

Title	高精度な複製終結点予測アルゴリズムによるバクテリアゲノム対称性の解析
Sub Title	Comparative study of compositional symmetry in bacterial genomes using a novel prediction algorithm for replication terminus
Author	河野, 暢明(Kono, Nobuaki) 富田, 勝(Tomita, Masaru)
Publisher	慶應義塾大学湘南藤沢学会
Publication year	2010-03
Jtitle	優秀修士論文
JaLC DOI	
Abstract	<p>多くのバクテリアは真核生物とは違い環状の染色体を持っており,その環状ゲノム内には様々な対称性が維持されている.例えば,リーディング鎖とラギング鎖における遺伝子の偏りや,オリゴ配列の方向性,さらには塩基組成バイアスなどが挙げられ,様々な種類の対称性を観察する事ができる.その中でも最も代表的な例の1つが複製開始・終結点の対称性である.環状ゲノムを持つバクテリアは複製開始・終結点をそれぞれ一対だけ持っており,ほぼすべての生物でこれらは対称に位置している.現在複製開始・終結点の決定はin silicoによる予測がde factoスタンダードとなっており,これらが実験的に同定されている生物種は少ない.この予測にはリーディング鎖とラギング鎖の複製機構の違いから生じる,グアニンとシトシンの塩基組成バイアスが用いられている.しかしながら,複製の産物として得られた塩基組成バイアスが対称的に維持されている謎はこれまで明確に解明されてきていない.さらに,この塩基組成バイアスは全ての種で正確に保存されているわけではなく,いくつかの種では極端に対称性が崩れていることが確認されており,その揺らぎが何によってもたらされているかも明らかではない.一方で,近年塩基組成バイアスに加えて複製終結点を特定するための新たな因子として4dif配列が注目を集め始めた.4ヴ配列とは28塩基の配列であり,複製終結時に二量体DNAを2つの娘DNAに分割するチロシンリコンビナーゼの標的配列となっている.そのため,この4dif配列こそが生物学的な根拠に基づいた複製終結点であると考えられるようになってきた.ところが4dif配列は未だすべての種で同定されておらず数種で解明されている程度である.</p> <p>そこで,本研究では塩基組成バイアスの対称性の揺らぎの原因,及び,その塩基組成バイアスと複製プロセスの関係を解明すべく,全バクテリアゲノムを対象とした網羅的ゲノム解析を行なった.まず第2章にて塩基組成バイアスの対称性の揺らぎの原因を探るべく,水平伝播と塩基組成の関係を解析した.その結果,水平伝播と対称性のゆらぎに強い相関関係はないという結果が得られ,さらに挿入された領域の長さや位置,遺伝子の数などにも対称性のゆらぎは依存していない事が解明された.そのため,この対称性のゆらぎは何か別の因子によって引き起こされている可能性を示唆することができた.次に第3章では塩基組成バイアスと複製終結点の関係を解明すべく,dif配列予測を網羅的にを行い,その得られたdif配列の位置と塩基組成の関係を解析した.このためには再帰的隠れマルコフモデリングという手法を開発し,前提解析として網羅的にdif配列の予測を行った.その上で,予測によって得られたdif配列と塩基組成バイアスの示す終結点との位置関係を比較解析した.その結果641種・714染色体でのdif配列を予測する事に成功し,dif配列と塩基組成バイアスはほぼ同じ場所に位置している事が判明した.塩基組成の示す終結点とは,複製が行われるにつれて蓄積されてきた塩基バイアスの痕跡と言えるため,バクテリアの複製は恒常的にdif配列近傍で複製を終結してきており,環状ゲノムに維持されている塩基組成バイアスのゆらぎはその結果であるという事が示唆された.</p>
Notes	先端生命科学プロジェクト2009年度
Genre	Thesis or Dissertation
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=0302-0000-0637

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the Keio Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

優秀修士論文

高

精度な複製終結点予測アルゴリズムによる
バクテリアゲノム対称性の解析

2009年度

河野 暢明 政策・メディア研究科 修士課程

先端生命科学プロジェクト

慶應義塾大学湘南藤沢学会

優秀修士論文推薦のことば

本論文ではバクテリアの複製終結点を予測するソフトウェアを開発し、それを用いて数多くのバクテリアゲノムを対象に、網羅的な複製終結点の予測を世界に先駆けて行った。これまで謎が多かったゲノム対称性に対する新たな考察と議論の展開は、今後の分子生物学のみならず、合成生物学やバイオテクノロジーへの貢献などに重要な意義があると考えられる。

そのため、本論文を優秀修士論文として推薦します。

慶應義塾大学
環境情報学部教授
富田 勝

修士論文 2009 年度（平成21年度）

高精度な複製終結点予測アルゴリズムによる バクテリアゲノム対称性の解析

慶應義塾大学大学院 政策・メディア研究科

河野 暢明

修士論文 2009 年度（平成21年度）

[題目]

高精度な複製終結点予測アルゴリズムによるバクテリアゲノム対称性の解析

[論文要旨]

多くのバクテリアは真核生物とは違い環状の染色体を持っており、その環状ゲノム内には様々な対称性が維持されている。例えば、リーディング鎖とラギング鎖における遺伝子の偏りや、オリゴ配列の方向性、さらには塩基組成バイアスなどが挙げられ、様々な種類の対称性を観察する事ができる。その中でも最も代表的な例の1つが複製開始・終結点の対称性である。環状ゲノムを持つバクテリアは複製開始・終結点をそれぞれ一対だけ持っており、ほぼすべての生物でこれらは対称に位置している。現在複製開始・終結点の決定は*in silico*による予測が*de facto*スタンダードとなっており、これらが実験的に同定されている生物種は少ない。この予測にはリーディング鎖とラギング鎖の複製機構の違いから生じる、グアニンとシトシンの塩基組成バイアスが用いられている。しかしながら、複製の産物として得られた塩基組成バイアスが対称的に維持されている謎はこれまで明確に解明されてきていない。さらに、この塩基組成バイアスは全ての種で正確に保存されているわけではなく、いくつかの種では極端に対称性が崩れていることが確認されており、その揺らぎが何によってもたらされているかも明らかではない。一方で、近年塩基組成バイアスに加えて複製終結点を特定するための新たな因子として*dif*配列が注目を集め始めた。*dif*配列とは28塩基の配列であり、複製終結時に二量体DNAを2つの娘DNAに分割するチロシンリコンビナーゼの標的配列となっている。そのため、この*dif*配列こそが生物学的な根拠に基づいた複製終結点であると考えられるようになってきた。ところが*dif*配列は未だすべての種で同定されておらず、数種で解明されている程度である。

そこで、本研究では塩基組成バイアスの対称性の揺らぎの原因、及び、その塩基組成バイアスと複製プロセスの関係を解明すべく、全バクテリアゲノムを対象とした網羅的ゲノム解析を行なった。まず第2章にて塩基組成バイアスの対称性の揺らぎの原因を探るべく、水平伝播と塩基組成の関係を解析した。その結果、水平伝播と対称性のゆらぎに強い相関関係はないという結果が得られ、さらに挿入された領域の長さや位置、遺伝子の数などにも対称性のゆらぎは依存していない事が解明された。そのため、この対称性のゆらぎは何か別の因子によって引き起こされている可能性を示唆することができた。次に第3章では塩基組成バイアスと複製終結点の関係を解明すべく、*dif*配列予測を網羅的に行い、その得られた*dif*配列の位置と塩基組成の関係を解析した。このためには再帰的隠れマルコフモデリングという手法を開発し、前提解析として網羅的に*dif*配列の予測を行った。その上で、予測によって得られた*dif*配列と塩基組成バイアスの示す終結点との位置関係を比較解析した。その結果641種・714染色体での*dif*配列を予測する事に成功し、*dif*配列と塩基組成バイアスはほぼ同じ場所に位置している事が判明した。塩基組成の示す終結点とは、複製が行われるにつれて蓄積されてきた塩基バイアスの痕跡と言えるため、バクテリアの複製は恒常的に*dif*配列近傍で複製を終結してきており、環状ゲノムに維持されている塩基組成バイアスのゆらぎはその結果であるという事が示唆された。

[キーワード]

環状染色体、ゲノム対称性、水平伝播、*dif*配列、隠れマルコフモデル

慶應義塾大学大学院 政策・メディア研究科

河野 暢明

Abstract of Master's Thesis Academic Year 2009

[Title]

Comparative study of compositional symmetry in bacterial genomes using a novel prediction algorithm for replication terminus.

[Summary]

Many bacteria have circular chromosomes unlike eukaryotes, exhibiting various symmetries and polarities including the strand preference of genes between the leading and lagging strands, oligonucleotide orientation, and base compositional bias. These symmetries are all defined by a symmetrically located pair of finite replication origin and terminus. However, experimental evidence of replication origin and terminus are still limited, and majority of the discussions currently rely on *in silico* predictions using compositional bias of guanine and cytosine formed by the difference in the replication mechanisms between the leading and lagging strands. Such prediction sometimes results in highly asymmetric pairs of replication origin and terminus, and thus a comprehensive study for the cause of compositional symmetry in bacteria is still lacking. On the other hand, a conserved 28bp sequence element targeted by a tyrosine recombinase upon the resolution of malformed chromosome dimers during cell division, named the *dif* sequence, is recently suggested as a new marker of the replication terminus. The *dif* sequences, however, are identified only in a limited number of organisms.

In light of these backgrounds, this thesis discusses the selectional and mutational forces of replication machinery driving the order and disorder of compositional symmetry in bacterial chromosomes, through comprehensive genome analyses. Firstly, the effects of genomic insertions through horizontal gene transfer in disrupting the symmetry of base compositional bias is studied in chapter 2. As a result for this chapter, genomic compositional symmetry is shown to be highly robust to horizontal gene transfer events, regardless of the inserted segment lengths, positions and the number of genes. Secondly, in chapter 3, accurate and comprehensive prediction of *dif* sequences is discussed, which is further utilized to study the order between the symmetry in base composition bias and the pair of replication origin and terminus. A novel algorithm named the Recursive Hidden Markov Modeling is discussed for this purpose. Using this novel method, *dif* sequences are identified in 714 chromosomes harbored by 641 organisms, and their positions are shown to be strongly correlated with the skew shift point of base compositional bias. Since the replication terminus represented by the skew shift point of base compositional bias is presumably the result of cumulative mutational bias in replication, replication in bacteria is suggested to be coordinately terminated around the *dif* sequence for effective cell division.

[Keywords]

Circular chromosome, Genome polarity, Horizontal gene transfer, *dif* sequence, Hidden Markov modeling

Keio University Graduate School of Media and Governance
Nobuaki Kono

目次

第1章 序論	4
1.1 複製	4
1.2 バクテリアの複製と対称性	5
1.3 染色体分裂	8
1.4 本研究のねらい	10
第2章 水平伝播による対称性秩序への影響	11
2.1 背景	11
2.2 対象と手法	13
2.3 結果	16
2.3.1 GIsのシフトポイントへの影響	16
2.3.2 GIs長及び位置との関係	16
2.3.3 GIs内遺伝子数との関係	17
2.4 考察	20
第3章 維持されている複製終結点	21
3.1 背景	21
3.2 対象と手法	22
3.2.1 <i>dif</i> 配列予測	22
3.2.2 <i>dif</i> 配列とGC skewの比較解析	26
3.3 結果	28
3.3.1 <i>dif</i> 配列の予測結果概要	28
3.3.2 門毎の予測結果	29
3.3.3 <i>dif</i> 配列とGC skewの位置関係	42
3.4 考察	44
3.4.1 <i>dif</i> 配列予測	44
3.4.2 複製終結点と <i>dif</i> 配列の関係	47
第4章 議論	49
4.1 GC skewのゆらぎ	49
4.2 アーキアにおける <i>dif</i> 配列	50
4.3 総括	50
謝辞	52
参考文献	53
付録	64
成果発表一覧	78

第 1 章

序論

1.1 複製

2000年にヒトの全ゲノムが解読 (Venter et al. 2000) されてから10年経った現在, 次世代シーケンサー (Shendure and Ji 2008; Medvedev et al. 2009) などの新たなシーケンシング技術により, 1100種ものゲノムが解読されている (Liolios et al. 2010). その結果, ポストゲノム時代の到来と共に計算機によるバイオインフォマティクスが生物学に必要不可欠となっており, 遺伝子機能や近縁種間の進化傾向, そして代謝経路など生物の様々な機構が比較ゲノムや*in silico*解析によって解明されるようになってきた (Kono et al. 2006; Ellegren 2008; Jiang 2008; Kono et al. 2009). このように飛躍的な進歩を遂げている分子生物学は, ヒトゲノムが解読される半世紀前の1953年の, 遺伝物質であるDNAの構造解明 (Watson and Crick 1953) に端を発している. そして二重螺旋構造の発見者の1人であるCrickはその後に, セントラルドグマという概念を提唱したが (Crick 1970), 遺伝情報の伝達や発現を担う複製, 転写, そして翻訳という3つのプロセスを含むこの全ての生物に共通する中心原理の中でも, 特にその根源に位置する機構がDNA複製である.

複製はDNAを半保存的に複製し, ゲノム情報を後世へ伝達するために必要不可欠な機構である. この複製は複製開始, DNA合成そして複製終結の3つのステップによって構成されている. まず複製関連酵素である, ヘリカーゼやDNAポリメラーゼなどが複製開始点で複製フォークを形成し, DNAの合成が開始される. DNA合成はプライマーゼと呼ばれるDNAプライマーゼによって作られたRNAプライマー (Frick and Richardson 2001) を起点として, 真核生物ではPCNA (Proliferating Cell Nuclear Antigen), 原核生物では (β) Sliding clampという共に環状のタンパクと共にDNAポリメラーゼが5'→3'方向に進めていく (Johnson and O'Donnell 2005; Kelly and Brown 2000; Stukenberg et al. 1991). DNAポリメラーゼは校

正機能によって、自身が合成する塩基より手前の塩基が鋳型鎖に対して正しく合成されているか、を確認しなければ次のDNA合成作業に入る事ができない (McCulloch and Kunkel 2008). そのためDNA合成地点の手前にDNAプライマーゼがRNAプライマーを生成することで、DNAポリメラーゼはそのRNAプライマーを起点としてDNA合成を行なう事ができる。またこのDNA合成はリーディング鎖では連続的に行われるが、ラギング鎖では不連続複製が行われる。その原因としてリーディング鎖に対してラギング鎖の塩基が逆方向であることが挙げられる。DNAポリメラーゼはデオキシリボースの3'水酸基にヌクレオチドのリン酸基を結合させる事で5'→3'方向へと進行するため、その塩基配列が逆方向であるラギング鎖ではリーディング鎖のように合成を進めることができない。そこでラギング鎖ではRNAプライマーを数カ所に形成し、それらRNAプライマーを起点として岡崎フラグメント (Okazaki et al. 1968; Ogawa and Okazaki 1980) と呼ばれる短い断片を次にあるRNAプライマーに出会うまで断続的に合成しなければならない (Stukenberg et al. 1994; Li and Marians 2000) (図1-1).

このような機構はすべての生物において共通しているが、細かいいくつかの点では真核生物と原核生物の間で違いが見られる。まず真核生物は複製開始点を潜在的に多く持っており、これら異なる場所から複製を開始する。それに対してバクテリアは1箇所の複製開始点から両方向に複製を進行していく。それに伴ない、複製の終結にもバクテリアと真核生物の間で違いが生じる。真核生物の場合、数カ所で開始された複製は、隣接する箇所から同じように始まった複製の進行と出逢うか、染色体末端のテロメアにたどり着くことでその終結を迎えるのに対し、原核生物の複製は両方向に進行した複製フォークが概ね開始点から対称の位置で衝突することで終結する。バクテリアの細かい複製機構に関しては次節にて詳細を述べる。

1.2 バクテリアの複製と対称性

*Escherichia coli*を例に挙げると、まず約245bpの複製開始領域に存在する3つの13-merリピートのATリッチ領域と、5つのDnaA-boxと呼ばれる9塩基の配列 (5'-TTATCCACA-3') (Schaper and Messer 1995; Messer 2002) にDnaAタンパクが結合し、近接のATリッチな領域の二本鎖DNAをほどく。ほどかれた一本鎖DNAにDnaBヘリカーゼやDnaGプライマーゼ、さらにはDNA Pol IIIなどといった多くの酵素が結合し、複製フォークが形成される (Schaeffer et al. 2005). 複製フォークは複製開始点を起点として、反対方向にそれぞれのReplichoreのDNA二本鎖をほどきながら複製を進めていく。また複製の終結は実際には単に逆方向に進行した複製フォーク同士の偶然の衝突によって引き起こされるわけではなく、複製方向とは逆向きに備えられているTer配列に結合したTusタンパクによって複製フォークの進行が妨げられ (Bussiere and Bastia 1999; Neylon et al. 2005), 開始点とはほぼ正反対の約5%の領域で終結すると考えられてきた (de Massy et al. 1987).

以上の事から、実際に環状ゲノムの複製開始点と終結点は多くの種で対称に位置している事が解明されているが (Rocha 2008), 複製開始・終結点の他にもバクテリアゲノムには様々な対称性が備わっており (Rocha 2004), 中でも複製機構と関わりが深いとされている対称性として塩基組成がある。そのため、これまでの複製開始・終結点予測は主にこの塩基組成に基づいて行われて来た (Worning et al. 2006; Frank and Lobry 2000). 特に複製開始点は塩基組成のみならず、DnaA-boxやrRNAなど関連遺伝子からほぼ同定する事ができるが (Hansen et al. 2006), 一方で複製終結点予測は塩基組成バイアスに依存してい

る事が多い (Hendrickson and Lawrence 2006). この塩基組成バイアスは $(G-C)/(G+C)$ という式で表され、GC skewと呼ばれており (Lobry 1996a; Lobry 1996b), 特に第三コドンに強く見られる事が知られている (McLean et al. 1998). 現在複製開始・終結点予測を行うツールの *de facto* スタンダードとなっている Oriloc (Frank and Lobry 2000) もこのGC skewを基盤としたソフトウェアの1つである.

塩基組成のバイアスが引き起こされる原因として、2つの仮説がこれまで議論されてきている. まず1つ目は、塩基の自発的な脱アミノ化による変異が原因であるという、シトシン脱アミノ化説である. ゲノム複製プロセスの最中、DNAポリメラーゼの進行方向に起因して、リーディング鎖では複製が連続的に行われるのに対し、ラギング鎖では断続的に複製が行われる. そのため、ラギング鎖に比べてリーディング鎖の方が一本鎖の状態で存在する時間が長くなってしまふ. すると一本鎖DNAの状態では塩基が自発的な脱アミノ化を起こしてしまい、CとAがTとGに変異してしまう (Lobry and Sueoka 2002). さらに、複製プロセスによって引き起こされるこの変異の蓄積は、複製開始点の方が終結点に比べて一本鎖の状態が長く続くため、より変異が起こり易く、グアニンとシトシンの比率に大きな差が生じてくる. 実際に、GC skewを描くとグアニンとシトシンの比率が極端に変化している2つのシフトポイントを観察する事ができるため、この2点が複製開始・終結点であると考えられ、既述のGC skewによる予測はこのシフトポイントを用いている. しかしながら、この塩基組成バイアスの対称性は全ての種で共通して保存されている訳ではなく、いくつかの種では極端にこの対称性が維持されていない. もう1つの仮説は複製プロセスではなく、転写プロセスでの修復機構による変異である. 転写プロセスの最中に起こるTranscription-coupled repairはアンチセンス鎖にあるピリミジン二量体を優先的に修復することが解明されており (Hanawalt 1991; Mellon and Hanawalt 1989), 実際にセンス・アンチセンス鎖の塩基変異を比較すると、相補的なC, G → T, A移行が見られる (Francino et al. 1996; Francino and Ochman 1997). 転写プロセスはリーディング鎖とラギング鎖の区別はせず、遺伝子方向の偏りに依存すると考えられているため、この仮説は次に述べる遺伝子の偏りとの関係が深いと思われる.

ゲノムの対称性は以上の塩基組成だけでなく、リーディング鎖とラギング鎖における遺伝子の偏りにも見られる. これは複製や転写が行われるにあたって、DNAポリメラーゼとRNAポリメラーゼの衝突を最小限に抑えるために、遺伝子の方向が偏るように分配されていると考えられている (Brewer 1988). さらに塩基組成や遺伝子の方向性の他にも、ゲノム中では配列方向にも対称的な傾向が維持されている事が知られている (Salzberg et al. 1998). 中でも複製方向に強い傾向を示すRAG配列が注目される (Corre and Louarn 2002; Capioux et al. 2001), その配列はFtsKの方向性を示す機能が解明されたことからFtsK-orienting polar sequence (KOPS) とされた (Bigot et al. 2005; Bigot et al. 2004). さらに、このKOPSの様にタンパクを複製終結点方向に動かす可能性のある配列としてAIMS (Architecture IMParting Sequences) がバクテリアに広く観察されている (Hendrickson and Lawrence 2006).

これら全ての対称性は複製機構に基づいて形成されているものであり、増殖速度の速いバクテリアにおいて複製が生活環のほぼ全てを占めている事実を合わせて考えると、複製機構がゲノム構造に強い影響を及ぼしている事が伺える (図1-2).

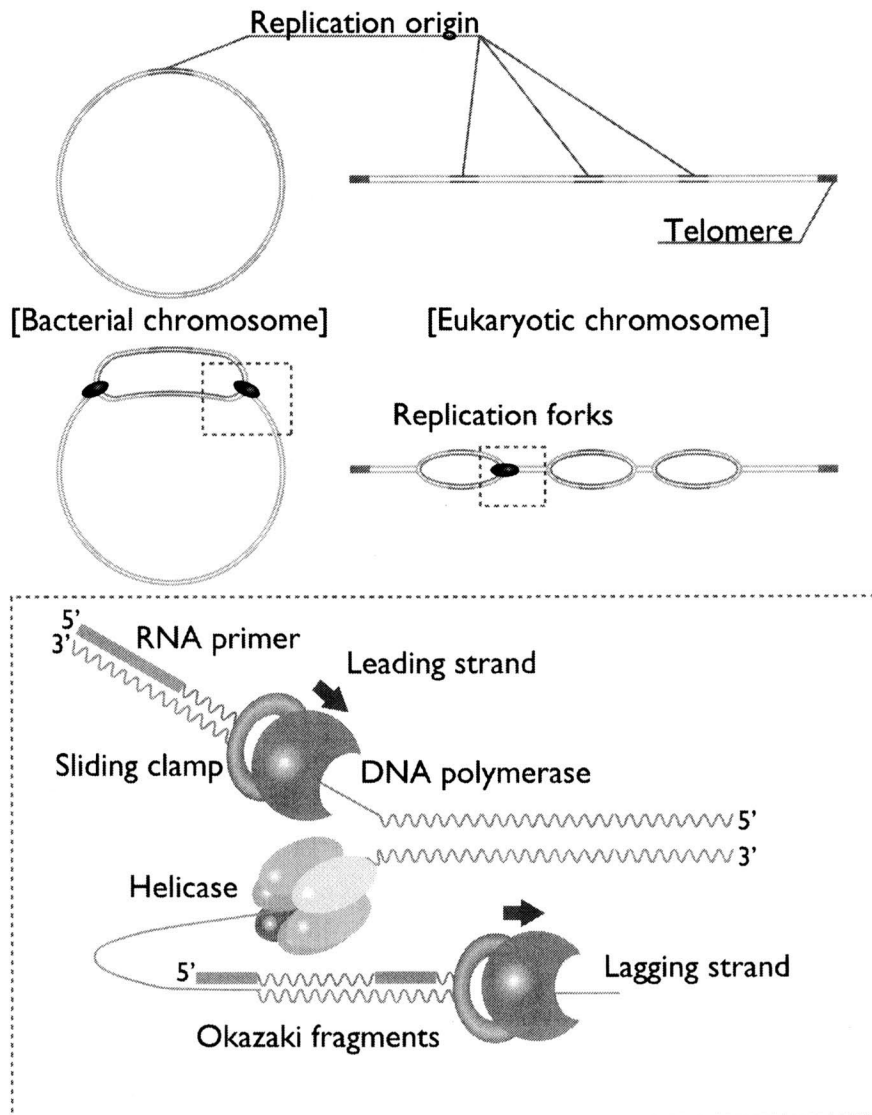


図1-1：複製機構

バクテリアは複製開始点 (Replication origin) を1つだけ持ち、真核生物は複製開始点を複数箇所に持っている。複製開始点に複製フォーク (Replication fork) が結合する事で複製が開始され、ヘリカーゼ (Helicase) によってほどかれた1本鎖DNAを鋳型にDNAポリメラーゼ (DNA polymerase) によってDNAが合成される。リーディング鎖では1カ所のRNAプライマー (RNA primer) から5'→3'方向にDNAポリメラーゼがSliding clampと共に進行していくのに対し、ラギング鎖では数カ所に付けられたRNAプライマーから岡崎フラグメント (Okazaki fragment) を生成しながらDNAを断続的に合成して行く。(Alberts et al. 2008 より改変)

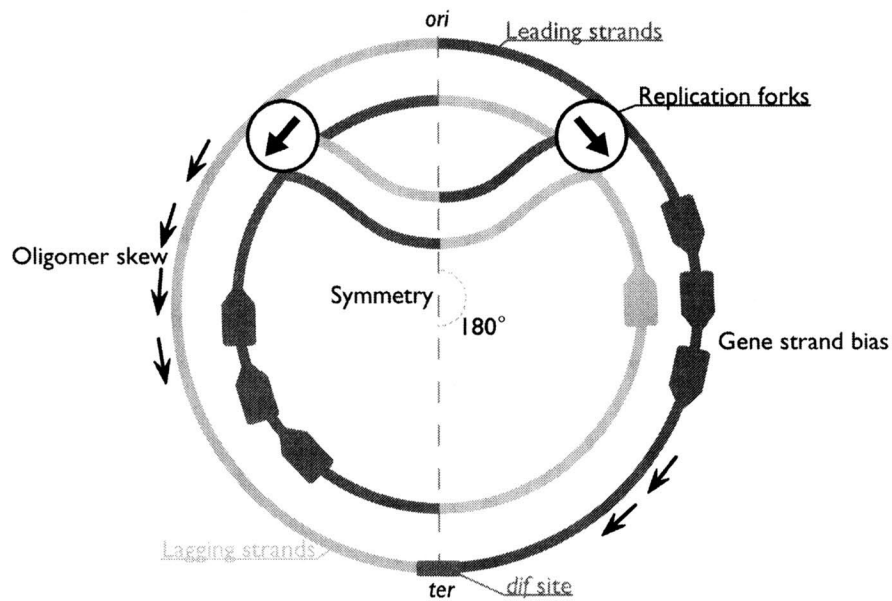


図1-2：ゲノム対称性

図中の青い線はリーディング鎖 (Leading strands) を、黄色い線がラギング鎖 (Lagging strands) を表わしている。多くのバクテリアの環状ゲノムには複製開始点 (*ori*) と複製終結点 (*ter*) の対称性をはじめ、KOPSの様に複製方向を向いた配列対称性 (Oligomer skew), リーディング鎖に偏った遺伝子の対称性 (Gene strand bias) が内在している。(Rocha 2008より改変)

1.3 染色体分裂

この様に塩基組成やTer/Tusシステムが複製終結点と密接に関連していると考えられて来たが、近年この複製の終結は染色体分裂機構のために備わっている *dif* 配列という28塩基の配列が適しているのではないかと注目され始めた (Hendrickson and Lawrence 2007; Higgins 2007). *dif* 配列は染色体分裂に不可欠な配列であり, *E. coli* や *Bacillus subtilis* で発見されている (Blakely and Sherratt 1996; Sciochetti et al. 2001). 複製プロセスの最中, 複製フォーク進行の障害が原因となり, 相同組換えによる複製フォークの修復・再開が起こる (Horiuchi et al. 1994). この相同組換えが奇数回起こると娘DNAは2つに分割されず, 1つの二量体DNAを形成してしまう. そこでこの二量体DNAを2つの一量体娘DNAへと切り離すために, *dif* 配列にチロシンリコンビナーゼであるXerCDが結合して複合体を形成し, FtsKがきっかけとなって染色体分裂 (Chromosome Dimer Resolution: CDR; 図1-3) を起こす (Blakely et al. 1993; Péralis et al. 2001; Li et al. 2003; Bigot et al. 2007; Strick and Quessada-Vial 2006; Yates et al. 2006).

FtsKはリング状の構造を取り, KOPS配列を読みながらDNA上を移動しているDNAトランスロカーゼであり (Aussel et al. 2002; Strick and Quessada-Vial 2006; Bigot et al. 2007), 膜貫通ヘリックスであるおよそ200残基のN末端ドメイン (FtsKN; Dorazi and Dewar 2000), プロリンとグルタミンが豊富なリンカー領域, そしてATPに依存するDNAトランスロカーゼである約500残基のC末端ドメイン (FtsKC) で構成さ

れており (Bigot et al. 2004), 特にFtsKNが細胞分裂に関わっているという報告がされている (Draper et al. 1998; Wang and Lutkenhaus 1998; Yu et al. 1998). このうちFtsKNが細胞分裂中に形成される隔膜と相互作用し, FtsKのC末端残基 (FtsK γ) がKOPS配列を認識しながら進行して*dif*配列に到着する (Ptacin et al. 2006; Sivanathan et al. 2006; Löwe et al. 2008; Dubarry et al. 2009). *dif*配列には2つのドメインが存在しており, それぞれXerCとXerDの標的ドメインとなっている (Val et al. 2008). *dif*サイトに到着したFtsKの γ サブユニットが, 結合したXerDのC末端と相互作用し (Yates et al. 2006), 二量体分裂が開始される (Kuempel et al. 1991). このFtsKやXerCDはバクテリアに広く保存されており, 実際にCDRは*E. coli*や*B. subtilis*の他にも*Haemophilus influenzae*, *Xanthomonas campestris*, *Caulobacter crescentus*, そして*Vibrio cholerae*などで確認されており (Neilson et al. 1999; Yen et al. 2002; Jensen 2006; Val et al. 2008; Recchia and Sherratt 1999), プラスミドでも観察されている (Trigueros et al. 2009).

