham の哲学思想（オッカムの剃刀）を理論的根拠としている (Nei and Kumar, 2006)。このほかにも、最尤法、UPGMA 法、最小二乗法などの系統樹作成方法が開発されているが、いずれの手法においても共通していることは、**全ての塩基サイトが同じ進化史を辿っており、二分岐型の系統樹で記述可能である**という仮定である。

理論的には、ある種や遺伝子が分岐する時や遺伝子重複が生じた時は、1つの塩基配列が2つへと分岐するため、進化史を反映した真の系統関係は二分岐型の系統樹で表現可能かもしれない。しかし、遺伝的組換え、遺伝子変換、遺伝子融合、水平伝播などが生じていると、塩基サイトの進化史が部分的に異なることとなり、**全ての塩基サイトが同じ進化史を**

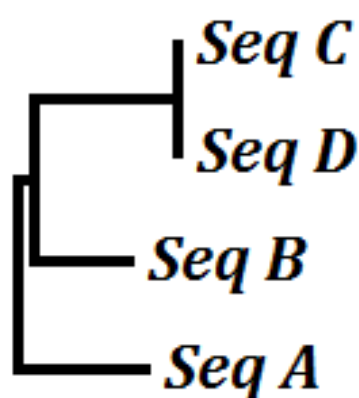
込っているという仮定を置いている系統樹作成方法では原理的に進化史を記述できない。

図 1.5(A)に示す遺伝的組換えを含む仮想的なデータセットから、近隣結合法によって系統樹を復元する場合について例として考えてみる。図 1.5(A)において、Seq A は 4 配列のうちで最も初期に分岐した配列(外群)、Seq C は Seq B と Seq D の間の遺伝的組換えによって生じた配列である。遺伝的組換えが生じた位置はサイト 9 以降なので、Seq C のサイト 8 よりも上流は青色で示す Seq D 由来で、9 サイト以降は赤色で示す Seq B 由来である。したがって、Seq C のサイト 8 よりも上流は図 1.5(B)に示す系統関係が正しく、サイト 9 以降は図 1.5(C)に示す系統関係が正しい。このように、**遺伝的組換えが生じた配列は、領域ごとに系統関係が異なる**。しかし、図 1.5(D)に示された、このデータセット全長を用いて近隣結合法で作成した系統樹の系統関係は、サイト 8 よりも上流で作成した系統樹と同様の系統関係しか示されておらず、サイト 9 以降の系統関係が遺伝的組換えによって異なっているということは図 1.5(D)の系統樹からは読み取ることができない。

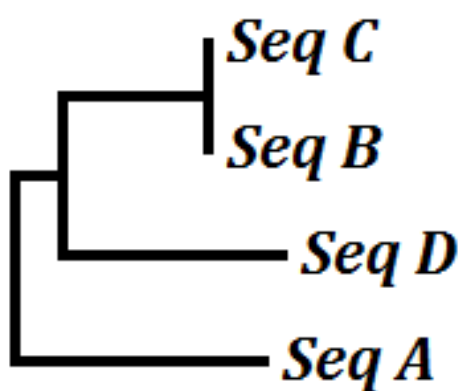
(A) 遺伝的組換えを含む仮想的なデータセット

	1	2	3	4	5	6	7	8	9	1	1	1	1	1	1	1	1
Seq A	G	T	A	A	A	A	A	T	A	C	A	T	T	T	A	T	T
Seq B	T	C	A	T	A	T	A	A	A	A	A	C	A	A	A	C	G
Seq C	T	T	C	T	T	T	C	T	A	A	A	C	A	A	A	C	G
Seq D	T	T	C	T	T	T	C	T	G	A	G	C	T	A	G	C	T

(B) サイト 1-8 領域の系統樹



(C) サイト 9-17 領域の系統樹



(D) 全領域の系統樹

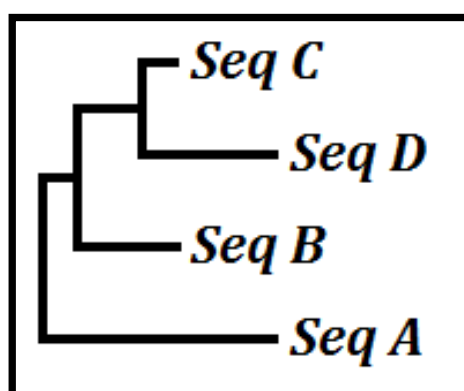


図 1.5 系統樹によって遺伝的組換えを検出できない原理的な理由。(A)組換えによって生じた配列 Seq C を含む仮想的なデータセット。最も初期に分岐した配列(外群)は Seq A で、Seq C は、Seq D(青色)のサイト 1-8 領域と Seq B(赤色)のサイト 9-17 領域が合わさった組換え体。(B) サイト 1-8 領域を用いて作成された系統樹。(C) サイト 9-17 領域を用いて作成された系統樹。(D) 全領域を用いて作成された系統樹。全ての系統樹は近隣結合法を用いて作成し、距離行列の計算には p-距離を用いた。

12. 複数通りの進化史を同時に表現できる系統ネットワーク法

このように、一つの遺伝子であっても部分的に異なる進化史をたどってきた遺伝子の進化を記述するためには、「ループ構造」を用いた表現が必要となる (Nei and Kumar, 2006)。Bandelt (1994)は、ループ構造をとることを許して節約原理に基づく系統樹を作成する手法「系統ネットワーク法」を考案した。SourdisとNei (1988)は、(1)配列間の違いが10%以下程度で、(2)配列間の進化速度がほぼ一定で、(3)調べられる塩基の数が多い場合、真の樹形を得るために最大節約法は優れていることを報告している。これらを、節約原理に基づく系統樹作成方法の適用範囲であると考え、Bandeltの系統ネットワーク法は節約原理に基づく系統樹の複合体と考えることができるので、配列間の違いが10%以下程度の配列間の比較に用いることができると考えられる。たとえば、Kitanoら(2012)は、ヒトのABO式血液型遺伝子の進化史における遺伝的組換えを含む進化史を、Bandeltの系統ネットワーク法を用いることによって明らかとしている。

Bandeltの系統ネットワーク法は節約原理に基づく手法であるが、距離行列から系統ネットワークを作成する手法も存在する。BryantとMoulton (2004)は、SaitouとNei (1987)によって開発された近隣結合法を改変することで、Neighbor-Net法を開発した。近隣結合法は以下のような過程で系統樹を作成する。

- (1) 種のクラスターが全く無いと仮定した星状系統樹を仮定する。
- (2) ある2つの配列を結合させたときの樹長の総和を、全ての2配列の組み合わせについて計算し、最小の樹長の総和を示す2配列を近隣として選ぶ。
- (3) 近隣として選ばれた2配列を合体配列として1つの配列とし、手順2)を二分岐型の系統樹が得られるまで繰り返す。

Neighbor-Net法は、手続き(2)において近隣として2配列が選ばれた際に、合体種として結合せずに、すでに近隣として選ばれている配列に2つ目の近隣が見つかるまで探し続ける。このような、近隣結合法に変更を加えた手続きをとることによって、Neighbor-Net法は系統ネットワークを作成する。

13. 系統ネットワーク法と系統樹について

近隣結合法や最尤法で作成される一般的な系統樹は二分岐型の樹形を持っており、データセットから一種類の進化経路を推定する手法である。従って、遺伝的組換えや水平伝播が生じているデータセットのように、領域的に系統関係が大きく食い違うような配列間の進化史を正確に記述することはできない。系統ネットワーク法は、二分岐型の樹形を仮定せずに、伝的組換えや水平伝播をループ構造により表現する手法である。以下に、モデルデータを用いて系統樹と系統ネットワーク法の違いについて説明する。

図 1.6 に示したデータセットを用いて最大節約法により系統樹を復元する場合について考えてみる。最大節約法は、与えられたデータセットを、最少の塩基置換数で説明できる二分岐型の系統樹を最も優れた系統樹として選ぶ手法である。したがって、最少の塩基置換数で説明できる系統樹が何通りか出現することもあり得る。図 1.7 の(A)と(B)は、いずれも図 1.6 に示したデータセットを元に最大節約法を用いて作成した系統樹である。系統樹の各枝に図 1.6 の各サイトで生じた塩基置換を記入した。どちらの系統樹も、図 1.6 のデータセットを 9 回の塩基置換で説明できる。赤で示したサイトは、2 回の塩基置換が生じたことを想定せざるを得なかったサイトである。

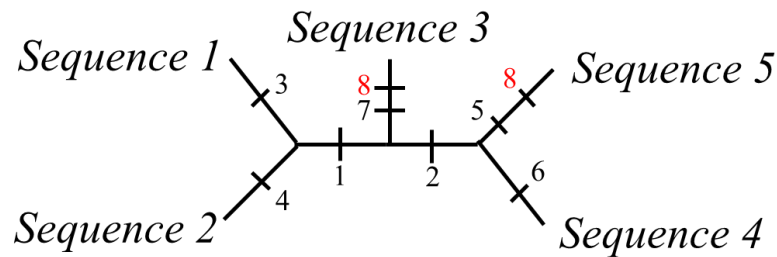
図 1.7(A)のサイト 8 や、図 1.7 (B)のサイト 2 のように、複数の系統で独立して同じ方向の塩基置換が同じサイトに生ずる現象を平行進化と言う。平行進化が存在することは、節約法による系統樹解析においては、ある樹形を想定した時に初めて発見できる。図 1.6 のデータセットの場合だと、サイト 2 の塩基置換数を 1 回と考える図 1.7(A)の系統樹に沿って考えると、サイト 8 は平行進化と考えられ、サイト 8 の塩基置換を 1 回と考える図 1.7(B)に沿って考えると、サイト 2 は平行進化ということになる。

Bandelt の系統ネットワーク法 (Bandelt, 1994)は、複数通りの節約系統樹を、ループ構造を許し組み合わせることによって系統ネットワークを作成する手法である。例えば、図 1.6 のデータセットを用いて、Bandelt の手法により系統ネットワークを作成すると、図 1.8(A)のようになる。図 1.8(A)には、図 1.6 のデータセットから推定された 2 通りの系統樹を、ループ構造を許すことで含んでおり、図 1.8 (B)で示した赤色の進化経路は、図 1.7(A)の系統樹によって示唆されたもので、図 1.8(C)で示した緑色の進化経路は、図 1.7(B)の系統樹が示唆する進化経路である。

	1 2 3 4 5 6 7 8
<i>Sequence 1</i>	AATAAACA
<i>Sequence 2</i>	AAAGAACA
<i>Sequence 3</i>	GAAAAAGG
<i>Sequence 4</i>	GTAAAGCA
<i>Sequence 5</i>	GTAACACG

図 1.6 系統ネットワーク法と系統樹の違いを説明するための 5 本の配列から成る仮想的なデータセット。上部に各サイトの番号を記した。また、多型が見られた塩基を赤色でハイライトした。

(A)



(B)

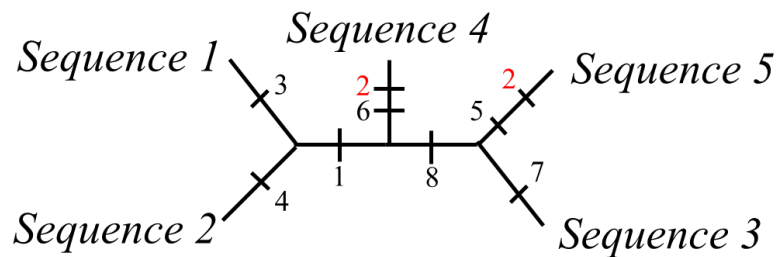
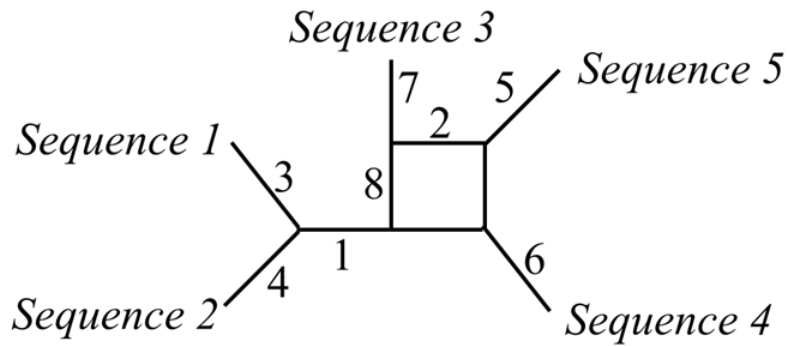
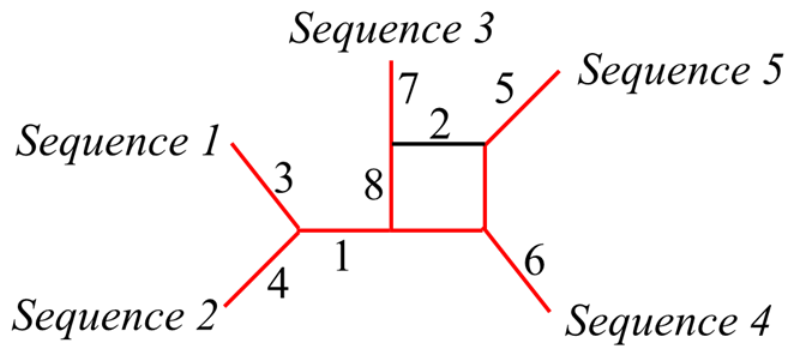


図 1.7 図 1.6 のデータセットを用いて、最大節約法によって作成した系統樹。(A)と(B)にそれぞれ示した 2 通りの樹形が最も節約的な樹形として選ばれた。図 1.6 のデータセットの各サイトにおける塩基置換を枝の近くに記した。赤色で示されたサイトは、各系統樹において平行進化と考えられるサイトである。

(A)



(B)



(C)

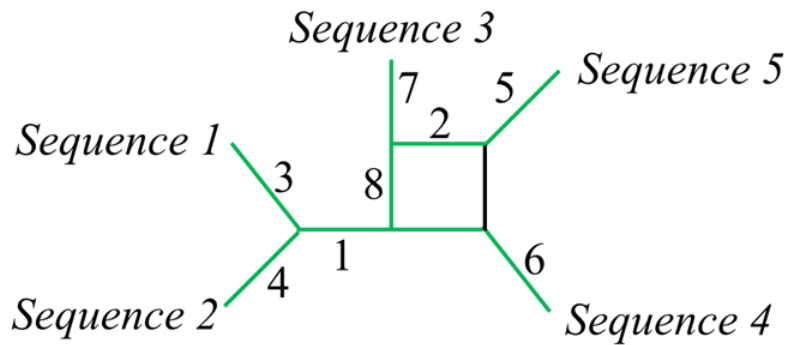


図 1.8 (A)節約的な方法(Bandelt 1994)を用いて作成した系統ネットワーク。(B)において赤色でハイライトした経路は、図 1.7 における系統樹(A)の経路で、(C)において緑色でハイライトした経路は、図 1.7 における系統樹(B)の経路である。このように、節約的な方法で作成された節約的な系統ネットワークは、あり得る全ての節約経路を、ループ構造をとることによって含んでいる。

14. Bandelt の方法を用いた系統ネットワークの作成方法

系統ネットワークを作成する手順を、図 1.9 に示したデータセットを例に説明する。サイト 1 と 13 は、同じパターンを示しており、図 1.10(A)に示す系統関係でそれぞれ 1 回の塩基置換で説明できる。同様に、サイト 2,7,8、サイト 5,12、そして、サイト 9,10 が同じパターンを示しており、図 1.10(A)に対して、さらにサイト 2,7,8 を 1 回の塩基置換で説明できる枝を加えると図 1.10(B)のようになる。平行な辺は同じ分岐を示唆するサイトをあらわす。サイト 1,13 は Sequence 1,2,3 / 4, 5 の分岐を示唆し、サイト 2,7,8 は Sequence 1,2,4 / 3, 5 の分岐を示しており、図 1.10(B)はそれぞれのサイトから示唆される節約的な系統関係をすべて含んでいる。図 1.10(B)のネットワークにさらに、サイト 5,12 とサイト 9,10 を加えることによって、図 1.10(C)の系統ネットワークが得られる。最後に各配列固有のシングルTONを外部枝として反映させることで、図 1.10(D)に示す系統ネットワークが得られる。

しかし、系統ネットワークへと反映することが難しいサイトも存在する。図 1.9 のデータセットのサイト 14 には 3 種類の塩基が存在しており、このようなサイトのパターンが実現するためには、最低でも 2 回はそのサイトに塩基置換が生じていなければならない。このサイトによって、図 1.11 の(A)(B)(C)に示す 3 通りの系統関係が節約的に示唆される。このように、節約的に系統関係を確定できないサイトを系統ネットワークに記述するためには、例えば、図 1.12 のように三角形で示すような方法が存在するが、三角形が存在すると系統ネットワークは視覚的に極めて複雑になる上に、実際この三角形の情報が進化史の解析に有効に働くことは少ない。したがって、本研究においては、このように系統関係を節約的に確定できないサイトは、進化史を推定するにあたり、有用である場合を除いて解析から取り除いた。進化史を推定するにあたり有用である場合とは、生物学的な理由により進化史を確定できる場合や、3 通りの系統関係のすべてを吟味することによって進化史を推定できる場合である。図 1.10 の系統ネットワークに反映されたサイトのうち、シングルTON以外のサイトは何らかの系統関係が節約的に示唆されている。これらのように、節約的に系統関係が確定するサイトは、「節約的に情報を持つサイト」と呼ばれる。

11111
12345678901234

Sequence 1 AATACACAACCAGC
Sequence 2 AAAGCACATGCAGG
Sequence 3 ATAAAAGTTGCGGC
Sequence 4 GAAAAGCAACGGCA
Sequence 5 GTAAAAGTTGGGCC

図 1.9 5本の配列から成る、節約的な系統ネットワークの作成方法を説明するための仮想的なデータセット。各サイトの上部にサイト番号を記した。異なる塩基は赤色でハイライトした。

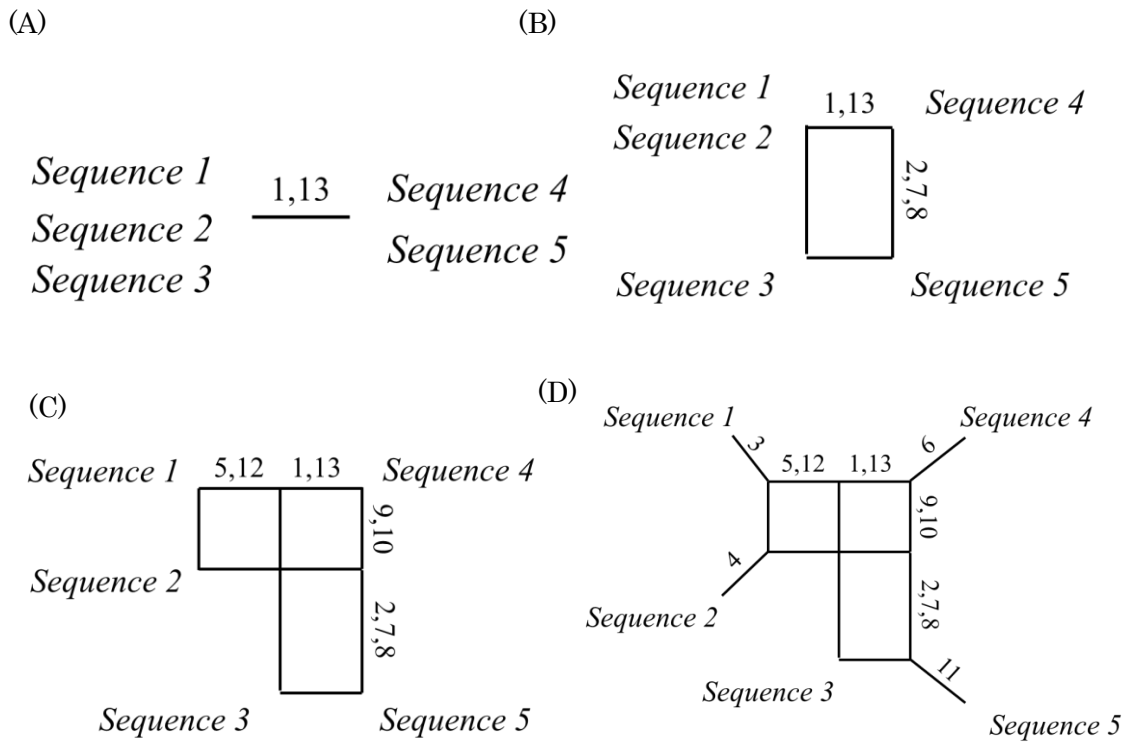


図 1.10 図 1.9 に示したデータセットから節約的な系統ネットワークを作成する手順。各枝のそばに対応するサイト番号を記した。平行する辺は同じサイトを表す。

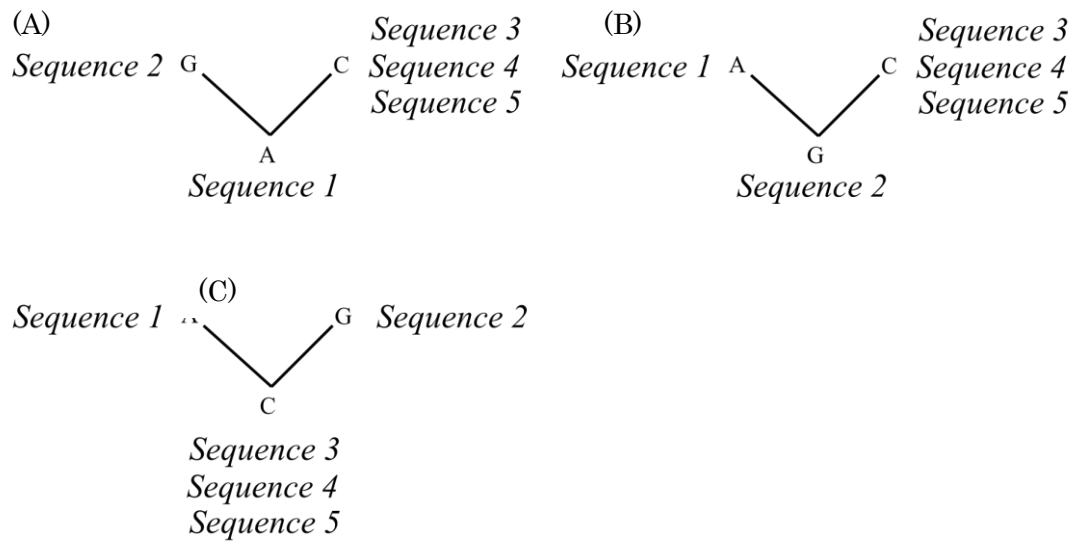


図 1.11 (A)(B)(C): 図 2.5 のデータセットのサイト 14 から考えられる 3 通りの系統樹

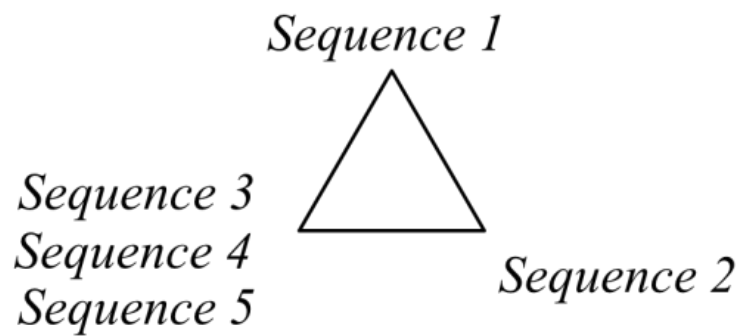


図 1.12 図 1.9 のデータセットのサイト 14 を節約的な系統ネットワークに反映させる方法。本研究においては、必要のない限り、このような 3 塩基以上の多型を示すサイトは取り除いた。

15. 系統ネットワーク法による水平伝播・遺伝的組換えの検出

古くから用いられている遺伝子の水平伝播や遺伝的組換えを検出する方法は、系統樹の比較である。例えば、Feil ら(2001)は、5つのハウスキーピング遺伝子についてそれぞれ系統樹を作成し、ある遺伝子を用いて作成した系統樹の系統関係が、他の4つの遺伝子から推定された系統樹の系統関係と異なるかどうかを検証することで、水平伝播を検出した。このように、異なる領域で系統樹を作成し、異なる系統関係が得られた場合、水平伝播や遺伝的組換えの存在が示唆される。しかし、これらの手法は、解析者が恣意的に遺伝子や領域を定めて、その間で少なくとも水平伝播が生じているか否かを判定する手法であり、ある塩基配列の範囲において水平伝播や遺伝的組換えが生じているのか否かを解析することができるわけではない。

たとえば、図 1.13 に示された遺伝的組み換えを含む進化史を辿り、長さ n の配列 Sequence 1-Sequence 4 が進化してきたとする。Sequence 4 は、Sequence 2 の $1..i$ 領域と Sequence 3 の $i..n$ 領域間に T1 Million Years Ago (MYA) に生じた遺伝的組み換えによって生成された配列である。また、各系統樹の枝にそって生じた塩基置換を A-G とし、配列全体に一樣に生じるものと仮定する。もし、領域 $1..i$ を用いて作成した節約系統樹 (図 1.14 (A)) と、領域 $i..n$ を用いて作成した節約系統樹 (図 1.14 (B)) の比較を行ったとしたら、系統関係の対立を発見し、遺伝的組換えや水平伝播の存在が示唆されるであろう。A_{1...i} のような表記は、図 1.13 において生じた塩基置換 A のうち、サイト $1..i$ の間に生じた塩基置換のサイトを表している。しかし、普通は何の根拠もなく遺伝的組換えが生じた位置を知ることは不可能であり、このような解析が可能な場合はごく限られている。

これに対して、Bandelt の方法によって作成された系統ネットワークは、全てのサイトの節約的な分岐情報を含んでいるので、言うなれば、全てのサイト同士を節約的に比較していると言える。Sequence 1-Sequence 4 を用いて Bandelt の系統ネットワークを作成すると、図 1.15(A) に示す系統ネットワークが得られる。図 1.15 の系統ネットワークには、図 1.14 における 2 通りの進化経路が両方とも含まれている。すなわち、図 1.14(A) の進化経路を、図 1.15(B) で赤色にハイライトした部分に、図 1.14(B) の進化経路を図 1.15(C) に緑色でハイライトした経路に含んでいる。この 2 通りの進化経路は、対立しているため、それぞれが、系統ネットワークにおいてループ構造の対立する辺を成している。このように、Bandelt の方法によって作成される系統ネットワークは、恣意的に領域を定めることなく水平伝播や遺伝的組換えによるキメラ化を発見できる。

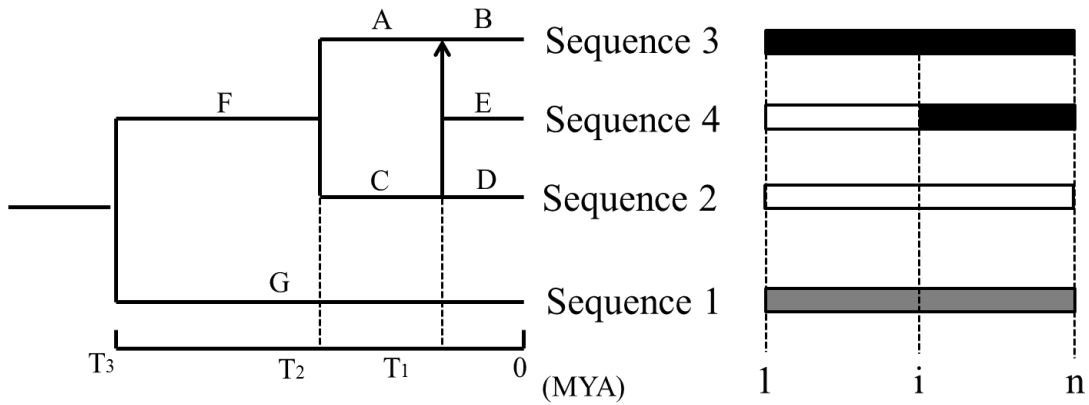


図 1.13 仮想的な進化史。Sequence 4 は Sequence 2 の 1...i 領域 Sequence 3 の i...n 領域の組換え体。各枝で生じた塩基置換をアルファベット A-G で表す。

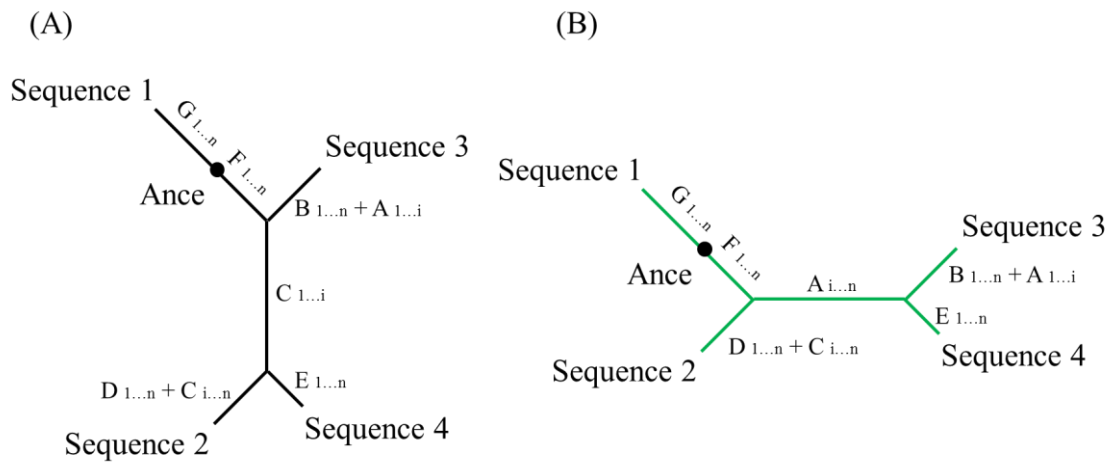


図 1.14 図 1.13 の進化史によって生じた配列、Sequence 1-Sequence 4 を用いて、最大節約法によって作成された系統樹。(A)は遺伝的組換えの前半領域(サイト 1-i)、(B)は遺伝的組換えによってキメラ化した領域(i-n)を用いて作成されている。A_{1...i} のような表記は、塩基置換 A のうち、サイト 1-i の間に生じたものであることを意味する。

16. Bandelt の方法で作成した系統ネットワークの特徴

図 1.15 (A)は以下のような特徴を持っている。

1. 遺伝的組換えが生じた領域と、そうでない領域が対立してループ構造を形成する。
2. 遺伝的組換えによって生成された組換え体は、外群の対角に位置する。また、組換え体の親配列は、組換え体を挟むように位置する。
3. 組換え体のシングルトン（固有の枝）は短い。
4. 親配列において、遺伝的組換えに関わった領域のシングルトンは、遺伝的組換えが生じた後に蓄積しているため、比較的少ない。
5. 外群には、配列全体に万遍なくシングルトンが見られる。