以上のようにCDRは複製の際にDNAが二量体になってしまうことを避ける機構であるため, *dif*配列の欠損が増殖効率を下げてしまう事がこれまで確認されている (Cornet et al. 1996). そのためこのCDRを効率良く機能させる為にも, 環状ゲノムの複製は*dif*配列付近で終結する事が望ましい. これまでの*dif*配列を予測する研究によって, *dif*配列は複製開始点とほぼ反対に位置しており, 一部がパリンドローム構造を取っている事が確認されたが (Carnoy and Roten 2009), 現状としてこの*dif*配列は実験的に*E. coli*や*B. subtilis*にて決定されている程度である.

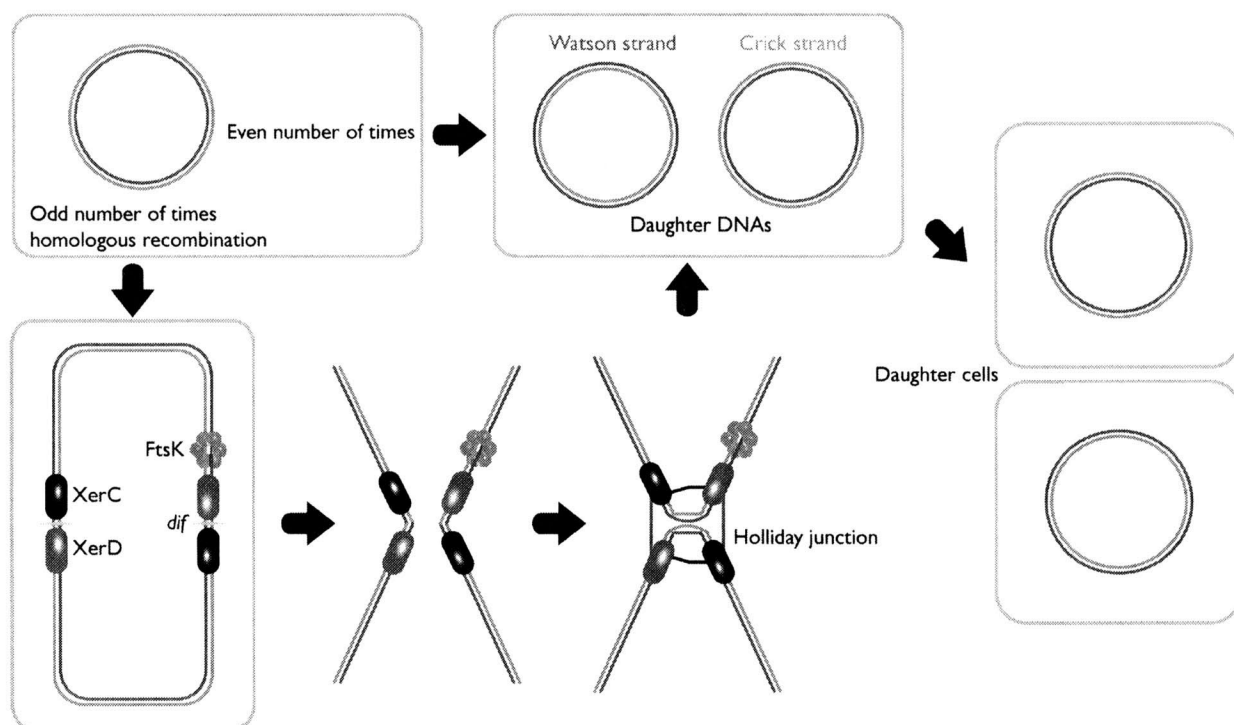


図1-3 : Chromosome Dimer Resolution

環状ゲノムが複製される際に奇数回相同組換えが起こると, 娘DNAに分割されずに二量体のDNAを形成してしまう. ここで, *dif*配列でXerCDやFtsKがHolliday junctionを形成して二量体DNAを分割し, 2つの娘DNAが生成される.

1.4 本研究のねらい

複製プロセスから塩基組成にバイアスがかかることによってゲノム上で塩基組成の対称性が観察されるが、全ての種でこの対称性が明確に観察されている訳ではない。いくつかの種では、GC skewの2つのシフトポイントが近傍に現れ、また、そもそもこのシフトポイントを一切確認する事ができない種も存在する。このような対称性の乱れは、水平伝播によって挿入された塩基配列の影響で塩基組成にゆらぎが生じ、その結果GC skewのシフトポイントに大きな変化をもたらした可能性があるため、本論文ではまず塩基組成バイアスの対称性のゆらぎの原因を探るべく水平伝播との関係を議論する（第2章）。

また、これまで一般的にはTer/Tusシステムによって、複製はTer領域で終結するとされて来たが、近年上述の様に*dif*配列という28塩基の配列がより正確な複製終結点なのではないかと言われる様になった。このように、新たに複製終結点として注目を集め出した*dif*配列はゲノムの対称性を議論するうえで必要不可欠であり、*dif*配列と対称性の関係を解析することで、対称性維持機構の解明につながると考えられる。そこで、第3章では塩基組成の対称性と複製機構の関係を解明する為に、まず網羅的に*dif*配列を予測し、その位置とこれまで複製終結点の予測に用いられてきた塩基組成バイアスであるGC skewのシフトポイントの位置を比較解析により議論する。

これら2つの解析によって、塩基組成バイアス対称性のゆらぎの原因を探ることができ、さらにその対称性を導きだしている複製が何を標的として終結を迎えているかという問題を解明する。

第2章

水平伝播による対称性秩序への影響

2.1 背景

バクテリアの環状ゲノムには複製開始点と終結点の対称性を筆頭に (Rocha 2008), 塩基組成や配列方向といった様々な対称性が維持されており, それら対称性は複製プロセスや転写プロセスと密接に関わっている (Rocha 2004). まずその1つ目の例として, リーディング鎖とラギング鎖における遺伝子の対称性が挙げられる. この遺伝子の対称性は複製や転写が行われるにあたって, DNAポリメラーゼとRNAポリメラーゼの衝突を最小限に抑えるために獲得したものだと考えられている (Brewer 1988). 実際に遺伝子はリーディング鎖に多く存在しているが, もしこの配分がリーディング鎖とラギング鎖で均等であった場合, 複製と転写が恒常的に行なわれているバクテリアにおいて, それぞれで機能しているDNAポリメラーゼとRNAポリメラーゼが衝突してしまい, どちらかのプロセスに支障が生じる可能性がある. そこで遺伝子をリーディング鎖に偏って配置する事でその衝突を最小限に抑える事ができ, この配置がその生物の進化において有利に機能してきたと考えられる. また対称性の他の例として, ゲノム中のオリゴ配列の方向性に関する対称性が報告されている (Salzberg et al. 1998). オリゴ配列の方向性もリーディング鎖とラギング鎖の分配によって観察する事ができ, 例えばあるオリゴ配列がリーディング鎖に極端に多く存在していた場合, このオリゴ配列は方向性があると見なされる. そのような配列の中で, *E. coli*にて特に複製方向に強い方向性を示す配列として最初に注目を集めた配列が, 5'-RRNAGGS-3'という組成から成るRAG配列である (Capiaux et al. 2001; Corre and Louarn 2002). このRAG配列は後にDNAトランスロカーゼであるFtsKの進行方向を決定する配列である事が解明され, KOPS (FtsK-Orienting Polar Sequence) 配列と命名された (Bigot et al. 2004; Bigot et al. 2005). さらにこのKOPSが*E. coli*にて観察された配列であるのに対して, 他の多くのバクテリアでも同じ様に方向性

を持つ配列としてAIMS (Architecture Imparting Sequences: Hendrickson and Lawrence 2006) と呼ばれる配列が確認される様になった。

ゲノムの対称性の中でも、以上の他に注目を集めている対称性の1つが塩基組成バイアスの対称性であり、広く一般に $(G-C)/(G+C)$ という式で表され、GC skewと呼ばれている (Lobry 1996a; Lobry 1996b)。この塩基組成のバイアスが引き起こされる原因として、これまでに2つの仮説が議論されてきている。まず1つ目にシトシン脱アミノ化説が挙げられる。これは、ゲノム複製プロセスの最中にリーディング鎖の方が一本鎖の状態が存在する時間が長くなり、塩基が自発的に脱アミノ化を起こしてしまい、CとAがTとGに変異してしまうという説である (Lobry and Sueoka 2002)。そしてもう1つの仮説は複製プロセスではなく、転写プロセスでの修復機構による変異である。転写プロセスの最中に起こる Transcription-coupled repairはアンチセンス鎖にあるピリミジン二量体を優先的に修復することが解明されており (Hanawalt 1991; Mellon and Hanawalt 1989)、実際にセンス・アンチセンス鎖の塩基変異を比較すると、相補的なC, G → T, A移行が見られる (Francino et al. 1996; Francino and Ochman 1997)。以上の2つの仮説により、GC skewを描くとグアニンとシトシンの比率が極端に変化している場所に2つのシフトポイントが観察される (図2-3)。

このように環状ゲノムを持つ細菌の多くは、塩基組成バイアスによってGC skewのシフトポイントが対称の位置に維持されているが、いくつかの種ではその限りではない。例えば、同じ属同士の *Shigella dysenteriae* Sd197 (NC_007606) と *Shigella flexneri* 2a str. 2457T (NC_004741) を比較しただけでも、それぞれのGC skewのシフトポイントは127.41°と175.84°となっており、大きな違いを示している (図2-1)。そもそもこのGC skewは先に示した式 $(G-C)/(G+C)$ の通り、塩基配列の並びに完全に依存しているため、なんらかの影響によって大規模に塩基配列組成が変化した場合、このGC skewは大きく変化すると考えられる。その影響の可能性として考えられる要因の1つにGenomic Islands (GIs) が挙げられる。GIsは外来の塩基配列が水平伝播などによって挿入された10~200Kbp程の領域と考えられており (Lawrence and Hendrickson 2003; Gogarten and Townsend 2005)、10Kbp未満の領域はGIsではなく Genomic isletsと呼ばれている (Hacker and Kaper 2000)。またGIsは属を越えて様々な種に挿入されている事が知られており (Vernikos and Parkhill 2008)、これは進化の過程において、独立して複数回GIsが移動を繰り返している事を示唆している (Juhas et al. 2007)。さらに、このGIsは様々な特徴的なファミリーによって構成されていると報告されている (Juhas et al. 2009)。例えば、増殖に影響を及ぼしていると考えられる様な表現型を持つGIsはPAIs (Pathogenicity islands) と呼ばれており、GIsの数多くのファミリーの中で最初に着目された特徴である (Hacker et al. 1990)。また他にも、代謝関連遺伝子を持つMetabolic islands (Gaillard et al. 2006) や、水銀耐性などを持つResistance islands (Larbig et al. 2002) などのファミリーが知られている。さらに、他種でオーソログ遺伝子を持たない遺伝子を持つGIsも知られており (Hsiao et al. 2005)、このGIsは進化の過程で、独自の機能を獲得するために適応してきたと考えられている。このように、様々な機能によって特徴づけられたファミリーによって構成されているGIsだが、生物界レベルで観察した場合でも、細菌とアーキアによってGIsの機能に偏りがある事が近年解明されてきている。例えば細菌のGIsには細胞内輸送や細胞運動、防御機構関連の遺伝子が多いのに対して、アーキアの場合は膜タンパクや二次代謝関連の遺伝子が多く、さらに翻訳、リボソーム構造そして生合成関連の遺伝子は細菌とアーキアのどちらでもあまり多くはない (Merkl 2006)。以上の事から、GIsは細菌に広く保持されており、宿主ゲノムの比較的多くの領域を占めている事がわかる。そこで本研究ではGIs領域を取り除くことによって、変異前の塩基組成観察を試み、GC skewの対称性のゆらぎの原因解明を目指した。先に述べた通りGC skewは複製によって引き起こされた塩基変異

が蓄積された結果であるが、外来から挿入されたGIsは宿主ゲノムに比べて複製による変異が蓄積されていない可能性がある。そのため、この挿入されたGIs領域がGC skewの対称性を乱している可能性があると考えられる。なお本研究は、慶應義塾大学環境情報学部4年池上慶太氏、同2年長谷部百合子氏らと共に行った研究であり、ソフトウェアの調査に関しては同氏らの協力によるものである。

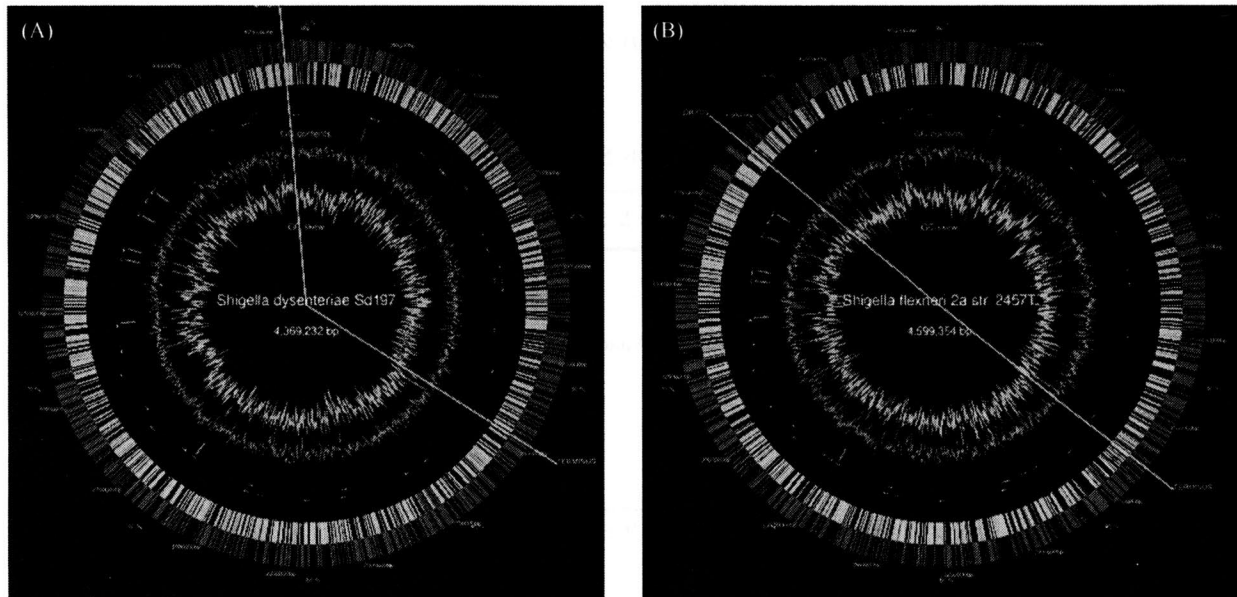


図2-1： *Shigella dysenteriae*と*Shigella flexneri*のGC skewシフトポイントの図 (Genome Projector: Arakawa et al. 2009b)
環状ゲノムを横断し、中心を經由して描かれている黄色い線がそれぞれGC skewのシフトポイントの位置を示しており、(A)が*Shigella dysenteriae*を、(B)が*Shigella flexneri*を示している。この様に比較すると(A)のGC skewシフトポイントは(B)に比べて対称の位置に存在していない事が解る。

2.2 対象と手法

本研究の解析にはG-language Genome Analysis Environment (Arakawa et al. 2003; Arakawa and Tomita 2006; Arakawa et al. 2008) を用い、GIsのデータは以下のそれぞれのソフトウェアを用いた (表2-1)。

・ゲノムデータ

対象生物には以下のGIs予測ソフトウェアが予測対象としているバクテリア全283種を対象とし、ゲノムデータはGenBankより利用した。

・GIs予測ソフトウェア

GIsを予測するに当たり、現在次の3つの手法が一般的に用いられている。まず1つ目がIslander (Mantri et al. 2004) というソフトウェアに用いられている、tRNA領域を考慮に入れた探索法である。水平伝播による配列の多くは、tRNAを標的とする酵素によってゲノムに挿入される。これはtRNAの発現量が高いことに起因しており、GIsが挿入された後に宿主ゲノムでより多く発現できるようになるためと考えられている (Ou et al. 2006)。2つ目は塩基組成の変化を解析する手法であり、代表的なソフトウェアとしてIslandPath-DIMOB (Hsiao et al. 2003) が挙げられる。ゲノム中の塩基組成をあるウィンドウ毎に

観察し、ゲノム全体の塩基組成とかけ離れた組成を持つ領域をGIsと推定する。さらに、遺伝子領域に限定して観察することで、そのコドン使用頻度などから外来の配列であると推定することも出来る (Karlin 2001)。3つ目は近縁種間の比較ゲノムによる予測であり、この手法を用いたソフトウェアとして SIGI-HMM (Merkel et al. 2004; Waack et al. 2006) がある。この手法は近縁種間の配列をマルチプルアライメントなどによって比較し、他種にはない領域を特定することでGIsの検出を行う (Langille et al. 2008)。そこで本研究ではそれぞれの手法における代表的な3つのソフトウェアと、それらを統合したデータベースである IslandViewer (Langille et al. 2009) のデータを用いた。

表2-1：用いたGIs予測ソフトウェア

ソフトウェア	リファレンス	予測手法	対象生物数
IslandViewer	Langille et al. 2009	†	84
SIGI-HMM	Merkel et al. 2004; Waack et al. 2006	種間比較	191
Islander	Mantri et al. 2004	tRNA	55
IslandPath-DIMOB	Hsiao et al. 2003	塩基組成	219

†：IslandViewerはSIGI-HMM, IslanderそしてIslandPath-DIMOBの結果をまとめたデータベースであり、独自の予測手法は用いていない。

・ *In silico*変異体の作成

これら4つのソフトウェアそれぞれによって予測されたGIsの領域を対象のゲノムから取り除き、*in silico*での変異体を作成した。例えば *E. coli* 536 (NC_008253) の場合、IslandPath-DIMOBによって予測されたGIs領域は3948Kbp～4025Kbp間の約76Kbpの領域であったため、*E. coli* 536のゲノムから3948Kbp～4025Kbp間の塩基配列を取り除き、NC_008253_DIMOB株という変異体を作成した。同様にSIGI-HMMでは2504Kbp～2513Kbp間の約9Kbpの領域を予測したため、同領域を取り除いてNC_008253_SIGI株という変異株を生成した (図2-2)。

・ GC skew

GC skewの計算はG-language GAEに含まれている `find_ori_ter`関数を用いて、シフトポイントを算出した。`find_ori_ter`関数は入力されたゲノム配列を元にGC skewを1bpずつ計算し、2つのシフトポイントを算出する (図2-3)。

・ ゲノム角度

本研究では環状ゲノムを持つバクテリア全種を対象としているため、塩基長を扱う際にそれぞれのゲノム長に依存して変化してしまう。そこで、塩基の長さや位置を表すときは全て角度での表現を採用した。例えばゲノムの長さが3600bpで、GC skewのシフトポイントが600Kbpと2700Kbpの位置にあるバクテリアの場合、その位置をGC skewの角度は150°となる (図2-4)。

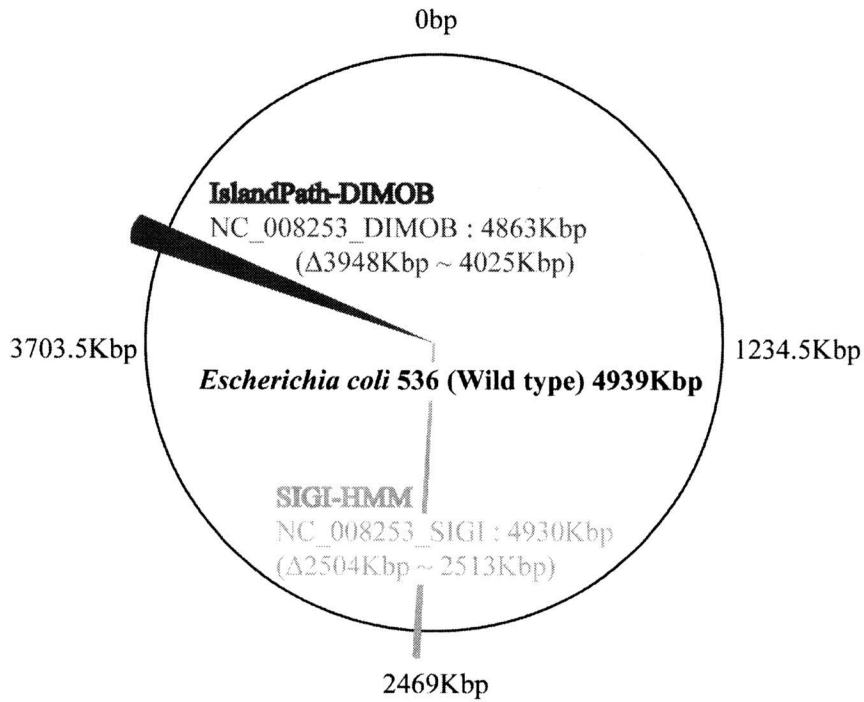


図2-2 : *in silico*変異株

図は*E. coli* 536のGIs領域を表している。IslandPath-DIMOBの示すGIs領域は3948Kbp~4025Kbpの領域であり、IslandPath-DIMOBの変異株を*in silico*で作成する為に同領域を取り除いた (NC_008253_DIMOB)。同様に、SIGI-HMMの示すGIs領域が2504Kbp~2513Kbpの領域であったため、この領域を取り除いたNC_008253_SIGI株を作成した。

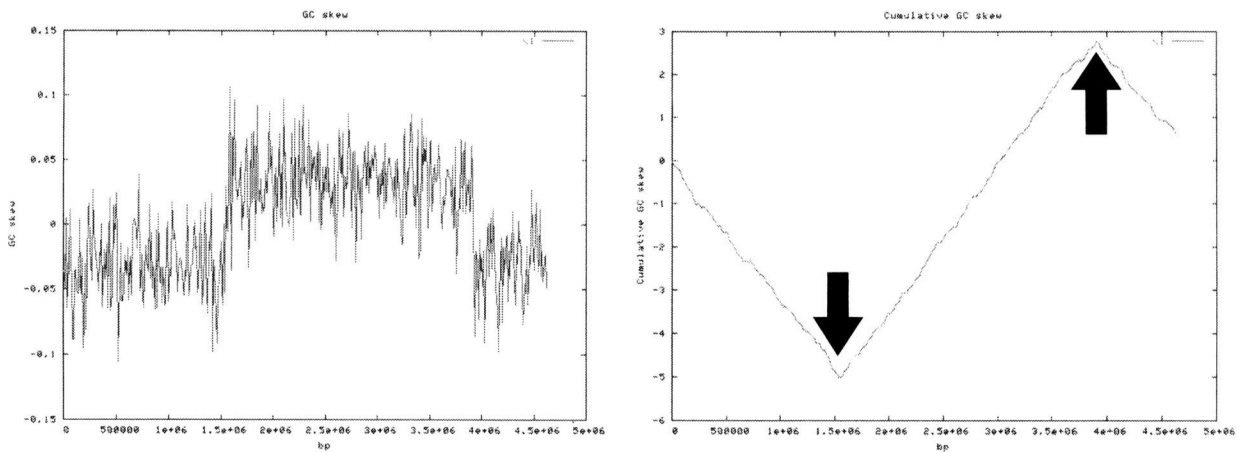


図2-3 : GC skewのシフトポイント

G-language GAEのgcskew関数によって描画された*Escherichia coli* K-12のGC skew。左図のGC skewのウィンドウ毎に足し合わせる事で、右図の累積GC skewが描画される。黒い矢印で示されている通り、累積GC skew (Grigoriev 1998)の方がシフトポイントを明確に観察することができる。ウィンドウサイズ、ステップ幅は共に10Kbp。

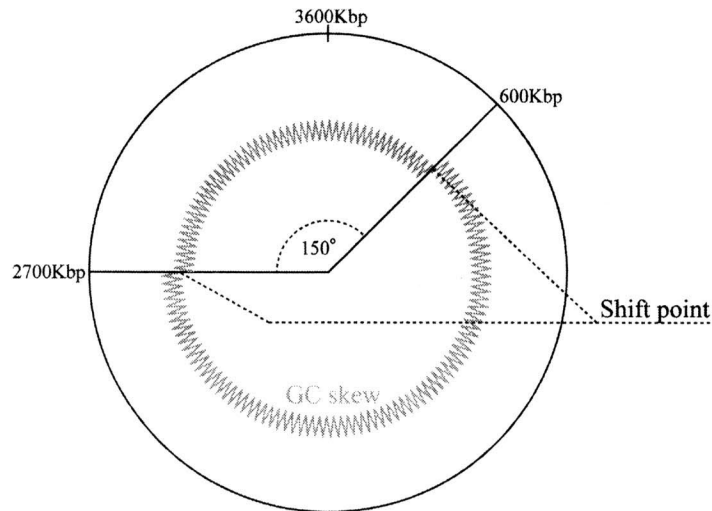


図2-4：ゲノム角度

図の中心に描かれている黄色い折れ線はGC skewを表している。角度 ($\theta = 150^\circ$) は「 $\theta = (360^\circ / \text{ゲノムの長さ}) \times \text{領域の長さ}$ 」にて計算されており、図の通り、領域の長さはGC skewのシフトポイント間の距離を示す。なおこの時のシフトポイント間の距離は短い方を用いる。

2.3 結果

2.3.1 GIsのシフトポイントへの影響

用いた4つのGIsデータから対象とする生物それぞれの*in silico*変異株を生成し、野生株のGC skewのシフトポイントからの変化量を算出した。その結果、いくつかの種では $30^\circ \sim 60^\circ$ の変化を示していたが、ほぼ全ての種でシフトポイントの変化量は $1^\circ \sim 3^\circ$ 程度に収まっており、シフトポイントの変化をほとんど観察することができなかった(図2-5)。

2.3.2 GIs長及び位置との関係

GC skewの変化に寄与している可能性のある因子として挙げられるのがGIsの長さとその位置である。野生株から削除する配列長が長ければ長いほどGC skewへの影響は強いと考えられ、またその位置も重要な意味を持つと思われる。つまり、野生株のGC skewのシフトポイントが 0° と 90° の位置に存在していた場合、取り除くGIsが $0^\circ \sim 90^\circ$ の領域とそれ以外の領域に同じ長さだけあれば、そのGC skewのシフトポイントの変化量は限りなく小さくなり、逆に $0^\circ \sim 90^\circ$ の領域のみにGIsが存在していればその変化量は大きくなる可能性がある。そこで、GIsの長さやGIsの位置それぞれとシフトポイントの変化量との相関係数を算出した(図2-6)。その結果、順位相関係数が $0.238 \sim 0.582$ であったため、ある程度の相関関係を見いだすことができた。そこで次に、GIsの長さや位置ではなくGIs内に含まれる遺伝子数との関係を解析した。

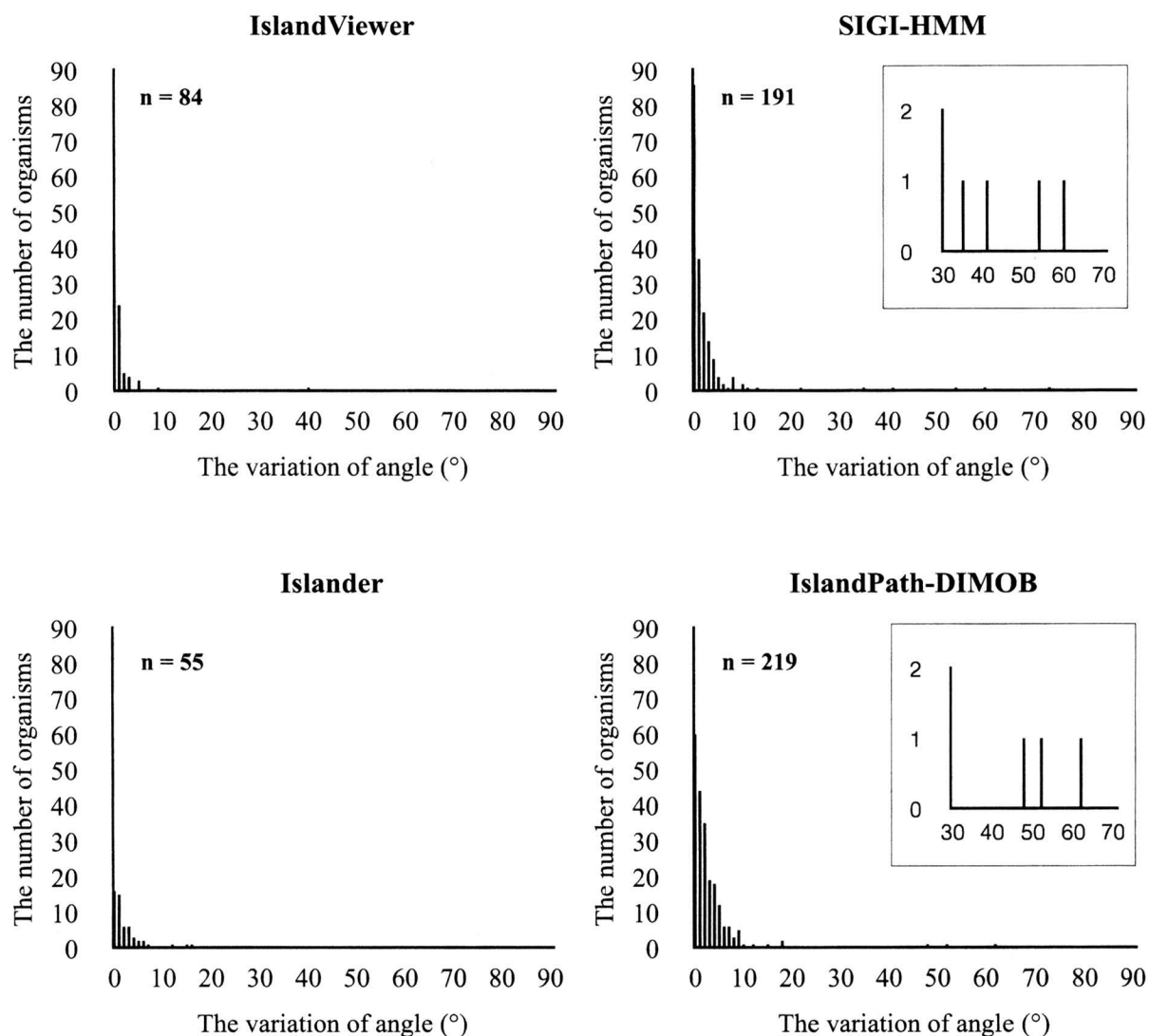


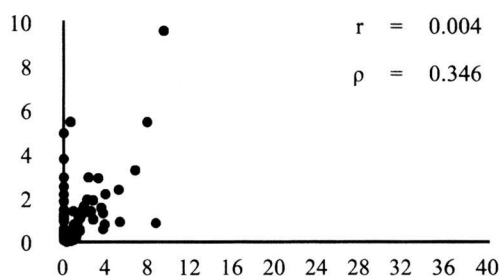
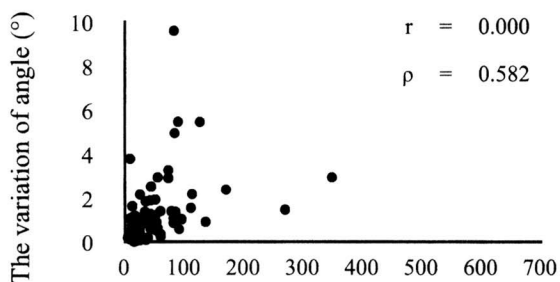
図2-5：GIsを取り除いた際のGC skewシフトポイントの変化

SIGI-HMM, IslandPath-DIMOB, IslanderそしてIslandViewerそれぞれが予測したGIs領域をゲノムから取り除き、GC skewのシフトポイントが取り除く前のゲノムに対してどれほど変化したかを計算した結果、X軸がその角度の変化量で、Y軸が生物種の数を表しており、nは対象とした全生物種数である。SIGI-HMMとIslandPath-DIMOBのグラフ内にある小さなグラフは、角度の変化(X軸)を30°~70°に限定した拡大図である。

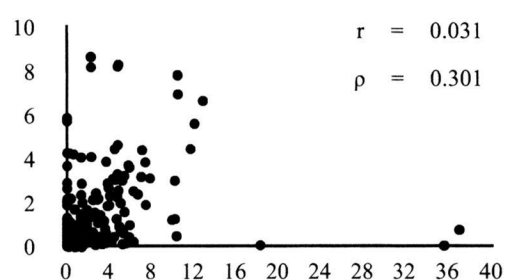
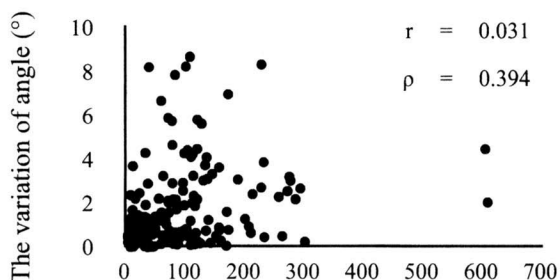
2.3.3 GIs内遺伝子数との関係

GIsの長さや位置にある程度相関があるということから、次に宿主ゲノムにGIsの塩基組成がどれほど順応してきているかを考慮に入れた解析を行った。水平伝播によって挿入されたGIs領域には宿主ゲノムとは違う塩基組成が保持されているが、もしそのGIs領域に遺伝子が多く含まれていれば、その外来の塩基組成は宿主ゲノムの塩基組成と似た変異が多く入って来ているかもしれない。そこでGIs内に含まれているCDS領域の数とGC skewシフトポイントの変化量との相関係数を算出した。ところが積率及び順位相関係数が0.001~0.298であったため、相関関係を見いだすことができなかった(図2-7)。

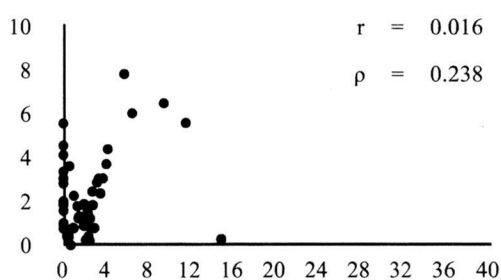
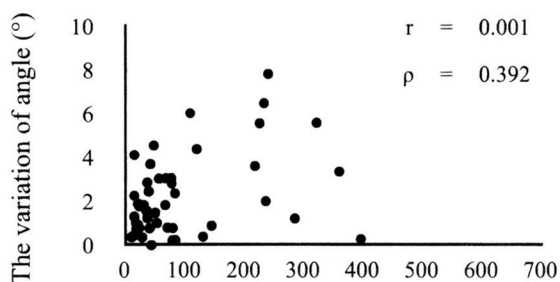
IslandViewer



SIGI-HMM



Islander



IslandPath-DIMOB

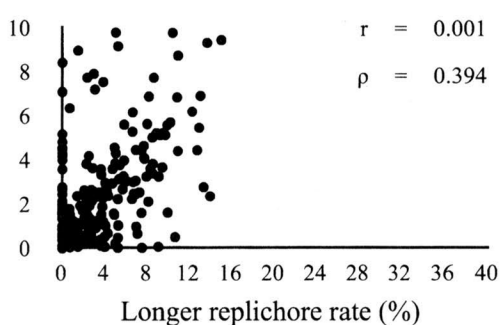
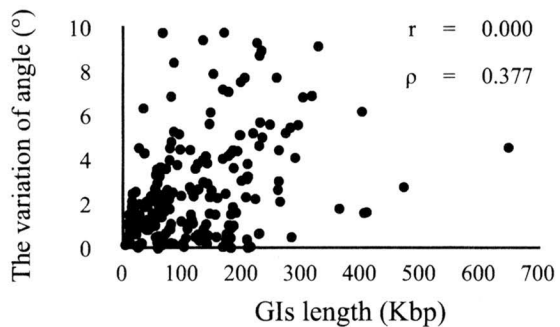


図2-6：GIsの長さやGIsの位置の関係

SIGI-HMM, IslandPath-DIMOB, IslanderそしてIslandViewerそれぞれが予測したGIs領域の長さ、あるいは位置とGC skewのシフトポイントの変化量の相関図。左図のX軸にはGIsの長さ (GIsの長さ / ゲノム全長), 右の図のX軸にはGIsの位置 (Longer replichoreにあるGIsの長さ / GIsの長さ) が示されており、それぞれのY軸にはGC skewシフトポイントの変化量が示されている。r値はピアソンの積率相関係数, p値はスピアマンの順位相関係数。

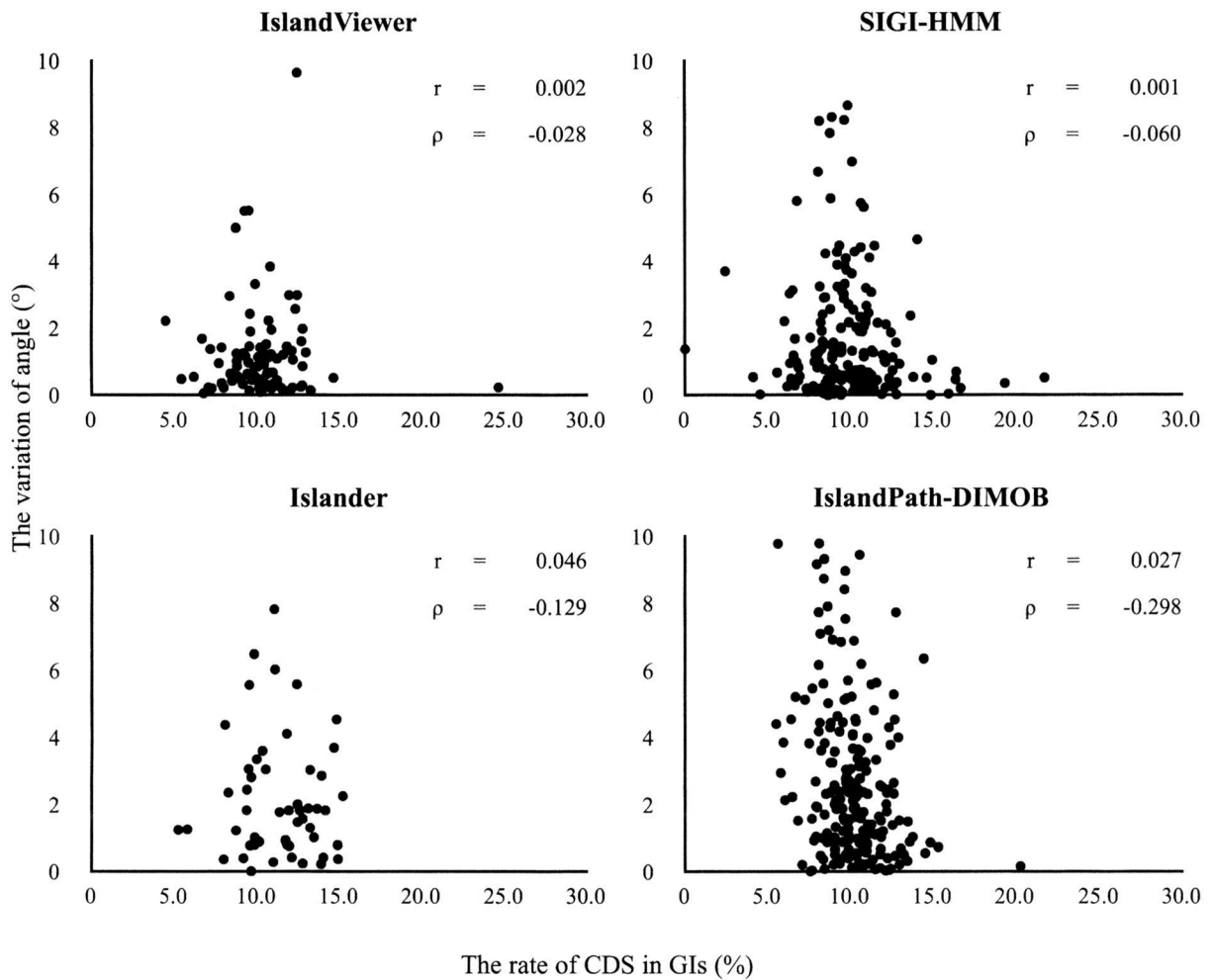


図2-7：GIsに含まれるCDSの割合とGC skewシフトポイントの変化量の関係

SIGI-HMM, IslandPath-DIMOB, IslanderそしてIslandViewerそれぞれが予測したGIs領域に含まれているCDSの割合をX軸に、GC skewのシフトポイントの変化量をY軸に置いた相関図。GIs内のCDSの割合は GIs内のCDS数 / ゲノム中のCDS数によって導きだされた。r値はピアソンの積率相関係数、 ρ 値はスピアマンの順位相関係数。

2.4 考察

図2-5より、GIsを取り除いた時のGC skewシフトポイントの変化量はあまり高くなかったという結果が得られた。またその対称性の変化量は図2-6の通り、全てのソフトウェアで予測されたGIsの長さ、及びその位置に対してある程度相関関係を見いだすことができたが、一方でGIsに含まれているCDS数との相関関係は検出されなかった(図2-7)。すなわちこれは、取り除くGIsが長ければ長いほど、また長いReplichoreに多く存在していればいるほど対称性に影響を与える、という当然の結果であると言え、GIsの塩基組成や機能などは対称性にあまり影響を与えていないということである。以上の事から水平伝播はGC skewのシフトポイントに影響を及ぼさないという事が示唆された。

そこでその原因として次の2点が考えられる。まず1つ目がGIs領域が宿主ゲノムの塩基組成に順応してしまっているという可能性である。そもそもバクテリアは長い歴史の中で数多くの水平伝播イベントによってGIsを獲得しているが、そのイベントはゲノムが解読されるより数年前に起こった訳ではなく、数万年以上前に獲得していると考えられる。そのため数万年以上前に獲得したGIsは、現在に至るまでに幾度も繰り返されてきた複製イベントによって引き起こされた変異によって、本来の塩基組成から宿主ゲノムの塩基組成に近づいて来たと考えられる。その結果、GIsが様々な種のゲノムに数多く観察されているにも関わらず、塩基組成バイアスの対称性に影響をあまりもたらさない可能性がある。しかしながらこの仮説は本研究で用いたGIs予測ソフトウェアの予測アルゴリズムを考慮することで反証される。用いた予測ソフトウェアの内、表2-1の通りSIGI-HMMやIslandPath-DIMOBは宿主ゲノムの塩基組成と異なる塩基組成を示す領域をGIsと推定している。そのため、宿主ゲノムに順応してしまった領域をGIsとして用いることはせず、純粋に塩基組成に変化を与えていると思われる領域だけがGIsとして採用されている。

そしてもう1つの可能性として塩基組成バイアスの対称性の偏りの原因にはまた別の因子が関わっているというものである。本章の冒頭で塩基組成バイアスの対称性の偏りは、宿主ゲノムとは異なった塩基組成を持つ配列の挿入によって引き起こされたものだと仮定した。しかしながら、実はこの偏りは宿主ゲノム自体に本来備わっている構造である可能性がある。塩基組成バイアスは繰り返し行なわれる複製プロセスによって形成されており、複製開始・終結点の位置と密接に関係している。そのためそもそもこの複製開始・終結点が対称ではなく、ある程度偏って存在していたとすれば、この対称性の偏りは乱れではなく、維持されている偏りであると考えることが出来る。

第3章

維持されている複製終結点

3.1 背景

近年複製が終結する地点は塩基組成から導きだされるGC skewのシフトポイントや、Ter配列などの配列方向性ではなく *dif*配列という28塩基の配列がより適しているのではないかと注目され始めた (Hendrickson and Lawrence 2007; Higgins 2007). *dif*配列は染色体分裂に不可欠な配列であり、*E. coli*や*B. subtilis*で発見されている (Blakely and Sherratt 1996; Sciochetti et al. 2001). 複製フォーク進行の阻害が原因となって相同組換えによる複製フォークの修復・再開が起こり (Horiuchi et al. 1994), この相同組換えが奇数回起こると娘DNAは2つに分裂されず、1つの二量体DNAを形成してしまう。そこでこの二量体DNAを2つの一量体娘DNAへと切り離すために、*dif*配列にチロシンリコンビナーゼであるXerCDが結合して複合体を形成し、FtsKがきっかけとなって染色体分裂 (Chromosome Dimer Resolution: CDR) を起こすと考えられている (Blakely et al. 1993; Péral et al. 2001; Li et al. 2003; Strick and Quessada-Vial 2006; Yates et al. 2006; Bigot et al. 2007). FtsKはリング状の構造を取り、複製方向に従って配置されているKOPS配列に沿ってDNA上を移動しているDNAトランスロカーゼである (Aussel et al. 2002; Strick and Quessada-Vial 2006; Bigot et al. 2007). そしてこのFtsKの目標となっている*dif*配列には2つのドメインが存在しており、それぞれXerCとXerDの標的ドメインとなっている (Val et al. 2008). *dif*配列に到着したFtsKの γ サブユニットが、結合したXerDのC末端と相互作用し (Yates et al. 2006), 二量体分裂 (CDR) が開始される (Kuempel et al. 1991). またこのCDR機構は*E. coli*や*B. subtilis*の他にもプラスミドをはじめとする数多くの種で確認されている (Neilson et al. 1999; Recchia and Sherratt 1999; Yen et al. 2002; Jensen 2006; Val et al. 2008; Trigueros et al. 2009). この様にCDRは複製の際にDNAが二量体になってしまうことを避

ける機構であるため、*dif*配列の欠損が増殖効率を下げってしまう事が確認されている (Cornet et al. 1996).

しかしながら、現状としてこの*dif*配列は実験的に*E. coli*や*B. subtilis*にて決定されている程度である。そこで、近年*in silico*による*dif*配列予測がいくつか行われるようになった。まず、Le Bourgeoisらが2007年にFirmicutesを対象とした*dif*配列予測を行った。彼らは*B. subtilis*の*dif*配列をクエリとして、Megablast (Zhang et al. 2000) を用いて22種の*dif*配列を報告している (Le Bourgeois et al. 2007)。またProteobacteriaを対象としたValらは2008年に、 γ -Proteobacteria 27種の*dif*配列を元にプロファイル隠れマルコフモデルを作成し、 β -と α -ProteobacteriaでHMMER 2 (Eddy 1998) によって84種の*dif*配列を予測した (Val et al. 2008)。さらに2009年には、Carnoyらが同じくProteobacteriaを対象として*E. coli*や*B. subtilis*の*dif*配列をクエリとし、BLAST (Basic Local Alignment Search Tool; Altschul et al. 1990; Altschul et al. 1997) やYASS DNA pairwise alignment tool (Noe and Kucherov 2005) といった解析ツールを用いて156種の*dif*配列予測に成功した (Carnoy and Roten 2009)。これらの予測により、*dif*配列は複製開始点とほぼ反対に位置しており、XerCDの結合ドメインがパリンδροーム構造を取っている事が確認されたが、これまでの*dif*配列予測はProteobacteriaやFirmicutesなど限られた門の域を出ず、網羅的にバクテリア全種で*dif*配列の予測は行われていない。さらに、配列相同性に頼った予測手法が主であり、*dif*配列がXerCDの結合サイトである事や、保存傾向などを考慮した予測は行なわれていない。

そこで本研究では、再帰的にHMMERを用いるRHMM (Recursive Hidden Markov Modeling) という手法を開発し、*dif*配列を標的とするXerCDの保存性を参考に予測を行った。さらに、バクテリアの進化の過程で幾度となく行われてきた複製の歴史が深く刻まれていると考えられるGC skewのシフトポイントを比較する事で、環状ゲノムの対称性と複製機構の関係を議論する。

3.2 対象と手法

以下に本解析で用いたゲノムデータと各手法について記述する。なおプログラムはPerl言語で作成しており、主なモジュールとして、ゲノムの統合解析環境であるG-language Genome Analysis Environment (Arakawa et al. 2003; Arakawa and Tomita 2006; Arakawa et al. 2008) を用いた。

3.2.1 *dif*配列予測

*dif*配列はXerCDの結合標的配列であるため、XerCDの構造に依存して*dif*配列は変化している可能性がある。ところが、XerCDは保存されているすべての種で高い配列相同性を示す訳ではない (Val et al. 2008)。そこで、まずはじめにKEGG (Kyoto Encyclopedia of Genes and Genomes) Orthologyデータベース (Kanehisa et al. 2009) を用いて、XerCDが保存されているバクテリアの選定を行ない、さらにその保存性を観察した。その結果、XerCDが門毎にある程度保存されている事が観察されたため、*dif*配列の予測は門毎に行なった。この際、独自の手法であるRHMMを開発し、門毎の予測を行った (図3-1)。以下に*dif*配列予測におけるデータと手法を述べる。

・ゲノムデータ

ゲノムデータはGenBankから取得しており、2009年11月の時点で解読されている全ゲノムの内、環状ゲノムとXerCDを共に持つ655種・731染色体を対象とした。

・既知の*dif*配列

プロファイル隠れマルコフモデルを作成するにあたり、利用した*dif*配列は以下の通りである。

E. coli 5'-GGTGCGCATAATGTATATTATGTTAAAT-3' (Blakely and Sherratt 1994)

B. subtilis 5'-ACTTCCTAGAATATATATTATGTAAACT-3' (Sciochetti et al. 2001)

Frankia alni 5'-CACGCCGATAATGCACATTATGTCAAGT-3' (Hendrickson and Lawrence 2007)

・XerCDの類似性

KEGG Orthology (KO) databaseよりXerCDを持つバクテリアを抽出し、この抽出された種のXerCDのアミノ酸配列の系統的距離を解析した。まず*dif*配列が解明されている*E. coli* (Proteobacteria), *B. subtilis* (Firmicutes), そして*F. alni* (Actinobacteria) のアミノ酸配列それぞれに対して、KO databaseより抽出した全種のXerCDのアミノ酸配列をCLUSTALW (Thompson et al. 1994) によってアライメントし、それぞれの系統的距離を計算した。次にXerCのアミノ酸配列の距離とXerDのアミノ酸配列の距離とを足し合わせ、対象の生物が持つXerCDが*E. coli* (Proteobacteria), *B. subtilis* (Firmicutes) あるいは*F. alni* (Actinobacteria) のどの門に類似しているかを観察した (図3-2)。

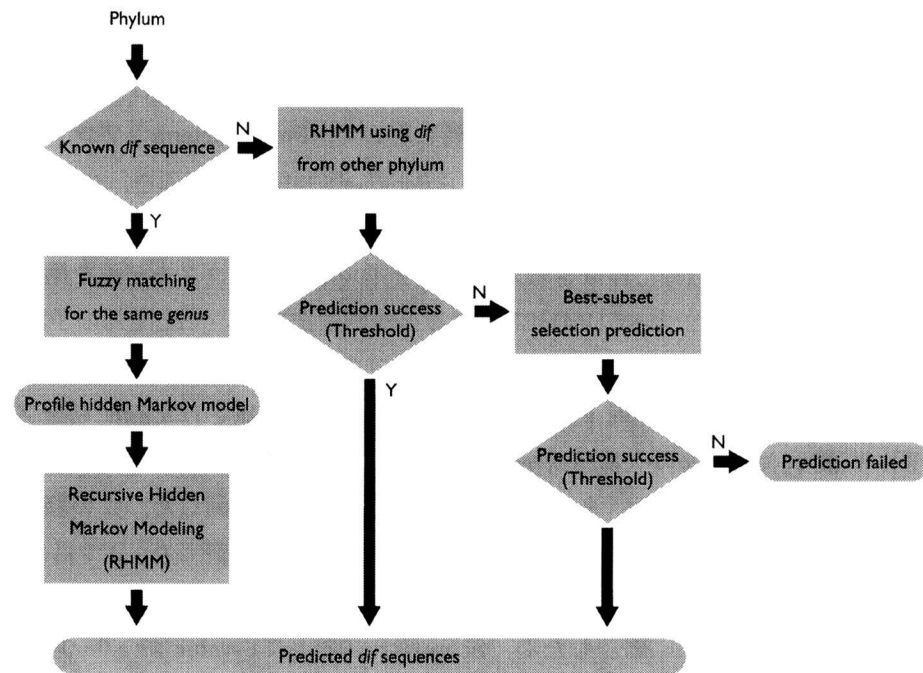


図3-1：予測ストラテジー

1) ある門に対して既知の*dif*配列があれば (Known *dif* sequence), その配列を元に同じ属の*dif*配列をまずはじめに曖昧検索 (Fuzzy matching) によって検出する (Fuzzy matching for the same genus). そして検出された*dif*配列と既知の*dif*配列を統合してプロファイル隠れマルコフモデルを生成する (Profile hidden Markov model). 生成されたプロファイル隠れマルコフモデルを用いて、他の属に対してXerCDの類似性に基づいた再帰的隠れマルコフモデリングによって予測する (Recursive Hidden Markov Modeling - RHMM). 2) 一方で既知の*dif*配列がない門の場合、RHMM法によって作成された他のいくつかの門のプロファイル隠れマルコフモデルを用いて予測を行う (RHMM using other phylum *dif* profile). 3) 最後に、他の門のプロファイル隠れマルコフモデルを用いても予測する事ができなかった種に関しては総当たり予測 (Best-subset selection prediction: 詳細は「総当たり予測」) を行なった。以上 (1)~(3) が*dif*配列予測における概括的な3つの手法である。

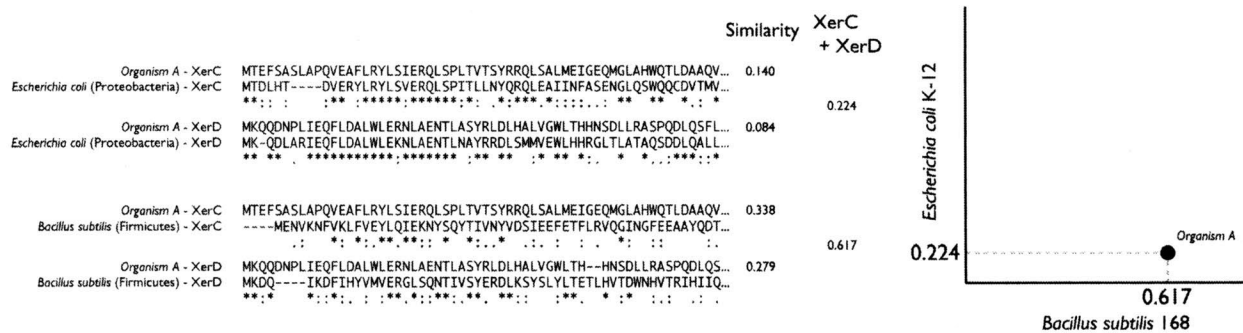


図3-2：XerCDの類似性

ある生物 (*Organism A*) のXerCのアミノ酸配列に対して*Escherichia coli* (Proteobacteria) と*Bacillus subtilis* (Firmicutes) それぞれのXerCのアミノ酸配列をアライメントし、その距離 (Similarity) を算出する。さらに、XerDでも同様に算出した距離を足し合わせ、*Organism A*のXerCDの距離とする。そしてその結果を右図の様にX軸にFirmicutes、Y軸にProteobacteriaを取ったグラフにプロットしていく。

・ Recursive Hidden Markov Modeling (RHMM)

*dif*配列の予測にはHMMER2による隠れマルコフモデルを再帰的に用いた。まず、プロファイル隠れマルコフを生成するために、*E. coli* K12の*dif*配列を元に他の*Escherichia*属28種の*dif*配列をPerlモジュールのString::Approxを用いた曖昧検索によって予測を行なった。この時のパラメータは報告されている*E. coli*の*dif*配列のゆらぎに基づき (Blakely and Sherratt 1994), 欠損と挿入は共に0とし、変異は8塩基のみとした。それによって得られた全29種の*dif*配列を用いて、Proteobacteriaのプロファイル隠れマルコフモデルを作成した。同様に*B. subtilis*の*dif*配列から*Bacillus*属24種の*dif*配列を、*F. alni*の*dif*配列から*Frankia* 2種の*dif*配列を予測し、それぞれFirmicutesとActinobacteriaのプロファイル隠れマルコフモデルを作成した。そして、それぞれの門毎に作成されたプロファイル隠れマルコフモデルを用いて、XerCDのアミノ酸配列の系統学的距離に元づいてHMMERによる再帰的手法を用いて*dif*配列の予測を行った。

はじめにProteobacteriaのプロファイル隠れマルコフモデルを用いて*E. coli*のXerCDに最も類似していた属 (この場合*Shigella*) の*dif*配列予測を行なった。その結果予測された*dif*配列をプロファイル隠れマルコフモデルに加え、*Shigella*の次に*Escherichia*に似ている属 (この場合*Salmonella*) の予測を行った。この様にして、XerCDの類似度の高い種から順に*dif*配列を再帰的に予測した。この際、プロファイル隠れマルコフモデルに*dif*配列を加えるかどうかを以下の3つの方法で検証した。1つ目はHMMERのScoreとE-value, そして候補数の閾値である。この値に満たない候補は明らかに*dif*配列とは違う配列である事が既知の*dif*配列を予測した際に確認した為、Scoreは ≥ 10 そしてE-valueは $< 1.0e-04$ という閾値を設けた。そして候補数とは、先の閾値によって得られた*dif*配列候補がいくつあるかという事だが、今回はこの候補が1つのみであった場合のみを採用した。2つ目に、予測された結果でクロスバリデーションを行ない、予測の再現性を確認した。3つ目は*dif*配列のパリンドローム構造を観察した。*dif*配列にはXerCとXerDが結合する為に2つの結合ドメインを持っており、その領域付近 (7~12塩基目と17~22塩基目) がパリンドロームの様な構造をとっている事が知られている (*E. coli* "--(--- ((((((-(-)))))) ---)---")。そこで、我々はパリンドロームの様な構造がとられている*dif*配列を採用した (図3-3)。

・ 網毎の予測

RHMMでは門毎の予測を基本的に行うが、図3-6の通りXerCDは完全に門毎に保存されているとは言いきれない。そこで、門毎の予測精度があまり高くない種に関してはさらに観察領域を狭め、網毎の予測に切り替えた。この際のプロファイル隠れマルコフモデルには、予測できた同じ網のdif配列を用いた。実際にこの手法で予測した種は複数染色体を持つProteobacteria60種・130染色体だけである。