注意が必要なのは、実践的には、系統ネットワークにおいて遺伝的組換えが示唆されたからと言って、直ちに、最も分岐が古い配列の対角を組換え体であるとは判断できないということである。たとえば、Sequence 1-4 が図 1.16 のような進化史を辿って生成された配列であるとする、Bandelt の系統ネットワークは図 1.17 のようになる。実際に図 1.15(A)で示されている系統ネットワークと図 1.17 で示されている系統ネットワークは、進化史が異なるにも関わらず、系統ネットワークの各辺のサイトの分布は同じである。Bandelt の系統ネットワークを作成して遺伝的組換えの存在が示唆された、という段階ではまだ遺伝的組換えの親子関係はわからないということである。図 1.15(A)、図 1.17 の系統ネットワークにおいて共通してみられる特徴は、組換え体のシングルトンは短いということである。従って、遺伝的組換えの存在が示唆され、かつシングルトンが極端に短い配列が存在した場合は、組換え体であると判断できる (Saitou and Yamamoto, 1997; Kitano et al., 2009, 2012)。

図 1.15(A)の系統ネットワークと図 1.17 の系統ネットワークで異なる点は、各配列のシングルトンの分布である。図 1.15(A)の系統ネットワークにおいては、各配列の進化速度を一定と仮定すると、Sequence 1 のシングルトンは他の配列のシングルトンよりも長くなる。しかし、図 1.17 の系統ネットワークにおいては、進化速度が一定であっても遺伝的組換えが生じたサイト i や遺伝的組換えが生じた時期によっては、Sequence 1 のシングルトンは Sequence 2 のシングルトンよりも短いかもしれない。このことは、最も分岐が古い配列のシングルトンの長さが他の配列と相違ないような場合は、遺伝的組換えの親子関係を再検証する必要があるということである。また、最も分岐が古い配列であるにも関わらず、シングルトンサイトの分布に偏りが見られた場合も同様にその配列が遺伝的組換えに関わっている可能性を検証する必要がある。これらの問題を解決するためには、さらに古い時代に分岐した配列を加えて解析してみるという方法がある。さらに分岐が古い配列が図 1.15(A)の Ancestor (Ance) の位置に結節したとしたら、Sequence 1 は組換えに関わっていないと判断できるし、図 1.18 のように Sequence 1 のシングルトンブランチ上にさらに分岐が古い配列が結節しなかったとしたら、Sequence 1 は組換えに関わっている可能性があるということ

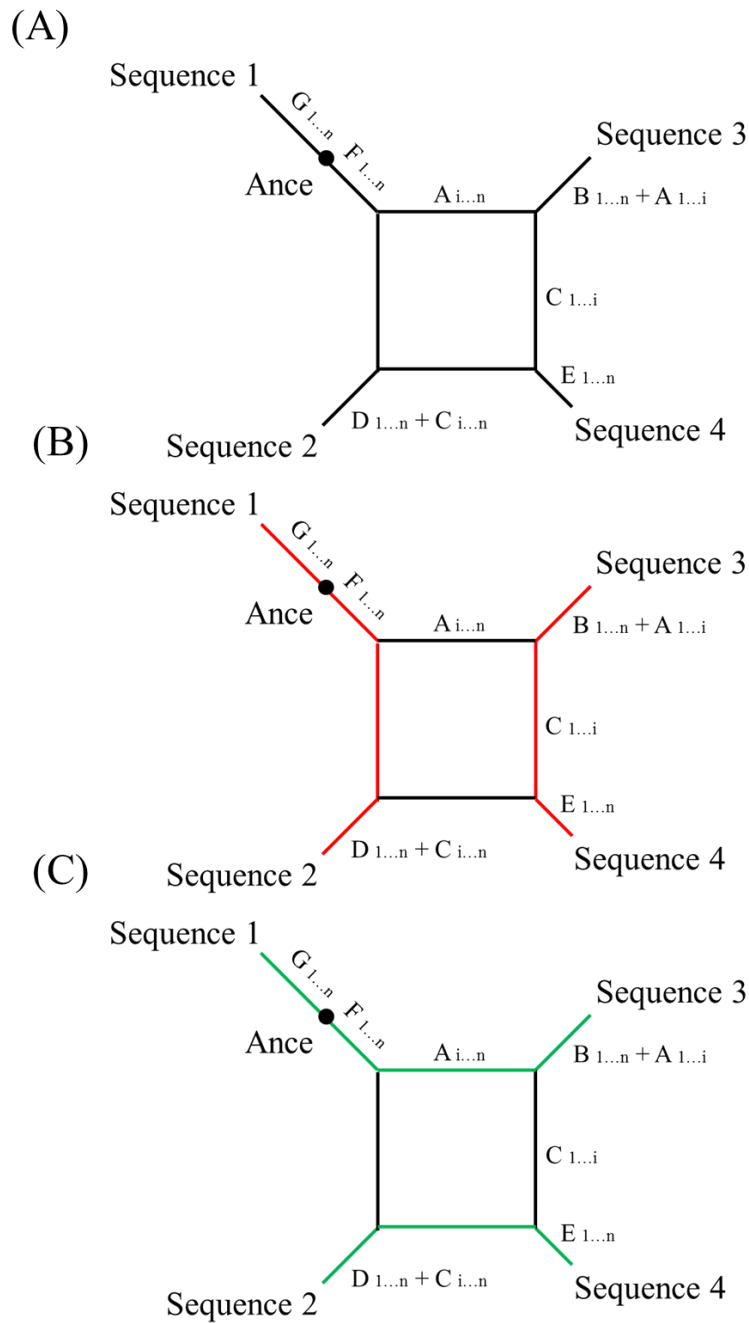


図 1.15 (A)図 1.13 の Sequence 1 – 4 を用いて作成した時に得られる節約的な系統ネットワーク。アルファベットは図 1.13 の各枝における塩基置換で、添え字はその塩基置換サイトが存在する領域をあらわしている。(B)の赤色で示した進化経路は、図 1.13 の(A)に示された系統樹の進化経路を、(C)の緑色で示した進化経路は、図 1.14 の(B)に示された系統樹の進化経路を、それぞれ系統ネットワークが含んでいることを表している。このように、節約的な系統ネットワークを作成することで、領域間での系統関係の対立を発見することができる。

である。

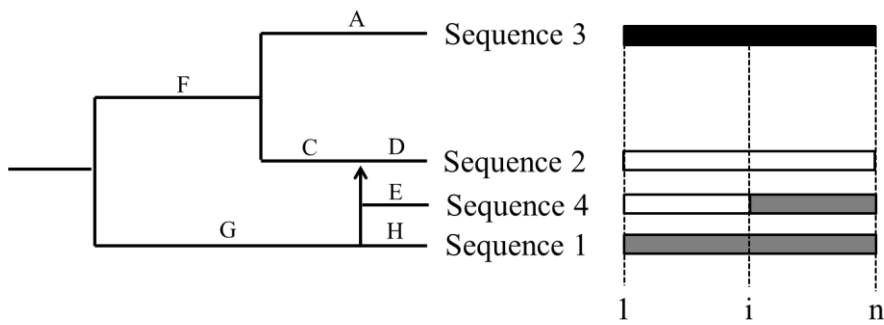


図 1.16 仮想的な進化史。Sequence 4 は Sequence 3 の 1... i 領域と Sequence 1 の i...n 領域のキメラ体。各枝で生じた塩基置換をアルファベット A – H で表す。図 1.13 では、Sequence 2 と Sequence 3 の間の遺伝的組換えによって Sequence 4 が生じたが、図 1.16 では、Sequence 1 と Sequence 2 の間の遺伝的組換えによって Sequence 4 が生じている。

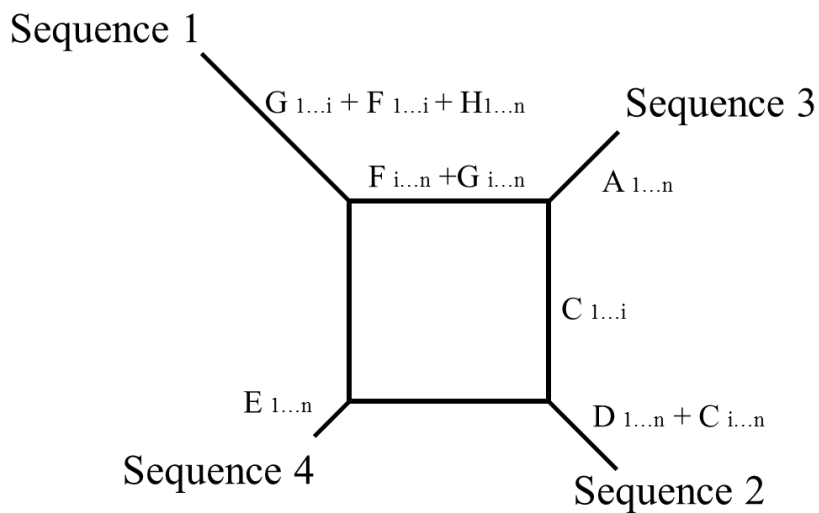


図 1.17 図 1.16 の進化史に従って進化してきた Sequence 1–4 を用いて作成した節約的な系統ネットワーク。アルファベットは図 1.16 の各枝における塩基置換で、添え字はその塩基置換サイトが存在する領域をあらわしている。図 1.15 に示された系統ネットワークとは異なる進化史で得られた系統ネットワークだが、ループ構造を構成するサイトの分布は同じなので、実際には見分けがつかない場合が多い。

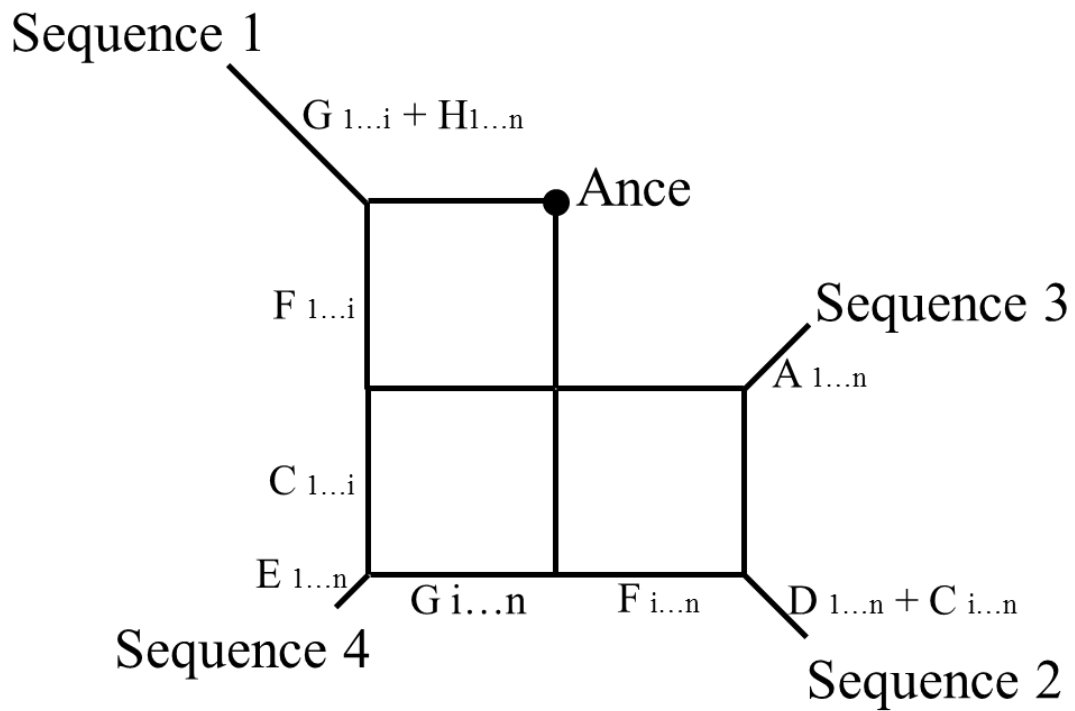


図 1.18 図 1.17 の系統ネットワークにさらに分岐が古い外群を導入した時の、ループ構造の変化。Ance(Ancestor)の位置にさらに古い外群は結節する。Sequence 1 のシングルトン上に Ance が結節しなかったことで、Sequence 1 が遺伝的組換えの親配列として関わったことが分かる。

17. 研究の目的

1) 16S rRNA遺伝子の部分的な水平伝播を系統ネットワーク法によって解析する

これまでの16S rRNA遺伝子の部分的な報告は、いずれも単純な配列比較によるものである。そのため、キメラ体がどの配列で、親配列の系統がどれであるか、などの具体的な進化史の記述はなされていない。また、部分的な水平伝播が実際にどの程度の多様性を16S rRNA遺伝子にもたらしているかなどの評価もなされていないため、本研究においては、系統ネットワーク法を用いることによりこれらの点を評価していく。

2) 16S rRNA遺伝子の水平伝播の分類群ごとの頻度を解析する

これまでの16S rRNA遺伝子の部分的な水平伝播の報告は、いずれも目より下位の分類群における固有の事例を紹介しているに過ぎず、分類群ごとに16S rRNA遺伝子の水平伝播が存在するの否か、現象として稀なことなのか普遍的なことのかなどが評価されていない。本研究において、分類群ごとの16S rRNA遺伝子の水平伝播の頻度を推定する。

3) 16S rRNA遺伝子の水平伝播がどの程度の分類群を超えて生じうるのかを検証する

これまでの報告によって、少なくとも属内・属間において16S rRNA遺伝子の部分的な水平伝播が生じていることが示唆された。しかしながら、目や綱、門などの属よりも上位の分類群を飛び越えて水平伝播が生じうるのかどうかについては不明である。本研究において、細菌の16S rRNA遺伝子の部分的な水平伝播が生じうる範囲を推定する。

4) 16S rRNA 遺伝子を用いた従来の系統解析や種同定の是非について考察する

Woese (1987)によって、16S rRNA を用いて全生物を3つのドメインへと系統分類されて以降、16S rRNA 遺伝子は細菌の系統分類に最も良く用いられる遺伝子である。また、Jain ら (1999)らによって Complexity Hypothesis が提唱されたことによって、16S rRNA 遺伝子を用いて作成した系統樹が種の系統関係を表すと考えられている。本研究は、もし16S rRNA 遺伝子が水平伝播する遺伝子なのであれば、16S rRNA 遺伝子を用いた系統樹解析によってどの程度の誤りが生じうるのかを評価する。

第一章 参考文献

- Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V., and Polz, M. F. (2004). Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. *J. Bacteriol.* 186(9):2629-2635.
- Allers, J., and Shamoo, Y. (2001). Structure-based analysis of protein-RNA interactions using the program ENTANGLE. *J. Mol. Biol.* 311(1):75-86.
- Ambler, R. P., Daniel, M., Hermoso, J., Meyer, T. E., Bartsch, R. G., and Kamen, M. D. (1979). Cytochrome C2 sequence variation among the recognised species of purple nonsulphur photosynthetic bacteria. *Nature* 278(5705):659-660.
- Bae, W., Xia, B., Inouye, M., and Severinov, K. (2000). *Escherichia coli* CspA-family RNA chaperones are transcription antiterminators. *Proc. Natl. Acad. Sci. USA* 97(14):7784-7789.
- Bandelt, H. J. (1994). Phylogenetic networks. *Verh Natwiss Ver Hambg.* 34:51-71.
- Bremer, H., and Dennis, P. P. (1996) In Neidhardt, F. C., Curtiss, R., III, Ingraham, J. L., Lin, E. C. C., Low, K. B. et al. (eds), *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology. Am. Soc. Microbiol. 1553-1569.
- Brodersen, D. E., Clemons, W. M., Jr, Carter, A. P., Wimberly, B. T., and Ramakrishnan, V. (2002). Crystal structure of the 30S ribosomal subunit from *Thermus thermophilus*: structure of the proteins and their interactions with 16S RNA. *J. Mol. Biol.* 316(3):725-768.
- Bryant, D., and Moulton, V. (2004). Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.* 21(2):255-265.
- Cannone J. J., et al. (2002). The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3(1):2.
- Cilia, V., Lafay, B., and Christen, R. (1996). Sequence heterogeneities among 16s ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level. *Mol. Biol. Evol.* 13(3):451-461.
- Cohen, O., Gophna, U., and Pupko, T. (2011). The complexity hypothesis revisited: Connectivity rather than function constitutes a barrier to horizontal gene transfer. *Mol. Biol. Evol.* 28(4):1481-1489.
- Condon, C., Philips, J., Fu, Z. Y., Squires, C., and Squires, C. L. (1992). Comparison of the expression of the seven ribosomal RNA operons in *Escherichia coli*. *EMBO. J.* 11(11):4175-4185.
- Conn, G. L., Draper, D. E., Lattman, E. E., and Gittis, A. G. (1999). Crystal structure of a conserved ribosomal protein-RNA complex. *Science* 284(5417):1171-1174.
- Dickerson, R. E. (1980). Structural conservatism in proteins over three billion years: cytochrome with a touch of collagen. (Pergamon, Oxford). 1:227-249.

- Eardly, B. D., Wang, F. S., and Vanberkum, P. (1996). Corresponding 16S rRNA gene segments in Rhizobiaceae and *Aeromonas* yield discordant phylogenies. *Plant. Soil.* 186(1):69-74.
- Errede, B., and Kamen, M. D. (1978). Comparative kinetic studies of cytochromes c in reaction with mitochondrial cytochrome c oxidase and reductase. *Biochemistry* 17(6):1015-1027.
- Feil, E. J., Holmes, E.C., Bessen, D. E., Chan, M. S., Day, N. P., Enright, M. C., Goldstein, R., Hood, D. W., Kalia, A., Moore, C. E., Zhou, J., and Spratt, B. G. (2001). Recombination within natural populations of pathogenic bacteria: Short-term empirical estimates and long-term phylogenetic consequences. *Proc. Natl. Acad. Sci. USA* 98(1): 1182-1187.
- Garcia, V. S., Simo, F. X., Montero, M. A. and Romeu, A. (2002). Simultaneous horizontal gene transfer of a gene coding for ribosomal protein l27 and operational genes in *Arthrobacter sp.* *J. Mol. Evol.* 55(6):632-637.
- Griffith, F. (1928). The Significance of *Pneumococcal* Types. *Journal of Hygiene (Cambridge University Press)* 27(2):113-159.
- Grinsted, J., Cruz, F., and Schmitt, R. (1990). The Tn21 subgroup of bacterial transposable elements *Plasmid* 24(3):163-189.
- Hashimoto, J. G., Stevenson, B. S., and Schmidt, T. M. (2003). Rates and consequences of recombination between rRNA operons. *J. Bacteriol.* 185(3):966-972.
- Head, I. M., Saunders, J. R., and Pickup, R. W. (1998). Microbial evolution, diversity, and ecology: a decade of ribosomal RNA analysis of uncultivated microorganisms. *Microb. Ecol.* 35(1):1-21.
- Hugenholtz, P., Goebel, B. M., and Pace, N. R. (1998). Impact of culture independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* 180(18):4765-4774.
- Jain, R., Rivera, M. C., and Lake, J. A. (1999). Horizontal gene transfer among genomes: The complexity hypothesis. *Proc. Natl. Acad. Sci. USA* 96(7):3801-3806.
- Jeffreys, A. (1979). DNA sequence variants in the G gamma-, A gamma-, delta-, and beta- globin genes of man. *Cell* 18(1):1-10.
- Kitahara, K., Yasutake, Y., and Miyazaki, K. (2012). Mutational robustness of 16S ribosomal RNA, shown by experimental horizontal gene transfer in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 109(47):19220-19225.
- Kitano, T., Blancher, A., and Saitou, N. (2012). The functional A allele was resurrected via recombination in the human ABO blood group gene. *Mol. Biol. Evol.* 29(7):1791-1796.
- Kitano, T., Noda, R., Takenaka, O., and Saitou, N. (2009). Relic of ancient recombinations in gibbon ABO blood group genes deciphered through phylogenetic network analysis. *Mol. Phylogenet. Evol.* 51(3):465-471.
- Kitano, T., and Saitou, N. (1999). Evolution of Rh blood group genes have experienced gene conversions and positive selection. *J. Mol. Evol.* 49(5):615-626.
- Klappenbach, J. A., Saxman, P. R., Cole, J. R., and Schmidt, T. M. (2001). rrndb: the ribosomal RNA

- operon Copy Number Database. Nucl. Acids Res. 29(1):181-184.
- Lan, R. T., and Reeves, P. R. (1998). Recombination between rRNA operons created most of the ribotype variation observed in the seventh pandemic clone of *Vibrio cholerae*. Microbiology 144(5):1213–1221.
- Lawrence, J. G., and Ochman, H. (1997). Amelioration of bacterial genome: rates of change and exchange. J. Mol. Evol. 44(4):383-397.
- Lederberg, J., and Tatum, E. L. (1946). Gene Recombination in *Escherichia Coli*. Nature 158 (4016):558.
- Lee, H., Popodi, E., Tang, H., and Foster P. L.(2012). Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. Proc. Natl. Acad. Sci. USA 109(41):2774-2783.
- Lercher, M. J., and Pal, C. (2007). Integration of horizontally transferred genes into regulatory interaction networks takes many million years. Mol. Biol. Evol. 25(3):559-567.
- Mylvaganam, S., and Dennis, P. P. (1992). Sequence heterogeneity between the two genes encoding 16S rRNA from the halophilic archaeobacterium *Haloarcula marismortui*. Genetics 130(3):399–410.
- Nei M and Kumar S. (2006) 分子進化と分子系統学. 株式会社 培風館.
- Nei, M., and Rooney, A. P. (2005). Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet. 39:121-152.
- Nelson, K. E., Clayton, R. A., Gill, S. R., and 22 other authors (1999) Evidence for lateral gene transfer between archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature 399(6734):323-329.
- Ohta, T. (1976). Simple model for treating evolution of multigene families. Nature 262(5572):74–76.
- Ohta, T. (1985). A model of duplicative transposition and gene conversion for repetitive DNA families. Genetics 110(3):513-524.
- Parker, M. A. (2001). Case of localized recombination in 23S rRNA genes from divergent *Bradyrhizobium* lineages associated with neotropical legumes. Appl. Environ. Microbiol. 67(5):2076–2082.
- Rivera, M. C., Jain, R., Moore, J. E. and Lake, J. A. (1998). Genomic evidence for two functionally distinct gene classes. Proc. Natl. Acad. Sci. USA 95(11):6239-6244.
- Rosenberg, H. F., Ackerman, S. J., and Tenen, D. G. (1989). Human eosinophil cationic protein. Molecular cloning of a cytotoxin and helminthotoxin with ribonuclease activity. J. Exp. Med. 170(1):163-176.
- Saitou, N., and Yamamoto, F. (1997). Evolution of primate ABO blood group genes and their homologous genes. Mol. Biol. Evol. 14(4):399-411.
- Saitou, N., and Nei, M. (1987). The neighbor-joining method: A new method for reconstructing

- phylogenetic trees. *Mol. Biol. Evol.* 4(4):406-425.
- Schouls, L. M., Schot, C. S., and Jacobs, J. A. (2003). Horizontal transfer of segments of the 16S rRNA genes between species of the *Streptococcus Anginosus* group. *J. Bacteriol.* 185(24):7241-7246.
- Slightom, J. L., Blechl, A. E., and Smithies, O. (1980). Human fetal G gamma- and A gamma-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* 21(3):627-638.
- Smith, G. P. (1976). Evolution of repeated DNA sequences by unequal crossover. *Science* 191(4227):528-535.
- Stackebrandt, E. and Goebel, B. M. (1994). Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* 44(4): 846-849.
- Stackebrandt, E., and Ebers, J. (2006). Taxonomic parameters revisited: tarnished gold standards. *Microbiol. Today.* 33(4):152-155.
- Tian, R. M., Cai, L., Zhang, W. P., Cao, H. L., and Qian, P. Y. (2015). Rare events of intragenus and intraspecies horizontal transfer of the 16S rRNA gene. *Genome Biol. Evol.* 7(8):2310-2320.
- Tsuda, M., and Iino, T. (1987). Genetic analysis of a transposon carrying toluene degrading genes on a TOL plasmid pWW0 *Mol. Gen. Genet.* 210(2):270-276.
- Yuri, Y. I., Aravind, L., Grishin, N. V. and Koonin, E. V. (1999). Evolution of aminoacyl-tRNA synthetases--analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome. Res.* 9(8):689-710.
- Ueda, K., T. Seki, T. Kudo, T. Yoshida, and M. Kataoka. (1999). Two distinct mechanisms cause heterogeneity of 16s rRNA. *J. Bacteriol.* 181(1):78-82.
- Voorhees, R. M., Weixlbaumer, A., Loakes, D., Kelley, A. C., Ramakrishnan, V. (2009). Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat. Struct. Mol. Biol.* 16(5):528-533.
- Wang, Y., and Zhang, Z. S. (2000). Comparative sequence analyses reveal frequent occurrence of short segments containing an abnormally high number of non-random base variations in bacterial rRNA genes. *Microbiology* 146(11):2845-2854.
- Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., Moore, L. H., Moore, W. E. C., Murray, R. G. E. and other authors (1987). Report of the adhoccommittee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Bacteriol.* 37(4):463-464.
- Wellner, A., and Gophna, U. (2008). Neutrality of foreign complex subunits in an experimental model of lateral gene transfer. *Mol. Biol. Evol.* 25(9):1835-1840.
- Wimberly, B. T., Guymon, R., McCutcheon, J. P., White, S. W., and Ramakrishnan, V. (1999). A

detailed view of a ribosomal active site: the structure of the L11-RNA complex. *Cell* 97(4):491–502.

Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* 51(2):221–271.

Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains archaea, bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* 87(12):4576-4579.

Yap, W. H., Zhang, Z. S., and Wang, Y. (1999). Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. *J. Bacteriol.* 181(17):5201–5209.

Young, J. D., Peterson, P. V., and Cohn, Z. A. (1986). Mechanism of membrane damage mediated by human eosinophil cationic protein. *Nature* 321(6070):613-616.

Zhang, J., Rosenberg, H. F., and Nei, M. (1998). Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. USA* 95(7):3708-3713.

第二章 *Escherichia coli*、*Shigella* 集団の 16S rRNA 遺伝子の解析

1. 序論

Escherichia coli と *Shigella* 属

Escherichia coli は Theodor Escherich によって 1885 年に発見された、グラム陰性の桿菌で通性嫌気性菌に属する、今日において最もよく知られたバクテリアの一種である (Escherich, 1885)。その 12 年後に志賀潔によって発見され、その名にちなんで *Shigella* 属と命名されたバクテリアは、当初、形態、生化学、病理学などの観点から *Escherichia coli* とは別種と考えられてきた。しかし、DNA-DNA 分子交雑法や、DNA の配列解析技術の向上によって、*Shigella* 属と *Escherichia coli* は近縁種であることが分かり、現代では *Shigella* 属は *Escherichia coli* の一系統であると考えられている (Milkman and Bridges., 1993; Gregory and Kim., 2011)。

Escherichia coli と *Shigella* 属の分類と系統関係

最も初期の、*Escherichia coli* と *Shigella* 属が近縁であることの指摘は数本の配列比較による Milkman と Bridges(1993)の報告による。最新の *Escherichia coli* の分類は、7つのハウスキーピング遺伝子 (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *recA*, *purA*) を用いた Wirth ら(2006)によるもので、A1、B1、B2、D、E の 5つの系統群(phylogroup)に分類されている。Wang ら(1997)は、ハウスキーピング遺伝子 (*icd*) を用いた系統解析から、phylogroup D が最も初期に分岐したグループであると結論づけている。後の解析では、Escobar ら(2003)や Gregory と Kim(2011)は、phylogroup B2 が phylogroup D よりも祖先的であるとし、Touchon ら(2009)は、phylogroup D が最も祖先的で、次に phylogroup B2 が祖先的であると結論付けている。いずれの報告においても(例えば、Wang ら(1997)、Escobar ら(2003) 、Gregory と Kim(2011) 、Touchon ら(2009))、phylogroup B2 か D が祖先的であることが示唆されている。

2. 材料と方法

データの取得と整理

Escherichia coli と *Shigella* 属の 16S rRNA 遺伝子配列は、2013 年 10 月 2 日に NCBI のアノテーションに従って、全ての *Escherichia coli* と *Shigella* 属のゲノム配列から取得した。*Escherichia coli* は 55 ゲノム、*Shigella* 属は 9 ゲノムの計 64 ゲノムから抽出した 16S rRNA 遺伝子を解析に用いた。表 2.1 に、用いた株の phylogroup を記載した。各株の phylogroup は、Wirth ら(2006)、および Bielecki ら(2014)に従って判断した。多重整列は、16S rRNA の二次構造を元にして多重整列を行うことが可能な、SINA のオンラインバージョン (<http://www.arb-silva.de/aligner/>) を利用した (Pruesse et al., 2012)。配列間の相同性の違いの計算、サイトあたりの塩基置換数の推定には MEGA 6.0 (Tamura et al., 2013) を用いた。

表 2.1 *Escherichia* と *Shigella* の各ゲノムにおける、各 16S rRNA 遺伝子グループの分布。
phylogroup (ハウスキーピング遺伝子による分類) は ¹Gregory and Kim 2011, ²Bielecki et al.,
2014 より

No.	アクセッション番号	株名	phylogroup	A	B	C	D	E
1	NC_004337.2	<i>Shigella flexneri</i> 2a str. 301	S ¹	5	2			
2	NC_004741.1	<i>Shigella flexneri</i> 2a str. 2457T	S ¹	5	2			
3	NC_007384.1	<i>Shigella sonnei</i> Ss046	S ¹	4	1		1	1
4	NC_007606.1	<i>Shigella dysenteriae</i> Sd197	S ¹			7		
5	NC_007613.1	<i>Shigella boydii</i> Sb227	S ¹	1				6
6	NC_008258.1	<i>Shigella flexneri</i> 5 str. 8401	S ¹	7				
7	NC_010658.1	<i>Shigella boydii</i> CDC 3083-94	S					7
8	NC_016822.1	<i>Shigella sonnei</i> 53G	S	5				2
9	NC_017328.1	<i>Shigella flexneri</i> 2002017	S	4	1			2
10	NC_000913.3	<i>Escherichia coli</i> str. K-12 substr. MG1655	A ¹	4	2			1
11	NC_002655.2	<i>Escherichia coli</i> O157:H7 str. EDL933	E ¹	2				5
12	NC_002695.1	<i>Escherichia coli</i> O157:H7 str. Sakai	E ¹	1				6
13	NC_004431.1	<i>Escherichia coli</i> CFT073	B2 ²	6				1
14	NC_007779.1	<i>Escherichia coli</i> str. K-12 substr. W3110	A ¹	4	2			1
15	NC_007946.1	<i>Escherichia coli</i> UTI89	B2 ¹	7				
16	NC_008253.1	<i>Escherichia coli</i> 536	B2 ¹	7				
17	NC_008563.1	<i>Escherichia coli</i> APEC O1	B2 ¹	6				1
18	NC_009800.1	<i>Escherichia coli</i> HS	A ¹	4			1	2
19	NC_009801.1	<i>Escherichia coli</i> E24377A	B1 ¹	1	2		3	1
20	NC_010468.1	<i>Escherichia coli</i> ATCC 8739	A ¹	3	1			3
21	NC_010473.1	<i>Escherichia coli</i> str. K-12 substr. DH10B	A ¹	4	2			1
22	NC_010498.1	<i>Escherichia coli</i> SMS-3-5	D ¹	1				6
23	NC_011353.1	<i>Escherichia coli</i> O157:H7 str. EC4115	E ¹	1				6
24	NC_011415.1	<i>Escherichia coli</i> SE11	B1 ¹	4	1			2
25	NC_011601.1	<i>Escherichia coli</i> O127:H6 str. E2348/69	B2 ²	7				
26	NC_011993.1	<i>Escherichia coli</i> LF82		5				2
27	NC_012759.1	<i>Escherichia coli</i> BW2952	A ²	4	2			1
28	NC_012892.2	<i>Escherichia coli</i> BL21(DE3)	A ¹	4				3
29	NC_012947.1	<i>Escherichia coli</i> 'BL21-Gold(DE3)pLysS AG'	A ¹	4				3
30	NC_012967.1	<i>Escherichia coli</i> B str. REL606	A ¹	4				3
31	NC_012971.2	<i>Escherichia coli</i> BL21(DE3)	A ¹	4				3
32	NC_013008.1	<i>Escherichia coli</i> O157:H7 str. TW14359	E ¹	1				6
33	NC_013353.1	<i>Escherichia coli</i> O103:H2 str. 12009	B1 ¹				4	3
34	NC_013361.1	<i>Escherichia coli</i> O26:H11 str. 11368	B1 ¹	3	2		2	
35	NC_013364.1	<i>Escherichia coli</i> O111:H- str. 11128	B1 ¹	3			3	1
36	NC_013654.1	<i>Escherichia coli</i> SE15	B2 ²	7				
37	NC_013941.1	<i>Escherichia coli</i> O55:H7 str. CB9615	E ²	1				6
38	NC_016902.1	<i>Escherichia coli</i> KO11FL	B1 ²	3	1	2	1	
39	NC_017625.1	<i>Escherichia coli</i> DH1	A ²	4	2			1
40	NC_017628.1	<i>Escherichia coli</i> IHE3034	B2 ²	7				
41	NC_017631.1	<i>Escherichia coli</i> ABU 83972	B2 ²	7				
42	NC_017632.1	<i>Escherichia coli</i> UM146	B2 ²	7				
43	NC_017633.1	<i>Escherichia coli</i> ETEC H10407	A ²	6	1			
44	NC_017634.1	<i>Escherichia coli</i> O83:H1 str. NRG 857C	B2 ²	5				2
45	NC_017635.1	<i>Escherichia coli</i> W	B1 ²	3	1	2	1	
46	NC_017638.1	<i>Escherichia coli</i> DH1	A ²	4	2			1
47	NC_017641.1	<i>Escherichia coli</i> UMNK88	A ²					7
48	NC_017644.1	<i>Escherichia coli</i> NA114	B2 ²	7				
49	NC_017646.1	<i>Escherichia coli</i> O7:K1 str. CE10	D ²					7
50	NC_017651.1	<i>Escherichia coli</i> str. 'clone Di2'	B2 ²	7				

表 2.1 続き

51	NC_017652.1	<i>Escherichia coli</i> str. 'clone Di14'	B2 ²	7				
52	NC_017656.1	<i>Escherichia coli</i> O55:H7 str. RM12579	E ²	1				6
53	NC_017660.1	<i>Escherichia coli</i> KO11FL	B1 ²	3	2	2		
54	NC_017663.1	<i>Escherichia coli</i> P12b	A ²		1		1	5
55	NC_017664.1	<i>Escherichia coli</i> W	B1 ²	3	2	1		
56	NC_017906.1	<i>Escherichia coli</i> Xuzhou21	E ²	1				6
57	NC_018650.1	<i>Escherichia coli</i> O104:H4 str. 2009EL-2050			6		1	
58	NC_018658.1	<i>Escherichia coli</i> O104:H4 str. 2011C-3493			7			
59	NC_018661.1	<i>Escherichia coli</i> O104:H4 str. 2009EL-2071			6			1
60	NC_020163.1	<i>Escherichia coli</i> APEC O78						6
61	NC_020518.1	<i>Escherichia coli</i> str. K-12 substr. MDS42		4	2			1
62	NC_022364.1	<i>Escherichia coli</i> LY180		3	3	1		
63	NC_022370.1	<i>Escherichia coli</i> PMV-1		7				
64	NC_022648.1	<i>Escherichia coli</i> JJ1886		7				
*1				85/78/84	38/44/39	9/11/9	16/11/16	62/56/61
*2				69/67/69	7/0/6	0/11/2	5/0/5	24/22/23
*3				229/51	56/13	15/3	18/4	128/29

*1: 各グループを保有するゲノムの割合(%) (*Escherichia coli*/*Shigella*/全体)

*2: 各グループを主要なコピーとして保有するゲノムの割合(%) (*Escherichia coli*/*Shigella*/全体)。主要なコピーとは、最もコピー数が多い 16S rRNA 遺伝子として、という意味。

*3: 全体の配列数(配列数/割合(%))

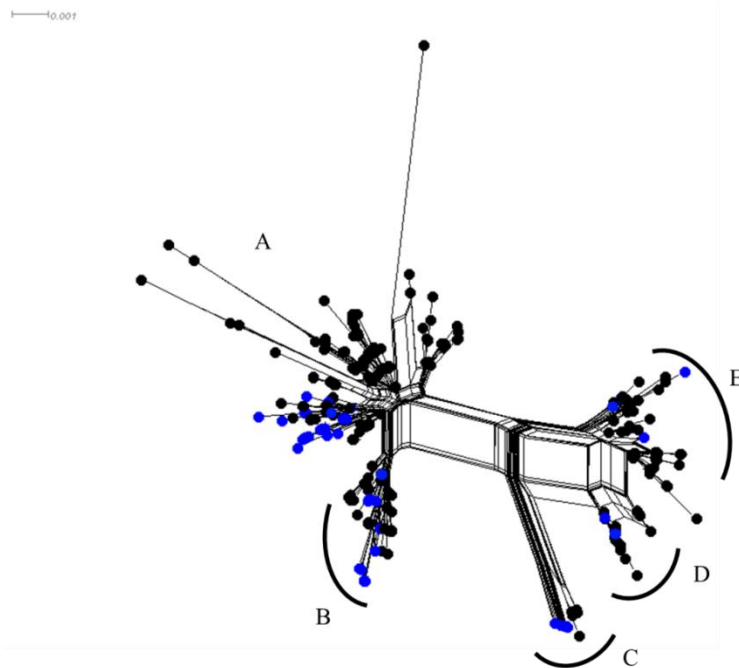


図 2.1 *Escherichia coli* と *Shigella* 属の 16S rRNA 遺伝子を用いて、Neighbor-Net 法により作成した系統ネットワーク。距離行列は p 距離を用い、5%以上のギャップを含むサイトは解析から取り除いた。*Escherichia coli* を黒丸、*Shigella* 属を青丸で示した。A-E の 5 つのグループに大きく分かれた。

系統ネットワーク解析

Neighbor-Net 法による系統ネットワークの作成には SplitsTree 4.0 を用いた(Huson and Bryant., 2006)。Neighbor-Net 法を用いた際の、距離行列の計算には p 距離を用い、5%以上のギャップを含むサイトは解析から削除した。節約的な系統ネットワークは、Bandelt (1994) や Saitou and Yamamoto (1997)の手法に従って作成した。3 塩基以上の多型が見られるサイトは解析から取り除いた。

3. 結果と考察

Escherichia coli, *Shigella* 集団の 5 つの 16S rRNA 遺伝子グループ

図 2.1 は *Escherichia* 属と *Shigella* 属のゲノム配列から抽出した 16S rRNA 遺伝子配列すべてを用いて作成した Neighbor-net である。大きく 4 つのグループ、A-E に分かれることがわかった。また、*Escherichia* (黒丸) と *Shigella* (青丸) の分布は各グループにおいて重複しており、16S rRNA 遺伝子配列を *Escherichia* や *Shigella* という分類上の属名により特徴づけることは不可能であった。

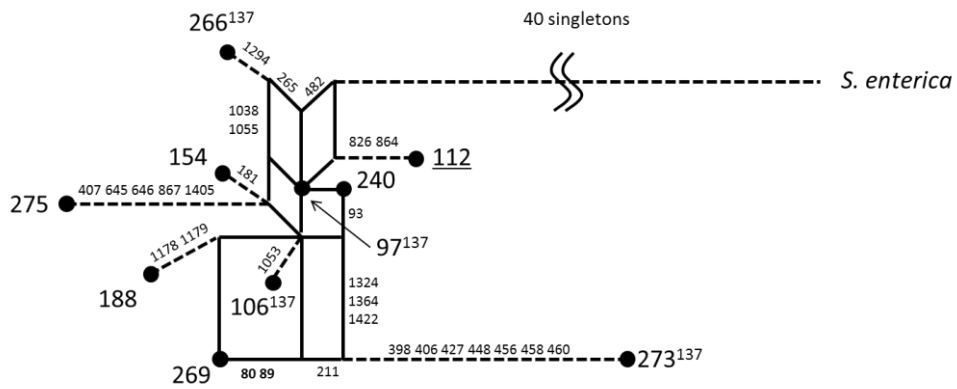
Escherichia coli, *Shigella* 集団のゲノム内多型

Escherichia も *Shigella* もゲノム中に 7 コピーの 16S rRNA 遺伝子を持つ。各ゲノムの 16S rRNA 遺伝子の各グループの分布を表 2.1 に示した。表 2.1 によると、*Escherichia*、*Shigella* の 64 本のゲノムのうち、グループ A-E の 16S rRNA 遺伝子のうち複数を持しているゲノムの割合は約 70%程度であり、*Escherichia* や *Shigella* にとってはゲノム内の 16S rRNA 遺伝子間で多型が見られることは普遍的なことであると言える。16S rRNA 遺伝子の各グループの割合では、グループ A が最も広く分布しており、*Escherichia*、*Shigella* のゲノムのうちおよそ 83%のゲノムに見られた。次いで広く分布しているのが、グループ E (59%)で最も共有されている割合が少なかったのはグループ C の 8%であった。コピー数においてもグループ A が 229 コピー (48%)、グループ E が 128 コピー (27%)と多く、グループ C はわずかに 15 コピーが見られたのみであった。グループ A を主要な 16S rRNA 遺伝子として保有しているゲノムは、*Escherichia* では 69%、*Shigella* では 67%と最も高く、グループ E を主要なグループとして保有しているゲノムの割合は、*Escherichia* で 24%、*Shigella* で 22%とグループ A に次いで高かった。コピー数、共有されている割合、主要な 16S rRNA 遺伝子として存在している割合はグループ A が最も高く、次いでグループ E が高いという結果になっており、頻度から考えると *Escherichia*、*Shigella* においてグループ A の 16S rRNA 遺伝子が最も祖先的なグループであると考えられる。

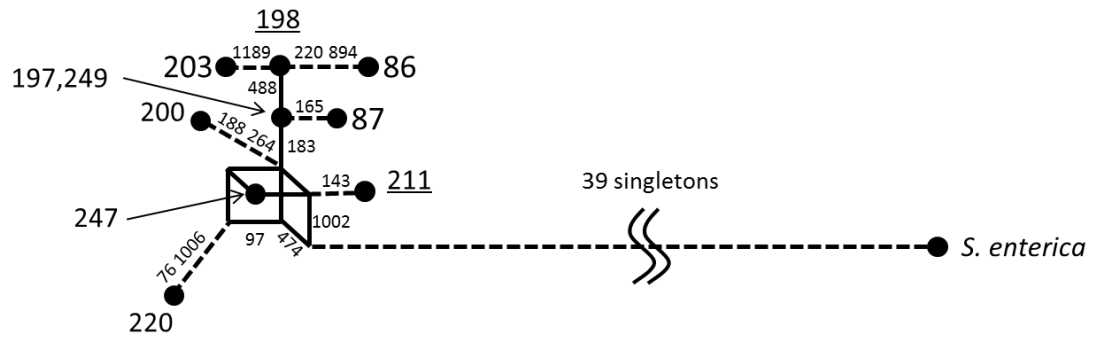
Escherichia coli, *Shigella* 集団の系統ネットワーク解析

A-E の 16S rRNA 遺伝子の各グループについて詳細な解析を行うべく、*Salmonella enterica*

(グループ A)



(グループ B)



(グループ C)

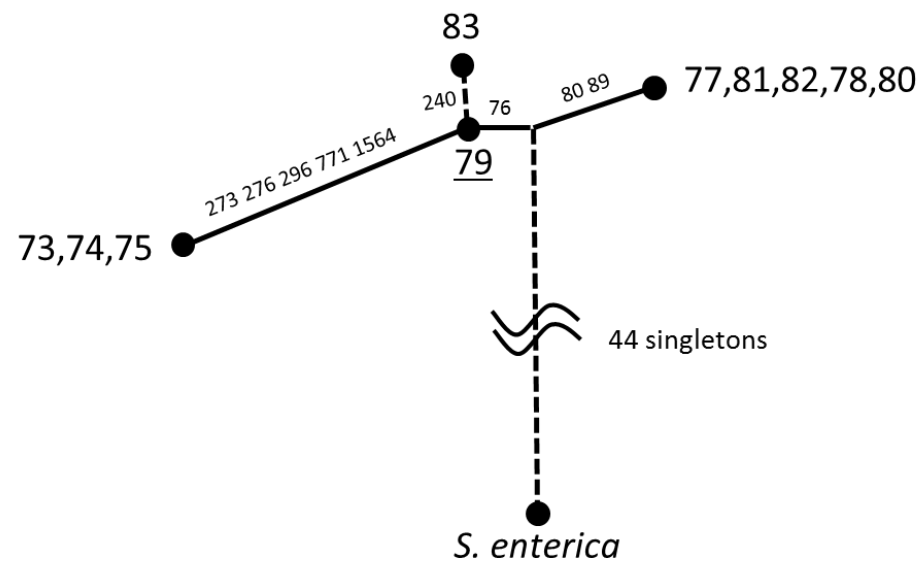
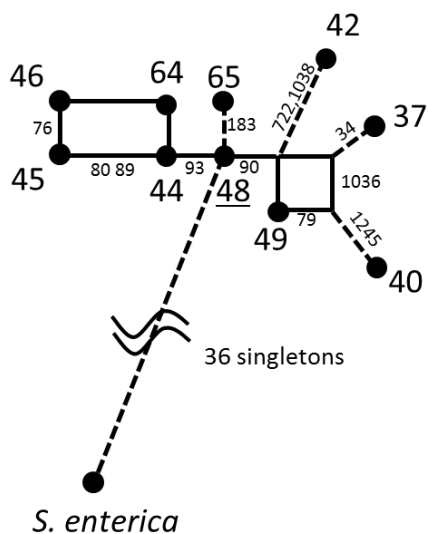


図 2.2 (次ページへ続く)

(グループ D)



(グループ E)

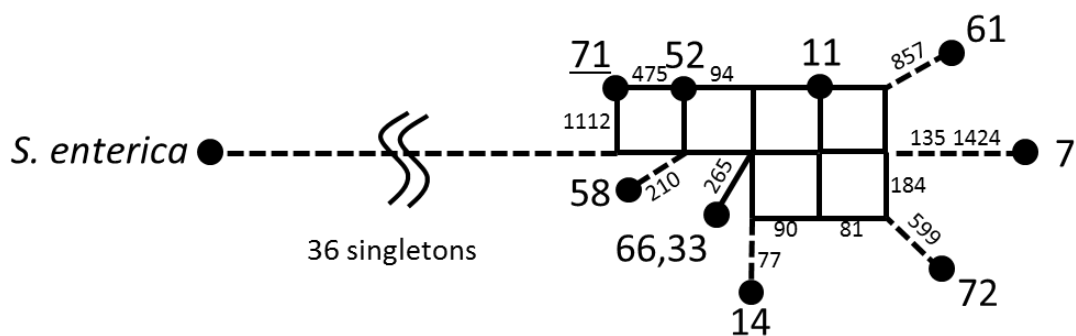


図 2.2 図 2.1 に示した、Neighbor-Net 法によって作成した系統ネットワークにおいて同定された 5 つのグループ(A-E)ごとの節約的な系統ネットワーク。外群として *Salmonella enterica* を導入した。各グループにおいて最も祖先的な配列と判断した配列を下線でハイライトした。各辺の数字は、大腸菌の 16S rRNA 遺伝子の残基番号に対応している。

表 2.2 Phylogroup ごとの、各 16S rRNA 遺伝子グループの保有率。

Phylogroup	16S rRNA 遺伝子のグループの頻度				
	16S-A	16S-B	16S-C	16S-D	16S-E
S (A)	0.52	0.12	0.17	0.02	0.17
E (E)	0.16	0.00	0.00	0.00	0.84
D (E)	0.07	0.00	0.00	0.00	0.93
B2 (A)	0.96	0.00	0.00	0.00	0.04
B1 (A)	0.37	0.18	0.11	0.23	0.11
A (A)	0.50	0.14	0.00	0.02	0.33

の 16S rRNA 遺伝子配列を外群として用いて、各グループの詳細な系統ネットワーク解析を行った (図 2.2)。解析に用いた配列は、各グループの Neighbor-Net 法によって作成した系統ネットワークから万遍なく選択した。16S rRNA 遺伝子グループ A、B、D、E の系統ネットワークにおいてループ構造が見られたため、遺伝的組換えか平行進化が生じていることがわかった。また、グループ A、C、D において、サイト 80, 89 が変化しており、異なるグループで平行して変化しているので、高頻度に変化しているサイトであるか、水平伝播によって他のグループから運ばれてきたサイトであると考えられる。

各 16S rRNA 遺伝子グループの進化史

各グループから選んだ祖先的な配列を用いて作成した系統ネットワークを図 2.3 に記した。表 2.1 に各 phylogroup における A-E の 16S rRNA 遺伝子グループの保有率を示した。最も祖先的と考えられている phylogroup B2 と phylogroup D においては、16S rRNA 遺伝子グループ A と E しか見られないため、16S rRNA 遺伝子グループ A と E は、グループ B、C、D に対して祖先的であると考えられる。

16S rRNA 遺伝子グループ A と E は、*Escherichia coli*、*Shigella* グループの祖先の段階から存在していた可能性がある。実際に、16S rRNA 遺伝子グループ A と E を分けている 1000 番台のサイトの多型は、同様のパターンが *Salmonella* 属においても見られ、*Salmonella* 属と *Escherichia* 属の共通祖先の段階から存在していた可能性は排除できない (Cilia, 1996)。このことから、16S rRNA 遺伝子グループ A と E が *Escherichia coli*、*Shigella* 集団の祖先の段階にすでに存在していたと仮定し、16S rRNA 遺伝子グループの進化史を図 2.3 の系統ネットワークから考えると以下の 2 通りの進化史が推定された。

- a) サイト 250,253,270 に塩基置換が生ずることによりグループ B がグループ A から分岐し、グループ B とグループ E の間の遺伝的組換えか、グループ E のサイト 250,253,270 に平行進化が生ずることでグループ D が生じた。
- b) 250,253,270 に塩基置換が生ずることによりグループ D がグループ E から分岐し、グループ D とグループ A の間の遺伝的組換えか、グループ A のサイト 250,253,270 に平行進化が生ずることでグループ B が生じた。

遺伝的組換えか平行進化か

サイト 253 とサイト 273 は 16S rRNA の二次構造上の helix11 において、塩基対を形成しており、Enterobacteriales 目集団において、すべてが A-U か U-A のいずれかの構造をとっていた。従って、A-U か U-A の塩基対しか認められない箇所であると考えられ、平行進化が生じやすい箇所であると考えられる。このことから、この 2 サイトを実質的にサイト 1 つぶんの多型であると考えれば、上記の進化史は、遺伝的組換え 1 回か、平行進化 2 回によってグループ B か D が生じたと考えることができる。*Escherichia coli*、*Shigella* 集団において、

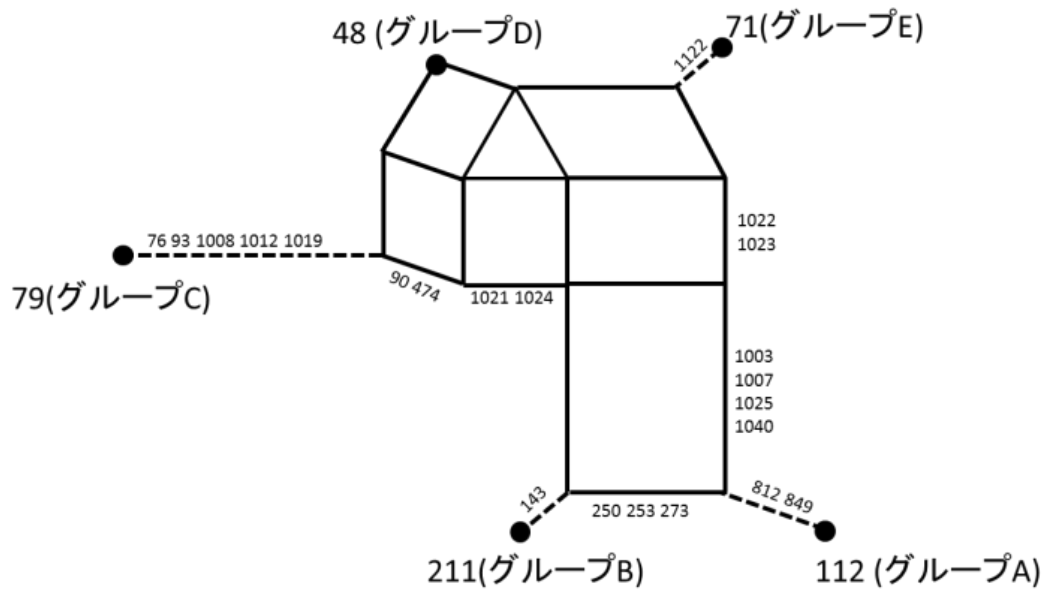


図 2.3 各 16S rRNA 遺伝子グループにおいて、最も祖先的と考えられる配列を用いて作成した節約的な系統ネットワーク。各グループにおいて最も祖先的と判断された配列を代表として用いた。

rRNA 遺伝子同士の平均のサイトあたりの塩基置換数は、Kimura's 2 parameter を用いて計算すると約 0.01 となった。最尤法を用いて計算したサイト 250, 253, 270 の相対進化速度は 1 で、全サイトに渡っての平均進化速度と差がないことが分かった(塩基置換モデル: General Time Reversible model (Nei and Kumar, 2000))。このことは、これらのサイトに平行進化が生ずる可能性は大変低いことを意味している。Hashimoto ら(2003)によって推定された、16S rRNA 遺伝子同士のゲノム内組換えが生ずる確率(世代あたりコピーあたり 5.7×10^{-9})が、Lee ら(2012)によって推定された *Escherichia coli* の自然突然変異率(世代あたりサイトあたり 2.2×10^{-10})よりも高いことなども考慮に入れると、平行進化というよりかは、遺伝的組換えによるものであると考えられる。

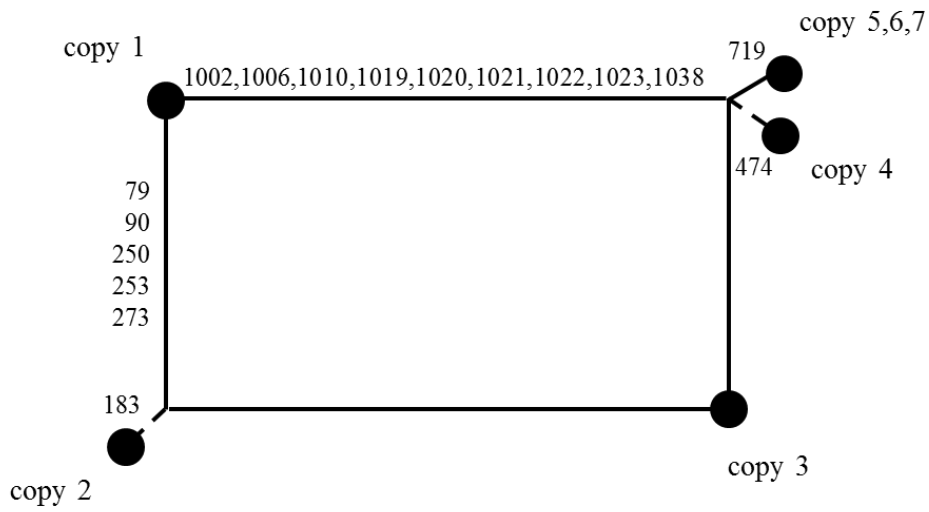
16S rRNA 遺伝子グループ C の進化

16S rRNA 遺伝子グループ C の進化は、はっきりとしなかった。図 2.3 の系統ネットワークにおいて、グループ C は、サイト 90, 474 の 2 つの塩基置換をグループ D と共有しており、比較的グループ D の近くに結節していた。グループ C は、1000 番台のサイトへの塩基置換の蓄積により、他の 16S rRNA 遺伝子グループに対して 5 サイトも異なっていた。もし、1000 番台のサイトを無視して考えると、グループ D へとサイト 76, 93 へと塩基置換が蓄積することによりグループ C は生じたと考えることができる。加えてグループ C の頻度は、*Escherichia coli*、*Shigella* 集団全体で最も低く、phylogroup の S と B1 しかグループ C 型の 16S rRNA 遺伝子を保有していないことなどを考えると、ごく最近に *Escherichia coli*、*Shigella* 集団に現れたグループであると考えられる。変異が集中している 1000 番台のサイトは、16S rRNA の二次構造上の helix33 にすべて含まれているため、helix33 が異種から水平伝播によって運ばれてくることによって、グループ C 型の 16S rRNA 遺伝子は生じた可能性がある。中には、7 コピーすべてが C 型 16S rRNA 遺伝子である株 (NC_007606.1 *Shigella dysenteriae* Sd197) も存在するため、生存に有利な点があるのかもしれない。

遺伝的組換えが生じた場

また、16S rRNA 遺伝子グループ A、B、D、E を、ゲノム中に保有する株が存在しており、(NC_007384.1 *Shigella sonnei* Ss046、NC_009801.1 *Escherichia coli* E24377A)、祖先の段階で生じた遺伝的組換えの特徴を今日まで引き継いでいるものと考えられる。このことは、遺伝的組換えが生じた段階で 16S rRNA 遺伝子グループ A、B、D、E がゲノム中に共存していた可能性を示唆し、グループ B かグループ D を生じさせた遺伝的組換えは、水平伝播というよりは、ゲノム内組換えであったと考えるのが妥当である。表 2.1 に示したように、*Escherichia coli*、*Shigella* 集団においては、ゲノム内の 16S rRNA 遺伝子コピーに複数のグループ由来のものを保有することは普遍的なこととして見られるため、複数系統の 16S rRNA 遺伝子の共存状態において、遺伝的組換えが生じ、組換え体を含めた多型が維持されることは不思議なことではない。

(A)



(B)

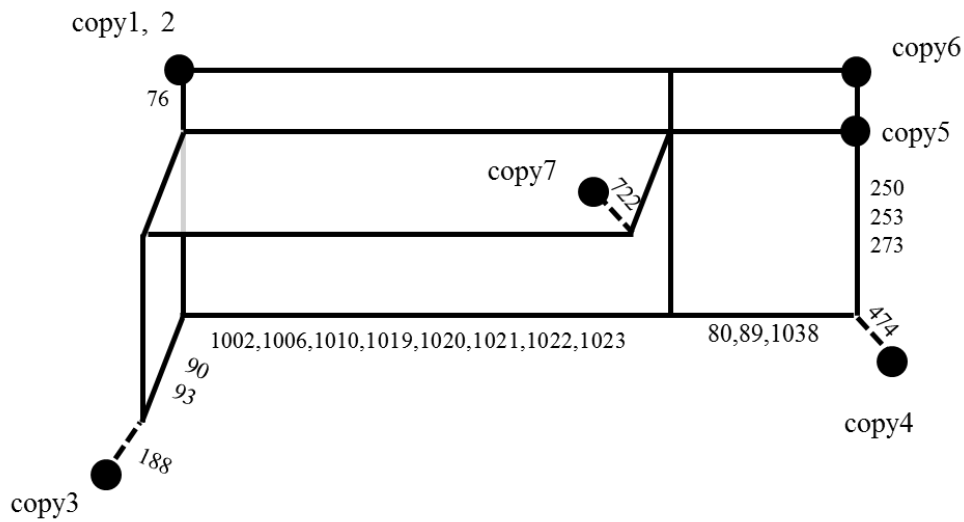


図 2.4 グループ A, B, D, E をゲノム中に保有する株の 16S rRNA 遺伝子を用いて作成した節約的な系統ネットワーク。(A) NC_007384.1 *Shigella sonnei* Ss046 の 7 コピーの 16S rRNA 遺伝子を用いて作成した節約的な系統ネットワーク。それぞれグループ A(copy 1)、B(copy 2)、D(copy 3)、E(copy 4,5,6,7)。(B) NC_009801.1 *Escherichia coli* E24377A の 7 コピーの 16S rRNA 遺伝子を用いて作成した節約的な系統ネットワーク。それぞれグループ A(copy 1,2)、B(copy 3)、D(copy 4)、E(copy 5,6,7)。

以上のことは、ゲノム内での 16S rRNA 遺伝子の遺伝的組換えが、16S rRNA 遺伝子の多様度に寄与した可能性を示す。すなわち、図 2.1 や図 2.3 で見られる *Escherichia coli*、*Shigella* 集団全体の 16S rRNA 遺伝子の多様度が、図 2.4 に示した NC_007384.1_ *Shigella sonnei* Ss046 や NC_009801.1 *Escherichia coli* E24377A といった一本のゲノム中にほぼ収まっているということである。このことから、16S rRNA 遺伝子の多様性が生み出される原動力の 1 つとして、ゲノム内組換えが関わっている可能性が指摘できた。また、今後の解析において、一見水平伝播によるモザイク化のように見えても、*Escherichia coli*、*Shigella* 集団のように実はゲノム内組換えによって生み出された多型が維持されているに過ぎないような場合を考慮せねばならないということが示唆された。

第二章 引用文献

- Bielecki, P., Muthukumarasamy, U., Eckweiler, D., Bielecka, A., Pohl, S., Schanz, A., Niemeyer, U., Oumeraci, T., von Neuhoff, N., Ghigo, J. M., and Häussler, S. (2014). In vivo mRNA profiling of uropathogenic *Escherichia coli* from diverse phylogroups reveals common and group-specific gene expression profiles. *MBio* 5(4):e01075-14.
- Escobar, P. P., Giudicelli, C., Parsot, C., and Denamur, E. (2003). The evolutionary history of *Shigella* and enteroinvasive *Escherichia coli* revised. *J. Mol. Evol.* 57(2):140-148.
- Gregory, E. S., and Kim, S. H. (2011). Whole-genome phylogeny of *Escherichia coli*/*Shigella* group by feature frequency profiles (FFPs). *Proc. Natl. Acad. Sci. USA* 108 (20) 8329-8334.
- Hashimoto, J. G., Stevenson, B. S., and Schmidt, T. M. (2003). Rates and consequences of recombination between rRNA operons. *J. Bacteriol.* 185(5):966-972.
- Lee, H., Popodi, E., Tang, H., and Foster P. L. (2012). Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Natl. Acad. Sci. USA* 109(41):2774-2783.
- Milkman, R., and Bridges, M. M. (1993). Molecular evolution of the *Escherichia coli* chromosome. IV. Sequence comparisons. *Genetics* 133(3):455-468.
- Nei, M., and Kumar, S. (2000). *Molecular evolution and phylogenetics*. Oxford University Press, New York.
- Pruesse, E., Peplies, J. and Glöckner, F. O. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28(14):1823-1829
- Saitou, N., and Yamamoto, F. (1997). Evolution of primate ABO blood group genes and their homologous genes. *Mol. Biol. Evol.* 14(4):399-411.

- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30(12):2725-2729.
- Touchon M, et al. (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS. Genet.* 5(1):e1000344.
- Wang, F. S., Whittam, T. S., and Selander, R. K. (1997). Evolutionary genetics of the isocitrate dehydrogenase gene (*icd*) in *Escherichia coli* and *Salmonella enterica*. *J. Bacteriol.* 179(21):6551–6559.
- Wirth T, et al. (2006). Sex and virulence in *Escherichia coli*: An evolutionary perspective. *Mol. Microbiol.* 60(5):1136–1151.

第三章 Enterobacteriales 目の 16S rRNA 遺伝子の解析

1. 序論

Enterobacteriales 目とは

Enterobacteriales 目は、Proteobacteria 門の Gammaproteobacteria 綱に属するバクテリアで、Enterobacteriaceae 科の一科から成り立つ分類群である。Enterobacteriales 目の和名は腸内細菌目といい、しばしば、腸内に生息しているバクテリアという意味での腸内細菌と混同されるが、Enterobacteriales 目は分類学上の系統群で、土壌や淡水においても幅広くみられる分類群である。また、Enterobacteriales 目には、*Escherichia coli* や *Shigella*、*Salmonella* といった食中毒菌が含まれており、ゲノム情報をはじめ、各種データベースが最も充実した分類群のひとつでもある。

2. 材料と方法

データの取得と整理

Enterobacteriales 目の 16S rRNA 遺伝子は、NCBI に登録されている Enterobacteriales 目のゲノム配列すべてを 2013 年 10 月 2 日に取得し、NCBI のアノテーションに従って 16S rRNA 遺伝子領域を抽出した。その際、AT リッチで解析が困難であった昆虫の細胞内共生菌である *Buchnera* 属と *Wigglesworthia* 属は取り除いた。表 3.1 に Enterobacteriales 目の解析に用いた各分類群とゲノム数、16S rRNA 遺伝子数のリストを示した。データセットは、属ごとに重複配列を削除して番号を振りなおし整理した。多重配列は、重複配列を削除後に、MAFFT version 7 (Katoh and Standley 2013) を用いて計算し、配列の編集には MEGA 6.0 (Tamura et al., 2013) を用いた。

系統ネットワーク解析

Neighbor-Net 法による系統ネットワークの作成には SplitsTree 4.0 を用いた (Huson and Bryant, 2006)。Neighbor-Net 法を用いた際の、距離行列の計算には p 距離を用い、5% 以上のギャップを含むサイトは解析から削除した。節約的な系統ネットワークは、Bandelt (1994) や Saitou と Yamamoto (1997) の手法に従って作成し、3 塩基以上の多型が見られるサイトは解析から取り除いた。

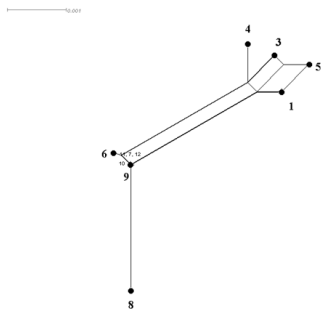
属ごとの代表配列の選択

属ごとに重複配列を削除したのちに、Neighbor-Net 法により属ごとに系統ネットワークを作成した (図 3.1)。各系統ネットワークから、万遍なく配列を選択し、各属の代表配列として解析に用いた。各属から選ばれた配列の番号は、太字でハイライトした (図 3.1)。

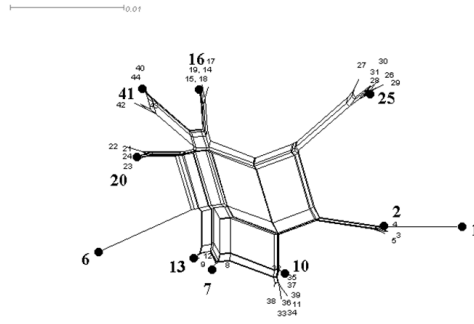
表 3.1 解析に用いた Enterobacteriales 目の 16S rRNA 遺伝子配列

属名	ゲノム数	16S rRNA 遺伝子数
<i>Escherichia</i>	54	377
<i>Shigella</i>	9	63
<i>Salmonella</i>	35	245
<i>Citrobacter</i>	1	7
<i>Klebsiella</i>	9	63
<i>Enterobacter</i>	9	70
<i>Cronobacter</i>	4	28
<i>Erwinia</i>	7	49
<i>Pantoea</i>	6	41
<i>Sodalis</i>	1	7
<i>Pectobacterium</i>	5	35
<i>Dickeya</i>	4	28
<i>Edwardsiella</i>	4	28
<i>Serratia</i>	12	72
<i>Yersinia</i>	19	123
<i>Providencia</i>	1	7
<i>Proteus</i>	2	14
<i>Photorhabdus</i>	2	14
<i>Xenorhabdus</i>	2	14
計	186	1285

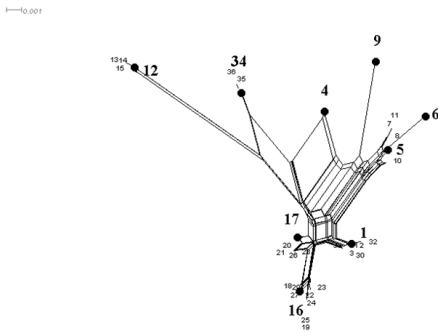
Edwardsiella



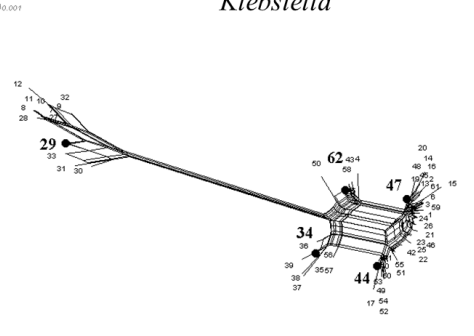
Enterobacter



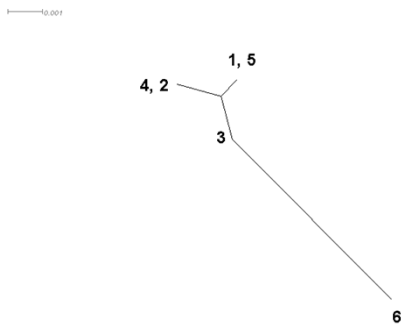
Erwinia



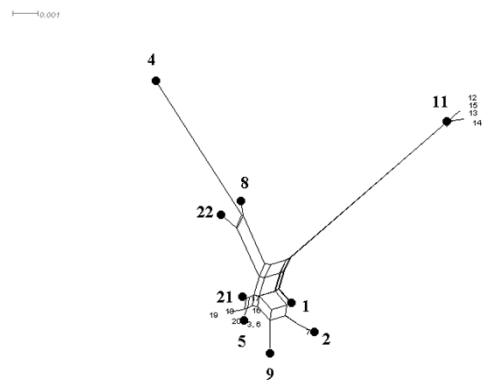
Klebsiella



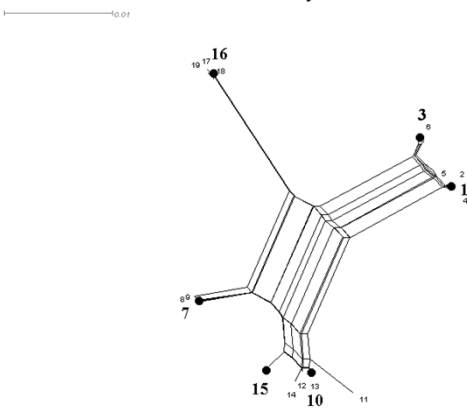
Citrobacter



Cronobacter



Dickeya



Escherichia

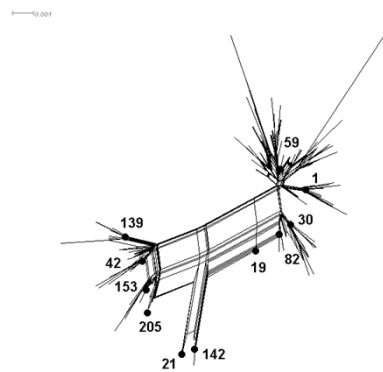


図 3.1 (1 ページ目 / 3 ページ)

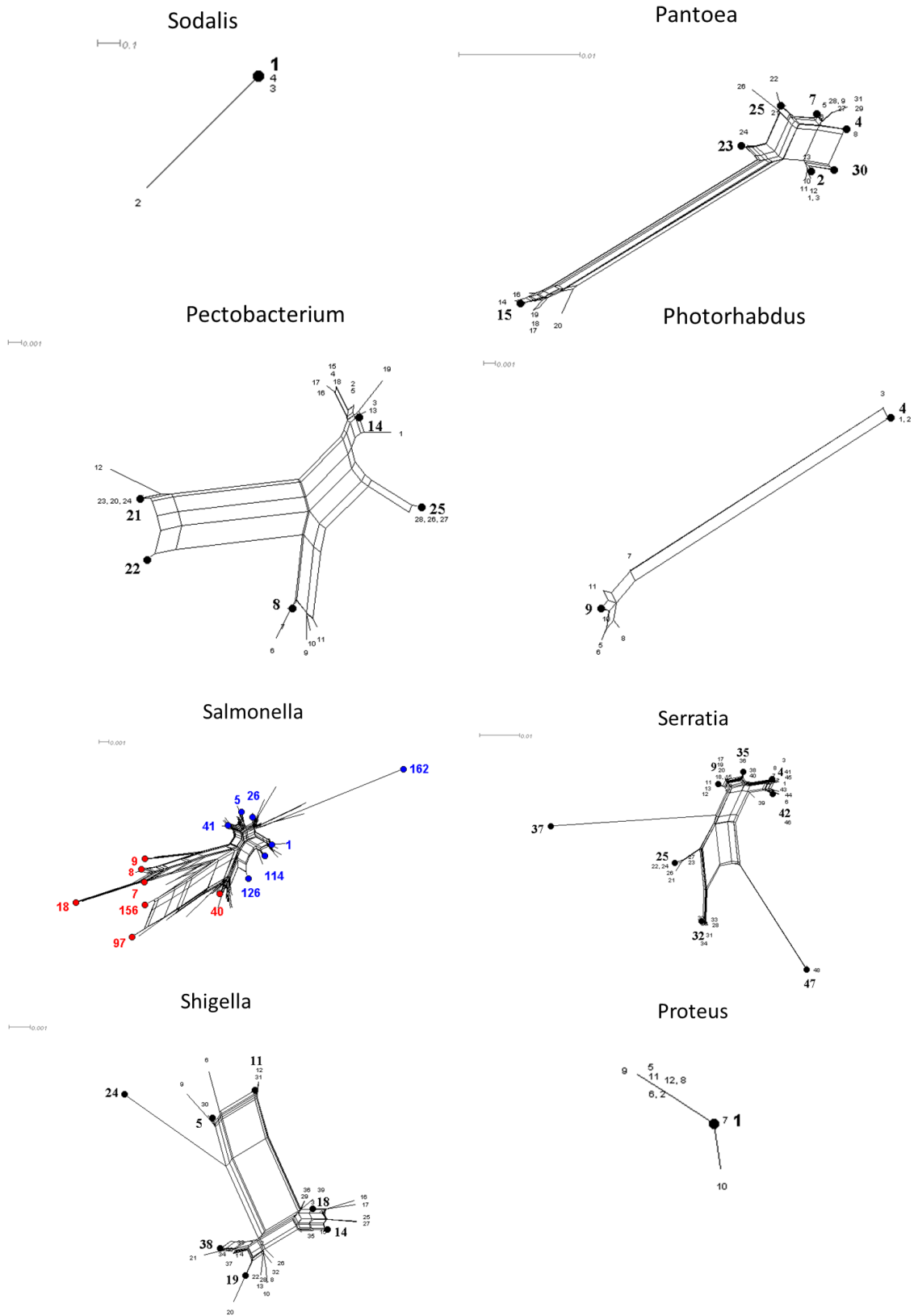


図 3.1 (2 ページ目 / 3 ページ)

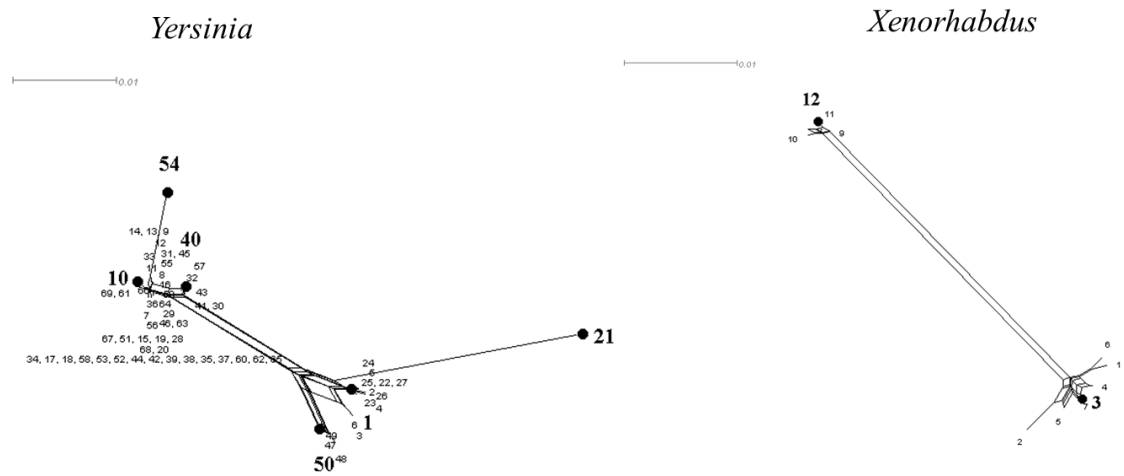


図 3.1 Enterobacteriales 目各属の 16S rRNA 遺伝子を用いて、Neighbor-Net 法によって作成した系統ネットワーク。距離行列は p 距離を用いて作成し、ギャップが 5%以上含まれるサイトは解析から除いた。各属から代表として選んだ配列を太字でハイライトした。各配列名は番号となっており、重複配列を削除後、番号を付け直したものである。*Providencia* 属は、配列が全て同じであったため、省略した。*Salmonella* 属は、多型のパターンが多すぎて代表配列全てを用いて系統ネットワークを作成することが困難であったため、2 グループ（赤、青）に分けた。

3. 結果と考察

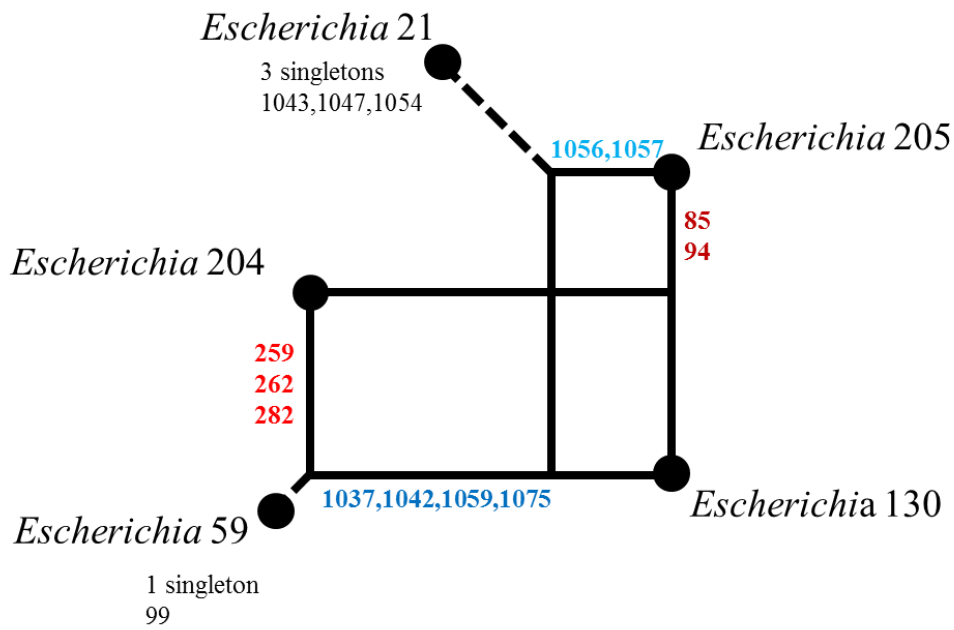
属内で 16S rRNA 遺伝子の水平伝播が見られた分類

a) *Escherichia* 属

Escherichia 属の 16S rRNA 遺伝子の代表配列を用いて作成した系統ネットワークを図 3.2(A)に示した。赤色サイト(259 – 282)に対して、青色サイト(1037-1075)、水色サイト(1056,1057)が対立していた。連続するサイト同士が対立しているため、遺伝的組換えによるものである可能性が高いが、第三章でも触れたように、水平伝播ではなく祖先の段階から多型が維持されている可能性が高い。

Escherichia 59, 204 と 130, 205 を分けているサイトは、Cannone ら(2002)のウェブサイトを参考にすると、*Escherichia* の *rrsB* の二次構造上において *helix33* の位置に相当した。この領域は、Cannone ら(2002)の二次構造に当てはめると、*Escherichia* 59, 204 と *Escherichia* 130, 205 で二次構造が異なり、*Escherichia* 59, 204 は 16S rRNA 遺伝子グループ A 型で *Escherichia* 130, 205 はグループ E 型であった(図 3.2(B))。グループ A とグループ E は、同一ゲノム内に普遍的に共存しているので、どちらも機能的であると考えられる。

(A)



(B)

(*Escherichia* 59) (*Escherichia* 130)

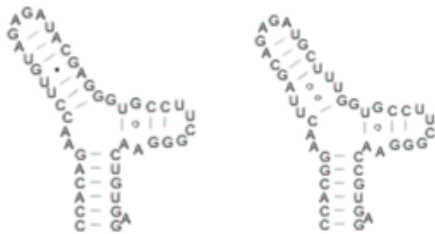


図 3.2 (A)図 3.1 の Neighbor-Net 解析によって代表として選ばれた、*Escherichia* の 16S rRNA 遺伝子を用いて作成した節約的な系統ネットワーク。ループ構造の縦辺は 259, 262, 282 に固まっており、横辺は 1037 – 1075 の領域に固まっているため、遺伝的組換えの存在が示唆される。(B)サイト 1037-1057 領域の二次構造。

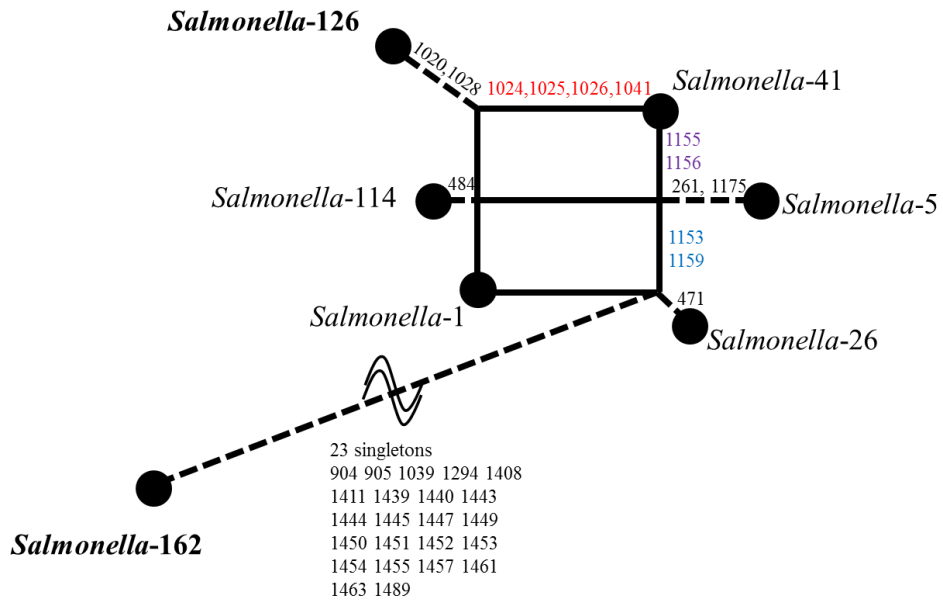
b) *Salmonella* 属

Salmonella 属 16S rRNA 遺伝子の代表配列を用いて作成した系統ネットワークを図 3.3 に示した。*Salmonella* 属は属内の多様度が著しく高く、代表配列を一つの系統ネットワークに示すことができなかつたため、さらに 2 つの集団に分けて解析した。

青グループ(図 3.3 (A))においては、対立するループの辺同士が短いが、*Salmonella* 5 と *Salmonella* 114 を無視して考えると、赤サイト、1024, 1025, 1026, 1041 に対して、青、紫サイト 1155, 1156, 1153, 1159 が対立しており、それぞれのサイトは連続しているため、一回の水平伝播と 4 回の平行進化で生じうるが、平行進化だけで説明しても平行進化 4 回で説明できるため、水平伝播が生じているかどうかは明確ではない。

一方で、図 3.3 (B) に示した赤グループの *Salmonella* 属の系統ネットワークにおいては、複数回の水平伝播が示唆された。同程度に節約的な水平伝播を含む経路は何通りか考えられ、平行進化が生じていないと仮定すると、どこの配列を起点に進化すると考えても最少の水平伝播回数は 4 回である。例えば、図 3.3 (B) において、*Salmonella*-18 を起点に考えると、緑の枝が塩基置換による経路で、赤の枝(1-4)が 4 回の考えられる水平伝播であった。

(A)



(B)

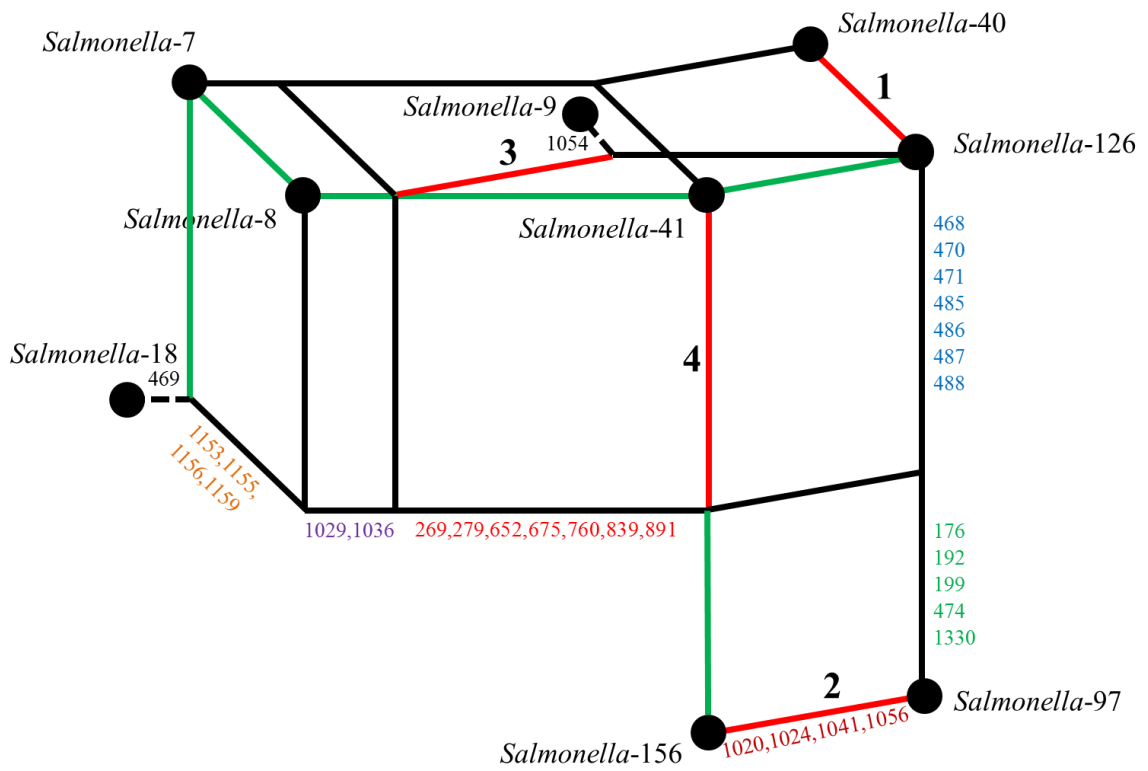


図 3.3 (1 ページ目 / 2 ページ)

図 3.3(A)から *Salmonella*-126,162 を、図 3.3 (B)から *Salmonella*-7, 18, 97 をそれぞれ選んで作成した *Salmonella* の節約的な系統ネットワークを図 3.3 (C)に示した。*Salmonella*-162 は、シングルトンサイトのほとんどが 1400 番台に集中しており、異なるグループからの、サイト 1400 番台を含む領域の水平伝播による可能性があった。したがって、*Salmonella*-162 の次にシングルトンが長い *Salmonella* 97 を起点に考えると、*Salmonella*-7 は *Salmonella*-18 の系統に、青色で示したサイト (サイト 468-488) が *Salmonella*-40 や 126, 162 の系統から水平伝播することにより生じた配列であると考えられる。

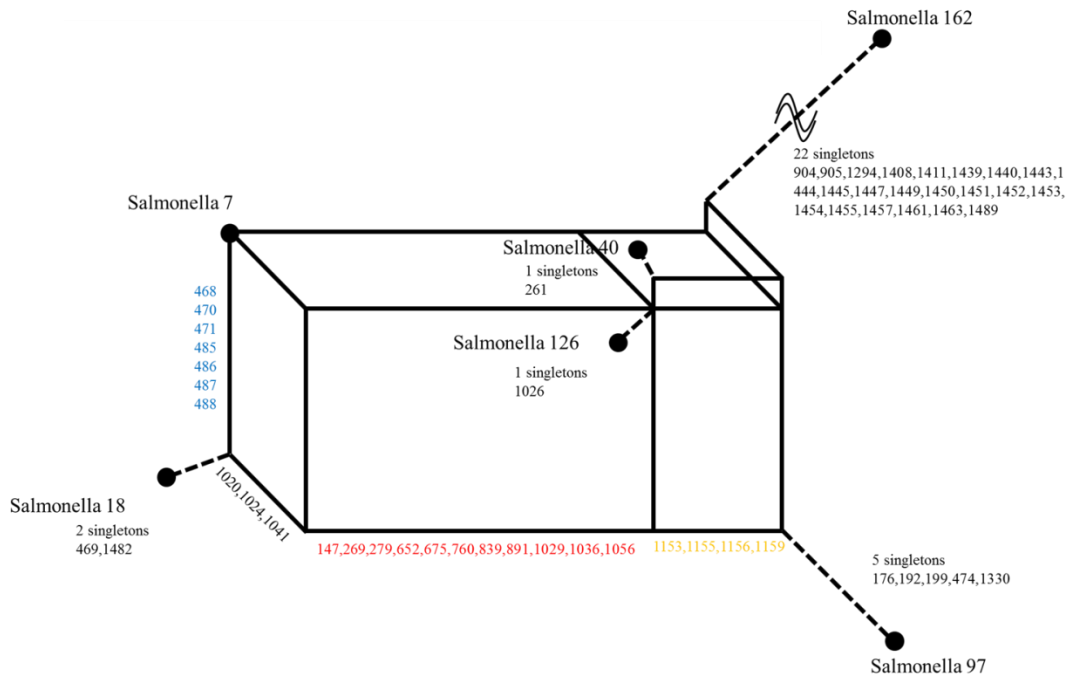


図 3.3 *Salmonella* 属の 16S rRNA 遺伝子を用いて作成した、節約的な系統ネットワーク。(A)は図 3.1 の *Salmonella* 属の Neighbor net において青色で示された分類群で、(B)は赤色で示された分類群、そして(C)は(A)(B)から配列を選んで作成された節約的な系統ネットワーク。

c) *Enterobacter* 属

Enterobacter 属の 16S rRNA 遺伝子の代表配列を用いて作成した系統ネットワークを図 3.4 に記した。緑色でハイライトしたサイト 666-893 領域と青色でハイライトした 1004-1239 領域が水平伝播領域と考えられ、2 回の部分的な水平伝播が生じていることが示唆された。シングルトンの分布が偏っているが、シングルトンを全く持たない *Enterobacter*-16, 20 がキメラ体であると考え、*Enterobacter*-2 が外群であり、親配列の系統は *Enterobacter*-6 と *Enterobacter*-25 である。*Enterobacter*-10 は *Enterobacter*-2 と *Enterobacter*-6 の間に位置しているが、*Enterobacter*-10 によって、*Enterobacter*-2 と *Enterobacter*-6 の間にある塩基置換が、666-893 と 1004-1239 に分けられていた。加えて、*Enterobacter*-10 はシングルトンを持っていないことを考えると、*Enterobacter*-6 の系統に *Enterobacter*-2 の 1004-1239 領域が水平伝播するか、*Enterobacter*-2 へと *Enterobacter*-6 の系統の 666-893 領域が水平伝播することによって *Enterobacter*-10 は生じたものと考えられる。