・ 総当たり予測

門毎に予測を行っていくと、プロファイル隠れマルコフモデルを生成する際に参考にした生物種のXerCDの配列から、遠く離れた種を予測しなければならない場合がある。この時本来その種のXerCD配列は、比較的他の門の生物のXerCD配列と類似している可能性があるにも関わらず、それを考慮して予測することができない。そこでその生物のXerCD配列に対して、dif配列の予測が成功している他の生物(531種)すべてのXerCD配列との距離をCLUSTALWで計算した。その結果、その生物のXerCD配列との類似性が得られるため、その内上位3種のdif配列を用いてプロファイル隠れマルコフモデルを作成し、HMMERによってその生物のdif配列予測を行った(図3-4)。

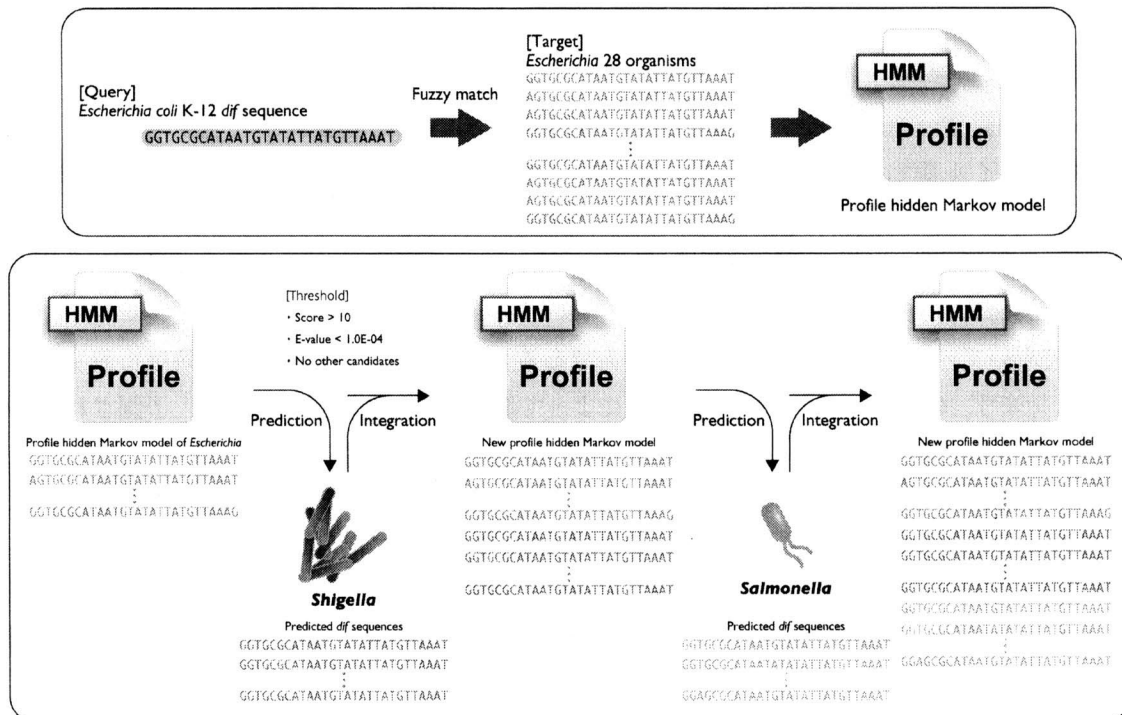


図3-3：再帰的隠れマルコフモデリング

Proteobacteriaでの再帰的隠れマルコフモデリング (RHMM法) の例。 *Escherichia coli* K-12の既知のdif配列をクエリとして同じ属の *Escherichia* 28種を対象に曖昧検索を行ない、全29種のdif配列を検出し、プロファイル隠れマルコフモデルを作成する。このプロファイル隠れマルコフモデルを用いて *Escherichia coli* のXerCDのアミノ酸配列と配列類似性が最も高い種 (*Shigella*) のdif配列を予測する。予測されたdif配列が閾値 (Threshold) に取まっていた場合、そのdif配列は正しいと見なして元のプロファイル隠れマルコフモデルに組み込む。次に、 *Escherichia* と *Shigella* のdif配列が組み込まれたプロファイル隠れマルコフモデルを用いて、 *Shigella* の次に *Escherichia* のXerCDのアミノ酸配列と配列類似性が高い種 (*Salmonella*) のdif配列予測を行う。

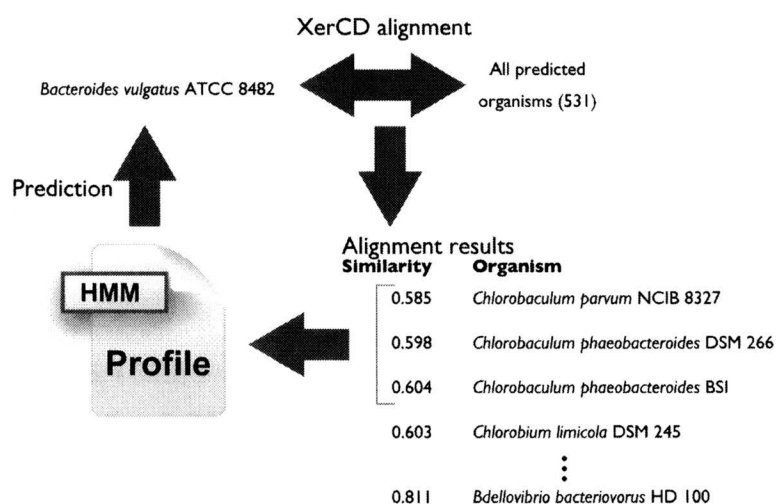


図3-4：総当たり予測の図

対象の生物 (*Bacteroides vulgatus* ATCC 8482) のXerCDに対して、これまで*dif*配列の予測に成功している全種のXerCDをCLUSTALWによってアライメントし、その結果の中で配列距離が最も近かった3種、*Chlorobaculum parvum* (0.585)、*Chlorobaculum phaeobacteroides* (0.598)、そして*Chlorobaculum phaeobacteroides* (0.604) の*dif*配列を用いてプロフィール隠れマルコフモデルを作成し、*Bacteroides vulgatus*の*dif*配列予測を行なった。

・他の門の予測

Proteobacteria, FirmicutesそしてActinobacteria以外の門は既知の*dif*配列が一切解明されていないため、プロフィール隠れマルコフモデルが生成できない。そこで、RHMMによって得られたProteobacteria, FirmicutesそしてActinobacteriaの*dif*配列をプロフィール隠れマルコフモデルとして用い、*dif*配列の予測を行った。その際予測精度の検証には再帰的予測時の3つの方法に加えて、ゲノム対称性を観察した。

本研究で用いているゲノムはすべて環状ゲノムであるため、ゲノム中の各位置を角度で表すことができる。そこでまず、GC skewのシフトポイントによって導き出された複製終結点と、予測された*dif*配列の位置の差を角度で表現した。このように角度で表すことによって、塩基長とは違い、ゲノムの長さに影響されずに他種間比較を行うことができる。今回考慮したゲノム対称性とは、この角度が 100° 以上であった場合は*dif*配列の候補から外すというものである。ほぼすべての種において、複製開始点と終結点は対称に近い場所に位置しているため、GC skewのシフトポイントと*dif*配列の位置の差が 100° 程ある場合、*dif*配列の位置は複製開始点の近くに存在することになる。そこでゲノム対称性という指標には 100° という閾値を用いた。

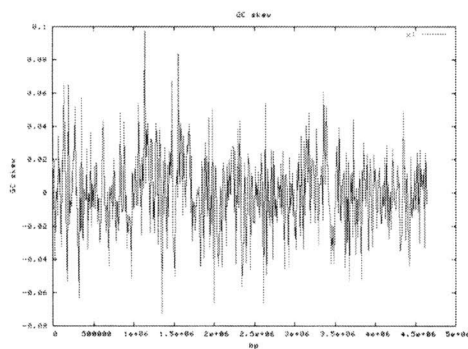
3.2.2 *dif*配列とGC skewの比較解析

・ゲノムデータ

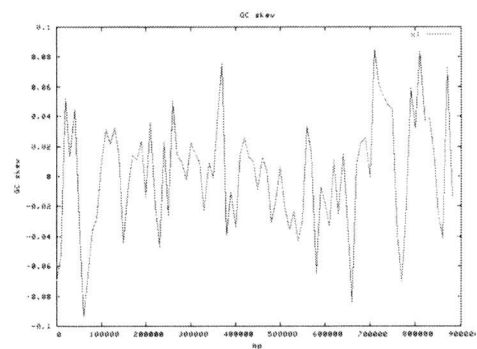
ゲノムデータはGenBankから取得しており、*dif*配列を予測する事ができた環状ゲノムバクテリアの内、予測時にゲノム対称性を考慮した25種と、GCSI (GC Skew Index: 後述) が0.05未満の71種を除いた種を対象とした。

・ GC skew

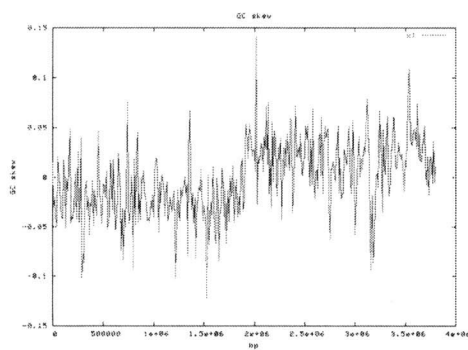
GC skewの計算はG-language GAEに含まれているfind_ori_ter関数を用い、シフトポイントを明確に確認するために累積GC skew (Grigoriev 1998) を算出した。その際のステップ幅は1bpで計算を行った。また、GC skewは広くバクテリアで形成されているが、全てのバクテリアで同等に観察できる訳ではない (Arakawa et al. 2007; Arakawa and Tomita 2007)。そこでArakawaらはGC skewのシフトポイントが明確に現れるかどうかを示すためにGCSIという指標を開発した (Arakawa et al. 2009a)。GC skewはある一定の周波数で表されるため、GC skewのグラフをフーリエ変換することでGC skewのスペクトル強度を算出した。その結果、GCSIが0.05以上であればGC skewのシフトポイントが明確であるという手掛かりを得ることができた (図3-5)。



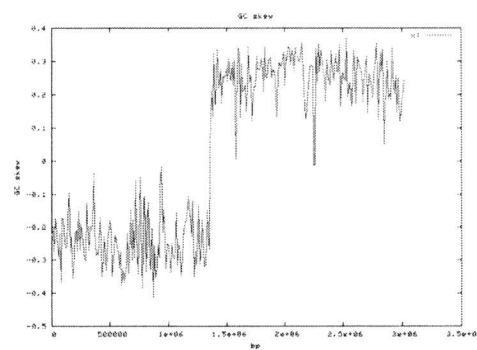
Gloeobacter violaceus PCC 7421 (GCSI = 0.010)



Mycoplasma hyopneumoniae 232 (GCSI = 0.019)



Geobacter sulfurreducens PCA (GCSI = 0.054)



Clostridium perfringens str. 13 (GCSI = 0.629)

図3-5 : GCSI

4種のGC skewの図。上段二つはGCSI値が0.05未満のGC skewの例として*Gloeobacter violaceus* PCC 7421 (GCSI = 0.010)と*Mycoplasma hyopneumoniae* 232 (GCSI = 0.019)を示している。そして下段の2つの図はGCSI値が0.05以上のGC skewの例として*Geobacter sulfurreducens* PCA (GCSI = 0.054)と*Clostridium perfringens* str. 13 (GCSI = 0.629)を示す。

・ ゲノム角度

本研究では環状ゲノムを持つバクテリア全種を対象としているため、塩基長を扱う際にそれぞれのゲノム長に依存して変化してしまう。特に、GC skewのシフトポイントと*dif*配列の位置の距離を議論する際は重要である。そこで、塩基の長さや位置を表すときは全て角度での表現を採用した。例えば、ゲノム長さが3.6Mbpのバクテリアの場合、GC skewのシフトポイントが1.8Mbp、*dif*配列が2.0Mbpの位置にあると、その差は0.2Mbpである事から差を20°と表現した。

3.3 結果

3.3.1 *dif*配列の予測結果概要

XerCDの系統学的距離を計算した結果、配列相同性の高いXerCDは門毎に保存されてる事が分かった(図3-6)。この結果から、今後の*dif*配列予測は門毎に行う事となった。

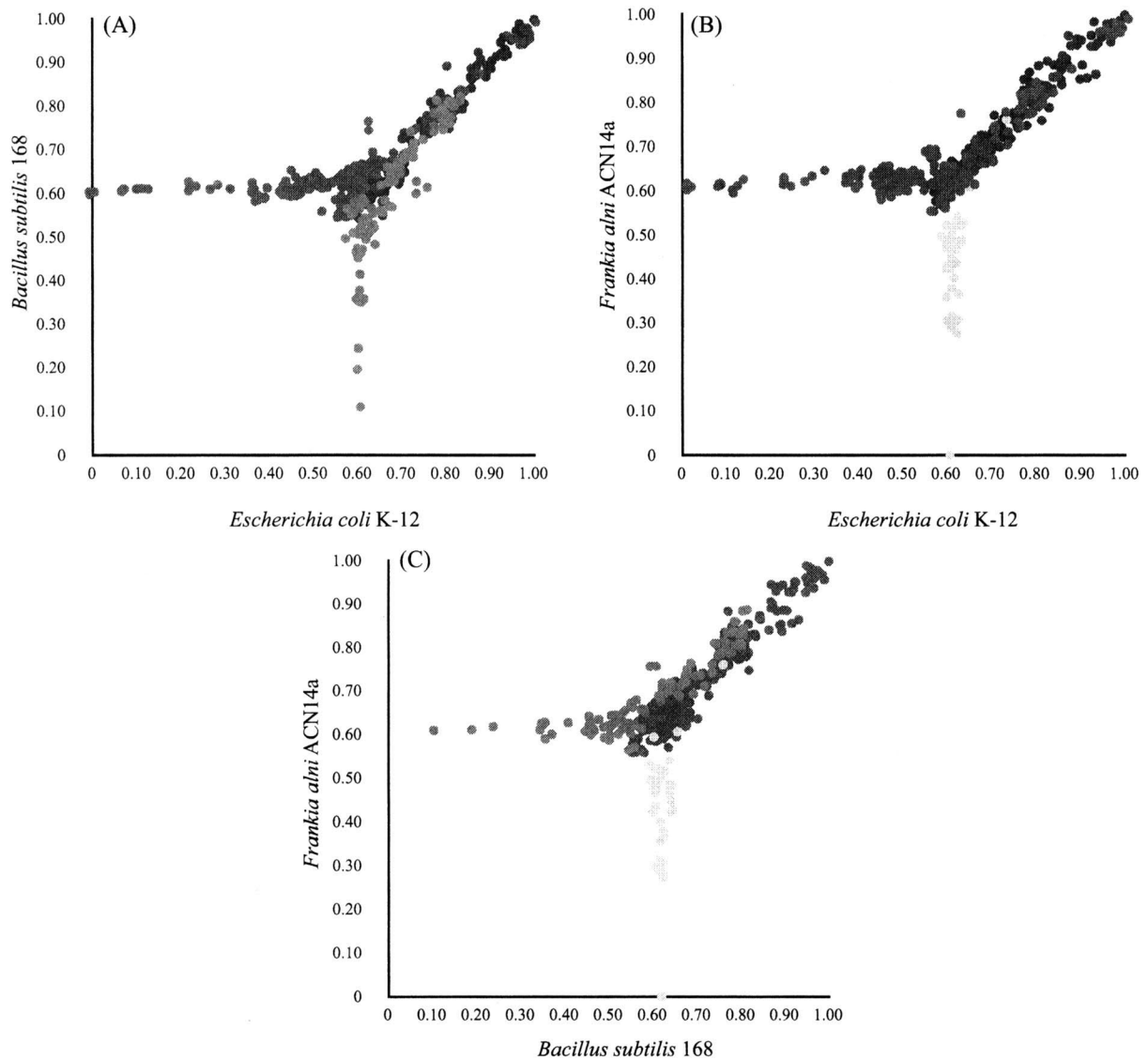


図3-6：XerCDの系統学的距離

任意のある生物の持つXerCDのアミノ酸配列と、*dif*配列が既知の*Escherichia coli* K-12、*Bacillus subtilis* 168そして*Frankia alni* ACN14aそれぞれのXerCDのアミノ酸配列とのアライメント結果の図。(A)は任意の生物のXerCDに対して、*Escherichia coli* K-12と*Bacillus subtilis* 168それぞれのXerCDのアミノ酸配列をアライメントした時の類似性を示しており、同様に(B)は*Escherichia coli* K-12と*Frankia alni* ACN14a、(C)は*Bacillus subtilis* 168と*Frankia alni* ACN14aに対する類似性である。青いプロットはProteobacteria、緑のプロットはFirmicutesそして黄色のプロットはActinobacteriaに属する種を表しており、灰色のプロットはそれ以外の種である。そこからわかる通り、それぞれの門が持つXerCDのアミノ酸配列はある程度独立している。

XerCDの保存性 (図3-6) の結果に基づいた予測によって599種を対象に予測を行った結果, 588種の*dif*配列予測に成功した (詳細データ: 表S1). さらに, 複数ゲノムを持つバクテリア66種・142染色体に対しても同様の予測を行ったところ, 63種・137染色体の予測に成功した (表3-1).

表3-1 : *dif*配列の予測結果

Single Chromosome	Organism	Prediction success	%
Proteobacteria	359	357	99.44
Firmicutes	100	97	97.00
Actinobacteria	66	66	100.00
Bacteroidetes	19	19	100.00
Chlamydiae	14	14	100.00
Chlorobi	11	11	100.00
Acidobacteria	3	3	100.00
Verrucomicrobia	3	3	100.00
Chloroflexi	3	3	100.00
Gemmatimonadetes	1	1	100.00
Nitrospirae	1	1	100.00
Elusimicrobia	1	1	100.00
Tenericutes	1	1	100.00
Spirochaetes	1	1	100.00
Cyanobacteria	5	0	0.00
Planctomycetes	1	0	0.00
Total	589	578	98.13

Multiple Chromosome	Organism (chr [†])	Success (chr)	% (chr %)
Proteobacteria	60 (130)	57 (125)	95.00 (96.15)
Spirochaetes	6 (12)	6 (12)	100.00 (100.00)
Total	66 (142)	63 (137)	94.45 (96.48)

† : chromosome

3.3.2 門毎の予測結果

Proteobacteria, FirmicutesそしてActinobacteriaに対して, ChrolobiやChlamydiaeなどの門でとった予測手法は異なるため, 本節では門毎の予測結果を述べる.

Proteobacteria

Proteobacteriaに属する*E. coli* K-12 (NC_000913) の*dif*配列が既知であったため, まずはじめにこの*dif*配列 (5'-GGTGCGCATAATGTATATTATGTAAAT-3': Blakely and Sherratt 1994) を元に*Escherichia*のプロファイル隠れマルコフモデルを作成した. 曖昧検索によって*E. coli* K12 (NC_000913) の*dif*配列をクエリとして検索を行ったところ, 28種すべてにおいて*dif*配列が検出された (表S2; 図3-7).

検出された*dif*配列に既知の*dif*配列も含めた*Escherichia*29種の*dif*配列を用いてプロファイル隠れマルコフモデルを作成し, RHMMによって*Escherichia*以外のProteobacteria全種 (390種・460染色体) の予測を行ったところ, 306種・306染色体の*dif*配列予測に成功した. 予測された*dif*配列はすべてHMMERのScore > 10かつE-value < 1.0E-4の閾値に収まり, かつ唯一の候補として検出された. さらに, クロスバリデーションの結果は0~3塩基の誤差範囲に収まった. 次に, RHMMのみでは予測できなかった84種・54染色体に関して, 網毎の予測を行った. その結果, 69種・137染色体にて*dif*配列の予測が成功した. この場合も上記と同様の閾値やクロスバリデーションの結果が得られた. 最後に門や網毎のRHMMに

よって予測できなかった種に関して、総当たり予測を行った。その結果10種・10染色体の*dif*配列の予測が成功し、この結果も同様に閾値やクロスバリデーションで確からしさが得られた。

以上の結果、Proteobacteriaでは新規の*dif*配列を413種・482染色体発見することができた (表3-1)。しかしながら、5種・7染色体、 ϵ -Proteobacteria の*Sulfurimonas denitrificans* DSM 1251 (NC_007575), *Nitratiruptor* sp. SB155-2 (NC_009662), β -Proteobacteriaの*Burkholderia phytofirmans* PsJN 1番染色体 (NC_010681), *Burkholderia* sp. 383 1・3番染色体 (NC_007510, NC_007509), α -Proteobacteriaの*Paracoccus denitrificans* PD1222 1・2番染色体 (NC_008686, NC_008687), の*dif*配列を予測することはできなかった。

Firmicutes

Firmicutesに属している*B. subtilis*の*dif*配列が既知であったため、*B. subtilis*の*dif*配列 (5'-ACTTCCTAGAATATATATTATGTAAACT-3': Sciochetti et al. 2001) を元にまず*Bacillus*のプロファイル隠れマルコフモデルを作成した。曖昧検索によって*B. subtilis* str. 168 (NC_000964) の*dif*配列をクエリとして検索を行ったところ、17種すべてにおいて*dif*配列が検出された (表S3; 図3-8)。

B. subtilis str. 168の既知の*dif*配列に加えて、検出された*Bacillus*17種の*dif*配列を用いてプロファイル隠れマルコフモデルを作成し、RHMMによって*Bacillus*以外のFirmicutes全種 (82種・82染色体) の予測を行ったところ、79種・79染色体の*dif*配列予測に成功した (表S1)。予測された*dif*配列はすべてHMMERのScore > 10かつE-value < 1.0E-4の閾値に収まり、かつ唯一の候補として検出された。さらに、クロスバリデーションの結果は0~3塩基の誤差範囲に収まった。

予測できなかった3種・3染色体に対してProteobacteriaと同様に網毎の予測や総当たり予測を行ったが、Clostridiaの*Clostridium perfringens* str. 13 (NC_003366), *Clostridium beijerinckii* NCIMB 8052 (NC_009617), Lactobacillalesの*Lactobacillus helveticus* DPC 4571 (NC_010080) の*dif*配列を予測する事はできなかった。

Actinobacteria

Actinobacteriaに属している種の中で、実験的に*dif*配列が証明されている種はいなかったが、先行研究で*F. alni*の*dif*配列は5'-CACGCCGATAATGCACATTATGTCAAGT-3'なのではないかと予測されていた (Hendrickson and Lawrence 2007)。そこでこの*F. alni*の*dif*配列をクエリとして、*F. alni*のXerCD配列と最も類似性が高かった*Nocardia farcinica* IFM 10152 (NC_006361)と*Mycobacterium avium* subsp. *paratuberculosis* K-10 (NC_002944) の*dif*配列を曖昧検索によって検出した。その結果、両方の種で相同性の高い*dif*配列を得る事ができたため、これらをActinobacteriaのプロファイル隠れマルコフモデルとして用いた (表S4; 図3-9)。

ProteobacteriaやFirmicutesと同様に上記のプロファイル隠れマルコフモデルを用いて、RHMMによってActinobacteria全種 (63種・63染色体) の予測を行ったところ、全種で*dif*配列の予測に成功した (表S1)。予測された*dif*配列はすべてHMMERのScore > 10かつE-value < 1.0E-4の閾値に収まり、かつ唯一の候補として検出された。さらに、クロスバリデーションの結果は100%一致した。

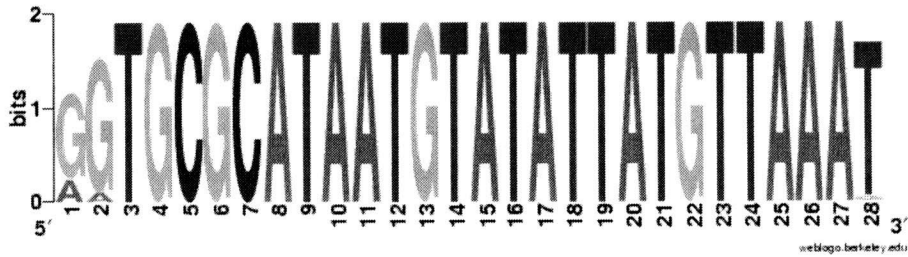


図3-7: *Escherichia*属の*di*配列シーケンスロゴ

WebLogo (Crooks et al. 2004) によって作成された*Escherichia*属28種 (*Escherichia coli* K-12 substr. W3110, O157:H7 EDL933, O157:H7 str. Sakai, CFT073, UTI89, 536, APEC O1, HS, E24377A, ATCC 8739, K-12 substr. DH10B, SECEC SMS-3-5, O157:H7 str. EC4115, SE11, O127:H6 str. E2348/69, IAI1, S88, ED1a, 55989, IAI39, UMN026, LF82, BW2952, BL21, BL21(DE3), REL606, O157:H7 str. TW14359, *Escherichia fergusonii* ATCC 35469) の全*di*配列シーケンスロゴ。

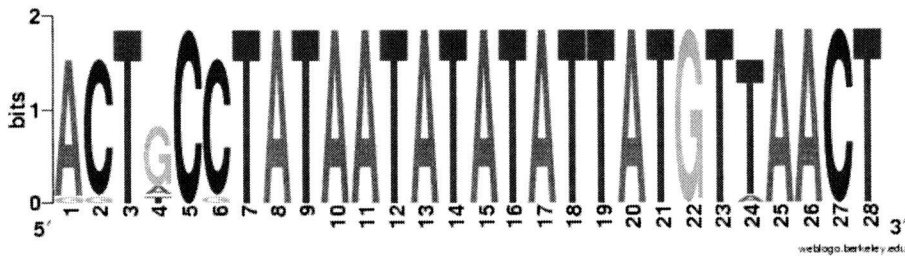


図3-8: *Bacillus*属の*di*配列シーケンスロゴ

*Bacillus*属17種 (*Bacillus anthracis* Ames, Sterne, Ames Ancestor, CDC 684, A0248, *Bacillus cereus* ATCC 10987, ATCC 14579, E33L, AH187, B4264, Q1, 03BB102, *Bacillus cytotoxicus* cytotoxis NVH 391-98, *Bacillus halodurans* C-125, *Bacillus thuringiensis* serovar konkukian str. 97-27, *Bacillus thuringiensis* Al Hakam, *Bacillus weihenstephanensis* KBAB4) の全*di*配列シーケンスロゴ。



図3-9: *Frankia alni*近縁種の*di*配列シーケンスロゴ

*Frankia alni*近縁の2種 (*Nocardia farcinica*, *Mycobacterium avium*) の*di*配列シーケンスロゴ。

Chlorobi

ChlorobiのXerCD配列はProteobacteria, Firmicutes, Actinobacteriaらと同様に門毎に独立していたが, Proteobacteria, Firmicutes, Actinobacteriaらの様に既知の*dif*配列をもっていないため, まずはじめにプロファイル隠れマルコフモデルを作成する上で必要となる数種の*dif*配列をProteobacteria, Firmicutes, Actinobacteriaそれぞれのプロファイルを用いて予測した(表3-2). その結果全8種・8染色体の内, *Chlorobaculum parvum* NCIB 8327 (NC_011027) および*Prosthecochloris aestuarii* DSM 271 (NC_011059) の2種でFirmicutesのプロファイルを用いた場合の予測結果が特に高く(表中太文字表記), 閾値に収まっているという結果が得られた(表3-3).

表3-2: Chlorobiに対する予測結果

Profile Organism	Proteobacteria		Firmicutes		Actinobacteria	
	Score	E-value	Score	E-value	Score	E-value
<i>Chlorobium phaeobacteroides</i> DSM 266	16.7	1.80E-05	14.8	2.10E-05	6.3	5.70E-04
<i>Prosthecochloris vibrioformis</i> DSM 265	14.9	6.80E-05	18.2	6.40E-06	4.2	1.40E-03
<i>Chlorobium limicola</i> DSM 245	19.3	3.00E-06	16.9	1.60E-05	6.8	3.40E-04
<i>Chlorobium phaeobacteroides</i> BS1	15.6	4.10E-05	17.2	1.30E-05	6.2	6.30E-04
<i>Chloroherpeton thalassium</i> ATCC 35110	14.0	5.70E-05	17.9	8.00E-06	7.1	7.30E-04
<i>Chlorobaculum parvum</i> NCIB 8327	18.2	6.70E-06	19.8	2.20E-06	8.8	4.20E-04
<i>Prosthecochloris aestuarii</i> DSM 271	18.2	6.70E-06	22.8	2.80E-07	14.9	3.50E-05
<i>Pelodictyon phaeoclathratiforme</i> BU-1	7.5	5.00E-04	11.3	1.60E-04	6.2	1.00E-03

表3-3: Chlorobiのプロファイル隠れマルコフモデルに用いる*dif*配列

Accession	Genus	Species	<i>dif</i> sequence	Palindrome
NC_011027	<i>Chlorobaculum</i>	<i>parvum</i>	'GTTGTATAATATAAATTATGTAAA--((- -((((-(-))))-- -))--	
NC_011059	<i>Prosthecochloris</i>	<i>aestuarii</i>	'GTTGTATAATATAAATTATGTAAA--((- -((((-(-))))-- -))--	

そこでこの2種を用いてChlorobiのプロファイル隠れマルコフモデルを生成し(表3-4), RHMMによって残り9種の*dif*配列予測を行った. その結果, 全ての種で閾値に収まっているという結果が得られた(表3-5).