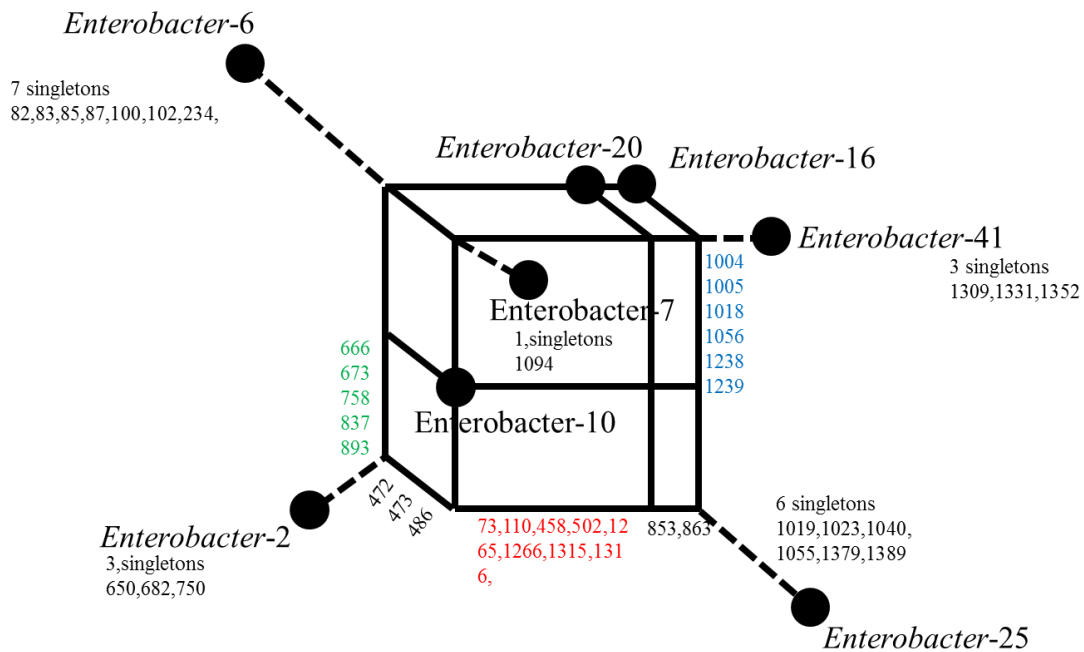


図 3.4 *Enterobacter* 属の 16S rRNA 遺伝子の代表配列を用いて作成した節約的な系統ネットワーク。

d) *Klebsiella* 属

Klebsiella 属の 16S rRNA 遺伝子の代表配列を用いて作成した系統ネットワークを図 3.5 に示した。*Klebsiella*-29 はシングルトンが多く、シングルトンサイトの分布も配列全体に及んでいるので遺伝的組換えに関わっていない配列であると考えられる。従って、平行進化を仮定しないで考えると、*Klebsiella*-44 は *Klebsiella*-47 と *Klebsiella*-34 の間の遺伝的な組換えによって生じた配列であると考えられる。

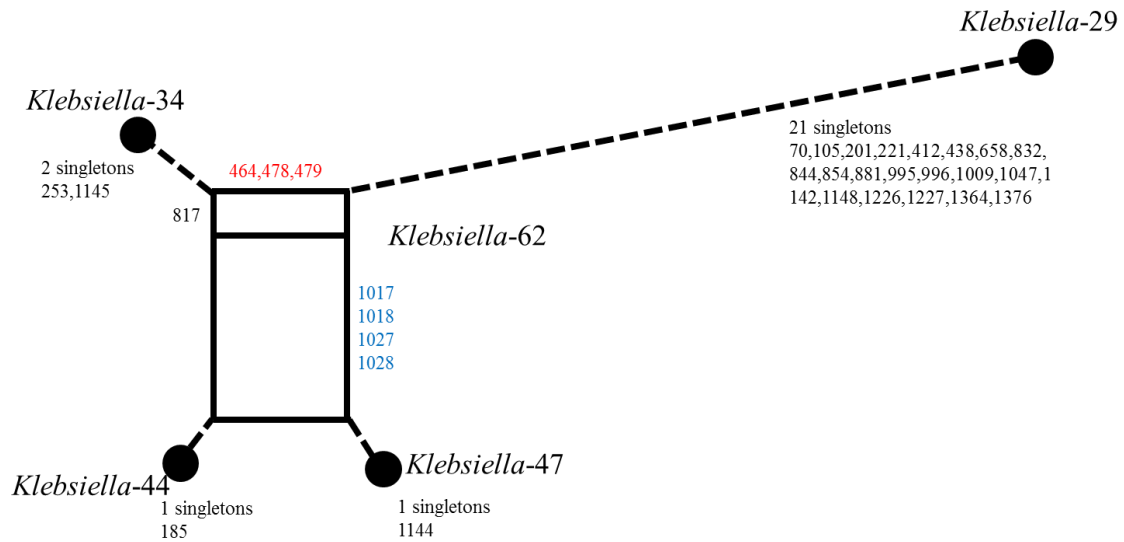


図 3.5 *Klebsiella* の代表配列の 16S rRNA 遺伝子の配列を用いて作成した節約的な系統ネットワーク。

e) *Erwinia* 属

Erwinia 属の 16S rRNA 遺伝子の代表配列を用いて作成した系統ネットワークを図 3.6 に示した。シングルトンが最も長く、また最もシングルトンサイトの分布に偏りが見られない *Erwinia*-12 が遺伝的組換えに関わっていない配列であると仮定すると、*Erwinia*-4 は *Erwinia*-5 と *Erwinia*-34 の間の組換え体である。*Erwinia*-1,17 は *Erwinia*-12 と *Erwinia*-9 の間の枝の中間に位置しているが、*Erwinia*-1,17 によって、赤サイト(78-100)と黄色サイト (1022,1033,1034,1035)が分けられていた。加えて、シングルトンも存在しないため、*Erwinia*-5 の系統に *Erwinia*-12 の系統のサイト 1022,1033,1034,1035 を含む領域が水平伝播するか、サイト 78,79,80,81,97,99,100 を含む領域が水平伝播することによって生じた配列であると考えられる。

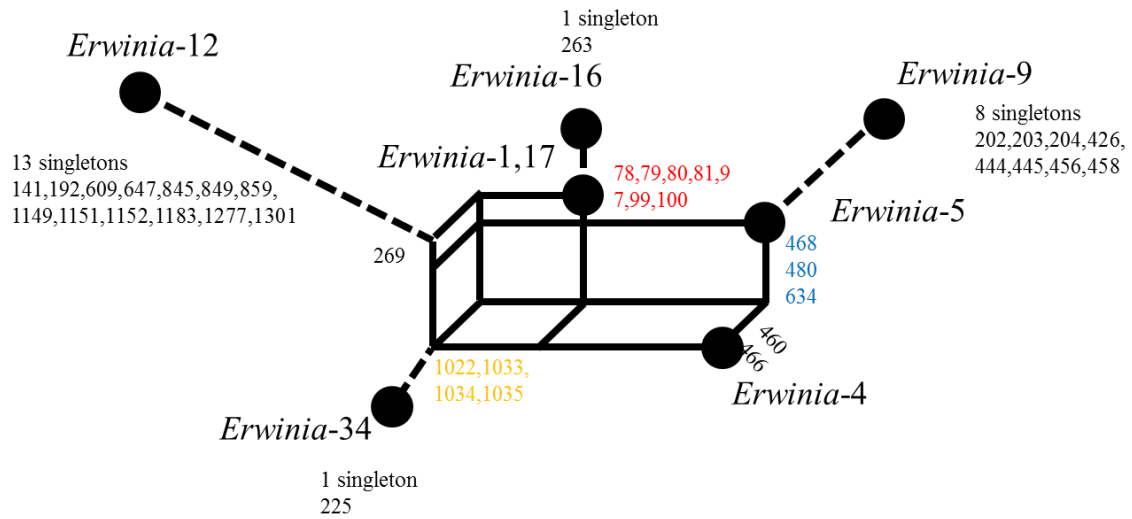


図 3.6 *Erwinia* 属の 16S rRNA 遺伝子の代表配列を用いて作成した節約的な系統ネットワーク。

f) *Dickeya* 属

Dickeya 属の 16S rRNA 遺伝子の代表配列を用いて作成した系統ネットワークを図 3.7 に示した。赤色で示した 19 サイトに及ぶ領域(70-1293)に対して、青色で示した 8 サイトの連続領域(839-1152)が対立する形でループ構造が形作られており、水平伝播の存在が強く示唆された。シングルトンが多く、シングルトンサイトが比較的全域に分布している *Dickeya*-16 を遺伝的組換えに関わっていない配列であると考え、*Dickeya*-10, 15 の系統は、*Dickeya*-7 の系統に *Dickeya*-1,3 の系統の青色で示したサイト(839-1152)を含む領域が水平伝播して生じた配列であると考えられる。

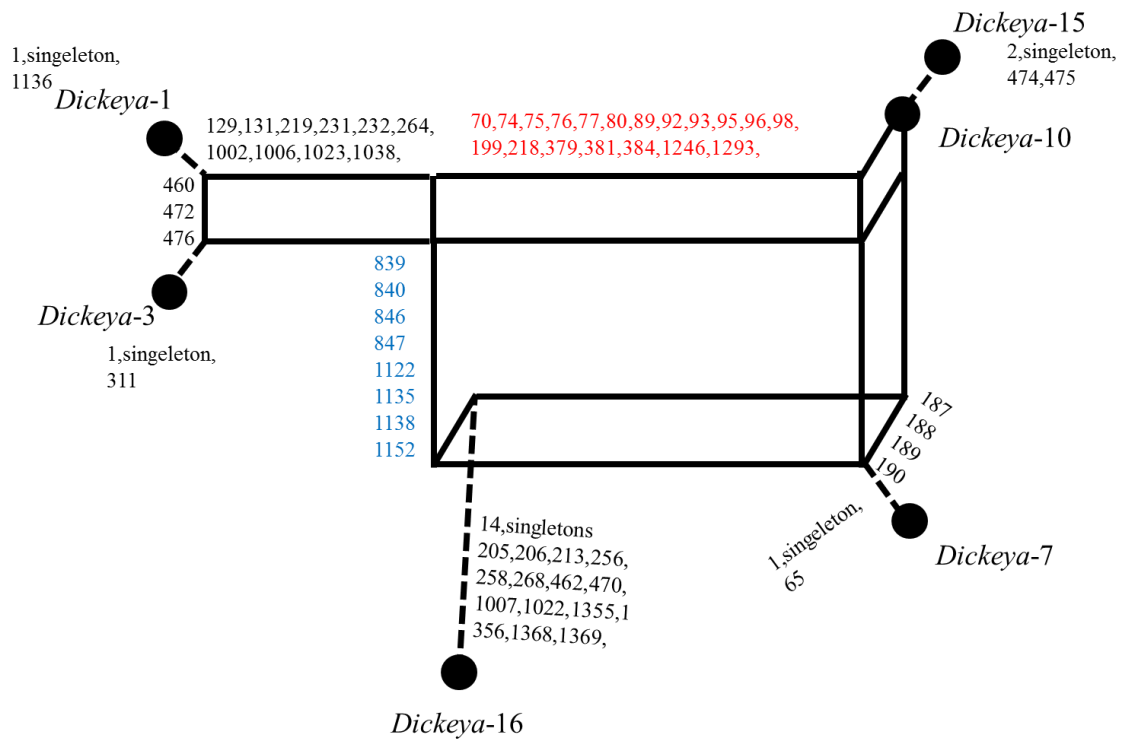
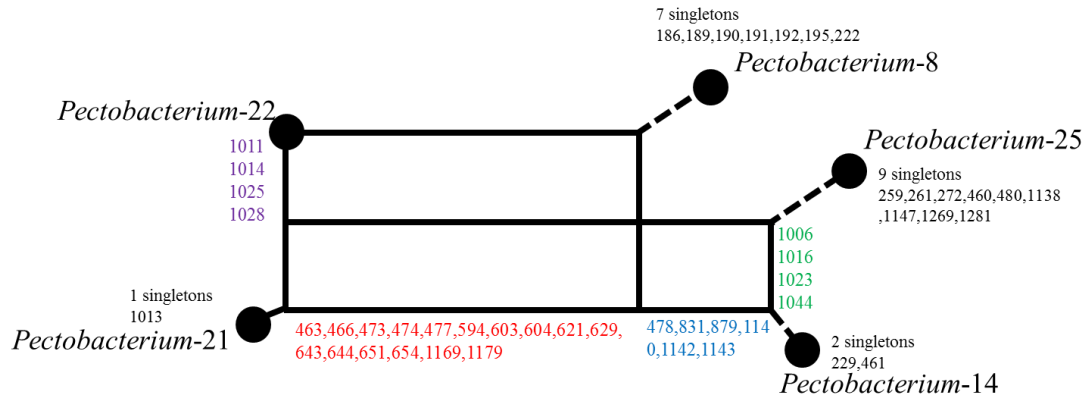


図 3.7 *Dickeya* 属の 16S rRNA 遺伝子の代表配列を用いて作成した節約的な系統ネットワーク。

g) *Pectobacterium* 属

Pectobacterium 属の 16S rRNA 遺伝子の代表配列を用いて作成した系統ネットワークを図 3.8 に示した。*Pectobacterium* の系統ネットワークには 3 つのループ構造が見られるが、これらは図 3.9 (A) に示した一回の部分的な水平伝播を含む進化史で構成されうる。図 3.9(A) のように進化してきた OTU1-OTU4 と組換えに関わっていない OUT GROUP を用いて Bandelt の系統ネットワークを作成すると、図 3.9(B) のようになる。*Pectobacterium* の系統ネットワークにおいても、サイト 1006-1044 までを ij 領域と考えると、図 3.9(B) の系統ネットワークと同様の形とサイトの分布を示す。加えて、*Pectobacterium*-25 はシングルトンが長く、シングルトンサイトの分布も比較的全域に及んでいるので組換えに関わっていない配列であると考えられる。従って、*Pectobacterium*-21 は、*Pectobacterium*-22 の系統へと *Pectobacterium*-14 のサイト 1006-1044 を含む領域が水平伝播することにより生じた配列であると考えられる。

(A)

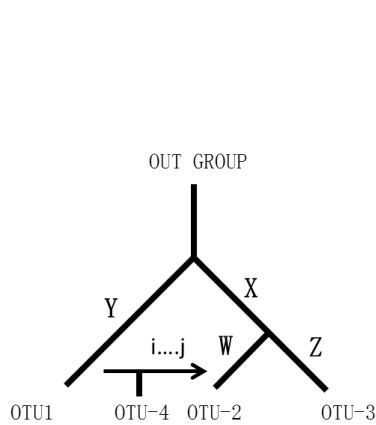


(B)



図 3.8 (A)*Pectobacterium* の 16S rRNA 遺伝子の代表配列を用いて作成した節約的な系統ネットワークと、(B)変異サイトの分布。i-j で示した領域は図 3.9 の進化モデルをあてはめた際に考えられる水平伝播領域。

(A)



(B)

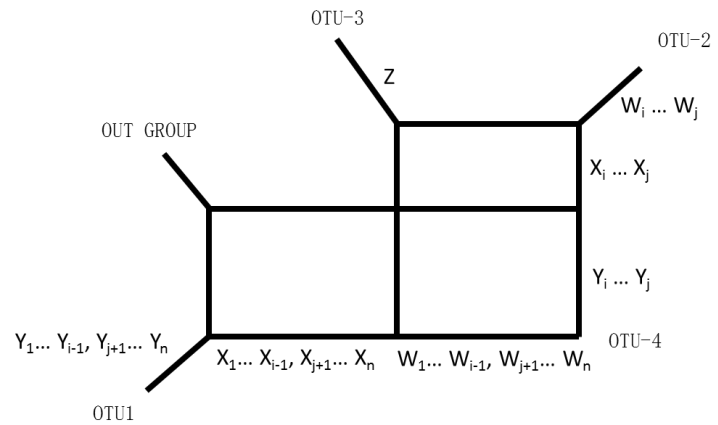


図 3.9 (A) OTU1 の系統の領域 i-j が OTU2 へと水平伝播することにより OTU4 が生じたという進化史。各枝のアルファベットはその枝で生じた塩基置換を意味し、配列全域に生じたものと仮定する。これらの進化史を辿って生じた OTU 1-4 に OUT GROUP を加えて作成した節約的な系統ネットワークを(B)に示した。

属を跨いで 16S rRNA 遺伝子の水平伝播が見られた分類

a) *Escherichia* 属と *Salmonella* 属、*Citrobacter* 属

図 3.10 に *Escherichia* 属と *Salmonella* 属の 16S rRNA 遺伝子を用いて作成した系統ネットワークを示した。*Enterobacter-6* を外群と考えると、*Citrobacter-3*, 6 は *Salmonella-44* の系統に、*Escherichia* 属のサイト 611-764(辺 C, F)を含む領域が水平伝播することにより生じたものと考えられる。本研究で用いた *Citrobacter* の 16S rRNA 遺伝子 7 本は、全て同一のゲノム由来であり、7 本すべてが *Escherichia* 属と *Salmonella* 属の間の組換え体であった。このことは、*Citrobacter* 属において水平伝播によって形作られた 16S rRNA 遺伝子が主要な 16S rRNA 遺伝子として機能していることを意味しており、ある分類群の 16S rRNA 遺伝子がまるごと組換え体であるという場合があり得るということを示している。

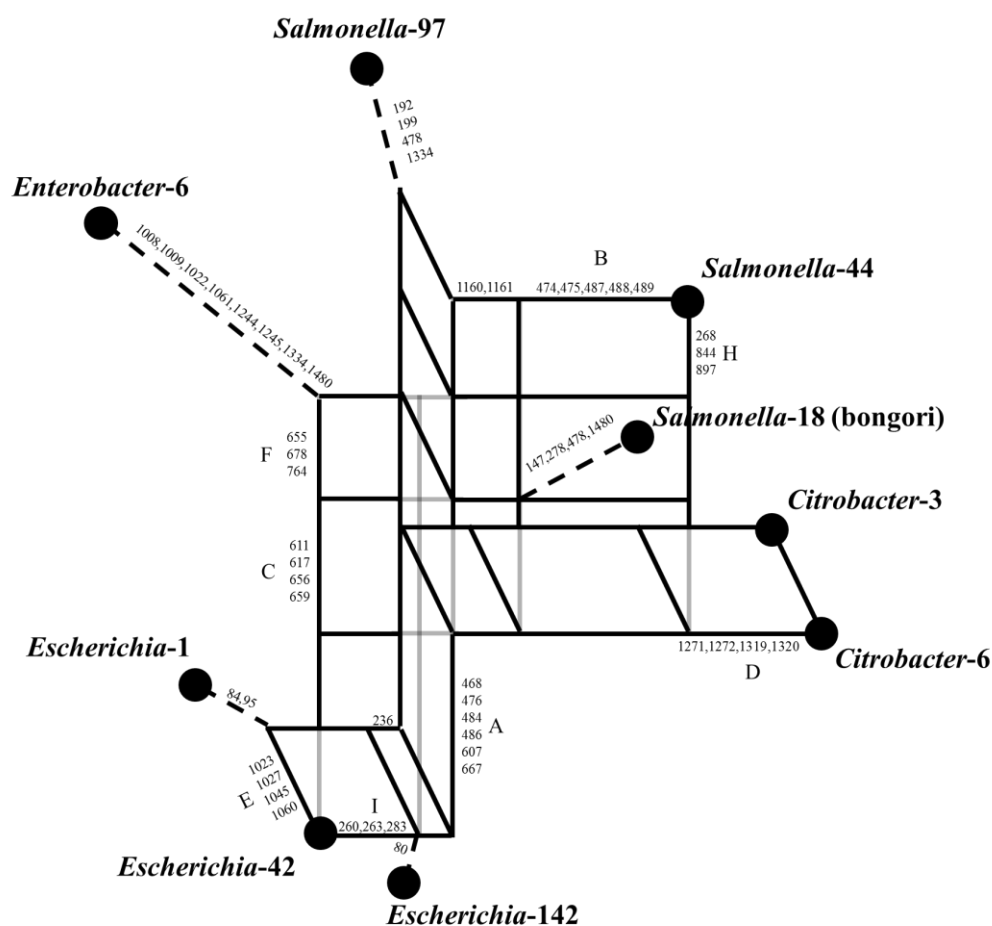


図 3.10 *Enterobacter-6* を外群として用いて作成した、*Escherichia* 属、*Salmonella* 属、*Citrobacter* 属の 16S rRNA 遺伝子を用いて作成した節約的な系統ネットワーク。*Enterobacter-6* を外群と考えると、*Citrobacter* 属は *Escherichia* 属、*Salmonella* 属の間の組換え体であると考えられる。

b) *Enterobacter* 属と *Klebsiella* 属

Enterobacter 属と *Klebsiella* 属の 16S rRNA 遺伝子を用いて作成した系統ネットワークを図 3.11 に示した。*Klebsiella* 属に固有の塩基置換が見られず、全て *Enterobacter* 属の 16S rRNA 遺伝子の多様度内に位置している。このことは、*Klebsiella* 属と *Enterobacter* 属の分類と、16S rRNA 遺伝子の分類が一致していないことを意味しているので、分類群間の水平伝播という観点がそもそも成り立たない。*Klebsiella* の配列を固有に議論すると、*Klebsiella*-29 と *Klebsiella*-44 は、A 辺 (サイト 829-1222) と D 辺 (サイト 70-220) が対立してできているループの対角に位置している。このループに焦点を当てるために、*Klebsiella*-29, 44 と *Enterobacter*-2, 20 にさらに遠い外群として *Salmonella* 属の配列を加えて系統ネットワーク解析を行った(図 3.12)。その結果、*Salmonella*-1 は *Enterobacter*-2 の位置に結節することが分かった。このことから、*Enterobacter*-20 は *Klebsiella*-29 と *Klebsiella*-44 の間の組換え体であることがわかる。したがって、*Klebsiella* 属同士の遺伝的組換えによって *Enterobacter* 属の中のある 16S rRNA 遺伝子は成り立っているということになる。また、ゲノム NC_015663.1 *Enterobacter aerogenes* KCT 2190 の 8 コピーの 16S rRNA 遺伝子のうち、組換え体である *Enterobacter*-20 の系統の 16S rRNA 遺伝子は 7 コピーで、組換え体 16S rRNA 遺伝子が主要な 16S rRNA 遺伝子となっている例である。

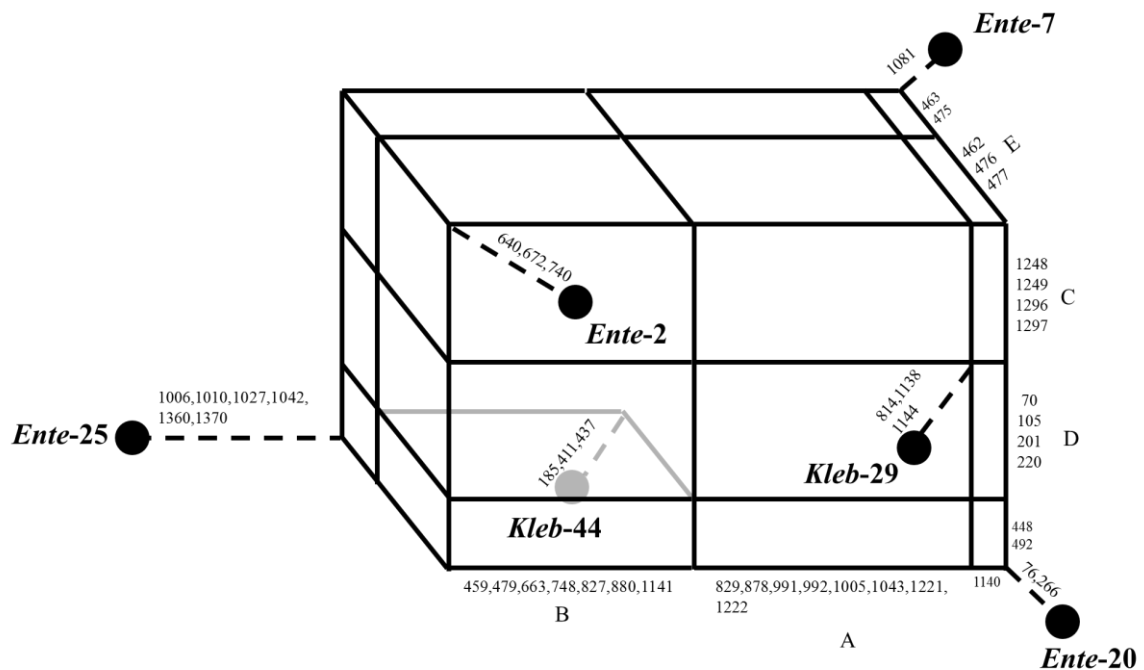


図 3.11 *Enterobacter* 属と *Klebsiella* 属の 16S rRNA 遺伝子を用いて作成した節約的な系統ネットワーク。

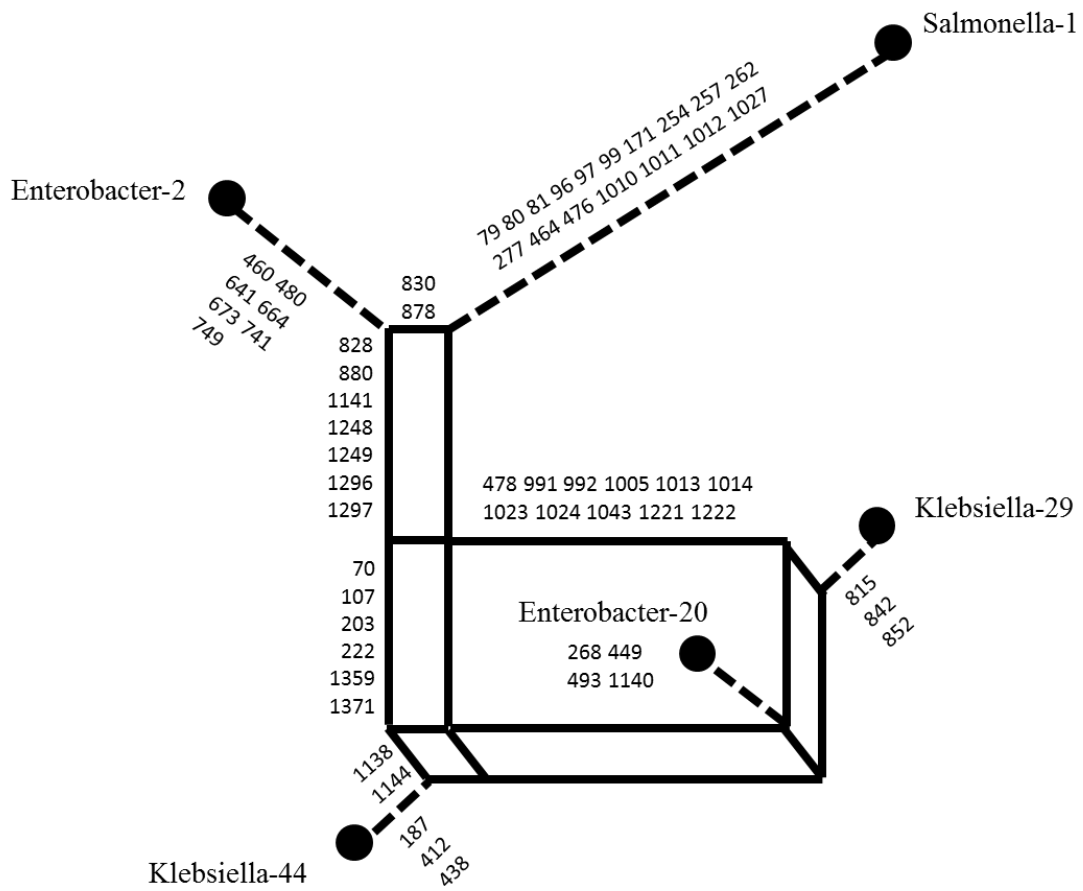


図 3.12 *Salmonella* を加えて作成した、*Enterobacter* 属と *Klebsiella* 属の 16S rRNA 遺伝子の節約的な系統ネットワーク。

c) *Erwinia* 属と *Pantoea* 属

Erwinia 属と *Pantoea* 属の 16S rRNA 遺伝子を用いて作成した系統ネットワークを図 3.13 に示した。連続したサイトで構成されている辺 A(サイト 133-238)や辺 C(サイト 79-99)が、辺 B(サイト 96-1380)と対立しているのが水平伝播の存在が示唆されている。系統群としての外群は *Cronobacter-1* だが、*Pantoea-2* のほうがシングルトンは長く、またシングルトンサイトも全域に分布しているため、実際に *Cronobacter-1* を遺伝的組換えに関わっていない配列であると判断ができない。そこで、さらに遠い系統群である *Sodalis* を加え、図 3.13 からループの外側の配列のみを用いて Bandelt の系統ネットワークを作成してみると、*Sodalis-1* は *Cronobacter-1* の近傍に結節しなかった(図 3.14)。加えて、*Sodalis-1* はシングルトンが極めて長く、シングルトンサイトの分布も広くみられるため、遺伝的組換えに関わっていない配列であると考えられる。このことは、*Cronobacter-1* と *Pantoea-15* の間に生じた水平伝播によって *Erwinia-5* が生じたことを意味する。

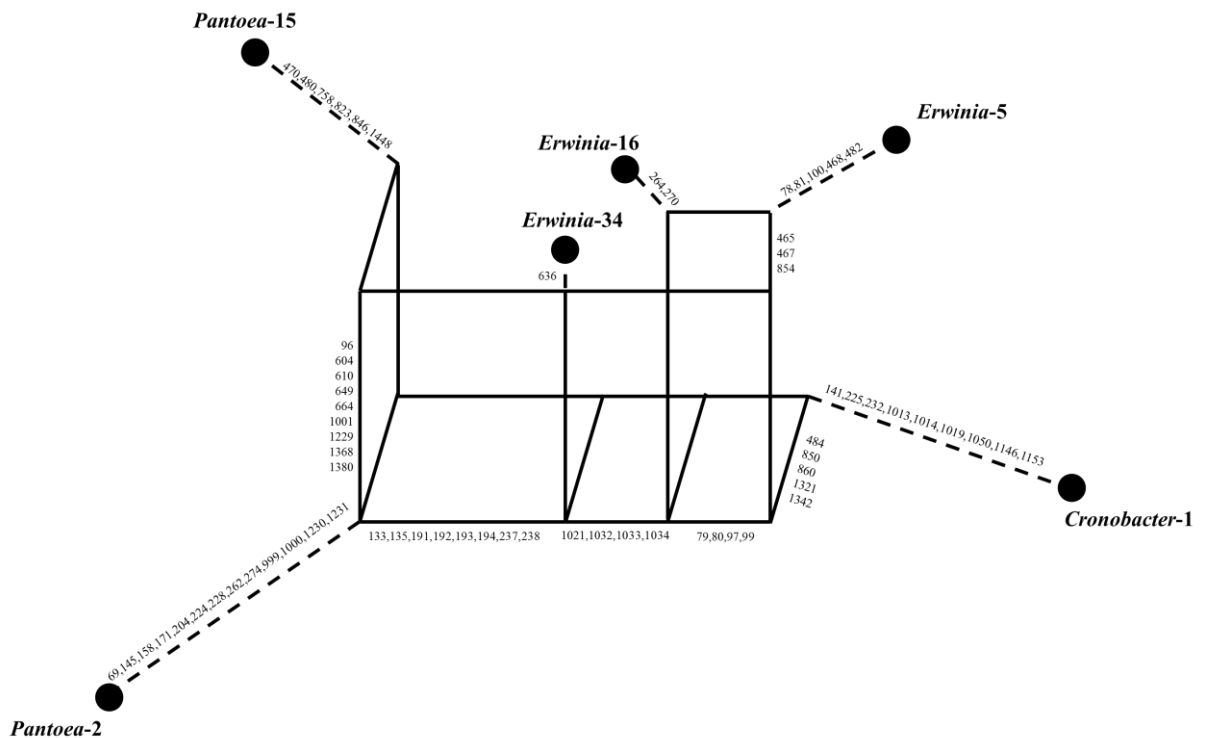


図 3.13 *Pantoea* 属と *Erwinia* 属の 16S rRNA 遺伝子を用いて、外群として *Cronobacter* の配列を加えて作成した節約的な系統ネットワーク

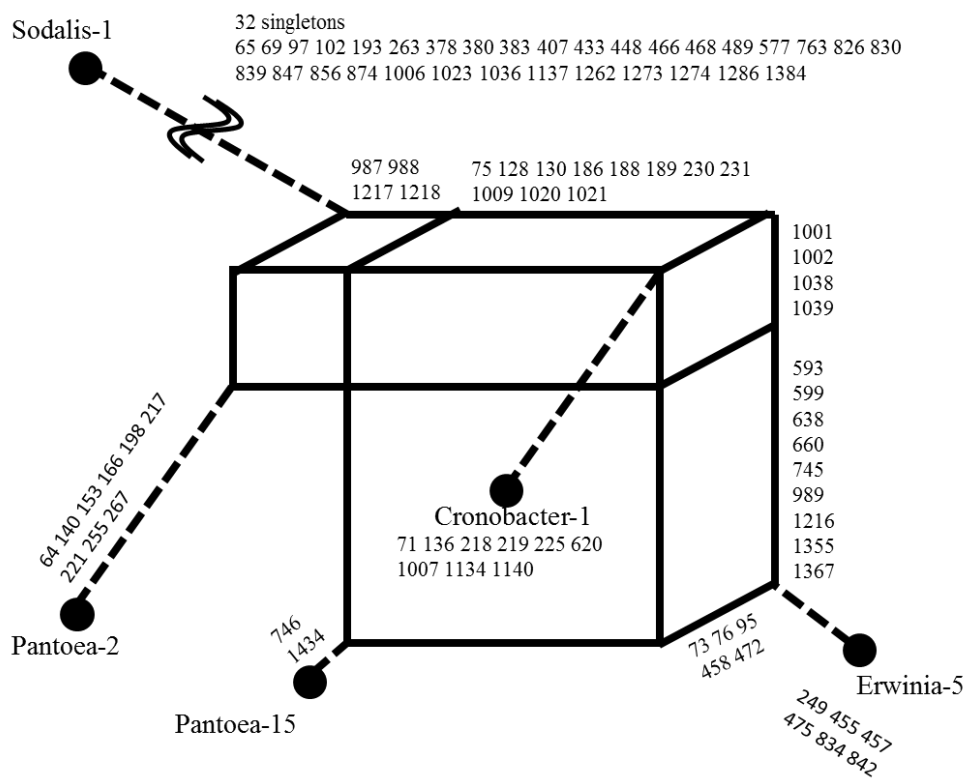


図 3.14 図 3.13 の系統ネットワークから網目の外側を構成する配列に外群として *Sodalis* の 16S rRNA 遺伝子を加えて作成した節約的な系統ネットワーク。

d) *Pectobacterium* 属と *Dickeya* 属

Pectobacterium 属と *Dickeya* 属の 16S rRNA 遺伝子を用いて作成した系統ネットワークを図 3.15 に示した。*Sodalis-1* はシングルトンが長く、シングルトンサイトの分布も配列全域に見られるので遺伝的組換えに関わっていない配列であると考えられる。サイト 1008-1154 は連続しており、かつ、サイト 70-1294 と対立してループ構造を示しているので水平伝播の存在が示唆される。*Sodalis-1* を遺伝的組換えに関わっていない配列であると考え、*Dickeya-10* は *Dickeya-1* と *Pectobacterium* 属の間の水平伝播によって生じた配列であると考えられる。

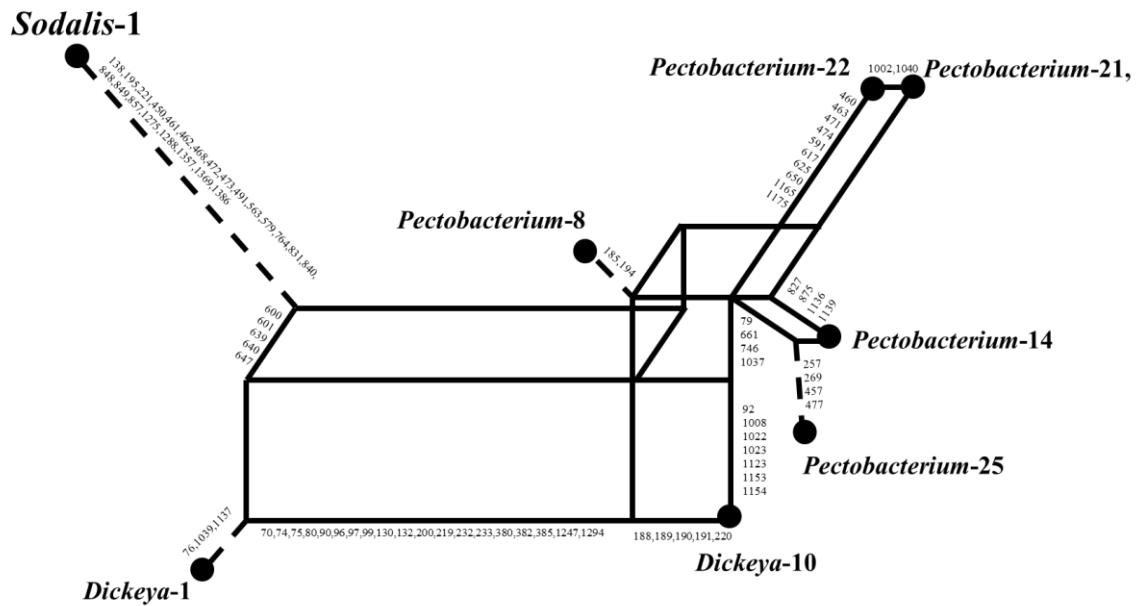


図 3.15 *Sodalis-1* を外群として作成した、*Dickeya* 属と *Pectobacterium* 属の 16S rRNA 遺伝子の節約的な系統ネットワーク。

e) *Yersinia* 属と *Escherichia* 属

ゲノム NC_0154224.1 *Yersinia enterocolitica* subsp. Palearctica 105.5R(r) のもつ 7 コピーの 16S rRNA 遺伝子のうち 1 コピーは他の 6 コピーとは大きく異なった配列を持っており、Enterobacteriales 目内において *Yersinia* 属以外では *Escherichia* 属に近い。図 3.16 は、NC_0154224.1 のもつ 7 コピーの 16S rRNA 遺伝子と *Escherichia coli* MG1655 株の *rrsB*、*Providencia* の 16S rRNA 遺伝子によって作成した系統ネットワークである。シングルトンが長く、シングルトンサイトの分布も広くみられる *Providencia* を遺伝的組換えに関わっていない配列と考えると、*Yersinia* NC_015224.1 copy1 は *Escherichia* と *Yersinia* NC_015224.1 copy 2-7 の間の水平伝播によって生じた配列であると考えられる。ループの大きさが特に大きく、対立する辺同士のサイトの分布も横辺が前半で連続しているのに対して、縦辺は後半で連続しているため、水平伝播を強力に支持する結果であり、*Yersinia* NC_015224.1 の 7 コピー中の 1 コピーに *Escherichia* の 16S rRNA 遺伝子が水平伝播したことを意味する。水平伝播によって生じた *Yersinia* NC_015224.1 copy1 は 7 コピー中 1 コピーに過ぎず、16S rRNA 遺伝子として機能を保っているかは分からないが、16S rRNA 遺伝子の水平伝播は、このようにして複数コピーの中の 1 コピーを異種由来の 16S rRNA 遺伝子が乗っ取ることにより始まるのかもしれない。また、*Yersinia* NC_015224.1 copy1 が、今後、他のコピーとの間で配列が均一化されてゆく過程で、*Escherichia* 由来の 16S rRNA 遺伝子が、部分的には *Yersinia* NC_015224.1 の 7 コピーの間で主要な遺伝型となってゆくような狭い領域でのモザイク化が進行してゆくかもしれない。

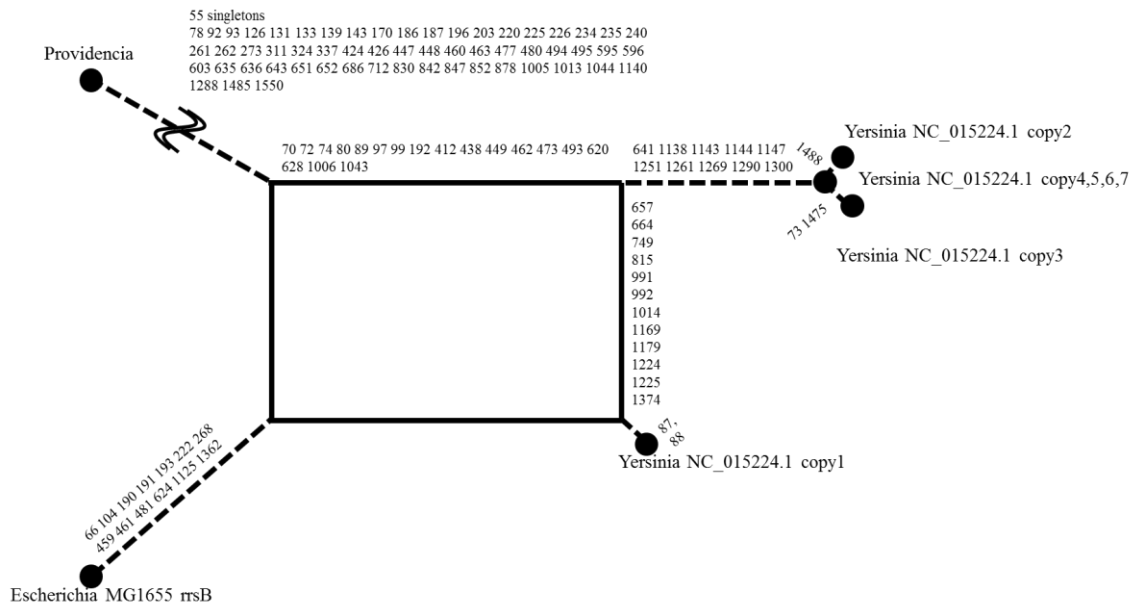


図 3.16 ゲノム NC_0154224.1 *Yersinia enterocolitica* subsp. Palearctica 105.5R(r) の 7 コピーの 16S rRNA 遺伝子に *Escherichia coli* MG1655 *rrsB* と外群として *Providencia* を加えて作成した節約的な系統ネットワーク。

Enterobacteriales 目において普遍的にみられる 16S rRNA 遺伝子の水平伝播

図 3.17 に、a-e において推定された水平伝播を、アミノ酸配列を元に最尤法を用いて作成された系統樹(Gillespie et al. 2011)に対して、矢印で示した。また、属内で水平伝播が観察された分類群には黒丸で記しを付けた。図 3.17 は、Enterobacteriales 目で 16S rRNA 遺伝子の水平伝播に約半数の分類群が関わっていることを表しており、過去に固有の事例として議論されていた 16S rRNA 遺伝子の水平伝播は、すくなくとも Enterobacteriales 目内では広くみられる現象であることが分かった。

水平伝播によって大きく変化する 16S rRNA 遺伝子

また、系統ネットワークのループ構造に対して、シングルトンが短いという特徴があった。例えば、翻訳に関わる遺伝子の 1 つであるエロンゲーションファクターG をコードしている *fus* 遺伝子を用いて作成した系統ネットワーク (図 3.18) と比べても、同じ系統群の範囲で比較した 16S rRNA 遺伝子の場合 (図 3.10) ループ構造の大きさに対してのシングルトンの長さが比較的短い。また、シングルトンが極めて短い配列がいくつもループ構造の淵や中に結節しているという特徴も見られる (図 3.3(B)(C)、図 3.11、図 3.13、図 3.15 など) このことは、16S rRNA 遺伝子の多様度が生み出される過程で、点突然変異と比較して、水平伝播のほうが受け入れられやすいことを意味しており、16S rRNA が二次構造の変化を受け入れにくい分子であることを考えると自然である。Cannone ら(2002)による *Escherichia coli rrsB* の二次構造に当てはめて考えると、本研究において推定された水平伝播は、いずれも二次構造に大きな変化はなかった。16S rRNA 遺伝子が機能的に点突然変異に対して脆弱である(Yassin ら, 2005)という結果や、二次構造が大きく変わらない異種の 16S rRNA 遺伝子ならば、大腸菌の生育を相補可能である(Kitahara ら, 2012)という報告とも合致する結果が得られた。

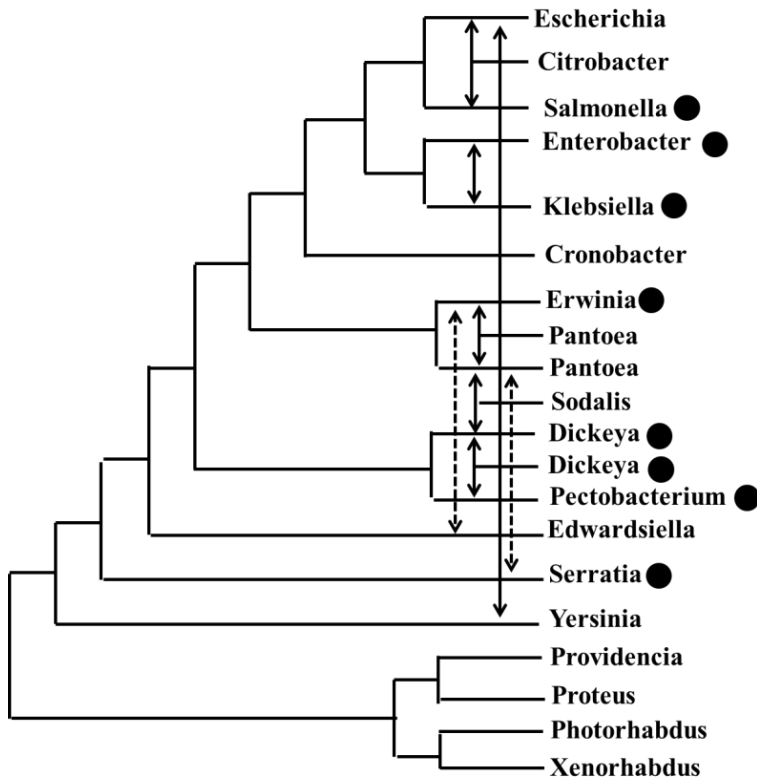


図 3.17 Enterobacteriales 目において示唆された水平伝播。●がついている分類群は属内で水平伝播が示唆された分類群。属を跨いだ生じた水平伝播は→で示した。水平伝播が生じていることは示唆されたが、親子関係が不明であったものについては、破線の→で示した。系統樹は最尤法を用いて、ゲノム上の全ての各遺伝子のアミノ酸配列を用いて作成されたものである(Gillespie et al., 2011)。

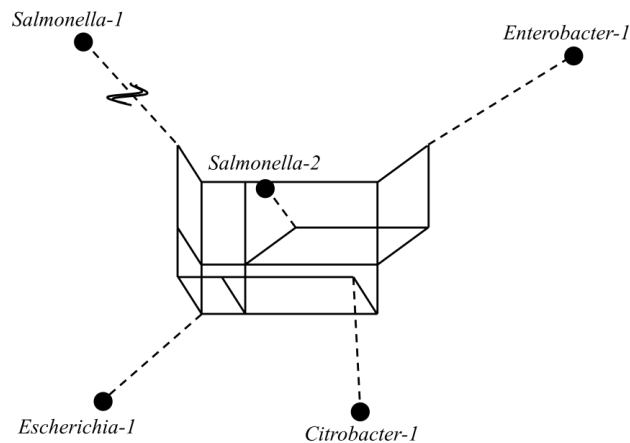


図 3.18 *Escherichia*, *Salmonella*, *Citrobacter*, *Enterobacter* のエロンゲーションファクターG (*fus* 遺伝子)を用いて作成した節約的な系統ネットワーク。同じ分類群の範囲で比較している、図 3.10 に比べ、網目の大きさに対してシングルトンの枝の長さが長い。

16S rRNA 遺伝子の水平伝播が与える影響

16S rRNA 遺伝子のゲノム内多型についてはこれまで知られている（例えば Mylvaganam と Dennis (1992)、Wang ら(1997)、Yap ら(1999)）が、ゲノム内多型が適応度に影響を与えているのかどうかについての報告は存在しない。しかし、水平伝播によって作られたキメラ 16S rRNA 遺伝子は、例えば *Citrobacter* 属の 7 コピーすべてがキメラ体であるように、十分に機能していると言える。このため、16S rRNA 遺伝子の水平伝播は、少なくとも中立に生ずることはあるということだ。

我々の研究室では、16S rRNA 遺伝子が部分的に水平伝播することによって、2 種類の薬剤に対して耐性を示す 16S rRNA 遺伝子をメタゲノム由来で取得した（宮崎、未発表）。図 3.19 の系統ネットワークを構成する 4 配列のうち、シングルトンの長さから、Spc1 は Spc6 と G418-4 の間に生じた水平伝播によって生じたキメラ体であった。G418-4 は G418 に対して耐性を示す 16S rRNA 遺伝子で、Spc6 はスペクチノマイシンへと耐性を示す 16S rRNA 遺伝子であった。そして、Scp1 は、G418、スペクチノマイシン双方への耐性を示すことがわかっている。このことは、水平伝播によって新たな機能が 16S rRNA 遺伝子に付与されうることを示している。ただし、Spc1 はメタゲノム由来なので、実際にホストの適応度に対する寄与は不明であり、16S rRNA 遺伝子として機能しているかどうか不明である。しかし、16S rRNA 遺伝子の水平伝播が新たな機能や性質をもたらすのではないかという問いは十分に追及する価値があるということだ。

加えて、16S rRNA 遺伝子には翻訳以外の機能も存在する。当研究室の Kitahara と Miyazaki (2011)は、16S rRNA の構造の一部である helix41 が、ペリプラズム酵素である RNase T2 を阻害することを発見した。Kitahara と Miyazaki (2011)の報告は、16S rRNA 遺伝子の水平伝播に対する防衛機構のようにも思われ、16S rRNA 遺伝子の水平伝播を阻む一種の壁である可能性はあるが、helix41 が部位特異的にこのような性質を持っているということなどから、16S rRNA の一部に埋め込まれた機能であると考えられることもできる。従って、16S rRNA 遺伝子の部分的な水平伝播によって、このような翻訳外機能がもたらされるようなことがあってもおかしくはない。

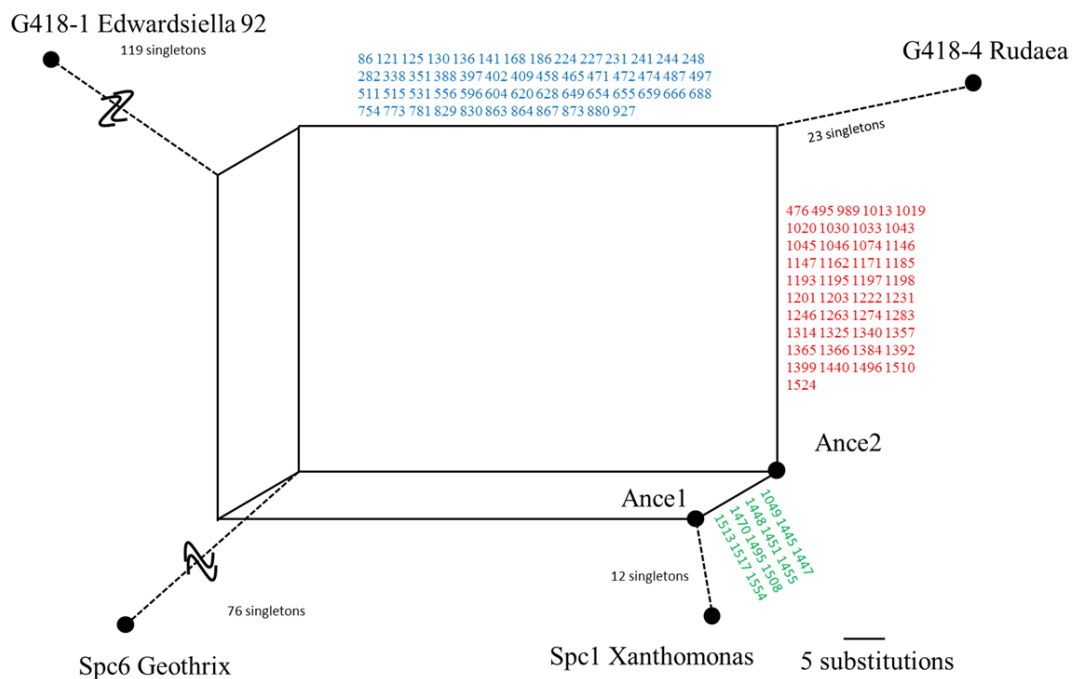


図 3.19 メタゲノム由来の薬剤耐性 16S rRNA 遺伝子の系統ネットワーク。G418-1、4 は G418 耐性でスクリーニングした 16S rRNA 遺伝子で、Spc-1 と Spc-6 はスペクチノマイシン耐性でスクリーニングした 16S rRNA 遺伝子。シングルトンの分布から、Spc1 は G418-4 と Spc-6 の間の遺伝的組換えによって生じた配列であると考えられる。また、Spc1 は、G418、スペクチノマイシン双方に耐性を示すため、遺伝的組換えによって親の 16S rRNA 遺伝子の特徴のハイブリッドとなっていると思われる。16S rRNA 遺伝子の水平伝播によって機能が付与される例である。

第三章 参考文献

- Bandelt, H. J. (1994). Phylogenetic networks. *Verh Natwiss Ver Hambg.* 34:51-71.
- Cannone et al. (2002). The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3:2.
- Gillespie et al., (2011). PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. *Infect. Immun.* 79(11):4286-4298.
- Huson, D.H., and Bryant, D., (2006). Application of phylogenetic networks in evolutionary studies, *Mol. Biol. Evol.* 23(2):254-267.
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30(4):772-780.
- Kitahara, K., and Miyazaki, K. (2011). Specific inhibition of bacterial RNase T2 by helix 41 of 16S ribosomal RNA. *Nat. Commun.* 2;2:549.
- Kitahara, K., Yasutake, Y., and Miyazaki, K. (2012). Mutational robustness of 16S ribosomal RNA, shown by experimental horizontal gene transfer in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 109(47):19220-19225.
- Mylvaganam, S., and Dennis, P. P. (1992). Sequence heterogeneity between the two genes encoding 16S rRNA from the halophilic archaeobacterium *Haloarcula marismortui*. *Genetics* 130(3):399-410.
- Saitou, N., and Yamamoto, F. (1997). Evolution of primate ABO blood group genes and their homologous genes. *Mol. Biol. Evol.* 14(4):399-411.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30(12):2725-2729.
- Wang, Y., Zhang, Z., and Ramanan, N. (1997). The actinomycete *Thermobispora bispora* contains two distinct types of transcriptionally active 16S rRNA genes. *J. Bacteriol.* 179(10):3270-3276.
- Yap, W. H., Zhang, Z. S., and Wang, Y. (1999). Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. *J. Bacteriol.* 181(17):5201-5209.
- Yassin, A., Fredrick, K., and Mankin, A. S. (2005). Deleterious mutations in small subunit ribosomal RNA identify functional sites and potential targets for antibiotics. *Proc. Natl. Acad. Sci. USA* 102(46):16620-16625.

第四章 網羅的な 16S rRNA 遺伝子の水平伝播の検出

1. 序論

遺伝的組換えの検出に有益な系統ネットワーク法

遺伝的組換えを検出するためには、領域間の系統関係の食い違いを検出する必要がある。Bandelt の方法によって作成する節約的な系統ネットワークが遺伝的組換えの検出に有益である理由は、サイトごとの情報を反映させて作成するため、どこの領域と、どこの領域の系統関係が対立しているかを発見できる点にある (Saitou and Yamamoto 1997; Kitano et al., 2012)。

節約ネットワーク法に基づいて遺伝的組換えを検出する PNarec

PNarec (Saitou と Kitano 2012)は、節約的に情報を持ったサイトのパターンの食い違いを検出することで遺伝的組換えと組換え体を検出するプログラムである。節約的に情報を持ったサイトとは、節約原理に基づいて系統関係を確定することができるサイトの総称である。PNarec の遺伝的組換え体を検出する原理は、Bandelt の方法で作成される系統ネットワークに基づいている。

PNarec は、データセットに含まれる配列から、全ての 3 配列の組み合わせに外群を加えて 4 配列のデータとし、4 配列間の節約的に情報を持ったサイトのパターンの食い違いを検出することで、組換え体を発見する手法を用いている。4 配列間の、節約的に情報を持ったサイトが示唆する系統関係は、図 4.1 に示した 3 通りだけである。遺伝的組換えが生ずると、この節約的に情報を持ったサイトの系統関係は、遺伝的組換えが生じたサイトを境に切り替わることになるので、系統ネットワークにループ構造が見られる。図 4.2 のモデルデータは、Sequence 2 が組換え体で、Sequence 1 と 3 が親配列、Sequence 4 は外群である。これらの配列を用いて作成した節約的な系統ネットワークはループ構造を示し、

- 1) 外群の対角に、組換え体が短いシングルトンで結節する。
- 2) 組換え体の両側に親配列が結節する。
- 3) 系統ネットワークにループ構造が見られ、縦辺と横辺を構成するサイトの分布がそれぞれ領域的に固まっている。

という特徴を持つので、遺伝的組換えの存在が示唆される。そして、節約的に情報を持ったサイトの系統関係のパターンは、遺伝的組換えが生じたサイト以前は 1 だが、それ以降は系統関係 3 を示している。PNarec は、この 4 配列のうち、1 配列は外群として固定した配列を用い、あとの 3 配列をデータセットから加えて、節約的に情報を持ったサイトごとに系統関係を見ていき、図 4.2 で示したような途中から系統関係が変化するような配列の組み合わせを探索することで組換え体を探す。

PNarec 法をバクテリア遺伝子に用いることの問題点

手法に関する議論においても述べたように、これまでの報告における節約的な系統ネットワークの解釈 (Saitou and Yamamoto 1997, Kitano et al., 2009, 2012) をバクテリア遺伝子に適用すると、組換えの親子関係について間違える恐れがある。すなわち、ある分類群内において遺伝的組換えが生じている時、その分類群よりも外の分類群 (外群) を加えて系統ネットワークを作成すると、組換え体は外群の対角に位置する。しかし、**離れた系統群同士であっても水平伝播しうるバクテリアにおいては、系統ネットワーク上で、外群の対角が常に組換え体とは限らない**。したがって、外群を固定して解析を行う PNarec は、バクテリア遺伝子の解析には適さない手法であると考えられる。

2. 材料と方法

遺伝的組換え体の網羅的な検出方法

外群を固定せずに解析を行うために、データセットから、全てのあり得る 4 配列の組み合わせを取り出し、図 4.2 に示したように、節約的に情報を持ったサイトの系統関係を記述していくプログラムを、Perl スクリプトを用いて作成した。もし、遺伝的組換えが生じているのであれば、節約的に情報を持ったサイトの系統関係は、遺伝的組換えが生じたサイトの上流と下流で変化するはずである。例えば、図 4.3 に示す 4 配列においては、上流 (サイト 709-1126) における系統関係 3 が、下流 (サイト 1364-4906) においては系統関係 2 へと変化している。

遺伝的組換えを判断する検定方法

サイトごとの系統関係の食い違いは、平行進化によってもたらされうる。例えば、図 4.3 における系統関係 1 は、先頭に 2 サイト、下流部分に 1 サイトが示されているのみで、3 回の平行進化によってもたらされうる。平行進化はランダムに生ずると考えられるため、平行進化によってもたらされたこのようなサイトの分布はランダムになると考えられる。そこで、本研究においては、4 配列が示しうる 3 つの系統関係の並び順のランダム性を検証した。ランダム性の検証には、Takahata (1996) の 2 sample runs test に基づいた、Ezawa ら (2006) の手法を用いた。

2 sample runs test の計算方法

2 sample runs test は、 α 個の文字 A、 β 個の文字 B から成る長さ n ($n=\alpha+\beta$) のランダムな並びの文字列について、'Run'数 (Z) が観察値 (R) を下回る確率 $P(Z \leq R)$ を求める。'Run'とは、異なる文字同士によって区切られた、同じ文字のブロックと定義され、例えば、ABAABBA という文字列であれば、

A|B|AA|BB|A

となり、 $R=5$ 、 $\alpha=4$ 、 $\beta=3$ となる。一般的には、ある Run 数 Z が観察される確率 $P(Z)$ は、 α 、 β 、 n を用いて、 $Z=2k$ が偶数の時には、

$$P(Z = 2k) = 2 \binom{\alpha - 1}{k - 1} \binom{\beta - 1}{k - 1} \binom{n}{\alpha}^{-1}$$

$Z=2k+1$ が奇数の時には、

$$P(Z = 2k + 1) = 2 \left\{ \binom{\alpha - 1}{k} \binom{\beta - 1}{k - 1} + \binom{\alpha - 1}{k - 1} \binom{\beta - 1}{k} \right\} \binom{n}{\alpha}^{-1}$$