表3-4: Chlorobiのプロファイル隠れマルコフモデルを用いた予測結果

Organism	Score	E-value
<i>Chlorobium phaeobacteroides</i> DSM 266	15.0	4.80E-08
<i>Prosthecochloris vibrioformis</i> DSM 265	18.5	1.50E-07
<i>Chlorobium limicola</i> DSM 245	22.5	3.40E-07
<i>Chlorobium phaeobacteroides</i> BS1	25.6	1.50E-09
<i>Chloroherpeton thalassium</i> ATCC 35110	19.8	1.20E-08
<i>Pelodictyon phaeoclathratiforme</i> BU-1	21.1	8.80E-07
<i>Chlorobium chlorochromatii</i> CaD3	21.6	6.40E-07
<i>Chlorobium tepidum</i> TLS	27.9	7.80E-09
<i>Chlorobium luteolum</i> DSM 273	29.1	3.50E-09

表3-5：予測されたChlorobiの*dif*配列

Accession	Genus	Species	<i>dif</i> sequence	Palindrome
NC_002932	<i>Chlorobium</i>	<i>tepidum</i>	TTGTTGTATAATATAAATTATGTAAAGT	---(((-((((-)))))-))---
NC_007512	<i>Pelodictyon</i>	<i>luteolum</i>	TTGTTGTATAATATAAATTATGTAAAGC	---(((-((((-)))))-))---
NC_007514	<i>Chlorobium</i>	<i>chlorochromatii</i>	TTGTTGTATAATATATATTATGTAAACT	---(((-((((())))))-))---
NC_008639	<i>Chlorobium</i>	<i>phaeobacteroides</i>	TTGTTGTATAATATAAATTATGTAAAGT	---(((-((((-)))))-))---
NC_009337	<i>Chlorobium</i>	<i>phaeovibrioides</i>	ATGTTGTATAATATAAATTATGTAAAGC	---(((-((((-)))))-))---
NC_010803	<i>Chlorobium</i>	<i>limicola</i>	TTGTTGTATAATATAAATTATGTAAAGC	---(((-((((-)))))-))---
NC_010831	<i>Chlorobium</i>	<i>phaeobacteroides</i>	TTGTTGTATAATATAAATTATGTAAAGC	---(((-((((-)))))-))---
NC_011026	<i>Chloroherpeton</i>	<i>thalassium</i>	AAGTTTATAATATAAATTATGTAAAGT	(--(((-((((-)))))-))--)
NC_011027	<i>Chlorobaculum</i>	<i>parvum</i>	TTGTTGTATAATATAAATTATGTAAAGT	---(((-((((-)))))-))---
NC_011059	<i>Prosthecochloris</i>	<i>aestuarii</i>	ATGTTGTATAATATAAATTATGTAAAGT	(--(((-((((-)))))-))--)
NC_011060	<i>Pelodictyon</i>	<i>phaeoclathratiforme</i>	TTGTTGTATAATATTAGTTATGTAAAGT	---(((-((((-)))))-))---

Chlamydiae

Chlamydiaeに含まれる14種・14染色体すべてにおいてこれまで*dif*配列は同定されていないため、プロファイル隠れマルコフモデルを生成することができない。さらに、Chlorobiの様にXerCDが独立している訳ではなかったため、Chlamydiae独自のプロファイル隠れマルコフモデルは生成せず、これまで本研究で得られた*dif*配列のデータ (Proteobacteria, Firmicutes, Actinobacteria) を用いた予測のみを行った。その結果、Firmicutesのプロファイルを用いた予測結果が最も高いScoreを示しており、クロスバリデーションでも100%の精度が得られた (表3-6)。

しかしながらこの2種、*Chlamydomphila pneumoniae* CWL029 (NC_000922), *Protochlamydia amoebophila* UWE25 (NC_005861), はScoreやE-valueが閾値未満であったため、パリンδροーム構造とゲノム対称性を観察した。その結果GC skewのシフトポイントと*dif*配列の場所が0.01~1.48°程度しかずれていないことから、正しく予測されていると考えられた (表3-7)。

この2種の*dif*配列をChlamydiaeのプロファイル隠れマルコフモデルとして残りの12種・12染色体の*dif*配列をRHMMによって予測した。その結果ScoreやE-valueが閾値を満たし、かつパリンδροーム構造を確認できた (表3-8)。

表3-7：C. pneumoniaeとP. amoebophilaの予測された*dif*配列の位置

Accession	Genus	Species	<i>dif</i> angle	GC angle	GCSI	Δ	<i>dif</i> position	GC%
NC_000922	<i>Chlamydomphila</i>	<i>pneumoniae</i>	5.27	5.28	0.24	0.02	208353	40.58
NC_005861	<i>Protochlamydia</i>	<i>amoebophila</i>	17.28	15.79	0.14	1.49	1091346	34.72

dif angle : *dif*配列の位置 (°)

GC angle : GC skewの示すシフトポイントの位置 (°)

Δ : *dif* angleとGC angleの差

dif position : bp

表3-6：Chlamydiaeに対する予測結果

Profile Organism	Proteobacteria		Firmicutes		Actinobacteria	
	Score	E-value	Score	E-value	Score	E-value
<i>Chlamydophila pneumoniae</i> CWL029	3.9	8.90E-04	8.7	1.80E-04	6.0	3.80E-04
<i>Protochlamydia amoebophila</i> UWE25	7.7	3.40E-04	9.2	3.90E-04	5.9	8.80E-04
<i>Chlamydia trachomatis</i> D/UW-3/CX	14.1	3.90E-05	11.4	9.10E-05	4.4	8.20E-04
<i>Chlamydophila pneumoniae</i> AR39	7.3	4.00E-04	12.5	1.00E-04	13.6	1.00E-04
<i>Chlamydophila pneumoniae</i> J138	7.3	3.40E-04	12.5	6.40E-05	13.6	9.20E-05
<i>Chlamydophila felis</i> Fe/C-56	12.2	7.70E-05	13.2	7.30E-05	12.6	1.10E-04
<i>Chlamydophila pneumoniae</i> TW-183	8.0	3.60E-04	13.6	4.80E-05	13.5	4.30E-05
<i>Chlamydophila caviae</i> GPIC	15.1	5.70E-05	14.6	8.30E-05	7.7	4.90E-04
<i>Chlamydia muridarum</i> Nigg	16.1	2.90E-05	20.4	1.50E-06	11.0	9.10E-05
<i>Chlamydia trachomatis</i> A/HAR-13	16.7	1.90E-05	21.8	5.50E-07	13.0	1.30E-04
<i>Chlamydia trachomatis</i> L2b/UCH-1/proctitis	17.0	1.50E-05	22.6	3.10E-07	14.7	7.70E-05
<i>Chlamydia trachomatis</i> 434/Bu	17.0	1.50E-05	22.6	3.10E-07	14.7	3.40E-05
<i>Chlamydia trachomatis</i>	17.0	1.60E-05	22.6	3.10E-07	14.7	7.70E-05
<i>Chlamydophila abortus</i> S26/3	17.6	1.00E-05	23.4	1.80E-07	12.9	9.10E-05

表3-8：予測されたChlamydiaeのdif配列

Accession	Genus	Species	dif sequence	Palindrome
NC_005861	<i>Protochlamydia</i>	<i>amoebophila</i>	ACGTCTGATAACATATGTTATGTTTCTT	(-(--- -(((((()))))))- ---)-)
NC_002620	<i>Chlamydia</i>	<i>muridarum</i>	AGGTCTGATAAATTTATTATGTTGAAT	(----- -(((((-))))))- -----)
NC_012686	<i>Chlamydia</i>	<i>trachomatis</i>	AGGTCTGATAAATTTATTATGTTGAAT	(----- -(((((-))))))- -----)
NC_007429	<i>Chlamydia</i>	<i>trachomatis</i>	AGGTCTGATAAATTTATTATGTTGAAT	(----- -(((((-))))))- -----)
NC_000117	<i>Chlamydia</i>	<i>trachomatis</i>	AGGTCTGATAAATTTATTATGTTGAAT	(----- -(((((-))))))- -----)
NC_010280	<i>Chlamydia</i>	<i>trachomatis</i>	AGGTCTGATAAATTTATTATGTTGAAT	(----- -(((((-))))))- -----)
NC_010287	<i>Chlamydia</i>	<i>trachomatis</i>	AGGTCTGATAAATTTATTATGTTGAAT	(----- -(((((-))))))- -----)
NC_003361	<i>Chlamydophila</i>	<i>caviae</i>	TGGTCTGATAAATGTATTATGTTGCAT	--(--- -(((((-))))))- ---)--
NC_007899	<i>Chlamydophila</i>	<i>felis</i>	AGGTCTGATAAATTCATTATGTTGCAT	-(--- -((((----)))))- ---)-)
NC_004552	<i>Chlamydophila</i>	<i>abortus</i>	ACGTCTGATAAATGTATTATGTTGAAT	(----- -(((((-))))))- -----)
NC_002179	<i>Chlamydophila</i>	<i>pneumoniae</i>	AAGTCTGATAATTGTTATTATGTTGCAT	-(--- -((((----)))))- ---)-)
NC_005043	<i>Chlamydophila</i>	<i>pneumoniae</i>	AAGTCTGATAATTGTTATTATGTTGCAT	-(--- -((((----)))))- ---)-)
NC_000922	<i>Chlamydophila</i>	<i>pneumoniae</i>	AAGTCTGATAATTGTTATTATGTTGCAT	-(--- -((((----)))))- ---)-)
NC_002491	<i>Chlamydophila</i>	<i>pneumoniae</i>	AAGTCTGATAATTGTTATTATGTTGCAT	-(--- -((((----)))))- ---)-)

Acidobacteria

Acidobacteriaに含まれる全3種のどの種のdif配列も、Chlorobiらと同様にこれまでに同定されてきてはいない。そのため本研究ではAcidobacteriaに含まれる全生物種に対して、上記の予測によって得られたProteobacteria, Firmicutes, Actinobacteria, ChlorobiそしてChlamydiaeの予測されたdif配列を元に予測を行った。その結果、Firmicutesのプロファイルを用いた予測精度が最もよく (太文字表記)、かつそのdif配列にはパルンドローム構造が確認された (表3-9)。また*Acidobacterium capsulatum* ATCC 51196 (NC_012483) は最も予測結果が悪かったが、Firmicutes, ChlamydiaeそしてChlorobiのプロファイルがすべて同じdif配列を示しており、配列自体にもパルンドローム構造が確認できた (表3-10)。

表3-9：Acidobacteriaに対する予測結果

Profile Organism	Proteobacteria		Firmicutes		Actinobacteria		Chlorobi		Chlamydiae	
	Score	E-value	Score	E-value	Score	E-value	Score	E-value	Score	E-value
<i>Acidobacteria bacterium</i> Ellin345	11.2	8.40E-05	21.3	7.60E-07	17.6	1.00E-05	5.5	2.00E-03	24.5	8.20E-08
<i>Solibacter usitatus</i> Ellin6076	16.5	1.50E-05	24.7	7.50E-08	22.5	3.40E-07	8.7	8.90E-04	24.1	1.10E-07
<i>Acidobacterium capsulatum</i> ATCC	-3.1	5.30E-03	-3.3	5.70E-03	1.4	3.00E-03	2.5	8.70E-03	0.0	2.00E-03

表3-10：予測されたAcidobacteriaのdif配列

Accession	Genus	Species	dif sequence	Palindrome
NC_012483	<i>Acidobacterium</i>	<i>capsulatum</i>	ACGTCTGATAATGAATATTCTGCATTTC	----- -(((((----))))-)- -----
NC_008536	<i>Solibacter</i>	<i>usitatus</i>	ACTTCCGATAACGAATATTATGTCAACT	(-((- - ((((- ---- -))) -))-)-)-
NC_008009	<i>Acidobacteria</i>	<i>bacterium</i>	ATGTCCGATAACAGATATTATGTTAACT	(-(- - - ((((- (- -)-))) - - -)-)-

Verrucomicrobia

Acidobacteria同様、Verrucomicrobiaの3種もこれまでdif配列が一切確認されていないため、上記同様に本研究で得られたdif配列データを用いた予測を行った(表3-11)。

最も高いScoreを出した*Methylacidiphilum infernorum* V4 (NC_010794) はProteobacteria, FirmicutesそしてChlorobiのプロファイルすべてにおいて同じ配列が得られた(表3-12)。さらに、*Opitutus terrae* PB90-1 (NC_010571) と *Akkermansia muciniphila* ATCC BAA-835 (NC_010655) もProteobacteriaとFirmicutesで同じ配列を得ることができた。またこれらすべてにおいて、クロスバリデーションで100%の結果が得られ、ゲノム対称性検証やパリンδροーム構造でも確からしさが得られた(表3-13)。

表3-11：Verrucomicrobiaに対する予測結果

Profile Organism	Proteobacteria		Firmicutes		Actinobacteria		Chlorobi		Chlamydiae	
	Score	E-value	Score	E-value	Score	E-value	Score	E-value	Score	E-value
<i>Opitutus terrae</i> PB90-1	2.4	2.80E-03	-0.2	8.90E-03	4.9	2.50E-03	2.0	9.30E-03	4.3	3.70E-03
<i>Akkermansia muciniphila</i> ATCC	5.9	5.90E-04	9.7	2.50E-04	5.5	4.10E-04	5.8	1.50E-03	7.4	7.90E-04
<i>Methylacidiphilum infernorum</i> V4	20.1	1.80E-06	23.3	1.90E-07	6.3	1.30E-03	17.4	1.20E-05	7.7	1.00E-03

表3-12：予測されたVerrucomicrobiaのdif配列

Accession	Genus	Species	dif sequence	Palindrome
NC_010655	<i>Akkermansia</i>	<i>muciniphila</i>	ACTTCGCATAATATATGAATGCCAAAT	(-((- (((((- (()) -)-))))-))-)-
NC_010794	<i>Methylacidiphilu</i>	<i>infernorum</i>	ACTTCGCACAATATATGTTATGTAAAAT	(-((- (((((- (()) -)-))) -))-)-
NC_010571	<i>Opitutus</i>	<i>terrae</i>	ACTCCGCATCATGCATGTGATGCGAACA	--(-((((((- ---- -))))-))-)-

表3-13 : *A. muciniphila*, *M. inferorum*, *O. terrae*の予測された*dif*配列の位置

Accession	Genus	Species	<i>dif</i> angle	GC angle	GCSI	Δ	<i>dif</i> position	GC%
NC_010655	<i>Akkermansia</i>	<i>muciniphila</i>	3.03	3.66	0.22	0.63	2312983	55.76
NC_010794	<i>Methylococcus</i>	<i>inferorum</i>	2.38	6.86	0.19	4.48	1128462	45.48
NC_010571	<i>Opitutus</i>	<i>terrae</i>	31.37	15.58	0.16	15.80	3459633	65.34

dif angle : *dif*配列の位置 (°)

GC angle : GC skewの示すシフトポイントの位置 (°)

Δ : *dif* angleとGC angleの差

dif position : bp

Chloroflexi

*Acidobacteria*らと同様, *Chloroflexi* 3種・3染色体でもこれまでに作成した *Proteobacteria*, *Firmicutes*, *Actinobacteria*, *Chlorobi*そして*Chlamydiae*のプロファイルを用いて*dif*配列予測を行った(表3-14).

表3-14 : *Chloroflexi*に対する予測結果

Profile Organism	Proteobacteria		Firmicutes		Actinobacteria		Chlorobi		Chlamydiae	
	Score	E-value	Score	E-value	Score	E-value	Score	E-value	Score	E-value
<i>Dehalococcoides ethenogenes</i> 195	4.7	8.40E-04	3.8	7.90E-04	14.0	4.40E-05	6.0	1.50E-03	8.1	3.10E-04
<i>Dehalococcoides</i> sp. CBDB1	6.8	4.20E-04	7.1	6.10E-04	5.7	9.20E-04	12.9	9.00E-05	8.0	3.20E-04
<i>Dehalococcoides</i> sp. BAV1	11.0	2.40E-04	15.9	2.60E-05	9.7	1.20E-04	15.7	3.70E-05	16.1	2.80E-05

その結果, *Dehalococcoides* sp. BAV1 (NC_009455) と*Dehalococcoides* sp. CBDB1 (NC_007356) は *Chlorobi*, *Dehalococcoides ethenogenes* 195 (NC_002936) は *Actinobacteria*によって予測されたScoreが高く, それぞれ閾値やクロスバリデーション, パリンドローム構造が確認された(表3-15). さらに*dif*配列をGC skewの示すシフトポイント近くに観察する事ができた(表3-16).

表3-15 : 予測された*Chloroflexi*の*dif*配列

Accession	Genus	Species	<i>dif</i> sequence	Palindrome
NC_007356	<i>Dehalococcoides</i>	sp.	TTTAATTCGTAAATTATTTAA(-(((((- (()) -)-))))-)	
NC_010794	<i>Dehalococcoides</i>	sp.	TATAATTCATAAAATTATTTAA(-- (((--(- () -)--))) --)	
NC_010571	<i>Dehalococcoides</i>	<i>ethenogenes</i>	TTTTATTCTATAAACTATTTAA((- ((--(- () -)--))- -))	

表3-16 : *Chloroflexi*の*dif*配列の位置

Accession	Genus	Species	<i>dif</i> angle	GC angle	GCSI	Δ	<i>dif</i> position	GC%
NC_007356	<i>Dehalococcoides</i>	sp.	7.37	2.65	0.14	4.72	671037	47.03
NC_009455	<i>Dehalococcoides</i>	sp.	21.25	31.67	0.11	10.43	750139	47.17
NC_002936	<i>Dehalococcoides</i>	<i>ethenogenes</i>	8.51	3.94	0.13	4.56	771319	48.85

dif angle : *dif*配列の位置 (°)

GC angle : GC skewの示すシフトポイントの位置 (°)

Δ : *dif* angleとGC angleの差

dif position : bp

Bacteroidetes

Bacteroidetes 19種・19染色体でも Acidobacteria らと同様に、これまでに作成した Proteobacteria, Firmicutes, Actinobacteria, Chlorobi, Chlamydiae のプロファイルを用いて *dif* 配列予測を行った (表3-17).

表3-17より14種・14染色体では、Proteobacteriaによる予測のScoreとE-valueが閾値に収まり、かつパリンドローム構造を取っている事が確認できた。そこで次に、残りの5種・5染色体に対して総当たり予測を行った。その結果、Bacteroidetesでは新規の*dif*配列を19種・19染色体発見することができ (表3-18)、またその配列はGC skewのシフトポイント付近に位置していた (表3-19).

表3-17：Bacteroidetesに対する予測結果

Profile Organism	Proteobacteria		Firmicutes		Actinobacteria		Chlorobi		Chlamydiae	
	Score	E-value	Score	E-value	Score	E-value	Score	E-value	Score	E-value
<i>Porphyromonas gingivalis</i> ATCC 33277	11.1	1.60E-04	3.7	1.40E-03	14.1	5.20E-05	18.1	7.10E-06	7.4	1.20E-03
<i>Porphyromonas gingivalis</i> W83	5.7	7.40E-04	6.2	2.60E-04	6.2	6.30E-04	7.0	7.40E-04	7.2	5.30E-04
<i>Bacteroides thetaiotaomicron</i> VPI-5482	7.9	3.10E-04	9.1	2.10E-04	11.7	1.70E-04	9.2	9.20E-04	15.2	5.50E-05
<i>Azobacteroides pseudotrichonymphae</i> genomovar. CFP2	5.0	8.60E-04	9.4	1.70E-04	8.6	5.30E-04	9.6	5.50E-04	6.8	6.80E-04
<i>Amoebophilus asiaticus</i> 5a2	16.7	1.90E-05	9.8	3.50E-04	6.6	8.60E-04	12.7	1.40E-04	13.1	1.90E-04
<i>Parabacteroides distasonis</i> ATCC 8503	11.8	9.50E-05	10.7	1.30E-04	7.6	8.50E-04	10.8	2.10E-04	10.0	2.60E-04
<i>Gramella forsetii</i> KT0803	14.7	7.70E-05	10.7	3.30E-04	8.5	6.80E-04	11.5	1.80E-04	14.2	7.90E-05
<i>Bacteroides fragilis</i> NCTC 9343	10.0	1.80E-04	13.0	1.50E-04	8.1	3.70E-04	10.5	5.50E-04	11.9	1.20E-04
<i>Cytophaga hutchinsonii</i> ATCC 33406	14.6	4.50E-05	13.3	4.00E-05	7.9	3.80E-04	11.7	4.00E-04	8.3	5.50E-04
<i>Pedobacter heparinus</i> DSM 2366	19.7	2.30E-06	14.4	5.40E-05	8.0	6.50E-04	9.7	5.10E-04	16.1	2.90E-05
<i>Bacteroides vulgatus</i> ATCC 8482	16.8	1.70E-05	14.8	2.60E-05	12.4	1.20E-04	10.8	3.10E-04	14.9	5.50E-05
<i>Capnocytophaga ochracea</i> DSM 7271	16.5	2.10E-05	14.9	6.60E-05	15.9	3.20E-05	9.5	9.90E-04	13.6	8.90E-05
<i>Dyadobacter fermentans</i> DSM 18053	20.0	1.90E-06	15.5	4.40E-05	11.1	3.20E-04	7.9	1.10E-03	16.3	2.50E-05
<i>Flavobacterium johnsoniae</i> UW101	20.5	1.30E-06	17.4	1.20E-05	11.6	4.10E-04	11.2	2.50E-04	18.9	4.20E-06
<i>Bacteroides fragilis</i> YCH46	12.6	1.20E-04	17.9	8.00E-06	11.4	1.30E-04	20.2	1.60E-06	16.6	1.90E-05
<i>Chitinophaga pinensis</i> DSM 2588	17.7	9.50E-06	18.4	5.70E-06	13.2	6.00E-05	11.9	4.20E-04	20.3	1.50E-06
<i>Flavobacterium psychrophilum</i> JIP02/86	13.4	5.90E-05	18.5	5.30E-06	10.6	1.70E-04	11.8	2.00E-04	21.3	7.80E-07
<i>Robiginitalea biformata</i> HTCC2501	22.8	2.70E-07	21.9	5.30E-07	17.2	1.30E-05	6.5	1.50E-03	21.7	5.80E-07
<i>Flavobacteriaceae bacterium</i> 3519-10	22.6	3.30E-07	22.9	2.60E-07	16.0	3.10E-05	9.3	9.90E-04	24.9	6.50E-08

表3-18：予測されたBacteroidetesのdif配列

Accession	Genus	Species	dif sequence	Palindrome
NC_009614	<i>Bacteroides</i>	<i>vulgatus</i>	GTGACGGATAATGACTATTATGTAACT	----- -(((((----)))-- -----
NC_004663	<i>Bacteroides</i>	<i>thetaitaomicron</i>	AATTCCTAGGATTTAGGTTATATCATAT	(--(-- ((--(- (-) -)--) --))--
NC_003228	<i>Bacteroides</i>	<i>fragilis</i>	ATTCGTATAAATTAGTTTATGTCATAT	((--(-- -(((((-) -))))-- --))--
NC_006347	<i>Bacteroides</i>	<i>fragilis</i>	ATTCGTATAAATTAGTTTATGTCATAT	((--(-- -(((((-) -))))-- --))--
NC_011565	<i>Azobacteroides</i>	<i>pseudotrichonymphae</i>	AATTCATAGAAATAATGGCTGTCATAT	(--(-(- ((--(- ---- -)--) -)--)--
NC_010830	<i>Amoebophilus</i>	<i>asiaticus</i>	TAGTTATATAATTTATATTATGTTAAAT	--(-(- -(((((-) -))))-- -)---
NC_013162	<i>Capnocytophaga</i>	<i>ochracea</i>	AATGGTTATAATGTATATTATGTCAAAT	(-(-(- -(((((-) -))))-- -)--)--
NC_013132	<i>Chitinophaga</i>	<i>pinensis</i>	AATACTTATAATTTATATTATGTTAAAT	(-(-(- -(((((-) -))))-- ---))--
NC_008255	<i>Cytophaga</i>	<i>hutchinsonii</i>	TTGTTATATAATTTACATTATGTTAAAT	-(-(-(- -(((((-) -))))-- -)--)--
NC_013037	<i>Dyadobacter</i>	<i>fermentans</i>	TTGTTGATAATTTATATTATGTTAAAT	-(-(-- -(((((-) -))))-- --))--
NC_013062	<i>Flavobacteriaceae</i>	<i>bacterium</i>	AAGTCCTATAATTTATATTATGTTAAAT	(--(-- -(((((-) -))))-- --))--
NC_009613	<i>Flavobacterium</i>	<i>psychrophilum</i>	ACGAACTATAATTTATATTATGTAAAAT	(----- -(((((-) -))))-- -----
NC_009441	<i>Flavobacterium</i>	<i>johnsoniae</i>	TTGACCTATAATTTATATTATGTAAAAT	(----- -(((((-) -))))-- -----
NC_008571	<i>Gramella</i>	<i>forsetii</i>	TTGACTATAATTTATATTATGTAAAAT	(-(-(-- -(((((-) -))))-- --))--
NC_009615	<i>Parabacteroides</i>	<i>distasonis</i>	TCTTGCCATAATGAATATCATGTTATGT	(-(-(-- ((((-(- ----))-))) --))--
NC_013061	<i>Pedobacter</i>	<i>heparinus</i>	TAGTCACATAATTAATATTATGTTAAGT	--(-(- (((((((----)))))))) -)---
NC_002950	<i>Porphyromonas</i>	<i>gingivalis</i>	AATTCGCAGATTTTATATCATGTTATGT	(--(-- ((---(- (-))---)) --))--
NC_010729	<i>Porphyromonas</i>	<i>gingivalis</i>	TAATCCCATATCGTATATACTGTTATAT	--((-- ((---- (-) ----)) --))--
NC_013222	<i>Robiginitalea</i>	<i>biformata</i>	TTGTCCTATAATTTATATTATGTTAAAT	(-(-(-- -(((((-) -))))-- --))--

表3-19：Bacteroidetesのdif配列の位置

Accession	Genus	Species	dif angle	GC angle	GCSI	Δ	dif position	GC%
NC_009614	<i>Bacteroides</i>	<i>vulgatus</i>	138.38	2.88	0.25	135.50	2136810	42.20
NC_004663	<i>Bacteroides</i>	<i>thetaitaomicron</i>	59.40	11.29	0.20	48.11	6132660	42.84
NC_003228	<i>Bacteroides</i>	<i>fragilis</i>	13.34	8.92	0.22	4.42	2795474	43.19
NC_006347	<i>Bacteroides</i>	<i>fragilis</i>	2.87	1.49	0.21	1.38	2680696	43.27
NC_011565	<i>Azobacteroides</i>	<i>pseudotrichony</i>	13.50	23.90	0.25	10.40	892745	32.66
NC_010830	<i>Amoebophilus</i>	<i>asiaticus</i>	5.29	5.29	0.14	0.00	1317733	35.05
NC_013162	<i>Capnocytophaga</i>	<i>ochracea</i>	17.21	23.18	0.08	5.98	2330633	39.59
NC_013132	<i>Chitinophaga</i>	<i>pinensis</i>	20.26	28.14	0.13	7.88	5077355	45.23
NC_008255	<i>Cytophaga</i>	<i>hutchinsonii</i>	11.42	10.51	0.13	0.92	1965461	38.85
NC_013037	<i>Dyadobacter</i>	<i>fermentans</i>	4.48	5.78	0.08	1.30	1212803	51.54
NC_013062	<i>Flavobacteriaceae</i>	<i>bacterium</i>	8.94	8.49	0.08	0.44	1452762	42.69
NC_009613	<i>Flavobacterium</i>	<i>psychrophilum</i>	40.81	64.55	0.12	23.74	1096230	32.54
NC_009441	<i>Flavobacterium</i>	<i>johnsoniae</i>	20.03	16.39	0.12	3.65	3387721	34.11
NC_008571	<i>Gramella</i>	<i>forsetii</i>	5.48	5.48	0.16	0.00	1648336	36.61
NC_009615	<i>Parabacteroides</i>	<i>distasonis</i>	28.59	4.94	0.22	23.65	822255	45.06
NC_013061	<i>Pedobacter</i>	<i>heparinus</i>	6.49	0.78	0.16	5.71	2490571	42.05
NC_002950	<i>Porphyromonas</i>	<i>gingivalis</i>	56.50	0.73	0.06	55.77	1544496	48.29
NC_010729	<i>Porphyromonas</i>	<i>gingivalis</i>	109.28	10.00	0.06	99.28	284425	48.36
NC_013222	<i>Robiginitalea</i>	<i>biformata</i>	9.38	9.31	0.06	0.07	2517214	55.29

dif angle : dif配列の位置 (°)

GC angle : GC skewの示すシフトポイントの位置 (°)

Δ : dif angleとGC angleの差

dif position : bp

Cyanobacteria

これまで同様に、Proteobacteria, Firmicutes, Actinobacteria, ChlorobiそしてChlamydiaeのプロファイル隠れマルコフモデルを用いてCyanobacteria5種・5染色体の予測を行った。しかしながら、Scoreはどれも低く、パリンドローム構造やゲノム対称性などの検証でも芳しい結果は得られなかった(表3-20)。