となる。 $P(Z \leq R)$ は $2 \leq Z \leq R$ における $P(Z)$ の総和、

$$P(Z \leq R) = \sum_{Z=2}^R P(Z)$$

である。

Ezawa ら (2006) の手法の計算例

図 4.3 における系統関係の文字列を用いて実際に計算してみる。2 sample runs test においては、2 種類の文字種から成る文字列について考えたが、節約的に情報を持ったサイトは 4 配列から 3 通り出現しうるため、実際には 3 種類の文字種から成る文字列について考える。

Ezawa ら (2006) の手法は、Takahata (1996) と同様に、ある文字に着目した場合、そのほかの文字種については、全て同一の文字種であると見做す。すなわち、例えば、図 4.3 における系統関係の文字列の系統関係 1 に着目した場合、2 と 3 を同一の文字種と見做すため、 $R=4$ 、 $\alpha=3$ 、 $\beta=17$ となり、 $P(Z \leq 4)=0.07$ である。

11 | 3333333222 | 1 | 222222

同様に、系統関係 2 に着目した場合は、 $P(Z \leq 4)=0.001$ 、系統関係 3 に着目した場合は、 $P(Z \leq 3)=0.0002$ である。本研究においては、有意水準を 1% と定めて検定を行った。また、組換えを示唆する節約的に情報を持ったサイトのうち、90% 以上が連続していることを組換えが生じていると判定する要件とした。なぜならば、2 sample runs test のみでは、配列のランダム性は評価できても、領域的な連続性は評価できないからである。

種名	シングルトン数
1 NC_019738.1 428308377:2627318-2628787_Microcoleus_sp._PCC_7113	72
2 NC_019684.1 427727289:2137947-2139413_Nostoc_sp._PCC_7524	17
3 NC_010628.1 186680550:2021439-2023027_Nostoc_punctiforme_PCC_73102	26
4 NC_003272.1 17227497:5945650-5947238_Nostoc_sp._PCC_7120	7

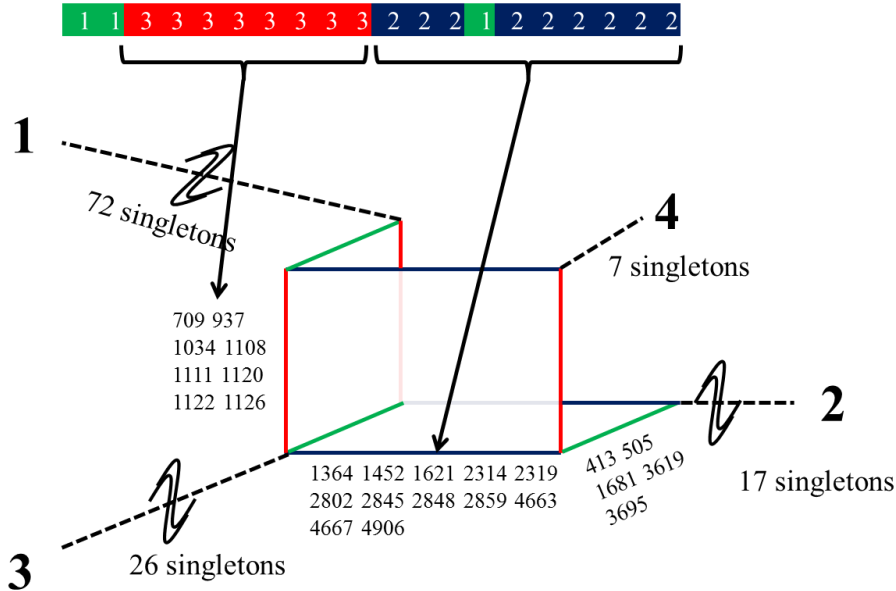


図 4.3 節約的に情報を持ったサイトを 5'末端側から記述した配列と、同じ配列データから作成した系統ネットワーク。赤領域で連続して 8 サイトほど系統関係 1 が示唆されており、後半部分では系統関係 2 が示唆されているため、遺伝的組換えの存在が示唆されている。

データの取得と整理

2014 年 11 月 1 日時点で、NCBI に登録されているバクテリアとアーキアのゲノム配列すべてから 16S rRNA 遺伝子領域を NCBI のアノテーションに従って抽出した。表 4.1 に解析に用いた分類群と 16S rRNA 遺伝子数を記載した。多重整列には MAFFT (version 7) を用い、バクテリア全体で 5% 以上のギャップを含むサイトは解析から取り除いた。抽出した配列は門ごとにまとめ、節約的に情報を持ったサイトの記載は、門ごとに行った。

遺伝的組換えが生じている領域の求め方

遺伝的組換えが生じている領域は、以下の手順で求めた

- (1) 全ての 4 配列の組み合わせのうち、節約的に情報を持ったサイトの系統関係のパターンが 2 sample runs test の結果が 1% 水準で有意であった組み合わせを取りだした。
- (2) 組換え体として判断された配列と同じ属に属する異なる配列が同様に組換え体である

と判断された場合、節約的に情報を持ったサイト量が最も多い4配列の組み合わせを代表として選んだ。

- (3) 異なる4配列の組み合わせにおいて、まったく同一のパターンが見られた場合は、同じ遺伝的組換えによって生じた組換え体であると判断し、1つを残して解析から除去した。
- (4) 複数領域が遺伝的組換え候補として選ばれた場合は、遺伝的組換えの回数を最少で説明可能な領域を選択した。

表 4.1 網羅的に遺伝的組換えの検出を試みた分類群と配列数。

分類群(門)	N
Actinobacteria	839
Aquificae	27
Armatimonadetes	2
Bacteroidetes	265
Chlorobi	19
Ignavibacteriae	2
Caldiserica	1
Chlamydiae	193
Verrucomicrobia	7
Chloroflexi	34
Chrysiogenetes	3
Cyanobacteria	168
Deferribacteres	7
Deinococcus-Thermus	51
Dictyoglomi	4
Elusimicrobia	1
Acidobacteria	6
Fibrobacteres	6
Firmicutes	3395
Fusobacteria	37
Gemmatimonadetes	1
Nitrospirae	7
Planctomycetes	18
Proteobacteria	4459
Spirochaetes	106
Synergistetes	11
Tenericutes	111
Thermodesulfobacteria	3
Thermotogae	29
Archaea (界)	858
合計	9812

どの分類群とどの分類群で遺伝的組換えが生じたのかを求める

本解析においては、細菌の 16S rRNA 遺伝子の遺伝的組換えが、どの水準の分類群を超えて生じるのかを以下の手順に従って調べた。

- (1) 遺伝的組換えが示唆された 4 配列の組み合わせにおいて、2 sample runs test の値が最も低い組み合わせを種ごとに選んだ。
- (2) 第二章の **Bandelt** の方法で作成した系統ネットワークの特徴で述べたとおり、4 配列から成る節約的な系統ネットワーク上で、親配列は組換え体を挟む位置に結節する。組換え体を挟む 2 配列を親配列として選び、両者の分類群を比較し、綱、目、属のどの分類群の水準で異なっているかを調べた。例えば、綱は同じだが目が異なる場合は、目を飛び越えて水平伝播が生じているということになり、属レベルでも異ならなかった場合は、属内で生じた遺伝的組換えであるということになる。

3. 結果と考察

分類群ごとの組換え体の頻度

図 4.4 は細菌において、20 本以上の 16S rRNA 遺伝子のエントリーが存在する分類群において検出された組換え体を含む種の割合である。細菌においては、13 門中 6 門で 16S rRNA 遺伝子の遺伝的組換えが生じていることが分かった。細菌とは界が異なる、アーキアにおいても、調べた 669 種のうち 170 種が組換え体 16S rRNA 遺伝子を保有していた。これらの結果は、細菌のみならず、アーキアにおいても 16S rRNA 遺伝子の遺伝的組換えが存在することが示唆している。

遺伝的組換えと 16S rRNA 遺伝子のコピー数

図 4.5 は、遺伝的組換えが検出された分類群それぞれにおける、組換え体 16S rRNA 遺伝子を保有するゲノムの平均 16S rRNA 遺伝子コピー数である。ゲノム内の 16S rRNA 遺伝子のコピー数と遺伝的組換えの頻度との間には相関が見られず、16S rRNA 遺伝子のコピー数に関係なく 16S rRNA 遺伝子の遺伝的組換えは生じていることが分かった。

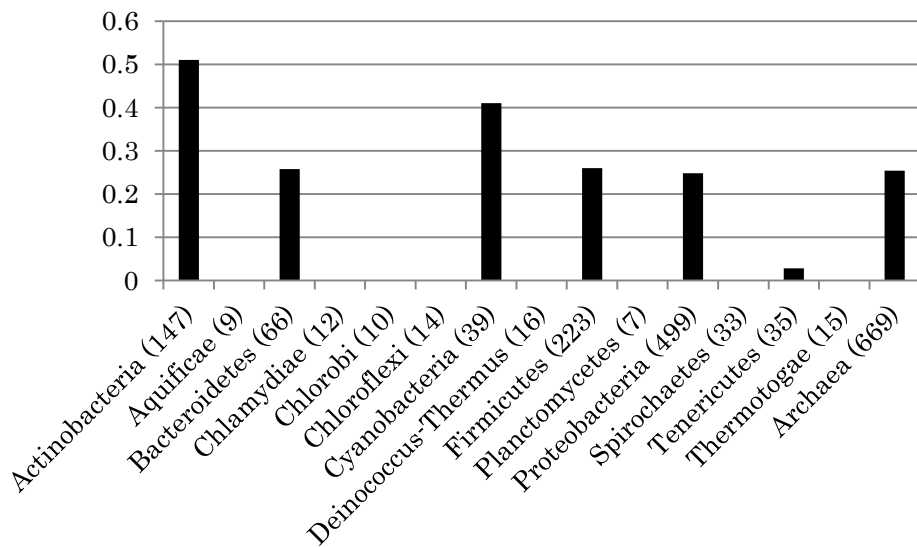


図 4.4 各分類群における、組換え体が検出された種の割合。配列名後のカッコ内に、各門における種の総数を記載した。

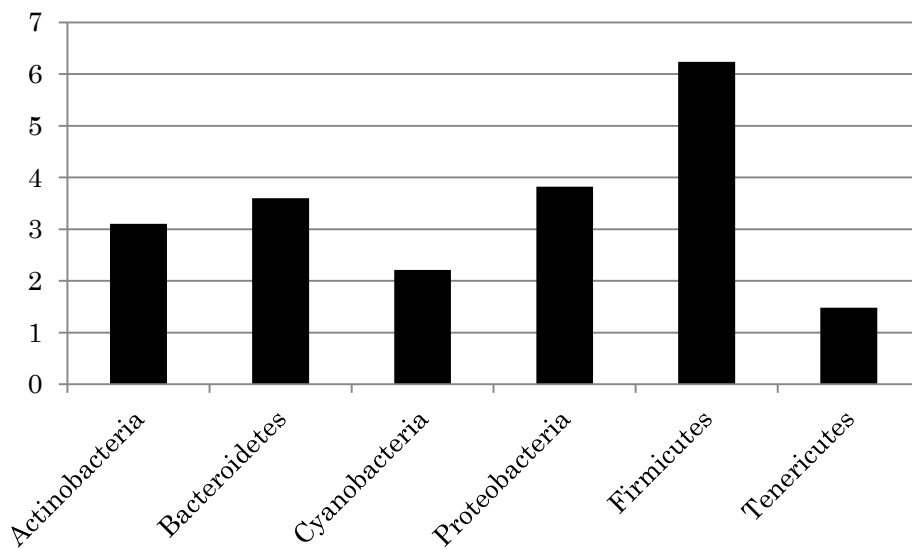


図 4.5 組換え体 16S rRNA 遺伝子を保有するゲノムの 16S rRNA 遺伝子のコピー数の、分類群ごとの平均。

遺伝的組換えが生じている領域

遺伝的組換えが生じている領域を、Proteobacteria 門 (図 4.6)、Cyanobacteria 門 (図 4.7)、Firmicutes 門 (図 4.8)、Actinobacteria 門 (図 4.9)、Bacteroidetes 門 (図 4.10) へとそれぞれ示した。記した配列は、2 sample runs test において、属ごとに最も低い P 値を示した代表配列である。遺伝的組換えによってもたらされたと考えられる領域を、ボックスで囲った。図横の矢印 (515f-806r) は、最も高頻度に種同定に用いられる領域である。

16S rRNA 遺伝子において種固有の領域は存在しない

Proteobacteria 門と、Actinobacteria 門においては配列全域に渡って、遺伝的組換えの影響を受けており、種固有の領域は存在しない (図 4.6、図 4.9)。そのほかの門においても、配列の大部分において遺伝的組換えが観察されており、16S rRNA 遺伝子において種固有の領域というものを見つけることは困難であることが分かる (図 4.7、図 4.8、図 4.10)。菌叢解析に良く用いられる領域 (515f-806r) も、遺伝的組換えの影響を受ける領域であり、この領域を用いて正確に種を同定することは不可能である。**16S rRNA 遺伝子の一部分のみを用いて種同定を行うことは、間違った結果をより得やすい**ということが言える。また、遺伝的組換えによってもたらされた断片の長さはまちまちで、15 ベース程度のものや、中には、配列の半分近くが組換えによってもたらされたものである場合もある (Yersinia) (図 4.6)。非常に短い領域に遺伝的組換えの痕跡が残っている場合については、一度異種より水平伝播が生じた後に、ゲノム内のホモジェナイゼーションが生ずることによって、極めて短い領域に遺伝的組換えの痕跡が見られるものと考えられる。

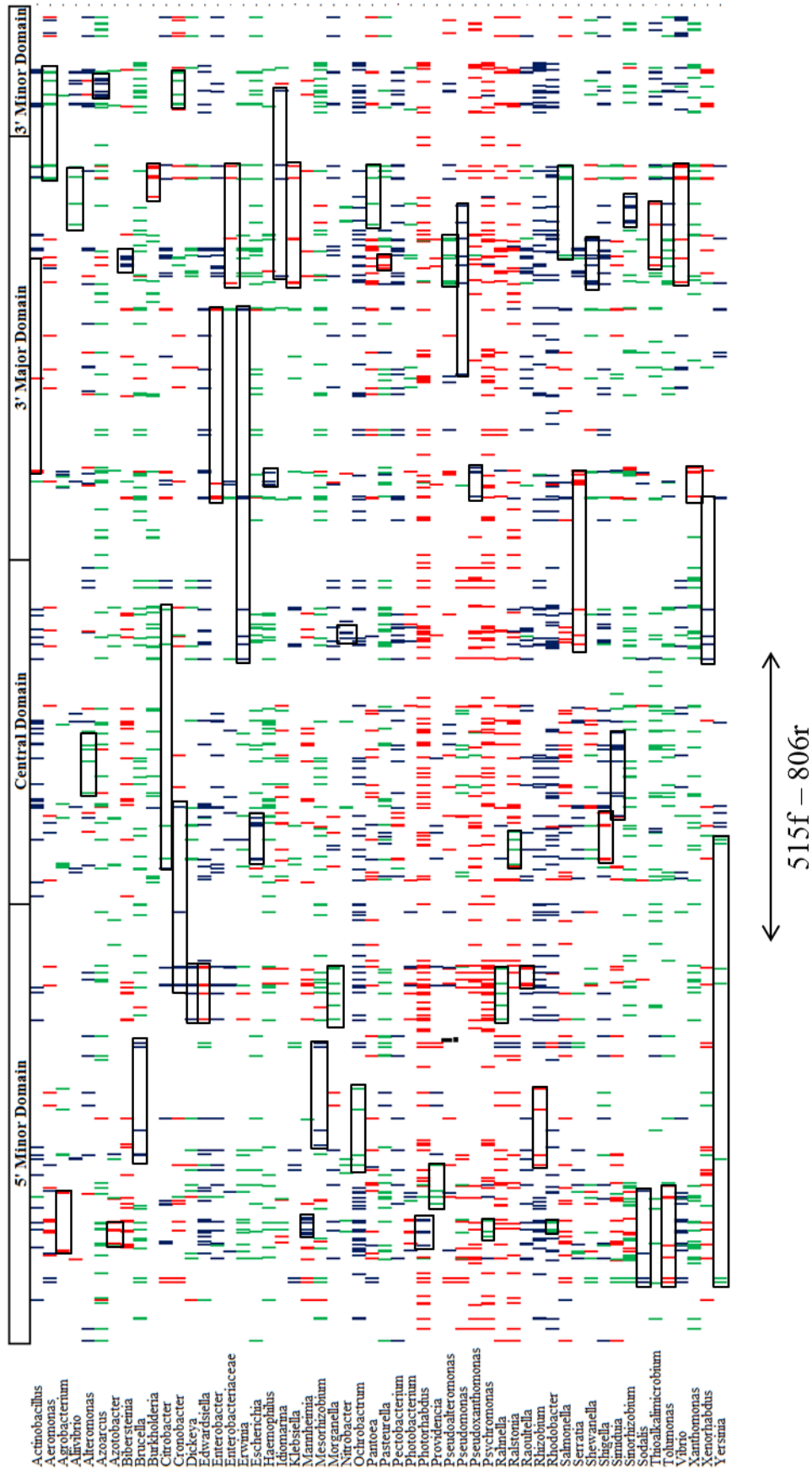


図 4.6 Proteobacteria 門において遺伝的組換えが生じていると示唆された領域。各属において組換え体と判定されたものうち、節
 約的に情報を持ったサイトが最も多く観察された 4 配列の組み合わせ計 59 通りについて、遺伝的組換えが生じていると考えられる領
 域を四角で囲った。

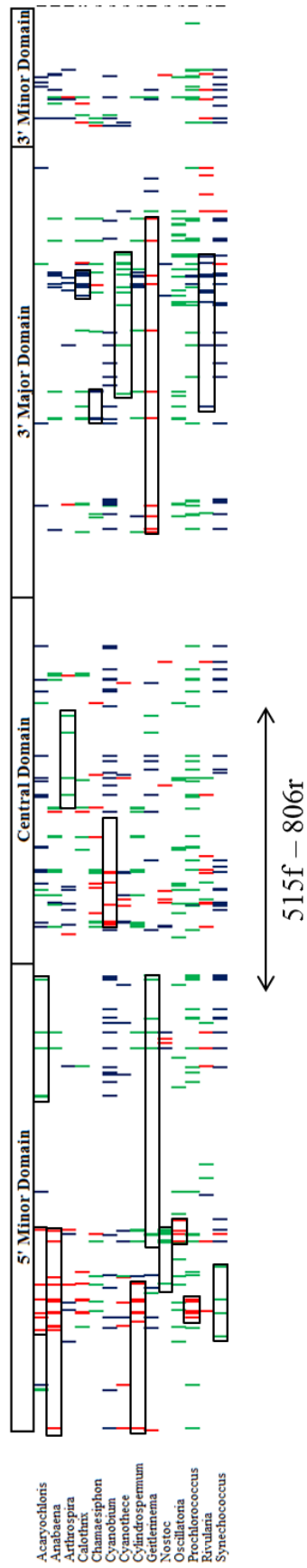


図 4.7 *Cyanobacteria* 門において遺伝的組換えが生じていると示唆された領域。

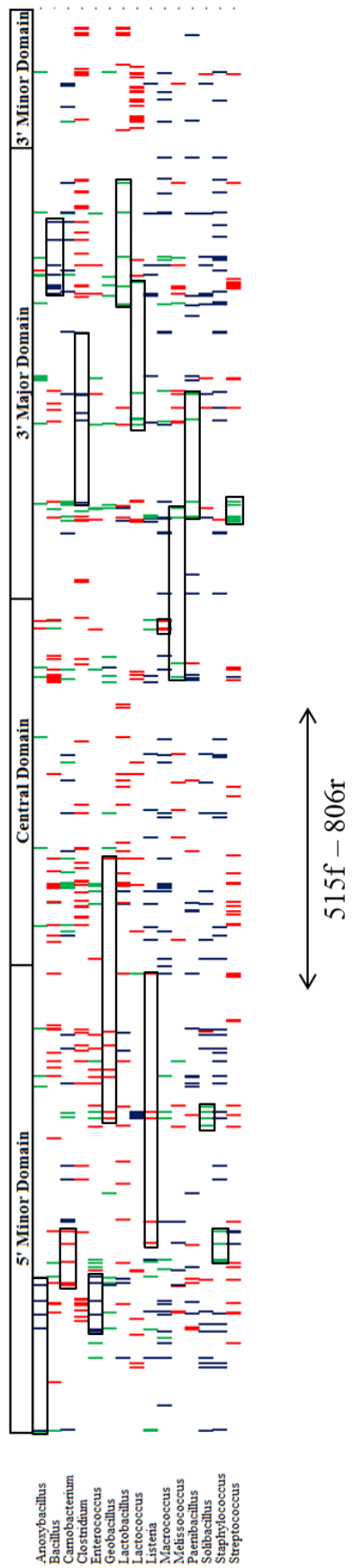


図 4.8 *Firmicutes* 門において遺伝的組換えが生じていると示唆された領域。

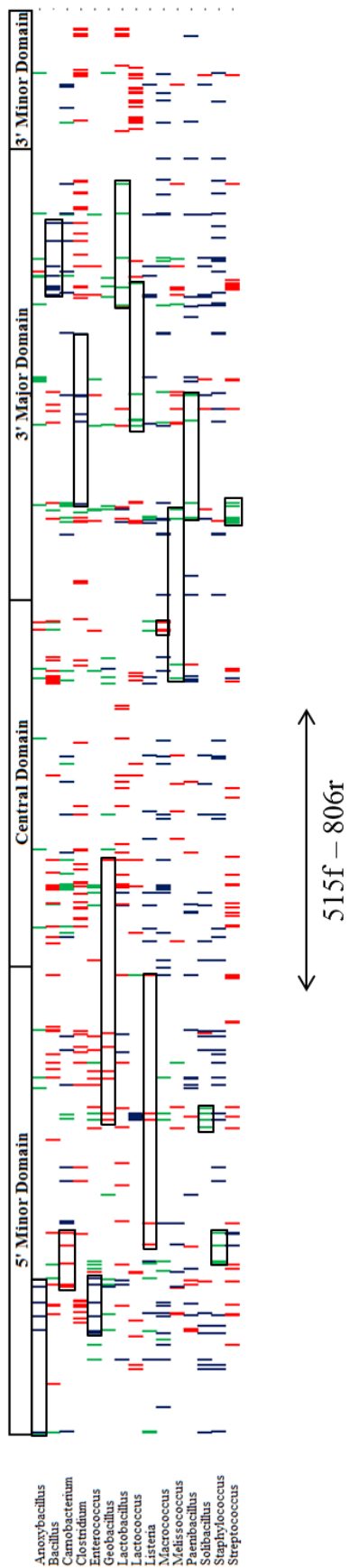


図 4.9 Actinobacteria 門において遺伝的組換えが生じていると示唆された領域。

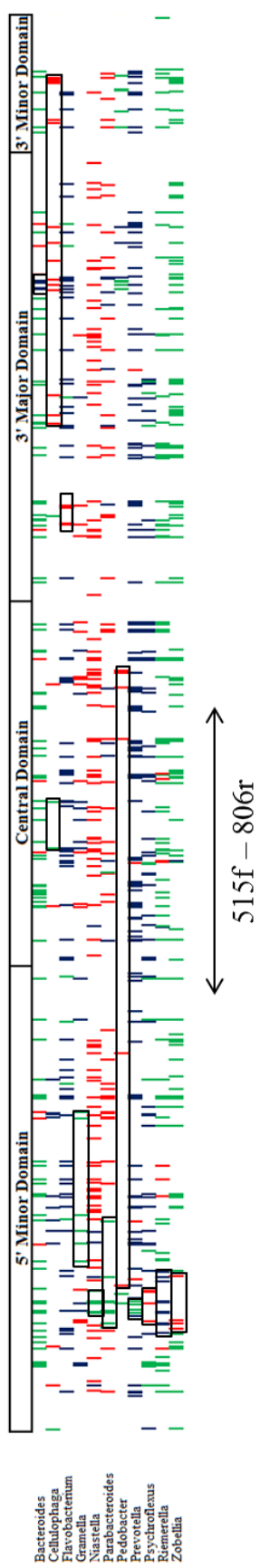


図 4.10 *Bacteroidetes* 門において遺伝的組換えが生じていると示唆された領域。

水平伝播で飛び越える分類群

16S rRNA 遺伝子の水平伝播が超える分類群を推定するために、各種内において、2 sample runs test の結果が最も低い種について、その親配列の分類群を比較した (図 4.11)。Proteobacteria 門においては、綱、目、属を超えた遺伝的組換えが、それぞれ 5、14、6 本検出された。この結果は、門以下ならば、分類群の水準を問わずに、同程度の頻度で 16S rRNA 遺伝子の遺伝的組換えは生じうることを示している。種間、属間で生じた遺伝的組換え数が、目間、綱間よりも少ない理由は、近縁な配列同士の比較となっているため、遺伝的組換えを検出するにあたっての情報量が少なかったためと考えられる。

綱を超えて生ずる 16S rRNA 遺伝子の水平伝播

第三章の *Escherichia coli*・*Shigella* 集団のように、祖先の段階のゲノム内多型が今日まで引き継がれることによって領域的な系統関係の矛盾が生じるような場合も存在する。しかし、綱や目レベルでの祖先の段階から多型が引き継がれ続けるとは考えづらいため、綱、目レベルで遺伝的組換えが生じているということは、その前段階として水平伝播が生じていると考えられる。

また、16S rRNA 遺伝子のコピー数が 1 や 2 のものも組換え体として検出されている。16S rRNA 遺伝子を 1 コピーしか持たない場合、16S rRNA 遺伝子の均一化が生じないため、16S rRNA 遺伝子が機能を失うような遺伝的組換えは受け入れられないと考えられるが、実際にはコピー数が 1 でも遺伝的組換えが生ずるということである。

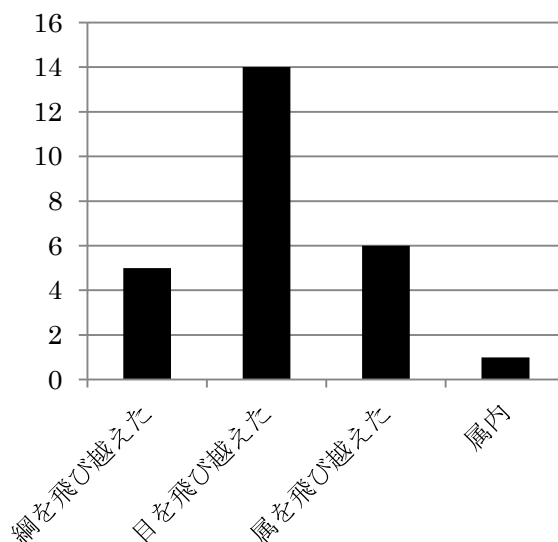


図 4.11 Proteobacteria 門において、遺伝的組換えが生じた分類群の範囲。

23S rRNA 遺伝子における遺伝的組換え

図 4.12 は、Proteobacteria 門の 23S rRNA 遺伝子において検出された遺伝的組換えの領域を示している。16S rRNA 遺伝子のみならず、23S rRNA 遺伝子にも遺伝的組換えが生じていることが示唆された。23S rRNA 遺伝子においては、16S rRNA 遺伝子よりも遺伝的組換えの領域が長く、よりダイナミックに遺伝的組換えによって 23S rRNA 遺伝子は進化しているということが示唆された。このように、バクテリアの rRNA 遺伝子にとって遺伝的組換え、延いては部分的な水平伝播は普遍的な現象であるということが考えられる。

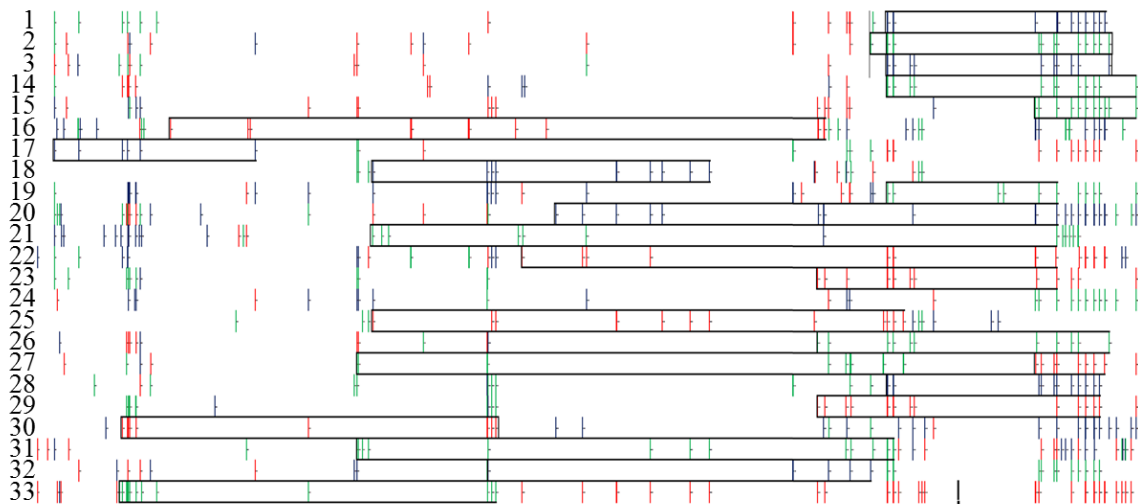


図 4.12 *Proteobacteria* 門の 23S rRNA 遺伝子において生じている遺伝的組換えの領域（一部）。

平行進化サイト数の期待値と実測値の比較

系統ネットワークのループ構造が水平伝播によってもたらされたものなのか、平行進化によってもたらされたものなのかを検定するためには、水平伝播が生ずる確率と平行進化が生ずる確率の比較が必要である。しかしながら、現状で水平伝播が生ずる確率は知られていない。しかし、もし遺伝的組換えが生じている場合、遺伝的組換えが生じていない場合に比べて見かけの平行進化サイト数は増大するはずである。そこで、本研究では、4配列間で見られる見かけの平行進化サイト数と、系統樹から計算した平行進化数の期待値との間の比較を行った。