表3-20：Cyanobacteriaに対する予測結果

Profile Organism	Proteobacteria		Firmicutes		Actinobacteria		Chlorobi		Chlamydiae	
	Score	E-value	Score	E-value	Score	E-value	Score	E-value	Score	E-value
<i>Gloeobacter violaceus</i> PCC 7421	5.8	3.40E-04	6.2	7.10E-04	6.0	6.80E-04	6.2	2.00E-03	6.6	5.70E-04
<i>Synechococcus</i> sp. JA-3-3Ab	-1.9	4.70E-03	2.6	1.70E-03	1.2	3.20E-03	5.3	3.50E-03	0.6	2.50E-03
<i>Synechococcus</i> sp. JA-2-3B'a(2-13)	6.9	4.60E-04	1.4	2.30E-03	1.4	3.20E-03	5.0	4.40E-03	0.3	5.00E-03
<i>Acaryochloris marina</i> MBIC11017	6.5	9.20E-04	6.7	4.40E-04	9.4	3.10E-04	10.5	5.60E-04	4.9	1.70E-03
<i>Synechococcus</i> sp. PCC 7002	7.6	5.90E-04	2.8	2.30E-03	6.5	1.10E-03	12.5	1.20E-04	3.9	1.80E-03

Other

Gemmatimonadetes (1種・1染色体: *Gemmatimonas aurantiaca* T-27) で Proteobacteria, Firmicutes, Actinobacteria, ChlorobiそしてChlamydiaeのプロファイル隠れマルコフモデルを用いて予測したところ、Firmicutesによる予測結果が最も高く、パリンドローム構造“(---(---(---(---)---)---)---)---)---)”やゲノム対称性などからも検証できた (AATGCGTATAATACATGTTATGTAACT)。さらに、ProteobacteriaやChlorobiによる予測結果とも一致していた(表3-21)。

表3-21：Gemmatimonadetesに対する予測結果

Proteobacteria		Firmicutes		Actinobacteria		Chlorobi		Chlamydiae	
Score	E-value	Score	E-value	Score	E-value	Score	E-value	Score	E-value
8.5	1.90E-04	20.0	1.90E-06	7.3	1.50E-04	11.3	2.60E-04	0.9	1.30E-02
		dif angle	GC angle	GCSI	Δ	Position	GC (%)		
Firmicutes result		30.139	30.128	0.071	0.011	1930284	64.274		

太文字下線：Scoreが高く、E-valueが低かったプロファイル

Δ : dif angleとGC angleの差

Nitrospirae (1種・1染色体: *Thermodesulfovibrio yellowstonii* DSM 11347) で Proteobacteria, Firmicutes, Actinobacteria, ChlorobiそしてChlamydiaeのプロファイル隠れマルコフモデルを用いて予測したところ、Actinobacteria以外はすべて閾値におさまり、かつProteobacteria, FirmicutesそしてChlamydiaeは全て同じ配列を予測した (ATGTCTGATAATGTATATTATGTAAAT)。また、パリンドローム構造“(---(---(---(---)---)---)---)---)---)”やゲノム対称性などからも検証できた(表3-22)。

表3-22：Nitrospiraeに対する予測結果

Proteobacteria		Firmicutes		Actinobacteria		Chlorobi		Chlamydiae	
Score	E-value	Score	E-value	Score	E-value	Score	E-value	Score	E-value
13.7	4.20E-05	10.4	9.90E-05	7.6	8.90E-05	10.7	3.00E-04	19.3	3.00E-06
Proteobacteria, Firmicutes, Chlamydiae result		dif angle	GC angle	GCSI	Δ	Position	GC (%)		
		2.814	15.403	0.104	12.588	1017567	34.129		

太文字下線：Scoreが高く，E-valueが低かったプロファイル

Δ：dif angleとGC angleの差

Elusimicrobia (1種・1染色体：*Elusimicrobium minutum* Pei191) で Proteobacteria, Firmicutes, Actinobacteria, ChlorobiそしてChlamydiaeのプロファイル隠れマルコフモデルを用いて予測したところ，全体的にScoreは高く，なかでもFirmicutesとChlamydiaeは同じ配列を予測しており (ACTAACGATAATTTTTATTATGTCAATT)，この配列はパリンδροーム構造を取っていた “(- (--- -(((((---)))))- ---)”)。さらに，ゲノム対称性もGC skewのシフトポイントとの差は約18°程度であった (表3-23)。

表3-23：Elusimicrobiaに対する予測結果

Proteobacteria		Firmicutes		Actinobacteria		Chlorobi		Chlamydiae	
Score	E-value	Score	E-value	Score	E-value	Score	E-value	Score	E-value
8.7	3.40E-04	9.1	1.10E-04	10.6	2.50E-04	11.5	1.10E-04	13.4	7.40E-05
Firmicutes, Chlamydiae result		dif angle	GC angle	GCSI	Δ	Position	GC (%)		
		20.894	2.001	0.178	18.893	917173	39.952		

太文字下線：Scoreが高く，E-valueが低かったプロファイル

Δ：dif angleとGC angleの差

Tenericutes (1種・1染色体：*Acholeplasma laidlawii* PG-8A) で Proteobacteria, Firmicutes, Actinobacteria, ChlorobiそしてChlamydiaeのプロファイル隠れマルコフモデルを用いて予測したところ，全てのプロファイルにおいてScoreが高く，その中でもFirmicutesとChlamydiaeのプロファイルは同じ配列を予測し (AGTGCCTATAATGTATATTATGTTAACC)，かつパリンδροーム構造 “(- (--- -(((((-))))))- ---)”)を取っていた (表3-24)。

表3-24：Tenericutesに対する予測結果

Proteobacteria		Firmicutes		Actinobacteria		Chlorobi		Chlamydiae	
Score	E-value	Score	E-value	Score	E-value	Score	E-value	Score	E-value
18.5	5.30E-06	22	4.70E-07	15.7	3.60E-05	12.5	2.10E-04	21.6	6.30E-07
Firmicutes, Chlamydiae result		<i>dif angle</i>	<i>GC angle</i>	<i>GCSI</i>	Δ	<i>Position</i>	<i>GC (%)</i>		
		7.680	7.718	0.256	0.038	716560	31.928		

太文字下線：Scoreが高く、E-valueが低かったプロファイル

Δ ：*dif angle*と*GC angle*の差

Spirochaetes (1種・1染色体：*Brachyspira hyodysenteriae* WA1) で Proteobacteria, Firmicutes, Actinobacteria, ChlorobiそしてChlamydiaeのプロファイル隠れマルコフモデルを用いて予測したところ、Tenericutesと同じく全てのScoreが高かった。中でもFirmicutesとChlamydiaeは同じ配列を示し (TATTCCGATAATGTGTATTATGTTAACT)，かつパリンδροーム構造 “--(((---)))--” を取っていた (表3-25)。

さらに*Spirochaetes*では6種・12染色体のゲノムが読まれているため、*B. hyodysenteriae*同様にFirmicutesのプロファイル隠れマルコフモデルを用いて予測を行ったところ、全種でScoreは閾値に収まっており、パリンδροーム構造もある程度確認された (表S1)。

表3-25：Spirochaetesに対する予測結果

Proteobacteria		Firmicutes		Actinobacteria		Chlorobi		Chlamydiae	
Score	E-value	Score	E-value	Score	E-value	Score	E-value	Score	E-value
21.3	7.80E-07	20.4	1.40E-06	18.8	4.30E-06	13.8	1.30E-04	22.5	3.30E-07
Firmicutes, Chlamydiae result		<i>dif angle</i>	<i>GC angle</i>	<i>GCSI</i>	Δ	<i>Position</i>	<i>GC (%)</i>		
		14.695	3.753	0.166	10.941	497170	27.063		

太文字下線：Scoreが高く、E-valueが低かったプロファイル

Δ ：*dif angle*と*GC angle*の差

Planctomycetes (1種・1染色体：*Rhodopirellula baltica* SH 1) でも同様に Proteobacteria, Firmicutes, Actinobacteria, ChlorobiそしてChlamydiaeのプロファイル隠れマルコフモデルを用いて予測したが、ScoreとE-valueが閾値に収まる候補は検出されず、どの配列もパリンδροーム構造を取っていなかった。また、総当たり予測を行った結果も芳しくなかった (表3-26)。

表3-26：Planctomycetesに対する予測結果

Proteobacteria		Firmicutes		Actinobacteria		Chlorobi		Chlamydiae	
Score	E-value	Score	E-value	Score	E-value	Score	E-value	Score	E-value
2.4	1.50E-03	-0.2	1.90E-03	10.2	9.60E-05	4.5	4.30E-03	6.2	5.10E-04

3.3.3 *dif*配列とGC skewの位置関係

GC skewのシフトポイントと*dif*配列の位置を比較したところ、ProteobacteriaやFirmicutesで強い相関が見られ、その他の門でもある程度の相関を観察する事ができた(図3-10).

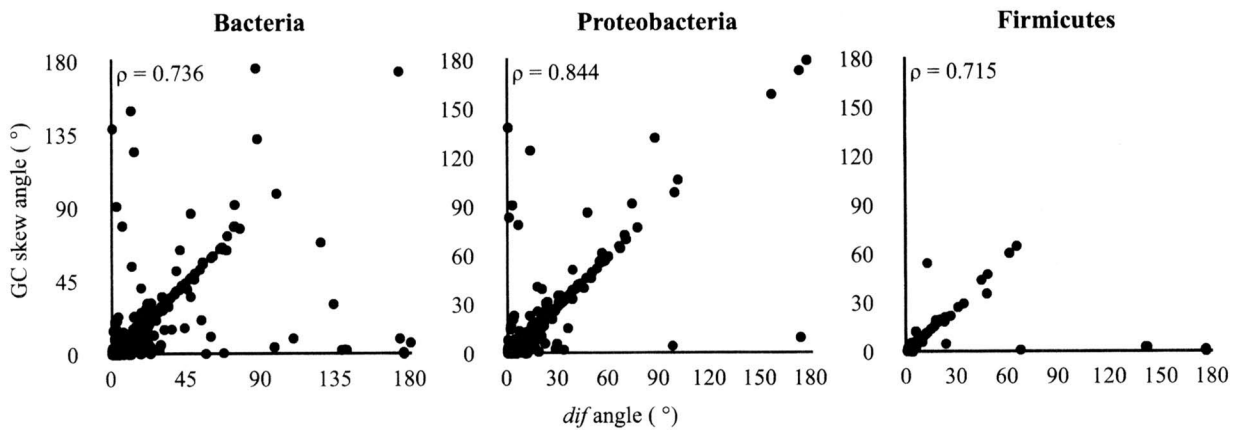


図3-10：*dif*配列とGC skewシフトポイントの関係

X軸に*dif*配列の位置、Y軸にGC skewシフトポイントの位置を示している。対象生物は*dif*配列が予測され、かつGCSIが0.05以上の種で、左図からBacteria全種(517染色体)、Proteobacteria(438染色体)、Firmicutes(97染色体)となっている。なお図中の ρ は順位相関係数を意味する。

またGC skewのシフトポイントと*dif*配列の位置の差のヒストグラムを作成した所、ほぼ全てが $0^{\circ}\sim 5^{\circ}$ 付近であったが、 10° 近い差がいくつかの種で観察された(図3-11)。さらに、この差が何によってもたらされているのかを解明すべく、GCSIとの相関を計算した(図3-12)。その結果、どの門においても相関を見いだすことができなかった。

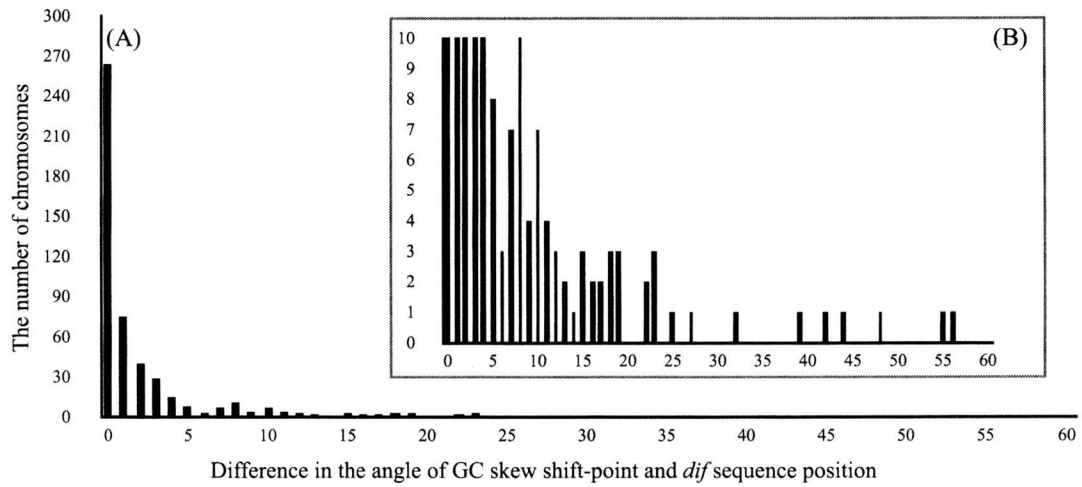


図3-11：*dif*配列とGC skewシフトポイントの差

(A) X軸に*dif*配列の位置，Y軸にその染色体数を示す。(B) X軸とY軸は同様で，角度の差を60°までに限定した(A)の拡大図。

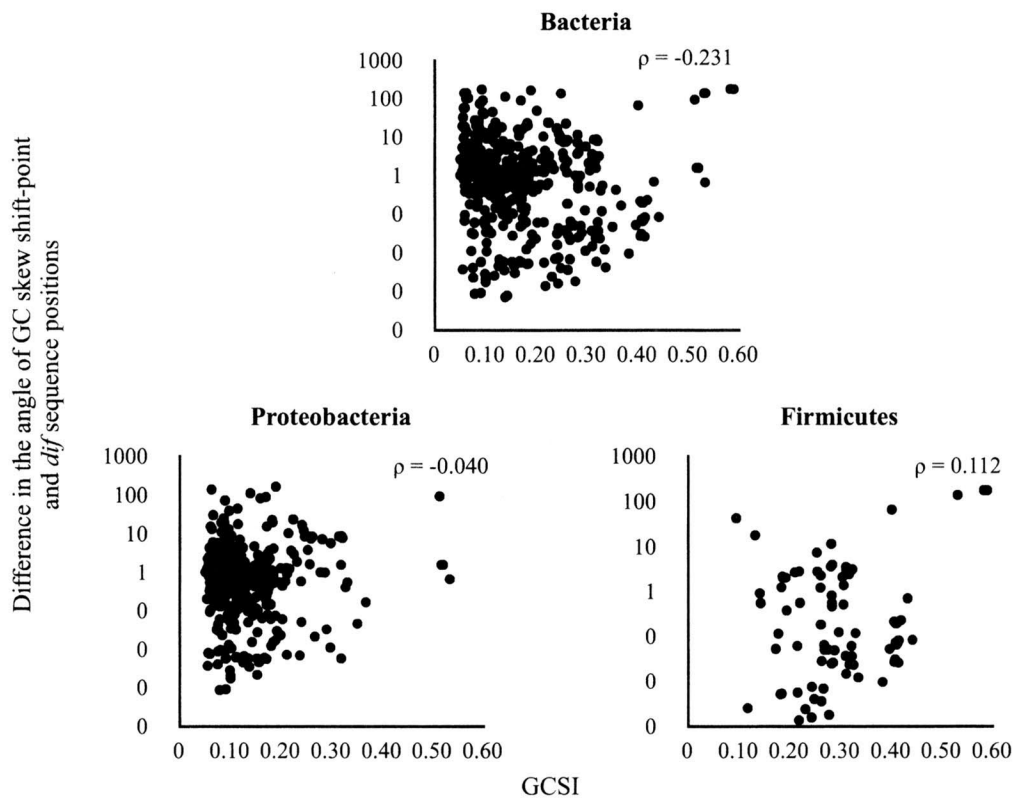


図3-12：GCSIとGC skewシフトポイントの関係

GCSI値に対してGC skewシフトポイントと*dif*配列の位置との差の相関を表した図。X軸にGC skewの強さを表すGCSI値，Y軸に*dif*配列とGC skewシフトポイントの差を示す。なお図中の ρ は順位相関係数を意味する。

3.4 考察

XerCDの門毎の保存はRHMMを用いるときの有用な指標となった。*dif*配列を標的とするXerCDの類似性は*dif*配列の類似性につながる可能性があり、RHMMは予測した配列を元に再帰的に次の予測を行うため、特に配列の似た集団で予測する際に効果を発揮する。

HMMERによって予測された*dif*配列の中でScoreが10以下、E-valueが1.0E-4以上の配列は構造や位置、組成がおよそ正しいとは思えない配列が多く、またこの閾値で*dif*配列は2つ以上検出されず、次の候補はScoreやE-valueが閾値に及ばなかった。しかしながら実際の予測において、閾値に収まる*dif*候補がひとつも検出されなかったケースもある。その場合はある程度ヒューリスティックな手法を用いた。*dif*配列は完璧ではないが、7-12番目と17-22番目付近にパリンドローム構造を持っており、予測に成功した*dif*配列のほぼすべてで観察された。そのため、閾値に達していない候補（とはいえScore \geq 7-9）の中からパリンドローム構造が観察できた配列を予測された*dif*配列とした。

さらに、RHMMによって予測された*dif*配列の位置はGC skewのシフトポイントと似た位置に存在している傾向が強かった。それだけでなく、*dif*-XerCDシステムを考慮すると、機能している*dif*配列が複製開始点付近にある場合、その生物にとって増殖効率を著しく低下させてしまう可能性がある (Cornet et al. 1996)。そのため、極端に複製開始点に近い位置 (例えば100°以上) にいる候補はたとえパリンドローム構造を持っていても候補からははずすようにした。なおGC skewと*dif*配列の位置を議論するうえで、この恣意的なバイアスがかかっている種の考察は避けた。

以上の結果、ヒューリスティックな部分も含め、本研究では上記の根拠により妥当であると考えられる*dif*配列を新規に648種・722染色体予測することに成功した。なお我々はこれら予測された*dif*配列のデータと、*dif*配列の位置を環状染色体上にマッピングした画像をデータベースとして公開しており、<http://www.g-language.org/data/repter/>にて自由に利用することができる (図3-13)。

3.4.1 *dif*配列予測

対象としたほぼ全ての種に置いて*dif*配列を新規に得ることができた結果は、ほぼ全てのXerCDを持つ種で*dif*-XerCDシステムが進化的に強く保存されてきていることを示唆した。さらに同じ生物であっても、複数の染色体を持つ場合、染色体ごとに*dif*配列は若干の違いを見せていた。この違いは*Escherichia*属間での*dif*配列の違いより大きかった (図3-14)。そのため染色体を2つ以上もつ生物における染色体毎の変異バイアスには差があるのかもしれない。

さらに予測された全ての*dif*配列を観察した所、特にFirmicutesの*dif*配列が広く保存されていることが示唆された。既知の*dif*配列がない門では、プロファイル隠れマルコフモデルを作成することができないため、その門での予測はProteobacteriaなど*dif*配列がすでに解明されている門のプロファイルを用いた。その結果Chlorobi, Acidobacteria, Gemmatimonadetes, Nitrospirae, Elusimicrobia, TenericutesそしてSpirochaetesらでFirmicutesのプロファイル隠れマルコフモデルによる予測が最も高いScoreを示していた。

しかしながら一方で、以下の種のようにXerCDを持っているにも関わらず、*dif*配列を予測できなかった種も存在する。Proteobacteriaで予測できなかった種は全5種・7染色体であり、単染色体の種では

Nitratiruptor sp. SB155-2 (NC_009662) 及び, *S. denitrificans* DSM 1251 (NC_007575) であり, 共にε-Proteobacteriaだった. 他のε-Proteobacteriaでは見つかったことからまだ定かではないが, ε-ProteobacteriaではXerCDとは違うCDR機構が保存されている可能性が示唆されているため (Carnoy and Roten 2009), それに起因する可能性がある. また複合染色体の種で予測できなかった種は*P. denitrificans* PD1222 1番染色体 (NC_008686), *P. denitrificans* PD1222 2番染色体 (NC_008687), *B. phytofirmans* PsJN 1番染色体 (NC_010681), *Burkholderia* sp. 383 1番染色体 (NC_007510) そして *Burkholderia* sp. 383 3番染色体 (NC_007509) であった. この内*B. phytofirmans* PsJN 1番染色体 (NC_010681), *Burkholderia* sp. 383 1番染色体 (NC_007510) そして*Burkholderia* sp. 383 3番染色体 (NC_007509) の他の染色体では*dif*配列が発見されているため, 通常は同種内の染色体であることから, 配列の相同検索によって類似した配列を見つけられる可能性があると考えられる. しかしながら図3-14からも, 同種内の染色体同士の*dif*配列は類似していないという結果が得られており, また実際に検索を行ったところ, 特に類似した配列を検出することはできなかった.

また, Firmicutesでは*L. helveticus* DPC 4571 (NC_010080), *C. perfringens* str. 13 (NC_003366) そして*C. beijerinckii* NCIMB 8052 (NC_009617) の*dif*配列を予測する事ができなかった. この内*L. helveticus* DPC 4571はXerCDの他にXerSを持っており, 先行研究でXerSをもつ種は*dif*-XerCDシステムとは違う特殊なCDR機構を持っている可能性が示唆されているため, それに起因すると考えられる. 一方で*C. perfringens* str. 13と*C. beijerinckii*ではGC含有量が低いという傾向が見られた (28.57%, 29.86%). この特徴も他の種のGC含有量に対して顕著に低いという訳ではないが, *dif*配列がATリッチであることから, この様な生物はRHMMでは予測しにくい可能性がある.

これら門の一部の種で*dif*配列を予測する事が出来なかった種に対して, Cyanobacteriaでは全く*dif*配列を予測する事ができなかった. その原因として考えられる仮説としてマルチコピー染色体の存在が考えられる. Cyanobacteriaの中でも*Synechocystis* sp. PCC 6803 (NC_000911) は1つの細胞中に, 12コピーの染色体がある事が確認されている (Labarre et al. 1989). そのため複数のコピーが既に存在している事から, CDR機構を持っていなかったとしてもその種の増殖効率に劇的な影響が出る可能性は低く, それにより*dif*-XerCDシステムを保存する進化的な圧力が比較的低くなった可能性がある.

本研究ではXerCDを持っている種を予測対象としたが, XerCDがない種などでは*dif*-XerCDシステム以外によってCDRが機能している可能性がある. Le Bourgeoisらの先行研究によると, StreptococciやLactococciではXerCDの代わりにXerSがCDRの機能に携わっており, 独自の*dif*配列が観察されている (Le Bourgeois et al. 2007). また同様に, Carnoyらもε-ProteobacteriaではXerHと彼らが名付けたXerCDやXerS以外の酵素が存在している可能性を示唆しており, それらは特有の*dif*配列を持っているのではないかと議論している (Carnoy and Roten 2009). そのためXerHの存在を確認すると共に, 新たにCDR機構に携わる酵素の存在の可能性を念頭に置いた上で, XerCDを持たない種での*dif*配列予測を行なう必要があるだろう.

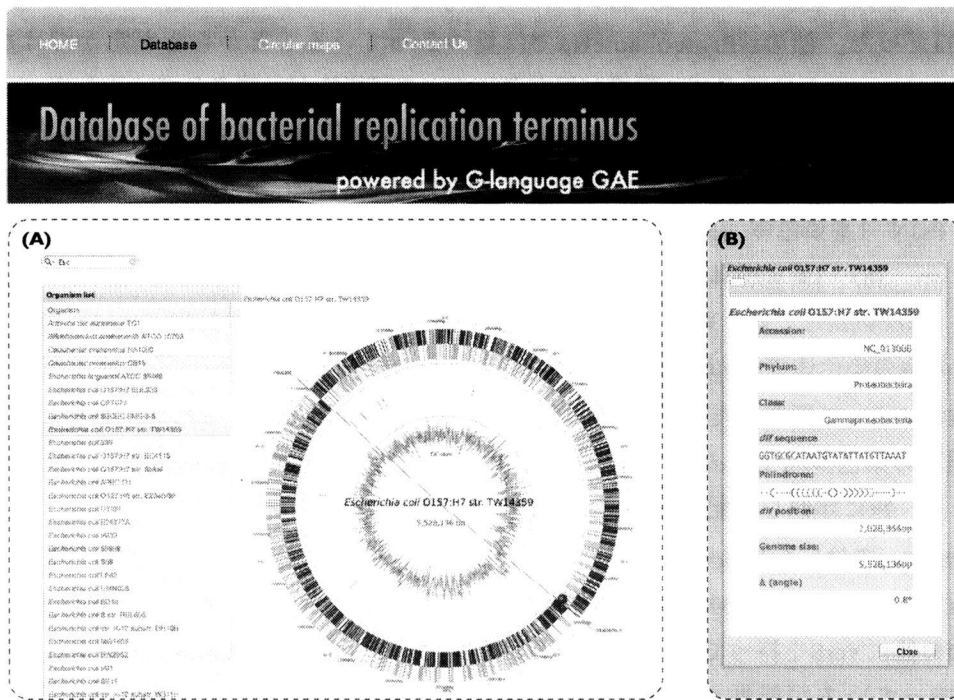


図3-13：データベースインターフェイス

公開しているデータベースの概略図。(A) 検索ボックスに種名を入力するとインクリメンタルサーチによってOrganism listに対象の種名が表示される。そして表示されている種名を選択すると、リストの右側に対象の環状マップが描画される。環状マップ上には種名や塩基座標の他に、GC skew (赤い線)、リーディング鎖上の遺伝子 (青)、ラギング鎖上の遺伝子 (オレンジ)、そして予測されたdif配列の位置 (赤いピン) が示されている。(B) 種を選択するとこのサブウィンドウが開き、そのウィンドウにはアクセッション番号、門名、綱名、dif配列、dif配列のパリンドローム構造、dif配列の位置、ゲノムサイズ、そしてdif配列の位置とGC skewの示す複製終結点の位置の差が載せられている。

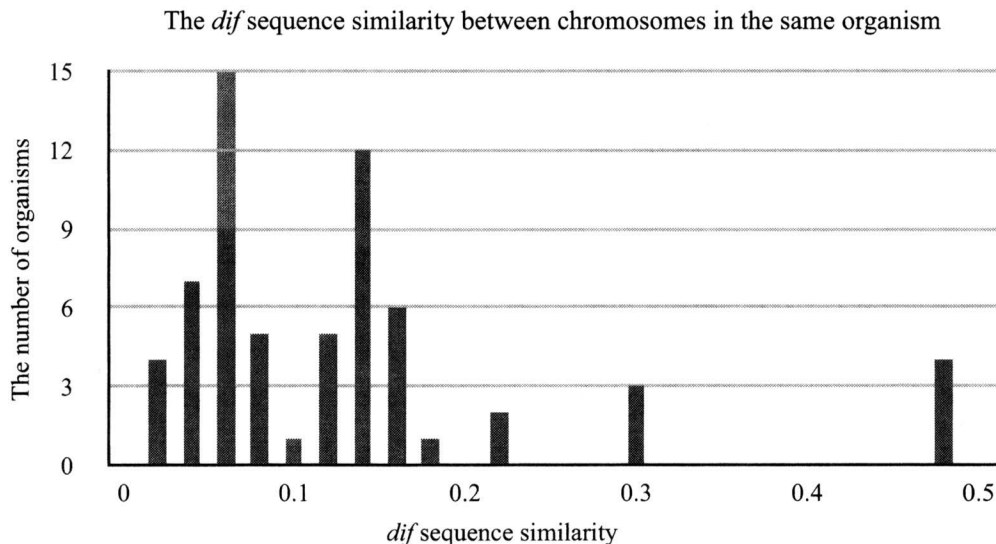


図3-14：複数染色体の配列距離とEscherichia属同士の配列距離

各種の1番染色体のdif配列と2番染色体のdif配列の距離をCLUSTALWによって求めたヒストグラム。青色のグラフはEscherichia属同士のdif配列の距離を示しており、全体的にEscherichia属同士のdif配列に比べて、複数染色体のdif配列の方がdif配列に多くの変異が見られた。

3.4.2 複製終結点と *dif*配列の関係

本研究の結果より、GC skewのシフトポイントと *dif*配列はほぼ同じ場所に位置している事が解明され、また比較的GC skewが正確に観察される種での *dif*配列予測精度が高かったことからGC skewと関係深い事が推察される (図3-15).

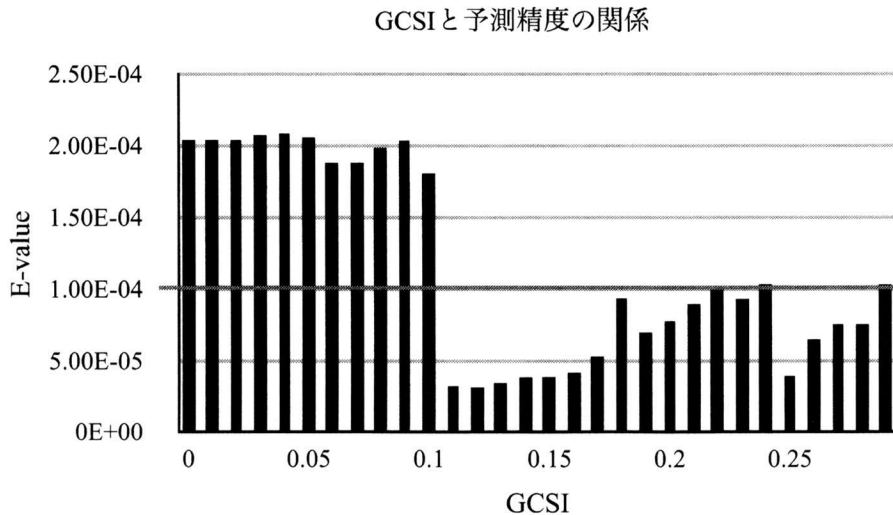


図3-15：GCSIと予測精度の関係

X軸にGCSI値、Y軸にE-valueを取ったグラフ。GCSI値が高くなるにつれてE-valueが低くなっている。GCSI値が0.1以上の全ての種では、今回設けた閾値 (E-value値1.00E-04) に収まっている (赤線).

GC skewは複製の歴史を表現しているという事から、複製の終結に関する以下の2つの仮説を立てることができる。まず1つ目は、複製の終結は単純に複製フォークの衝突によるものであり、ある程度先行研究によって複製開始点と対称の領域付近に絞られてきたが、ほぼランダムであるという仮説である。前述の通り複製が繰り返される事で塩基に変異が蓄積され、グアニンとシトシンの組成に偏りが生じ、GC skewのような塩基バイアスが観察されると考えられている。そのため、塩基バイアスが強く観察される種は複製が同じ様な場所で終結を迎えていると推測できる。またその塩基バイアスはGC skewで表す事ができ、その強さはGCSIという指標で定量的に求める事ができる。一方で、複製の終結がランダムに引き起こされていると仮定すると、中心極限定理によりある程度その終結点の平均的な位置が次第に確定してくると考えられる。すると、複製が繰り返されれば繰り返されるほどGC skewのシフトポイントと *dif*配列の位置の差は小さくなり、複製があまり繰り返されていない種ではGC skewのシフトポイントと *dif*配列の位置の差が大きくなっていくはずである。つまりこの仮説は、GC skewのシフトポイントと *dif*配列の位置の差は、GCSIと相関関係にあるという命題と同義であると考えられる。しかしながら図3-12の通りその関係に相関を見いだすことができなかつたため、この命題は偽であり、同時に複製はランダムに終結している訳ではないと言えるだろう。

もう1つ考えられる仮説は先のものとは対照的で、複製は *dif*配列付近で終結しているという仮説である。環状ゲノムを持つ細菌は複製時に二量体DNAを避ける為にCDR機構を持っており、CDRに関わる *dif*配列の欠損が増殖効率を下げてしまう事や、このCDRを効率良く機能させる為にも複製は *dif*配列付近で終結する事が望ましいと考えられる。これまでの研究で具体的な議論はされてきていないが、GC skewシフトポイントが複製開始点と対称に位置している事は言われて来ていた。しかしながら

この現象は複製開始点と対称を目指して複製が終結して来ていたというより、*dif*配列が対称の位置にある事を示していただけなのかもしれない。それはこれまで網羅的にバクテリアの*dif*配列予測が行われていなかったため検証できなかったためとも考えられる。実際、本論文で初めて網羅的な*dif*配列予測を行ない得られたデータを用いた解析により、*dif*配列とGC skewのシフトポイントの相関関係を見出す事に成功した。以上のことから、本論文ではバクテリアの長い歴史の中で幾度と繰り返されている複製は、*dif*-XerCDシステムでCDRを行うために、恒常的に*dif*配列周辺で複製を終結してきた可能性が示唆できた。

第4章

議論

4.1 GC skewのゆらぎ

第2章ではバクテリア環状ゲノムに現れる、塩基組成バイアスの対称性のゆらぎの原因解明を目的とした解析を行った。外来の塩基組成を原因として仮定し、水平伝播によって挿入された領域を取り除いたGC skewなどを、バクテリア300種近くで観察した結果、水平伝播はGC skewに影響を与えていない事が確認された。この際塩基組成に影響を与えているGIsとして、含むCDSの数とGC skewシフトポイントの変化量との関係も解析したが、それらに相関関係を見いだすことはできなかった。しかしながらこの解析ではCDSの数しか考慮していないため、例えばコドン使用頻度から導きだされるCAI (Codon Adaptation Index: Sharp and Li 1987) 値や発現量データを用いる事で、宿主ゲノムとかけ離れた塩基組成を持つGIsの重み付けができるだろう。また水平伝播によって移動して来た時期を系統的にクラスタリングすることで、同じ由来のGIsを持つ生物をカテゴライズすることができる (Gogarten et al. 2002)。そのため、カテゴライズされた生物を属毎に観察し、バクテリア全体での傾向ではなく、種ごとに対称性の変化量を観察することで、より局所的なケーススタディを議論する事が可能となるだろう。

また一方でこのゆらぎの原因としてもう1つ、逆位変異という要素が考えられる。逆位変異イベントはバクテリアの進化において度々起こっており、*in vitro*での逆位実験で増殖効率との関係が数多く示唆されている (Segall et al. 1988; Rebollo et al. 1988; Campo et al. 2004; Lesterlin et al. 2005; Kuroki et al. 2008)。また近年この逆位変異と増殖効率の関係は、環状ゲノム上のマクロドメイン (MacroDomain: MD) という領域と関係している事が示唆される様になった (Esnault et al. 2007)。例えば*E. coli*の染色体は複製開始点を含むOri MD, Replichoreそれぞれに存在するTer領域を含むRightとLeft, そして複製終点を含むTer MDという4つのマクロドメイン領域に分割する事ができ (Niki et al. 2000; Valens et al.

2004), このマクロドメインを跨いだ逆位変異が特に増殖効率に影響を与えていると言われている (Esnault et al. 2007). このため, バクテリアはある程度自然発生的に逆位変異が引き起こされたとしても, 増殖効率に影響が出る事から選択的な進化圧によって淘汰されているはずである. つまりGIsとは違い, 逆位変異は塩基組成の変化以外にあまり影響を与えていない可能性がある. そこで同種の株同士でドットプロット法を用いて逆位変異領域を推定し, その領域を取り除いた *in silico* 変異株を生成し, その変異株のGC skewのシフトポイントを野生株と比較してどれ程の変化量が現れるかを解析する事ができるだろう. またこれは *E. coli* や *Salmonella enterica* の様に同種株のゲノムが読まれている種は現在30種以上いるため, 検証する上で必要なだけの解析を行なう事は可能であると思われる. この解析により, 塩基組成の揺らぎの原因として逆位変異であるという結論を示唆できる可能性があり, また反対に原因でないとすれば, 後に記述する塩基組成の揺らぎは *dif* 配列に完全に依存しているという考察の裏付けにもなるだろう.

4.2 アーキアにおける *dif* 配列

第3章では塩基組成の対称性と複製機構の関係を解明する事を目的として, まず網羅的に *dif* 配列を予測し, その位置と塩基組成バイアスであるGC skewのシフトポイントの位置を比較解析した. その結果, XerCDを持つほぼ全ての環状ゲノムを持つバクテリアで未知の *dif* 配列予測に成功し, 複製が *dif* 配列付近で終結して来ている事を示唆できた.

本研究での *dif* 配列予測では, バクテリアだけでなくアーキアであるCrenarchaeotaやElusimicrobiaでも数種発見することができた. これら *dif* 配列が発見されたアーキアのGCSIは0.04~0.9と分散が高く, 発見された *dif* 配列の位置はGC skewのシフトポイントに近いわけではなかった. つまり, 特徴としてはバクテリアと酷似している訳ではないが, 特別な組成であったという訳でもない. しかしながら, パリンドローム構造やScoreが *dif* 配列である確からしさを示していた. アーキアはバクテリア, 真核生物と並ぶドメインの1つであるが (Woese et al. 1990), 複製機構に関しては, 真核生物に似た酵素やバクテリアに似た酵素を持っており (Lindås et al. 2008; Samson et al. 2008), その具体的な機構は明確に解明されていない (Vas and Leatherwood 2000). そのため, *dif* 配列に似た配列が本研究で得られた事は, アーキアの染色体分離機構の存在を発見する手がかりになるかもしれない.

4.3 総括

本論文では塩基組成バイアスのゆらぎと複製終結点と *dif* 配列の関係という主に2つのテーマを議論してきた. 第2章によって, 塩基組成バイアスの対称性のゆらぎは水平伝播によって挿入されたと考えられているGIsとはあまり関係がないという結果が得られた. GIsの長さやその位置との相関を見いだすことはできたが, それは塩基組成に影響を与えた結果ではなく, 単純にゲノムの長さや位置の問題であると考えられる. 一方で塩基組成に影響を与えている可能性のあるGIsとして, 含まれるCDSの数を考慮に入れて相関関係を解析したが, 対称性のゆらぎとの相関関係を見いだすことはできず, さらにそもそもGIsを取り除いた時のGC skewシフトポイントの変化量はあまり高くなかったという結果が得られた. その結果第2章ではこの塩基組成バイアスのゆらぎは水平伝播によるものではなく, あるいは何か

別の因子によるものであるという可能性が示唆された。第2章の解析はここで終わったが、この可能性の答えとして考えられる説が後の解析(第3章)によってもたらされる事となった。

第3章の解析で網羅的に予測された*dif*配列の位置と、塩基組成バイアスを表すGC skewシフトポイントの関係を解析したところ、両者間に強い相関を観察する事ができた。これは複製の終結点はほぼ*dif*配列付近を目指している可能性を示唆する結果であると言え、つまりGC skewのシフトポイントは*dif*配列の位置に依存して形成される事と同義であると言える。実際に、GC skewのシフトポイントが極端に偏って観察された*Geobacter* sp. FRC-32 (NC_011979) や*Haemophilus parasuis* SH0165 (NC_011852) のGC skewシフトポイントが77.672°と118.368°であったのに対して、*dif*配列の位置は77.105°と118.850°となっており、ほぼ同じ位置で*dif*配列が予測されていた。したがって、GC skewのシフトポイントに揺らぎが生じた場合、それは外来のDNAによる影響と考えるのではなく、*dif*配列の位置のゆらぎに影響を受けていると考えることができ、第2章の結果はこのような考えを支持することになるだろう。

以上第2章と第3章の解析から導き出された結果より、以下のモデルを提唱する。環状ゲノムを持つ細菌は複製を終結させる際には複製開始点と対称の位置を目標としている訳ではなく、実はCDR機構をより効率良く行うために*dif*配列付近を目標としており、塩基組成の対称性のゆらぎはこの*dif*配列と複製プロセスによって蓄積された結果である。

謝辞

本論文は大変多くの方々のお力添えによって完成を迎えることができ、まずはじめに常に応援してくださった河野家、中嶋家、遠藤家、花田家そして加藤家の人々に感謝申し上げます。

慶應義塾大学環境情報学部の野崎慎氏、同藤本誠也氏、同海老敬行氏、同浜島聖文氏にはユーモア溢れる研究生生活をサポートして頂きました。またアドバイザーである同木戸信博氏や、本文の細部にわたるチェックをして頂いたアドバイザーの同池上慶太氏と同小川（旧姓：中川）真菜氏、さらに彼らをはじめとするグループメンバーの同大下和希氏、同長谷部百合子氏、同小峰栞氏らとは常に活気溢れた議論をさせて頂きました。さらにコテージメンバーの方々をはじめとするプログレスセミナーの方々の意欲、同政策・メディア研究科博士課程の菅原潤一氏のリーダーシップ、同西野泰子氏の人柄、同先端生命科学研究所の谷内江望氏の才知に富んだ考察、同政策・メディア研究科修士課程の中村浩之氏の多岐に渡る技術知識、同武内麻里亜氏による心温まるサポート、同小川隆氏の秀逸なひらめきと発展的な議論、同環境情報学部の斎藤輪太郎講師の幅広い知識とアドバイス、そして同先端生命科学研究所の内藤泰宏準教授の機知に溢れた意見には大変助けて頂きました。特に同環境情報学部の田中美穂氏による支えは本論文を書き上げる上で必要不可欠であり、特に彼女の熱意と行動力には常に刺激されておりました。また同板谷光泰教授にはバクテリアを扱う上で、実験生物学者の立場から様々な有益かつ斬新なアドバイスを頂き、さらに板谷教授との議論は大変興味深く、常に大きな刺激と衝撃を与えて頂きました。そして、同政策・メディア研究科の荒川和晴講師には私が学部生の頃からアドバイザーとして面倒をみていただき、常に研究の細部に渡ると的確なアドバイスを賜りました。さらに日頃から様々な議論をさせて頂き、研究に対する好奇心を常に掻き立てられました。誠に僭越ながら研究者としての在り方のお手本とさせて頂いており、公私共に多大なる影響を与えられました。

最後に、本研究を行うに当たりあらゆる面から多大なご支援をしてくださった慶應義塾大学環境情報学部富田勝教授には特に感謝の意を表したいと思います。この恵まれた環境や知慮深い教員と学生の方々との議論は何物にも代え難く、この論文の根底に深く刻まれております。

参考文献

- Alberts B, Johnson A, Walter P, Lewis J, Raff M, Roberts K (2008) *Molecular Biology of the Cell*. 5th Revised edition. Garland Publishing Inc, New York
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403-410
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389-3402
- Arakawa K, Mori K, Ikeda K, Matsuzaki T, Kobayashi Y, Tomita M (2003) G-language Genome Analysis Environment: a workbench for nucleotide sequence data mining. *Bioinformatics* 19:305-306
- Arakawa K, Tomita M (2006) G-language System as a platform for large-scale analysis of high-throughput omics data. *Journal of Pesticide Science* 31:282-288
- Arakawa K, Saito R, Tomita M (2007) Noise-reduction filtering for accurate detection of replication termini in bacterial genomes. *FEBS Lett* 581:253-258
- Arakawa K, Tomita M (2007) The GC Skew Index: a measure of genomic compositional asymmetry and the degree of replicational selection. *Evol Bioinform Online* 3:159-168

- Arakawa K, Suzuki H, Tomita M (2008) Computational Genome Analysis Using The G-language System. *Genes, Genomes and Genomics* 2:1-13
- Arakawa K, Suzuki H, Tomita M (2009a) Quantitative analysis of replication-related mutation and selection pressures in bacterial chromosomes and plasmids using generalised GC skew index. *BMC Genomics* 10:640
- Arakawa K, Tamaki S, Kono N, Kido N, Ikegami K, Ogawa R, Tomita M (2009b) Genome Projector: zoomable genome map with multiple views. *BMC Bioinformatics* 10:31
- Aussel L, Barre FX, Aroyo M, Stasiak A, Stasiak AZ, Sherratt D (2002) FtsK is a DNA motor protein that activates chromosome dimer resolution by switching the catalytic state of the XerC and XerD recombinases. *Cell* 108:195-205
- Bigot S, Corre J, Louarn JM, Cornet F, Barre FX (2004) FtsK activities in Xer recombination, DNA mobilization and cell division involve overlapping and separate domains of the protein. *Mol Microbiol* 54:876-886
- Bigot S, Saleh OA, Lesterlin C, Pages C, El Karoui M, Dennis C, Grigoriev M, Allemand JF, Barre FX, Cornet F (2005) KOPS: DNA motifs that control E. coli chromosome segregation by orienting the FtsK translocase. *EMBO J* 24:3770-3780
- Bigot S, Sivanathan V, Possoz C, Barre FX, Cornet F (2007) FtsK, a literate chromosome segregation machine. *Mol Microbiol* 64:1434-1441
- Blakely G, May G, McCulloch R, Arciszewska LK, Burke M, Lovett ST, Sherratt DJ (1993) Two related recombinases are required for site-specific recombination at dif and cer in E. coli K12. *Cell* 75:351-361
- Blakely G, Sherratt D (1996) Determinants of selectivity in Xer site-specific recombination. *Genes Dev* 10:762-773
- Blakely GW, Sherratt DJ (1994) Interactions of the site-specific recombinases XerC and XerD with the recombination site dif. *Nucleic Acids Res* 22:5613-5620
- Brewer BJ (1988) When polymerases collide: replication and the transcriptional organization of the E. coli chromosome. *Cell* 53:679-686
- Bussiere DE, Bastia D (1999) Termination of DNA replication of bacterial and plasmid chromosomes. *Mol Microbiol* 31:1611-1618

- Campo N, Dias MJ, Daveran-Mingot ML, Ritzenthaler P, Le Bourgeois P (2004) Chromosomal constraints in Gram-positive bacteria revealed by artificial inversions. *Mol Microbiol* 51:511-522
- Capiaux H, Cornet F, Corre J, Guijo MI, Pérals K, Rebollo JE, Louarn JM (2001) Polarization of the *Escherichia coli* chromosome. A view from the terminus. *Biochimie* 83:161-170
- Carnoy C, Roten CA (2009) The dif/Xer recombination systems in proteobacteria. *PLoS One* 4:e6531
- Cornet F, Louarn J, Patte J, Louarn JM (1996) Restriction of the activity of the recombination site dif to a small zone of the *Escherichia coli* chromosome. *Genes Dev* 10:1152-1161
- Corre J, Louarn JM (2002) Evidence from terminal recombination gradients that FtsK uses replicore polarity to control chromosome terminus positioning at division in *Escherichia coli*. *J Bacteriol* 184:3801-3807
- Crick FH (1970) Central dogma of molecular biology. *Nature* 227:561-563
- Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14:1188-1190
- de Massy B, Béjar S, Louarn J, Louarn JM, Bouché JP (1987) Inhibition of replication forks exiting the terminus region of the *Escherichia coli* chromosome occurs at two loci separated by 5 min. *Proc Natl Acad Sci U S A* 84:1759-1763
- Dorazi R, Dewar SJ (2000) Membrane topology of the N-terminus of the *Escherichia coli* FtsK division protein. *FEBS Lett* 478:13-18
- Draper GC, McLennan N, Begg K, Masters M, Donachie WD (1998) Only the N-terminal domain of FtsK functions in cell division. *J Bacteriol* 180:4621-4627
- Dubarry N, Barre FX (2009) Fully efficient chromosome dimer resolution in *Escherichia coli* cells lacking the integral membrane domain of FtsK. *EMBO J* [Epub ahead of print]
- Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14:755-763
- Ellegren H (2008) Comparative genomics and the study of evolution by natural selection. *Mol Ecol* 17:4586-4596
- Esnault E, Valens M, Espéli O, Boccard F (2007) Chromosome structuring limits genome plasticity in *Escherichia coli*. *PLoS Genet* 3:e226
- Francino MP, Chao L, Riley MA, Ochman H (1996) Asymmetries generated by transcription-coupled repair in enterobacterial genes. *Science* 272:107-109

- Francino MP, Ochman H (1997) Strand asymmetries in DNA evolution. *Trends Genet* 13:240-245
- Frank AC, Lobry JR (2000) Oriloc: prediction of replication boundaries in unannotated bacterial chromosomes. *Bioinformatics* 16:560-561
- Frick DN, Richardson CC (2001) DNA primases. *Annu Rev Biochem* 70:39-80
- Gaillard M, Vallaeyts T, Vorhölter FJ, Minoia M, Werlen C, Sentschilo V, Pühler A, van der Meer JR (2006) The *clc* element of *Pseudomonas* sp. strain B13, a genomic island with various catabolic properties. *J Bacteriol* 188:1999-2013
- Gogarten JP, Doolittle WF, Lawrence JG (2002) Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* 19:2226-2238
- Gogarten JP, Townsend JP (2005) Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol* 3:679-687
- Grigoriev A (1998) Analyzing genomes with cumulative skew diagrams. *Nucleic Acids Res* 26:2286-2290
- Hacker J, Bender L, Ott M, Wingender J, Lund B, Marre R, Goebel W (1990) Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extraintestinal *Escherichia coli* isolates. *Microb Pathog* 8:213-225
- Hacker J, Kaper JB (2000) Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* 54:641-679
- Hanawalt PC (1991) Heterogeneity of DNA repair at the gene level. *Mutat Res* 247:203-211
- Hansen FG, Christensen BB, Nielsen CB, Atlung T (2006) Insights into the quality of DnaA boxes and their cooperativity. *J Mol Biol* 355:85-95
- Hendrickson H, Lawrence JG (2006) Selection for chromosome architecture in bacteria. *J Mol Evol* 62:615-629
- Hendrickson H, Lawrence JG (2007) Mutational bias suggests that replication termination occurs near the *dif* site, not at *Ter* sites. *Mol Microbiol* 64:42-56
- Higgins NP (2007) Mutational bias suggests that replication termination occurs near the *dif* site, not at *Ter* sites: what's the *Dif*? *Mol Microbiol* 64:1-4
- Horiuchi T, Fujimura Y, Nishitani H, Kobayashi T, Hidaka M (1994) The DNA replication fork blocked at the *Ter* site may be an entrance for the RecBCD enzyme into duplex DNA. *J Bacteriol* 176:4656-4663

- Hsiao W, Wan I, Jones SJ, Brinkman FS (2003) IslandPath: aiding detection of genomic islands in prokaryotes. *Bioinformatics* 19:418-420
- Hsiao WW, Ung K, Aeschliman D, Bryan J, Finlay BB, Brinkman FS (2005) Evidence of a large novel gene pool associated with prokaryotic genomic islands. *PLoS Genet* 1:e62
- Jensen RB (2006) Analysis of the terminus region of the *Caulobacter crescentus* chromosome and identification of the dif site. *J Bacteriol* 188:6016-6019
- Jiang Z (2008) Protein function predictions based on the phylogenetic profile method. *Crit Rev Biotechnol* 28:233-238
- Johnson A, O'Donnell M (2005) Cellular DNA replicases: components and dynamics at the replication fork. *Annu Rev Biochem* 74:283-315
- Juhas M, Power PM, Harding RM, Ferguson DJ, Dimopoulou ID, Elamin AR, Mohd-Zain Z, Hood DW, Adegbola R, Erwin A, Smith A, Munson RS, Harrison A, Mansfield L, Bentley S, Crook DW (2007) Sequence and functional analyses of *Haemophilus* spp. genomic islands. *Genome Biol* 8:R237
- Juhas M, van der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW (2009) Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* 33:376-393
- Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2009) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 38:D335-360
- Karlin S (2001) Detecting anomalous gene clusters and pathogenicity islands in diverse bacterial genomes. *Trends Microbiol* 9:335-343
- Kelly TJ, Brown GW (2000) Regulation of chromosome replication. *Annu Rev Biochem* 69:829-880
- Kono N, Arakawa K, Tomita M (2006) MEGU: pathway mapping web-service based on KEGG and SVG. *In Silico Biol* 6:621-625
- Kono N, Arakawa K, Ogawa R, Kido N, Oshita K, Ikegami K, Tamaki S, Tomita M (2009) Pathway projector: web-based zoomable pathway browser using KEGG atlas and Google Maps API. *PLoS One* 4:e7710
- Kuempel PL, Henson JM, Dircks L, Tecklenburg M, Lim DF (1991) dif, a recA-independent recombination site in the terminus region of the chromosome of *Escherichia coli*. *New Biol* 3:799-811

- Kuroki A, Toda T, Matsui K, Uotsu-Tomita R, Tomita M, Itaya M (2007) Reshuffling of the *Bacillus subtilis* 168 genome by multifold inversion. *J Biochem* 143:97-105
- Labarre J, Chauvat F, Thuriaux P (1989) Insertional mutagenesis by random cloning of antibiotic resistance genes into the genome of the cyanobacterium *Synechocystis* strain PCC 6803. *J Bacteriol* 171:3449-3457
- Langille MG, Hsiao WW, Brinkman FS (2008) Evaluation of genomic island predictors using a comparative genomics approach. *BMC Bioinformatics* 9:329
- Langille MG, Brinkman FS (2009) IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics* 25:664-665
- Larbig KD, Christmann A, Johann A, Klockgether J, Hartsch T, Merkl R, Wiehlmann L, Fritz HJ, Tümmler B (2002) Gene islands integrated into tRNA(Gly) genes confer genome diversity on a *Pseudomonas aeruginosa* clone. *J Bacteriol* 184:6665-6680
- Lawrence JG, Hendrickson H (2003) Lateral gene transfer: when will adolescence end? *Mol Microbiol* 50:739-749
- Le Bourgeois P, Bugarel M, Campo N, Daveran-Mingot ML, Labonté J, Lanfranchi D, Lautier T, Pagès C, Ritzenthaler P (2007) The unconventional Xer recombination machinery of *Streptococci/Lactococci*. *PLoS Genet* 3:e117
- Lesterlin C, Mercier R, Boccard F, Barre FX, Cornet F (2005) Roles for replicohores and macrodomains in segregation of the *Escherichia coli* chromosome. *EMBO Rep* 6:557-562
- Li X, Mariani KJ (2000) Two distinct triggers for cycling of the lagging strand polymerase at the replication fork. *J Biol Chem* 275:34757-34765
- Li Y, Youngren B, Sergueev K, Austin S (2003) Segregation of the *Escherichia coli* chromosome terminus. *Mol Microbiol* 50:825-834
- Lindås AC, Karlsson EA, Lindgren MT, Ettema TJ, Bernander R (2008) A unique cell division machinery in the Archaea. *Proc Natl Acad Sci U S A* 105:18942-18946
- Liolios K, Chen IM, Mavromatis K, Tavernarakis N, Hugenholtz P, Markowitz VM, Kyrpides NC (2010) The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 38:D346-354

- Lobry JR (1996a) A simple vectorial representation of DNA sequences for the detection of replication origins in bacteria. *Biochimie* 78:323-326
- Lobry JR (1996b) Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol Biol Evol* 13:660-665
- Lobry JR, Sueoka N (2002) Asymmetric directional mutation pressures in bacteria. *Genome Biol* 3:RESEARCH0058
- Löwe J, Ellonen A, Allen MD, Atkinson C, Sherratt DJ, Grainge I (2008) Molecular mechanism of sequence-directed DNA loading and translocation by FtsK. *Mol Cell* 31:498-509
- Mantri Y, Williams KP (2004) Islander: a database of integrative islands in prokaryotic genomes, the associated integrases and their DNA site specificities. *Nucleic Acids Res* 32:D55-58
- McCulloch SD, Kunkel TA (2008) The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res* 18:148-161
- McLean MJ, Wolfe KH, Devine KM (1998) Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. *J Mol Evol* 47:691-696
- Medvedev P, Stanciu M, Brudno M (2009) Computational methods for discovering structural variation with next-generation sequencing. *Nat Methods* 6:S13-20
- Mellon I, Hanawalt PC (1989) Induction of the *Escherichia coli* lactose operon selectively increases repair of its transcribed DNA strand. *Nature* 342:95-98
- Merkel R (2004) SIGI: score-based identification of genomic islands. *BMC Bioinformatics* 5:22
- Merkel R (2006) A comparative categorization of protein function encoded in bacterial or archeal genomic islands. *J Mol Evol* 62:1-14
- Messer W (2002) The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiol Rev* 26:355-374
- Neilson L, Blakely G, Sherratt DJ (1999) Site-specific recombination at dif by *Haemophilus influenzae* XerC. *Mol Microbiol* 31:915-926
- Neylon C, Kralicek AV, Hill TM, Dixon NE (2005) Replication termination in *Escherichia coli*: structure and antihelicase activity of the Tus-Ter complex. *Microbiol Mol Biol Rev* 69:501-526

- Niki H, Yamaichi Y, Hiraga S (2000) Dynamic organization of chromosomal DNA in *Escherichia coli*. *Genes Dev* 14:212-223
- Noe L, Kucherov G (2005) YASS: enhancing the sensitivity of DNA similarity search. *Nucleic Acids Res* 33:W540–543
- Ogawa T, Okazaki T (1980) Discontinuous DNA replication. *Annu Rev Biochem* 49:421-457
- Okazaki R, Okazaki T, Sakabe K, Sugimoto K, Sugino A (1968) Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc Natl Acad Sci U S A* 59:598-605
- Ou HY, He X, Harrison EM, Kulasekara BR, Thani AB, Kadioglu A, Lory S, Hinton JC, Barer MR, Deng Z, Rajakumar K (2007) MobilomeFINDER: web-based tools for in silico and experimental discovery of bacterial genomic islands. *Nucleic Acids Res* 35:W97-W104
- Pérals K, Capiiaux H, Vincourt JB, Louarn JM, Sherratt DJ, Cornet F (2001) Interplay between recombination, cell division and chromosome structure during chromosome dimer resolution in *Escherichia coli*. *Mol Microbiol* 39:904-913
- Ptacin JL, Nöllmann M, Bustamante C, Cozzarelli NR (2006) Identification of the FtsK sequence-recognition domain. *Nat Struct Mol Biol* 13:1023-1025
- Rebollo JE, François V, Louarn JM (1988) Detection and possible role of two large nondivisible zones on the *Escherichia coli* chromosome. *Proc Natl Acad Sci U S A* 85:9391-9395
- Recchia GD, Sherratt DJ (1999) Conservation of xer site-specific recombination genes in bacteria. *Mol Microbiol* 34:1146-1148
- Rocha EP (2004) The replication-related organization of bacterial genomes. *Microbiology* 150:1609-1627
- Rocha EP (2008) The organization of the bacterial genome. *Annu Rev Genet* 42:211-233
- Salzberg SL, Salzberg AJ, Kerlavage AR, Tomb JF (1998) Skewed oligomers and origins of replication. *Gene* 217:57-67
- Samson RY, Obita T, Freund SM, Williams RL, Bell SD (2008) A role for the ESCRT system in cell division in archaea. *Science* 322:1710-1713