平行進化サイト数の期待値の計算方法

平行進化サイト数の期待値の計算方法を、図 4.14 の系統樹に沿って説明する。各枝のサイトあたりの平均の塩基置換数(p 距離)を b_1 - b_4 で示した。平行進化の検定に関係のない枝の値は省略した。平行進化とは、異なる系統で同じ方向の塩基置換が生ずることを言う。例えば、図 4.14 に示したように、Sequence 1 と Sequence 3 の系統において、たまたま同じサイトに同じ A→T という塩基置換が生ずるような場合を言う。したがって、Sequence 1 と Sequence 3 に生じた平行進化数の期待値 $r'_{1,3}$ は、多重進化のうちの平行進化の割合を a 、配列の長さを l と考えると、 b_1 と b_3 を用いて

$$r'_{1,3} = lab_1b_3 \cdot \cdot \cdot \cdot \cdot 4.1 \text{ 式}$$

で求めることができる。 a の値は多重進化のうち平行進化である割合を表すが、この値は、塩基置換モデルによって決定される値であるため、定数と考えることができる。したがって、 r' と $b_1 \times b_3$ は線形の関係となる。平行進化は、全ての 2 配列の組み合わせに生じうるが、このうち系統ネットワークに網目構造をもたらすような 2 配列の組み合わせは、Sequence 1 と 3、Sequence 1 と 4、Sequence 2 と 3、Sequence 2 と 4、の組み合わせである。なぜならば、例えば、Sequence 1 と 2 の系統で平行進化が生じたとしても、そのサイトが示唆する系統関係は図 4.14 の系統関係と同じだからである。したがって、4.1 式より、全ての観察可能な平行進化数の期待値 r' を図 4.14 の系統樹にそって計算すると

$$\begin{aligned} r' &= la(b_1b_3 + b_1b_4 + b_2b_3 + b_2b_4) \\ &= la(b_1 + b_2)(b_3 + b_4) \cdot \cdot \cdot \cdot \cdot 4.2 \text{ 式} \end{aligned}$$

となる。

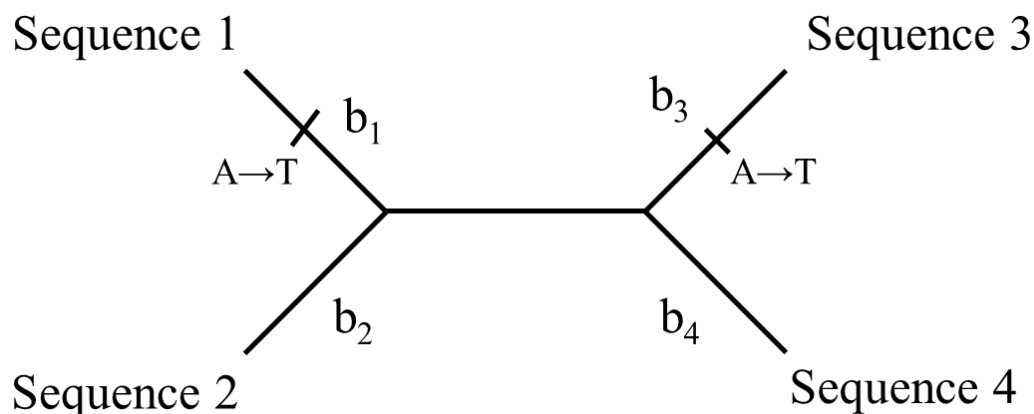


図 4.14 4 配列から成る系統樹。b₁-b₄ は各枝のサイトあたりの塩基置換数 (p 距離) を表す。平行進化数の見積もりに関係のない枝の値は省略した。平行進化の例として、Sequence 1 と Sequence 3 の系統で独立して同じサイトに生じた同じ方向の塩基置換 (A→T) を示した。

実際の r' の計算

図 4.15 は、Catarrhini 下目のミトコンドリア 12S rRNA 遺伝子と Enterobacteriales 目の 16S rRNA 遺伝子について、全ての 4 配列の組み合わせについて 4.2 式を用いて計算した r' と r の相関図である。Catarrhini 下目とは、狭鼻下目と呼ばれる分類群で、ヒトやニホンザルを含む旧世界サルと呼ばれる分類群である。r' の値の範囲が Enterobacteriales 目と比較しやすかったために用いた。Enterobacteriales 目において、本研究で用いた水平伝播判定の基準に従って、水平伝播が検出された 4 配列のペアから計算した r/r' については緑色で、水平伝播が検出されなかったペアについては青色で示した。Catarrhini 下目の r/r' の分布は赤色で示した。計算の過程で用いた系統樹は近隣結合法を用いて作成し、距離行列は p 距離を用いて計算した。

図 4.15 から、12S rRNA 遺伝子 (図中赤丸) が比較的線形的に分布しているのに比べて 16S rRNA 遺伝子 (図中青、および緑丸) の分布は線形的ではないことが分かる。12S rRNA 遺伝子は、16S rRNA 遺伝子と相同な遺伝子で、Alphaproteobacteria 綱の微生物がオルガネラ化することに起源をもつと考えられている。水平伝播の影響を受けない遺伝子であるため、みかけの平行置進化サイト数は全て平行進化によるものである。一方で、Enterobacteriales 目において水平伝播の存在が示唆された 4 配列の分布は 12S rRNA 遺伝子の分布よりも高めに分布しており、Enterobacteriales 目内の分布においても、比較的高い値の位置に分布している。このことから、みかけの平行置進化サイトは水平伝播によって多めにもたらされているということが分かった。しかし、水平伝播が生じていると判定が出なかった 4 ペアにおいても高い r/r' の値を示すものが見られた。これらのペアは、水平伝播以外の要因によっ

て平行進化が多めにもたらされているか、今回本研究で用いた水平伝播の検出基準によって検出できなかったペアであると考えられる。

図 4.16 は、高い r/r' 値を示すにも関わらず遺伝的組換えの判定が出なかった 2 組の例である。どちらの組み合わせにおいても、遺伝的組換えによってもたらされたと考えることができる多型パターンが見られたが、本研究で用いた遺伝的組換えの判定基準は満たしていない。したがって、本研究で用いた遺伝的組換えの判定基準は、精査の余地がある。

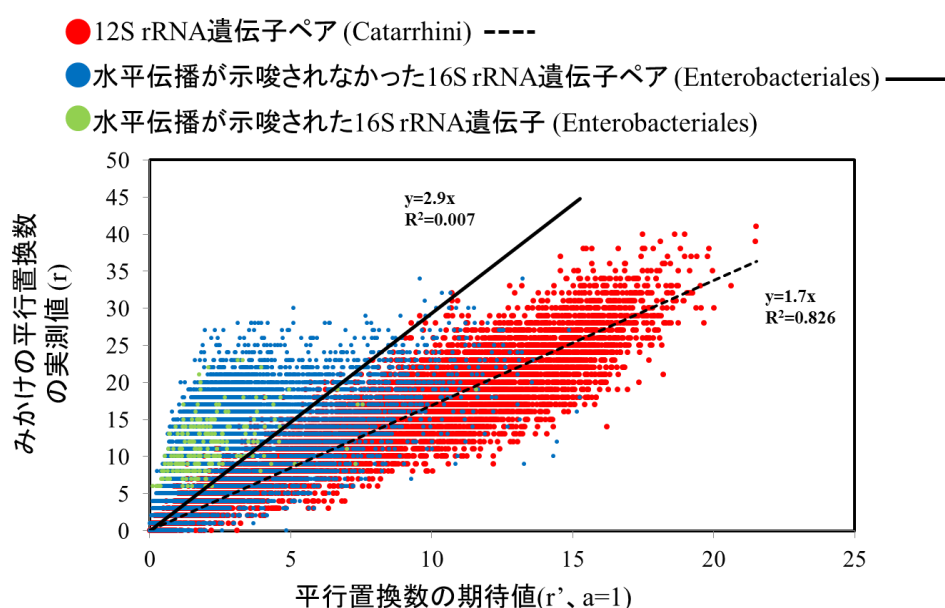


図 4.15 Catarrhini 下目 (赤)、水平伝播の判定が出なかった Enterobacteriales 目 (青)、水平伝播の判定が出た Enterobacteriales 目 (緑) の、平行進化数の期待値 (r') と実際に観察されたみかけの平行進化数 (r) の相関。 r' の計算には 2 式を用い、 $a=1$ とした。 Catarrhini 下目 (赤) と、水平伝播の判定が出なかった Enterobacteriales 目 (青) については、それぞれ破線と実線で近似直線を示した。 R^2 は決定係数を意味する。

第五章 結論と展望

結論

1. 過去の 16S rRNA 遺伝子の水平伝播の報告と本研究の成果の総括

a) 16S rRNA 遺伝子の水平伝播は普遍的な現象である

表 1.1 にまとめた、自然界における 16S rRNA 遺伝子の部分的な水平伝播の報告同様に、本研究においてもバクテリアにおいて 16S rRNA 遺伝子の部分的な水平伝播が存在することが示された。本研究により、自然界においてバクテリアの 16S rRNA 遺伝子の部分的な水平伝播が存在する証左がより確実に得られた。また、本研究においてバクテリアの門の過半数において普遍的に 16S rRNA 遺伝子の水平伝播が見られたことは、これまでの 16S rRNA 遺伝子の部分的な水平伝播の報告が、単なるレアケースを検出しているに過ぎないわけではないことを意味する。

b) ゲノム内組換えと水平伝播による進化モデル

Yap ら(1999)は 10%以上も配列が異なる同目異亜目のリボソーマル RNA オペロン丸ごとの水平伝播を報告している。しかし、Yap ら(1999)の報告における異種由来のリボソーマル RNA オペロンは 6 コピー中 1 コピーであるため、機能的なリボソームを構成できているかは定かではない。これまでに知られている機能的なリボソームを構成していると思われる大規模な水平伝播の例は、Lan と Reeves(1998)による *Vibrio cholerae* 内の極めて近縁な同種間におけるもののみである。Lan と Reeves(1998)や Yap ら(1999)の報告から考えられることは、少なくともリボソーマル RNA オペロン単位での大規模な水平伝播が、生き残るかどうかはともかく、現象としては生じうるということである。

本研究においても、第二章において、16S rRNA 遺伝子のコピー間に差異が存在することは普遍的なことであることが示された。また、*Citrobacter* の 7 つ全ての 16S rRNA 遺伝子が *Escherichia* と *Salmonella* 間の遺伝的組換えによって生じたものであったように、ゲノム内組換えが新たな遺伝型を生み出し、生み出された遺伝型は駆逐されずにコピー数を増やしているような例が存在することも示された。このことは、遺伝的組換えが、ある座を乗取った異種由来の 16S rRNA 遺伝子と、在来の 16S rRNA 遺伝子との間で生じることによって、異種由来の 16S rRNA 遺伝子の痕跡が断片的に残るようなことがあり得ることを意味する(図 6.1)。つまり、水平伝播が生じた当初の段階では Yap ら(1999)の報告にあるように大規模なゲノム内多型が存在していたとしても、その後を生ずるゲノム内組換えによる均一化の過程において、異種由来の 16S rRNA 遺伝子の機能相補的な領域は痕跡として残りうるのではなかろうか。本研究で用いたデータセットにおいては、プラスミド由来の 16S rRNA 遺伝子は見つからなかったため、この水平伝播は、現状は自然形質転換によるものと考えられる。このように考えると、第五章において検出された、水平伝播単位として考えるには

あまりにも短く見える数十塩基の領域は、水平伝播後のゲノム内組換えを免れた領域であると考えられることもできる。

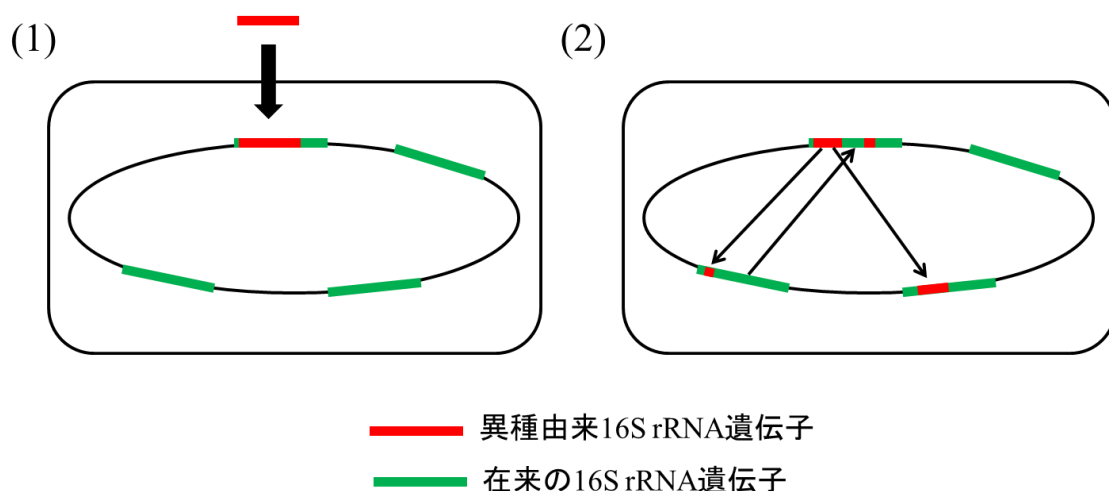


図 6.1 16S rRNA 遺伝子の水平伝播モデル。(1) 異種由来の 16S rRNA 遺伝子が水平伝播によって複数座ある 16S rRNA 遺伝子のうち、一つを乗っ取る。(2) 16S rRNA 遺伝子間の遺伝的組換えによって、断片的に在来 16S rRNA 遺伝子に取り込まれ異種由来の 16S rRNA 遺伝子と在来 16S rRNA 遺伝子のハイブリッドが形成される。

c) 16S rRNA 遺伝子の水平伝播と他の遺伝子の水平伝播の違い

このような、水平伝播-ゲノム内組換えによる変異-安定化のプロセスがあると考え、16S rRNA 遺伝子の水平伝播は、他の遺伝子よりも、断片的にはかなり受け入れられる余地はあるように思われる。微生物において、16S rRNA 遺伝子のような重複遺伝子の例は極めて稀で、ほとんどの遺伝子は単一のコピーしかゲノム上に存在せず、したがって、他の多くの遺伝子は 16S rRNA 遺伝子のようにゲノム内組換えによって機能相補的な領域のみを残すようなことはできないと考えられるからだ。すなわち、16S rRNA 遺伝子は、二次構造の保守が機能相補の重要な要件であるため一次配列の重要性は低く、ゲノム内組換えによる断片的な取り込みが可能であるという特徴を持っているため、複雑な相互作用の中におかれている遺伝子ではあっても、水平伝播の痕跡が非常に残りやすい遺伝子であると考えられることができる。

d) 水平伝播によって進化しやすい 16S rRNA 遺伝子

第三章では、Enterobacteriales 目の系統ネットワーク解析の結果から、他の情報系遺伝子 Elongation factor G と比べ 16S rRNA 遺伝子は、点突然変異よりは、むしろ水平伝播によってより進化しているという結果が示唆された。また、例えば、図 3.3(B)や図 3.11 の系統ネットワークを構成する配列のほとんどが、他の配列同士のキメラで説明できる。これらのことは、点突然変異よりも水平伝播の方がより 16S rRNA 遺伝子に多様度をもたらしている

ということを意味する。これらの知見は、二次構造が大きく変化しないな変異ならば 20% 程度まで 16S rRNA 遺伝子は受け入れるという Kitahara ら(2012)の結果や、点突然変異に対して 16S rRNA 遺伝子が脆弱であるという Yassin ら(2005)の報告とも合致する結果である。

e) Complexity Hypothesis を再解釈する

情報系遺伝子に属する 16S rRNA 遺伝子は、水平伝播を受け入れない遺伝子であると考えられており、系統群マーカー遺伝子として扱われてきた (Complexity Hypothesis, Jain et al., 1999)。しかし、Kitahara ら(2012)によって、16S rRNA 遺伝子の機能にとって配列そのものと言うよりは二次構造が重要であることが示された段階で、二次構造が大きく変化しなければ、30S サブユニットの中心骨格として機能するということが示されている。このことは、16S rRNA 遺伝子に生ずる突然変異に対して二次構造を保つような選択圧がかかるであろうことを意味しているので、二次構造に大きな変化がないのであれば、16S rRNA 遺伝子が水平伝播を受け入れたとしても不思議なことではない。確かに、Rivera ら(1998)の結果において、情報系遺伝子は実際に近縁な系統群に対してより近い距離を示したため、操作系遺伝子に比べて水平伝播の影響は受けていないと考えられる。しかし、これらのほとんどはアミノ酸配列であって、実際に一次元の配列の中身が機能や相互作用に影響を与えるものばかりであった。これに対して、16S rRNA は配列そのものではなく、二次構造が機能的に重要である。このように考えると、本研究の成果や Kitahara ら(2012)の結果によって 16S rRNA は水平伝播によって相互作用を失わないことが分かっているのだから、Complexity Hypothesis を根拠として 16S rRNA 遺伝子が水平伝播しないと考えることはできない。

2. 16S rRNA 遺伝子の水平伝播について新たに分かったこと

a) 16S rRNA 遺伝子の水平伝播は領域を問わずに生じている

本研究によって得られた新たな知見は、16S rRNA 遺伝子の断片的な水平伝播は領域を問わずに生じているという点である。表 1.1 に上げた、これまでの 16S rRNA 遺伝子の水平伝播領域の同定は、いずれも目視によるものである。水平伝播の痕跡を発見するためには、少なくとも組換え体と親配列 2 つの 3 配列がデータセット中に存在しなければならない上に、目視で親-組換え体の関係を持つ配列を同定可能なデータセットの規模はおのずと限られてくるし、非常に狭い領域で生じている水平伝播しか見つけることはできない。これらに対して、本研究はデータセットに含まれる配列全ての 4 配列ごとの組み合わせについてプログラムを作成し検証を行ったため、目視では検出不可能だった大規模な領域に渡る 16S rRNA 遺伝子の水平伝播を多く発掘することができた。16S rRNA 遺伝子の断片的な水平伝播が領域を問わずに生じているということは、16S rRNA 遺伝子に種固有の領域は存在しないことを意味する。このことから、16S rRNA 遺伝子は、これまで考えられてきたような種固有の遺伝子ではなく、少なくとも領域的な互換性は存在する遺伝子であると考えられる。

このように考えると、むしろ 16S rRNA 遺伝子は系統解析には不向きな遺伝子であるということが分かる。

b) 分類群ごとに水平伝播が高頻度に生じる領域が異なる

また、分類群ごとに 16S rRNA 遺伝子の水平伝播が生じやすい領域とそうでない領域が存在することが分かった。系統群ごとの特性として、16S rRNA 遺伝子に対して高い互換性を許容する領域とそうでない領域が存在するということであろう。そのような要因としては、16S rRNA は 30S サブユニットの中心骨格を形成する分子であるということを考えると、リボソームの構造上の特性である可能性が考えられる。我々の研究室においては、*Escherichia coli* の 16S rRNA 遺伝子を異種由来の 16S rRNA 遺伝子と置換することによって、翻訳の傾向が変化することが確認されている。このような点を加味すると、16S rRNA 遺伝子の互換によって、翻訳系は何らかの影響を受けうると考えられるため、系統群ごとに水平伝播を受け入れない 16S rRNA 遺伝子の領域が存在するものと考えられる。

3. 16S rRNA 遺伝子の水平伝播は何かをもたらしているのか？

先行研究と本研究によって、水平伝播-ゲノム内組換えの変異-安定化プロセスが 16S rRNA 遺伝子に高い互換性をもたらしている可能性が示唆された。このことは、いったい何を意味するのか？ 基本的には、16S rRNA 遺伝子同士のゲノム内組換えは、配列を均一化させることが目的であると考えられる。すなわち、多様度を創成する仕組みというよりは、多様度を低減する仕組みとして機能することが圧倒的に多い。16S rRNA 遺伝子の領域的な水平伝播の痕跡は、機能的に有利な点があったからゲノム内組換えを免れているのか、それとも、機能に影響を与えない中立領域で、ある一定の確率で生じる現象が単に観察されているだけなのか。第二章において明らかとなった通り、*Escherichia* と *Shigella* において 16S rRNA 遺伝子の複数の系統が同一のゲノムに高い割合で保持されている場合が存在する。従って、ゲノム内の 16S rRNA 遺伝子同士に多型が存在することで何らかの利益がもたらされている場合がありうることも可能性として十分に考えられる。しかし、ゲノム内組換えで生じた 4 つのグループを現在でも保有している株は 64 株中 2 株のみであった。このことは、ゲノム内組換えが生じた祖先の段階では 4 つのグループが存在するのだが、その後、各株においていずれかのグループが失われていくような Birth and Death モデルで説明することができる。すなわち、確かにゲノム内の 16S rRNA 遺伝子同士の多型は多くの株で保たれているのだが、方向性としてはあくまで均一化に向かっているということである。このような中において、例えば、第二章でふれたように、*Escherichia* と *Shigella* 集団内に存在する helix33 領域の多型が *Escherichia* と *Enterobacter* や *Klebsiella* の祖先の段階から保たれているなど、積極的に集団に保たれている多型も存在し、16S rRNA 遺伝子の部分的な水平伝播が機能的に有利な点をもたらしている可能性が示唆される。このように、16S rRNA 遺伝子の水平伝播がもたらすものについては不明であるが、多型や水平伝播の痕跡が積極的

に保たれているのか、それともある一定の確率で偶然残っているだけなのかを検証するための統計的な解析や、実験による機能的な変化の同定などを行っていけば明らかとなってゆくと思われる。

4. 16S rRNA 遺伝子を用いた種の同定について

本研究において、16S rRNA 遺伝子の水平伝播は、バクテリア全体で広く見られる現象で、機能的な 16S rRNA 遺伝子を形成しうることが示された。バクテリアにおいて、16S rRNA 遺伝子の水平伝播は、綱レベルの分類群を超えて生ずることが分かった。ただし、本論文の中では門や界を超えた水平伝播が生じるか否かについては解析していないため、より高次の系統間での水平伝播の可能性は現時点では不明である。これらのことは、**16S rRNA 遺伝子で作成した系統樹は、16S rRNA 遺伝子の系統樹に過ぎず、種の系統関係を反映するとは言えない**ということの意味している。特に、16S rRNA 遺伝子のごく一部の領域を読むことで、菌叢を解析するようなやり方は、16S rRNA 遺伝子は領域を問わず水平伝播するので、間違った種を同定しやすいということが分かった。例えば、イルミナ社のウェブサイトでも菌叢解析の例として紹介されている、Caporaso ら(2014)の方法では、16S rRNA 遺伝子の V4 領域 (16S rRNA 遺伝子において、9 つある多様な領域の 4 番目)を、ユニバーサルプライマー(515F、806R)を用いて増幅・解析している。しかしながら、図 4.6 – 図 4.10 において示す 515F、806R プライマーが増幅する領域は、いずれの分類群においても遺伝的組換えが生じている領域で、特に Proteobacteria 門や Actinobacteria 門では高頻度に遺伝的組換えが生じている。しかし、16S(12S、18S) rRNA 遺伝子は全ての生物が共通して持っている遺伝子であり、ユニバーサルプライマーの設計も比較的容易であるため、系統解析において便利な遺伝子であることは間違いない。本研究の結果から言えることは、16S rRNA 遺伝子を用いて系統解析や菌叢解析を行うような場合は、分類群ごとに水平伝播が起こりにくい領域を選択すべきであるということである。

展望

1. 16S rRNA 遺伝子の領域的な水平伝播がもたらす機能的変化の解明

今後は、16S rRNA 遺伝子の水平伝播の生物学的な意義の解明に興味もたれる。16S rRNA は翻訳の中心的な担い手であり、その遺伝子の変化は細胞全体に影響を与えるはずである。16S rRNA 遺伝子の水平伝播は、個体レベルでの機能を変えるという点で、自然環境中でも適応度に変化を与えると予想され、微生物の進化において極めて重要な変化をもたらしているかもしれない。我々のグループでは、16S rRNA 遺伝子に水平伝播が生じた際のゲノムの進化に着目した研究がすでに走り出している。

また、16S rRNA 遺伝子の翻訳以外の機能についても、明らかになりたい。これまでには、Kitahara と Miyazaki(2011)によって 16S rRNA の helix 41 領域が RNase T2 の阻害活性を有す

ることが知られている。この機能は、rRNAがRNaseに分解されないための防衛機構であり、16S rRNA 遺伝子の水平伝播を阻む機構として提案されている。このように、リボソーム以外の要素との相互作用が16S rRNAには領域的にもっと埋まっているかもしれない。これらのような知見が蓄積すれば、16S rRNA 遺伝子において多様度が高く、一見中立的な進化をしているように思われる領域に生じている水平伝播の意義が解明できるのではなかろうか。

2. 遺伝的組換えを検出する基準の精査

系統ネットワーク法は、遺伝的組換えを含む進化史をループ構造で表現するが、系統ネットワークにループ構造が見られたからと言って遺伝的組換えが生じているとは直ちに判断できない。例えば、節約的な系統ネットワーク法である Bandelt(1995)の手法ならば、異なる系統で同じ方向の塩基置換が生ずる平行進化や、異なる塩基がたまたま同じ塩基へと異なる系統で変化するような収斂進化が生じている場合もループ構造を示すし、Neighbor-Net法などの距離行列から作成される系統ネットワーク法においては、平行進化や収斂進化のほかに同じサイトに複数回の塩基置換が生ずるような多重置換によってもループ構造が形成されうる。したがって、系統ネットワークにループ構造が生じた時、それが遺伝的組換えによるものなのか、平行進化や収斂進化によるものなのかを判断するための確率論的な比較が必要である。現状では、水平伝播が生ずる確率は未知であるため、このような確率の議論は不可能である。打開策として、第四章において、平行進化によって系統ネットワークに構成されうるループ構造の大きさの期待値と、実測値の比較を行った。その結果、平行進化とは考え難い水平伝播が生じていると考えられる配列ペアを、今回用いた水平伝播の判定基準ではとり逃してしまっていることが示唆された。また、系統ネットワークのループ構造は平行進化以外にも、収斂進化によって構成されうるし、3塩基以上の多型を示すサイトは計算結果に含まれていない。実際に、遠い距離を比較すればするほど、収斂進化や3塩基以上の多型を示す多重置換サイトは増えてゆくと、何らかの補正や計算のやり方の改編を将来的には行ってゆかなければならない。

3. バクテリアにおける水平伝播の親子関係

本研究はこれまで、系統ネットワーク法を応用することにより遺伝的組換え、および水平伝播の検出を行ってきた。系統ネットワーク法において（あるいは、他の組換え検出アルゴリズムにおいても）、直接の親となったハプロタイプを推定できるわけではなく、あくまで親の系統に属するハプロタイプであることが分かるだけであることには留意せねばならない。すなわち、遠い種間での水平伝播が生物学的にあり得ないような生物（例えば、ヒト）においては、直接の親となったハプロタイプについて、少なくともその種程度ならば推定が可能であるが、遠縁な種からも水平伝播の影響を受けうるバクテリアにおいてはこの限りではない。本研究の第4章で示されたような高頻度に生じている16S rRNA 遺伝子の部分的な水平伝播は、あくまで、異なる種同士が、本来の系統関係を越えてある16S rRNA

遺伝子の領域を‘共有’しているに過ぎないのである。従って、本研究の結果から水平伝播頻度と生息域の間に関連があるのかどうかを精査することは極めて困難である。これらの問題を解決するためには、メタゲノムデータ由来のデータのような、生息域ごとにソートされたデータセット内・間の比較を行う必要がある。

第五章 引用文献

- Caporaso et al., (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6(8): 1621–1624.
- Hugenholtz, P., Pitulle, C., Hershberger, K., and Pace, N. R. (1998). Novel division level bacterial diversity in a Yellowstone hot spring. *J. Bacteriol.* 180(2): 366–376.
- Jain, R., Rivera, M. C., and Lake, J. A. (1999). Horizontal gene transfer among genomes: The complexity Hypothesis. *Proc. Natl. Acad. Sci. USA* 96(7):3801–3806.
- Kitahara, K. and Miyazaki, K. (2011). Specific inhibition of bacterial RNase T2 by helix 41 of 16S ribosomal RNA. *Nat. Commun.* 2(2): 549.
- Kitahara, K., Yasutake, Y., and Miyazaki, K. (2012). Mutational robustness of 16S ribosomal RNA, shown by experimental horizontal gene transfer in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 109(47):19220–19225.
- Yassin, A., Fredrick, K., and Mankin, A. S. (2005). Deleterious mutations in small subunit ribosomal RNA identify functional sites and potential targets for antibiotics. *Proc. Natl. Acad. Sci. USA* 102(46):16620–16625.
- Sawyer, S. (1989). Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* 6(5): 526–538.
- Woese, C., Kandler, O., and Wheelis, M. (1990). Towards a natural system of organisms: proposal for the domains archaea, bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* 87(12): 4576–4579.
- Woese, C. R. (1987). Bacteria l evolution. *Microbiol. Rev.* 51(2):221–271.
- Yassin, A., Fredrick, K., and Mankin, A. S. (2005). Deleterious mutations in small subunit ribosomal RNA identify functional sites and potential targets for antibiotics. *Proc. Natl. Acad. Sci. USA* 102(46):16620–16625.

謝辞

本研究、および学位論文の執筆に際し、多くの方々からご支援をいただきましたのでこちらでお礼を申し上げたく思います。

指導教官である、東京大学大学院新領域創成科学研究科客員教授 宮崎健太郎先生に、3年間ご指導ご鞭撻を頂戴いたしましたことを深く感謝しております。本研究、および、学位論文の執筆につきましても、たくさんの指導を頂いたことを大変感謝しております。

東京大学大学院新領域創成科学研究科教授で、産業技術総合研究所生命情報工学研究センター長を兼務されている浅井潔先生には、産業技術総合研究所ゲノム情報研究センター(CBRC)におきまして学習の機会を頂戴いたしました。また、中間審査、予備審査の折にご助言を頂きましたこと、心よりお礼申し上げます。

産業技術総合研究所ゲノム情報研究センター副研究センター長、藤博幸先生（現：関西学院大学理工学部生命医化学科教授）には、浅井潔先生より頂戴いたしました、産業技術総合研究所ゲノム情報研究センター(CBRC)における学習の機会におきまして、様々のご助言を頂きました。心よりお礼申し上げます。

東京大学大学院新領域創成科学研究科教授 上田卓也先生には、中間審査、予備審査においてご指導・ご助言を頂きましたことに心よりお礼申し上げます。

東京大学大学院新領域創成科学研究科教授 伊藤耕一先生には、予備審査におきまして、ご助言を頂きましたことに、心よりお礼申し上げます。

茨城大学工学部生体分子機能工学科准教授 北野誉先生には、予備審査におきまして、ご助言を頂きました。また、系統ネットワーク解析についてご指導、ご助言を頂きましたことを心よりお礼申し上げます。

国立遺伝学研究所集団遺伝研究部門教授で、総合研究大学院大学生命科学研究科遺伝学専攻教授、東京大学大学院理学系研究科生物科学専攻教授を兼任されている、斎藤成也先生には、遺伝的組換えの有意性を評価する手法について様々にご助言を頂きました。心よりお礼申し上げます。