- Schaeffer PM, Headlam MJ, Dixon NE (2005) Protein--protein interactions in the eubacterial replisome. *IUBMB Life* 57:5-12
- Schaper S, Messer W (1995) Interaction of the initiator protein DnaA of *Escherichia coli* with its DNA target. *J Biol Chem* 270:17622-17626
- Sciochetti SA, Piggot PJ, Blakely GW (2001) Identification and characterization of the dif Site from *Bacillus subtilis*. *J Bacteriol* 183:1058-1068
- Segall A, Mahan MJ, Roth JR (1988) Rearrangement of the bacterial chromosome: forbidden inversions. *Science* 241:1314-1318
- Sharp PM, Li WH (1987) The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 15:1281-1295
- Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26:1135-1145
- Sivanathan V, Allen MD, de Bekker C, Baker R, Arciszewska LK, Freund SM, Bycroft M, Löwe J, Sherratt DJ (2006) The FtsK gamma domain directs oriented DNA translocation by interacting with KOPS. *Nat Struct Mol Biol* 13:965-972
- Strick TR, Quessada-Vial A (2006) FtsK: a groovy helicase. *Nat Struct Mol Biol* 13:948-950
- Stukenberg PT, Studwell-Vaughan PS, O'Donnell M (1991) Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. *J Biol Chem* 266:11328-11334
- Stukenberg PT, Turner J, O'Donnell M (1994) An explanation for lagging strand replication: polymerase hopping among DNA sliding clamps. *Cell* 78:877-887
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673-4680
- Trigueros S, Tran T, Sorto N, Newmark J, Colloms SD, Sherratt DJ, Tolmasky ME (2009) mwr Xer site-specific recombination is hypersensitive to DNA supercoiling. *Nucleic Acids Res* 37:3580-3587
- Val ME, Kennedy SP, El Karoui M, Bonné L, Chevalier F, Barre FX (2008) FtsK-dependent dimer resolution on multiple chromosomes in the pathogen *Vibrio cholerae*. *PLoS Genet* 4:e1000201

Valens M, Penaud S, Rossignol M, Cornet F, Boccard F (2004) Macrodomain organization of the *Escherichia coli* chromosome. *EMBO J* 23:4330-4341

Vas A, Leatherwood J (2000) Where does DNA replication start in archaea? *Genome Biol* 1:REVIEWS1020

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X (2001) The sequence of the human genome. *Science* 291:1304-1351

- Vernikos GS, Parkhill J (2008) Resolving the structural features of genomic islands: a machine learning approach. *Genome Res* 18:331-342
- Waack S, Keller O, Asper R, Brodag T, Damm C, Fricke WF, Surovcik K, Meinicke P, Merkl R (2006) Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics* 7:142
- Wang L, Lutkenhaus J (1998) FtsK is an essential cell division protein that is localized to the septum and induced as part of the SOS response. *Mol Microbiol* 29:731-740
- Watson JD, Crick FH (1953) A structure for deoxyribose nucleic acid. *Nature* 171:737-738
- Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* 87:4576-4579
- Worning P, Jensen LJ, Hallin PF, Staerfeldt HH, Ussery DW (2006) Origin of replication in circular prokaryotic chromosomes. *Environ Microbiol* 8:353-361
- Yates J, Zhekov I, Baker R, Eklund B, Sherratt DJ, Arciszewska LK (2006) Dissection of a functional interaction between the DNA translocase, FtsK, and the XerD recombinase. *Mol Microbiol* 59:1754-1766
- Yen MR, Lin NT, Hung CH, Choy KT, Weng SF, Tseng YH (2002) oriC region and replication termination site, dif, of the *Xanthomonas campestris* pv. *campestris* 17 chromosome. *Appl Environ Microbiol* 68:2924-2933
- Yu XC, Tran AH, Sun Q, Margolin W (1998) Localization of cell division protein FtsK to the *Escherichia coli* septum and identification of a potential N-terminal targeting domain. *J Bacteriol* 180:1296-1304
- Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. *J Comput Biol* 7:203-214

Accession	Phylum	Class	Genus	Species	Strain	dif	Palindrome	dif angle	GC angle	GCSI	A	dif Position	GC%
NC_000962	Actinobacteria	Actinobacteridae	Mycobacterium	tuberculosis	H37Rv	TAAGCCGATAAGCGACATTATGTCAGT	----- (CCCC- - - -) - - - -	2.193	13.129	0.081	10.936	2232640	65.614
NC_012943	Actinobacteria	Actinobacteridae	Mycobacterium	tuberculosis	KZN 1435	TAAGCCGATAAGCGACATTATGTCAGT	----- (CCCC- - - -) - - - -	0.714	14.758	0.084	14.044	2190403	65.614
NC_009565	Actinobacteria	Actinobacteridae	Mycobacterium	tuberculosis	F11	TAAGCCGATAAGCGACATTATGTCAGT	----- (CCCC- - - -) - - - -	2.333	12.945	0.082	10.611	2240894	65.618
NC_008611	Actinobacteria	Actinobacteridae	Mycobacterium	ulcerans	Agly99	TGTGCCGATAAGCGACATTATGTCAGT	- (- - - - (CCCC- - - -) - - - -)	21.346	21.343	0.075	0.002	2481886	65.473
NC_010545	Actinobacteria	Actinobacteridae	Corynebacterium	werehlyticum	DSM 7109	TCCGATAATACATATAGCATTGTC	- (- - - - (C - - - -) - - - -)	10.241	4.260	0.079	5.981	1248850	64.185
NC_008726	Actinobacteria	Actinobacteridae	Mycobacterium	vanbaalenii	PYR-1	TTACCGATAAGCCGACATTATGTCAGT	----- (CCCC- - - -) - - - -	7.910	7.731	0.047	0.179	3103284	67.793
NC_013159	Actinobacteria	Actinobacteridae	Saccharomonospora	viridis	DSM 43017	AAGCGCGATAATGATCATTATGTAAGT	----- (CCCC (C - - -) - - - -)	2.934	1.656	0.127	1.278	2102887	67.315
NC_004551	Actinobacteria	Actinobacteridae	Tropheryma	whipplei	str. Twist	ACTGATAATATAATATGACAAITTTAA	----- (C - - - - (C - - - -) - - - -)	73.543	78.862	0.059	5.319	478943	46.314
NC_004572	Actinobacteria	Actinobacteridae	Tropheryma	whipplei	str. Twist	ACTGATAATATAATATGACAAITTTAA	----- (C - - - - (C - - - -) - - - -)	73.799	79.123	0.054	5.324	414339	46.332
NC_010830	Bacteroidetes	Candidatus Amoebohilus	Amoebophilus	asiaticus	5a2	TAGTTATAAATTTATATATGTAAGT	----- (CCCC- (C - - -) - - - -)	5.293	5.294	0.142	0.001	1317733	35.045
NC_013062	Bacteroidetes	Flavobacteriaceae	Flavobacterium	bacterium	3519-10	AAGTCTATAAATTTATATGTAAGT	- (- - - - (CCCC- (C - - -) - - - -)	8.936	8.493	0.077	0.443	1452762	42.693
NC_013222	Bacteroidetes	Flavobacteria	Robiginitalea	biformata	HTCC2501	TTCCTATAAATTTATATGTAAGT	- (- - - - (CCCC- (C - - -) - - - -)	9.376	9.309	0.058	0.067	2517214	55.293
NC_009615	Bacteroidetes	Bacteroidia	Parabacteroides	distansis	ATCC 8503	TCTTGCCATAAGTAAATCATGTATGAT	- (- - - - (CC- (C - - - -) - - - -)	28.588	4.935	0.224	23.653	822255	45.060
NC_013037	Bacteroidetes	Sphingobacteria	Dyadobacter	fermentans	DSM 18053	TGTGTATAAATTTATATGTAAGT	- (- - - - (CCCC- (C - - -) - - - -)	4.483	5.784	0.078	1.301	1212803	51.545
NC_008571	Bacteroidetes	Flavobacteria	Gramella	forsetii	KT0803	TGTACTATAAATTTATATGTAAGT	- (- - - - (CCCC- (C - - -) - - - -)	5.480	5.477	0.157	0.003	1648336	36.611
NC_003228	Bacteroidetes	Bacteroidia	Bacteroides	fragilis	NCTC 9343	ATTTGCTAAATTTAGTTTATGTCATAT	- (- - - - (CCCC- (C - - -) - - - -)	13.342	8.924	0.217	4.417	2795474	43.191
NC_006347	Bacteroidetes	Bacteroidia	Bacteroides	fragilis	YCH46	ATTTGCTAAATTTAGTTTATGTCATAT	- (- - - - (CCCC- (C - - -) - - - -)	2.869	1.487	0.207	1.382	2680696	43.266
NC_002950	Bacteroidetes	Bacteroidia	Porphyromonas	gingivalis	W83	AATTGCCAGATTTTATATGTAAGT	- (- - - - (C- (C - - -) - - - -)	56.497	0.725	0.056	55.772	1544496	48.288
NC_010729	Bacteroidetes	Bacteroidia	Porphyromonas	gingivalis	ATCC 33277	TATCCCATATGATGATCTGATATAT	- (- - - - (C- (C - - -) - - - -)	109.282	9.997	0.061	99.284	284425	48.359
NC_013061	Bacteroidetes	Sphingobacteria	Pedobacter	heparimus	DSM 2366	TAGTCACATAAATTTATATGTAAGT	- (- - - - (CCCC- - - -) - - - -)	6.487	0.777	0.164	5.710	2490571	42.049
NC_008255	Bacteroidetes	Sphingobacteria	Cytophaga	hutchinsonii	ATCC 33406	TGTGTATAAATTTACATATGTAAGT	- (- - - - (CCCC- (C - - -) - - - -)	11.422	10.506	0.127	0.915	1965461	38.846
NC_009441	Bacteroidetes	Flavobacterium	johnsoniae	UW101	TTGACTATAAATTTATATGTAAGT	- (- - - - (CCCC- (C - - -) - - - -)	20.034	16.388	0.120	3.645	3387721	34.113	
NC_013162	Bacteroidetes	Flavobacteria	Capnocytophaga	ochracea	DSM 7271	AATGGTATAATGATATGTCAGT	- (- - - - (CCCC- (C - - -) - - - -)	17.209	23.185	0.077	5.975	2330633	39.588
NC_013132	Bacteroidetes	Sphingobacteria	Chitinophaga	pinensis	DSM 25878	AATACTATAAATTTATATGTAAGT	- (- - - - (CCCC- (C - - -) - - - -)	20.261	28.141	0.134	7.880	5077355	45.227
NC_011565	Bacteroidetes	Bacteroidia	Azobacteroides	pseudotrichonophae	pseudotrichonophae genomovar. CFP2	AATTCTAGAAATGATGCTGCTCATAT	- (- - - - (C- (C - - - -) - - - -)	13.495	23.898	0.247	10.403	892745	32.658
NC_009613	Bacteroidetes	Flavobacteria	Flavobacterium	psychrophilum	JIP02/86	ACGAACATAAATTTATATGTAAGT	- (- - - - (CCCC- (C - - -) - - - -)	40.814	64.551	0.119	23.737	1096230	32.539
NC_004663	Bacteroidetes	Bacteroidia	Bacteroides	thetiaotomicon	VPI-5482	AATTTCTAGGATTTAGGTTATGTCATAT	- (- - - - (C- (C - - -) - - - -)	59.400	11.286	0.201	48.114	6132660	42.837
NC_009614	Bacteroidetes	Bacteroidia	Bacteroides	vulgatus	ATCC 8482	GTGACGGATAAGCTATGTCAGT	- (- - - - (CCCC- - - -) - - - -)	138.376	2.879	0.249	135.497	2136810	42.201
NC_004552	Chlamydiae	Chlamydiales	Chlamydia	abortus	S26/3	ACGCTGATAATGATGTAAGT	- (- - - - (CCCC- - - -) - - - -)	17.211	0.611	0.240	16.600	626899	39.866
NC_005861	Chlamydiae	Chlamydiales	Protochlamydia	amoebophila	amoebophila UWE25	ACGCTGATAAGCTATGTAAGT	- (- - - - (CCCC- (C - - -) - - - -)	17.279	15.792	0.138	1.487	1091346	34.717
NC_003361	Chlamydiae	Chlamydiales	Chlamydia	caeviae	GPIC	TGGCTGATAAATGTAATGTAAGT	- (- - - - (CCCC- (C - - -) - - - -)	20.881	20.838	0.243	0.043	654754	39.223
NC_007899	Chlamydiae	Chlamydiales	Chlamydia	felis	Fe/C-56	AGGCTGATAAATTTATATGTAAGT	- (- - - - (CCCC- - - -) - - - -)	23.105	0.655	0.258	22.450	508270	39.375
NC_002620	Chlamydiae	Chlamydiales	Chlamydia	muridarum	Nigg	AGGCTGATAAATTTATATGTAAGT	- (- - - - (CCCC- - - -) - - - -)	0.015	0.439	0.357	0.424	536430	40.337
NC_002179	Chlamydiae	Chlamydiales	Chlamydia	pneumoniae	AR39	AAGTCTGATAAATTTATATGTAAGT	- (- - - - (CCCC- - - -) - - - -)	5.263	5.295	0.240	0.031	632907	40.573
NC_005043	Chlamydiae	Chlamydiales	Chlamydia	pneumoniae	TW-183	AAGTCTGATAAATTTATATGTAAGT	- (- - - - (CCCC- - - -) - - - -)	5.885	5.901	0.240	0.016	205760	40.576
NC_000922	Chlamydiae	Chlamydiales	Chlamydia	pneumoniae	CWL029	AAGTCTGATAAATTTATATGTAAGT	- (- - - - (CCCC- - - -) - - - -)	5.266	5.281	0.241	0.016	208353	40.578
NC_002491	Chlamydiae	Chlamydiales	Chlamydia	pneumoniae	J138	AAGTCTGATAAATTTATATGTAAGT	- (- - - - (CCCC- - - -) - - - -)	5.730	5.746	0.242	0.016	207976	40.581
NC_012686	Chlamydiae	Chlamydiales	Chlamydia	trachomatis	B/Jali20/OT	AGGCTGATAAATTTATATGTAAGT	- (- - - - (CCCC- - - -) - - - -)	1.284	1.288	0.337	0.004	197103	41.299
NC_007429	Chlamydiae	Chlamydiales	Chlamydia	trachomatis	A/HAR-13	AGGCTGATAAATTTATATGTAAGT	- (- - - - (CCCC- - - -) - - - -)	1.275	1.279	0.337	0.004	197180	41.303
NC_000117	Chlamydiae	Chlamydiales	Chlamydia	trachomatis	D/UW-3/CX	AGGCTGATAAATTTATATGTAAGT	- (- - - - (CCCC- - - -) - - - -)	1.453	1.457	0.337	0.004	194922	41.307
NC_010280	Chlamydiae	Chlamydiales	Chlamydia	trachomatis	L2b/UCh-1/proctitis	AGGCTGATAAATTTATATGTAAGT	- (- - - - (CCCC- - - -) - - - -)	1.850	1.854	0.337	0.004	514097	41.328
NC_010287	Chlamydiae	Chlamydiales	Chlamydia	trachomatis	434/Bu	AGGCTGATAAATTTATATGTAAGT	- (- - - - (CCCC- - - -) - - - -)	1.859	1.863	0.337	0.004	514056	41.329
NC_011059	Chlorobi	Chlorobia	Prosthecochloris	aestuarii	DSM 271	ATGTTGATAAATAAATATGTAAGT	- (- - - - (CCCC- (C - - -) - - - -)	5.939	4.468	0.234	1.471	1215004	50.095
NC_007514	Chlorobi	Chlorobia	Chlorobium	chlorochromatii	Cad3	TGTGTGATAAATAAATATGTAAGT	- (- - - - (CCCC- (C - - -) - - - -)	24.405	24.404	0.138	0.001	1119635	44.278
NC_010803	Chlorobi	Chlorobia	Chlorobium	limicola	DSM 245	TGTGTGATAAATAAATATGTAAGT	- (- - - - (CCCC- (C - - -) - - - -)	6.950	4.431	0.182	2.519	1328244	51.317
NC_007512	Chlorobi	Chlorobia	Pelodictyon	luteolum	DSM 273	TGTGTGATAAATAAATATGTAAGT	- (- - - - (CCCC- (C - - -) - - - -)	6.575	2.661	0.123	3.914	1225612	57.330
NC_011027	Chlorobi	Chlorobia	Chlorobaculum	parvum	NCIB 8327	TGTGTGATAAATAAATATGTAAGT	- (- - - - (CCCC- (C - - -) - - - -)	2.542	2.449	0.138	0.093	1163191	55.796
NC_008639	Chlorobi	Chlorobia	Chlorobium	phaeobacteroides	DSM 266	TGTGTGATAAATAAATATGTAAGT	- (- - - - (CCCC- (C - - -) - - - -)	12.673	15.930	0.225	3.257	1456627	48.351
NC_010831	Chlorobi	Chlorobia	Chlorobium	phaeobacteroides	BS1	TGTGTGATAAATAAATATGTAAGT	- (- - - - (CCCC- (C - - -) - - - -)	14.423	15.681	0.216	1.258	1477835	48.927
NC_011060	Chlorobi	Chlorobia	Pelodictyon	phaeoclastratiforme	BU-1	TGTGTGATAAATGTAATGTAAGT	- (- - - - (CCCC- - - -) - - - -)	5.804	5.852	0.168	0.048	1460459	48.078
NC_009337	Chlorobi	Chlorobia	Chlorobium	phaeovibrioides	DSM 265	ATGTTGATAAATAAATATGTAAGT	- (- - - - (CCCC- (C - - -) - - - -)	2.420	0.565	0.219	1.855	970205	52.992
NC_002932	Chlorobi	Chlorobia	Chlorobium	tepidum	TLS	TGTGTGATAAATAAATATGTAAGT	- (- - - - (CCCC- (C - - -) - - - -)	9.447	8.478	0.131	0.969	1020925	56.530
NC_011026	Chlorobi	Chlorobia	Chloroherpeton	thalassium	ATCC 35110	AAGTTTTATAAATAAATATGTAAGT	- (- - - - (CCCC- (C - - -) - - - -)	0.964	0.124	0.124	0.840	2638363	45.039
NC_002936	Chloroflexi	Dehalococcoidetes	Dehalococcoides	ethenogenes	195	CTTTTTTCTATAAACTATTTAACT	- (- - - - (C- (C - - -) - - - -)	8.505	3.945	0.128	4.560	771319	48.850
NC_007356	Chloroflexi	Dehalococcoidetes	Dehalococcoides	sp.	CBDB1	TTTTTAAATCTGTAATAATTTAAACT	- (- - - - (CC- (C - - -) - - - -)	7.371	2.655	0.141	4.716	671037	47.026
NC_009455	Chloroflexi	Dehalococcoidetes	Dehalococcoides	sp.	BAV1	TTTTTAAATCTATAAATTTAAACT	- (- - - - (C- (C - - -) - - - -)	21.246	31.672	0.111	10.426	750139	47.173
NC_007181	† Crenarchaeota	Thermoprotei	Sulfolobus	acidocaldarius	DSM 639	ATCTGTATACGATATTAAGTAAGG	- (- - - - (C- (C - - -) - - - -)	53.715	21.443	0.055	32.272	2032387	36.708
NC_008818	† Crenarchaeota	Thermoprotei	Hyperthermus	butylicus	DSM 5456	ATAGGGTATAAAGCTGTGATTAAGT	- (- - - - (CC- (C - - -) - - - -)	64.986	27.180	0.030	37.806	1594073	53.737
NC_009776	† Crenarchaeota	Thermoprotei	Ignicoccus	hispidus	KLIN4/1	ATATGCTTCAATTTATGTCGACAT	- (- - - - (C- (C - - -) - - - -)	6.383	58.696	0.036	52.313	645125	56.520
NC_009033	† Crenarchaeota	Thermoprotei	Staphylothermus	marinus	F1	ATTTTCTACTAATAACTATATGTAAGT	- (- - - - (CC- (C - - -) - - - -)	86.755	175.831	0.094	89.076	1172365	35.727

†:アーキア

Accession	Phylum	Class	Genus	Species	Strain	dif	Palindrome	dif angle	GC angle	GCSI	Δ	dif Position	GC%
NC_000854	† Crenarchaeota	Thermoprotei	<i>Aeropyrum</i>	<i>pernix</i>	K1	ATATGTGATAAAATCTATATATACAGC	---(- (CCCC (-) ---) ---) ---	106.758	94.363	0.044	12.395	1046426	56.311
NC_009440	† Crenarchaeota	Thermoprotei	<i>Metallosphaera</i>	<i>sedula</i>	DSM 5348	ATATAGATACAGATAGATATATGTAACAA	(-(-(- (CCCC (-) ---) ---) ---)	11.645	149.980	0.057	138.335	1052001	46.225
NC_002754	† Crenarchaeota	Thermoprotei	<i>Sulfolobus</i>	<i>solfataricus</i>	P2	ATCTTTTATAAATTCATATATGTAACAA	(-(-(- (CCCC (-) ---) ---) ---)	133.478	31.003	0.066	102.475	1136600	35.787
NC_003106	† Crenarchaeota	Thermoprotei	<i>Sulfolobus</i>	<i>tokodaii</i>	str. 7	ATGTGTGACACATCTTCTTAAGTAAMT	(-(-(- (-(- (---) ---) ---) ---)	62.714	98.199	0.039	35.485	2296888	32.787
NC_009925	Cyanobacteria	Acaryochloris	<i>Acaryochloris</i>	<i>marina</i>	MBIC11017								47.254
NC_010475	Cyanobacteria	Chroococcales	<i>Synechococcus</i>	sp.	PCC 7002								49.631
NC_007776	Cyanobacteria	Chroococcales	<i>Synechococcus</i>	sp.	JA-2-3B'a(2-13)								58.450
NC_007775	Cyanobacteria	Chroococcales	<i>Synechococcus</i>	sp.	JA-3-3Ab								60.237
NC_005125	Cyanobacteria	Gloeobacteria	<i>Gloeobacter</i>	<i>violaceus</i>	PCC 7421								61.998
NC_010644	Elusimicrobia	Elusimicrobia (class)	<i>Elusimicrobium</i>	<i>minutum</i>	Pei191	ACTAACGATAATTTTATATGTAACAT	(-(-(- (CCCC (-) ---) ---) ---)	20.894	2.001	0.178	18.893	917173	39.952
NC_009637	† Euryarchaeota	Methanococci	<i>Methanococcus</i>	<i>maripaludis</i>	C7	ACAATCCATAAGTATATTTGTAACCT	(-(-(- (-(- (---) ---) ---) ---)	97.442	38.476	0.050	58.966	592578	33.284
NC_009975	† Euryarchaeota	Methanococci	<i>Methanococcus</i>	<i>maripaludis</i>	C6	ACAATCCATAAATATATATCTGTAACCT	(-(-(- (-(- (---) ---) ---) ---)	69.126	64.364	0.056	4.762	1237826	33.423
NC_013216	Firmicutes	Clostridia	<i>Desulfotomaculum</i>	<i>acetoxidans</i>	DSM 771	AGTCCGATAAAGATATATGTAACAT	(-(-(- (CCCC (-) ---) ---) ---)	3.422	3.473	0.275	0.051	2316017	41.555
NC_013205	Firmicutes	Bacillales	<i>Alicyclobacillus</i>	<i>acidocaldarius</i>	subsp. acidocaldarius DSM 446	ACTGCCGATAATCGATATATGTAACAT	(-(-(- (CCCC (-) ---) ---) ---)	19.899	19.847	0.172	0.053	1676243	62.335
NC_006814	Firmicutes	Lactobacillales	<i>Lactobacillus</i>	<i>acidophilus</i>	NCFM	TCTGCGGATAATTTGATATATGTAAGT	(-(-(- (CCCC (-) ---) ---) ---)	19.796	19.819	0.325	0.023	1106402	34.715
NC_009725	Firmicutes	Bacillales	<i>Bacillus</i>	<i>amyloliquefaciens</i>	FZB42	ACTCTTCAGAAATATATATGTAACAT	(-(-(- (-(- (---) ---) ---) ---)	8.954	8.948	0.184	0.005	1861835	46.476
NC_003997	Firmicutes	Bacillales	<i>Bacillus</i>	<i>anthracis</i>	str. Ames	ACTGCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	7.295	7.376	0.413	0.081	2507728	35.377
NC_007530	Firmicutes	Bacillales	<i>Bacillus</i>	<i>anthracis</i>	str. 'Ames Ancestor'	ACTGCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	7.290	7.371	0.413	0.081	2507852	35.377
NC_012659	Firmicutes	Bacillales	<i>Bacillus</i>	<i>anthracis</i>	str. A0248	ACTGCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	7.297	7.378	0.413	0.081	2507752	35.377
NC_005945	Firmicutes	Bacillales	<i>Bacillus</i>	<i>anthracis</i>	str. Sterne	ACTGCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	7.336	7.417	0.414	0.081	2507781	35.379
NC_012581	Firmicutes	Bacillales	<i>Bacillus</i>	<i>anthracis</i>	str. CDC 684	ACTGCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	60.837	60.754	0.441	0.083	1731214	35.381
NC_010424	Firmicutes	Clostridia	<i>Desulforudis</i>	<i>audaxviator</i>	audaxviator MP104C	ACTCCGATAAGATATATGTAACCT	(-(-(- (CCCC (-) ---) ---) ---)	1.029	1.144	0.177	0.115	1181453	60.847
NC_007793	Firmicutes	Bacillales	<i>Staphylococcus</i>	<i>aureus</i>	subsp. aureus USA300	ACTCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	4.542	4.518	0.317	0.024	1400139	32.751
NC_010079	Firmicutes	Bacillales	<i>Staphylococcus</i>	<i>aureus</i>	subsp. aureus USA300_TCH1516	ACTCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	2.803	0.295	0.310	2.509	1414086	32.756
NC_007622	Firmicutes	Bacillales	<i>Staphylococcus</i>	<i>aureus</i>	RF122	ACTCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	2.986	0.180	0.318	2.806	1348517	32.779
NC_002952	Firmicutes	Bacillales	<i>Staphylococcus</i>	<i>aureus</i>	subsp. aureus MRSA252	ACTCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	0.458	2.570	0.303	2.112	1447617	32.807
NC_002951	Firmicutes	Bacillales	<i>Staphylococcus</i>	<i>aureus</i>	subsp. aureus COL	ACTCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	2.463	5.632	0.323	3.168	1423936	32.818
NC_003923	Firmicutes	Bacillales	<i>Staphylococcus</i>	<i>aureus</i>	subsp. aureus MW2	ACTCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	3.238	0.087	0.318	3.151	1384863	32.830
NC_002745	Firmicutes	Bacillales	<i>Staphylococcus</i>	<i>aureus</i>	subsp. aureus N315	ACTCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	3.138	0.050	0.310	3.087	1382875	32.843
NC_002953	Firmicutes	Bacillales	<i>Staphylococcus</i>	<i>aureus</i>	subsp. aureus MSSA476	ACTCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	1.738	4.912	0.325	3.174	1413417	32.850
NC_007795	Firmicutes	Bacillales	<i>Staphylococcus</i>	<i>aureus</i>	subsp. aureus NCTC 8325	ACTCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	11.550	11.526	0.322	0.024	1320163	32.868
NC_009782	Firmicutes	Bacillales	<i>Staphylococcus</i>	<i>aureus</i>	subsp. aureus Mu3	ACTCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	2.565	2.528	0.309	0.037	1460604	32.878
NC_002758	Firmicutes	Bacillales	<i>Staphylococcus</i>	<i>aureus</i>	subsp. aureus Mu50	ACTCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	2.494	2.457	0.310	0.037	1459204	32.878
NC_009641	Firmicutes	Bacillales	<i>Staphylococcus</i>	<i>aureus</i>	subsp. aureus str. Newman	ACTCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	2.417	2.393	0.322	0.024	1420118	32.888
NC_009487	Firmicutes	Bacillales	<i>Staphylococcus</i>	<i>aureus</i>	subsp. aureus JH9	ACTCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	6.701	6.665	0.322	0.037	1507456	32.949
NC_009632	Firmicutes	Bacillales	<i>Staphylococcus</i>	<i>aureus</i>	subsp. aureus JH1	ACTCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	6.698	9.785	0.322	3.087	1507329	32.951
NC_009617	Firmicutes	Clostridia	<i>Clostridium</i>	<i>beijerinckii</i>	NCIMB 8052								29.856
NC_010723	Firmicutes	Clostridia	<i>Clostridium</i>	<i>botulinum</i>	E3 str. Alaska E43	GTTTCTATAAACCTAATTAAGTATCT	---(- (-(- (---) ---) ---) ---)	175.778	1.607	0.589	174.172	3616729	27.361
NC_010674	Firmicutes	Clostridia	<i>Clostridium</i>	<i>botulinum</i>	B str. Eklund 17B	ATTTCTATAAACCTAATTAAGTATCT	(-(-(- (-(- (---) ---) ---) ---)	175.983	0.455	0.583	175.528	3757923	27.511
NC_012658	Firmicutes	Clostridia	<i>Clostridium</i>	<i>botulinum</i>	Ba4 str. 657	AGTTTCTGTATATATTAAGTAATCT	(-(-(- (-(- (---) ---) ---) ---)	140.391	3.134	0.531	137.257	3540141	28.220
NC_009495	Firmicutes	Clostridia	<i>Clostridium</i>	<i>botulinum</i>	A str. ATCC 3502	AGTTTCTATAAATATATTAAGTAATCT	(-(-(- (-(- (---) ---) ---) ---)	141.289	3.053	0.534	138.236	3468951	28.245
NC_012491	Firmicutes	Bacillales	<i>Brevibacillus</i>	<i>brevis</i>	NBRC 100599	GAGTCCGATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	3.444	3.441	0.230	0.002	3208448	47.274
NC_008497	Firmicutes	Lactobacillales	<i>Lactobacillus</i>	<i>brevis</i>	ATCC 367	ACTTTCCGATAAATATATATGTAAGT	(-(-(- (CCCC (---) ---) ---) ---)	47.802	47.677	0.296	0.124	841377	46.219
NC_012121	Firmicutes	Bacillales	<i>Staphylococcus</i>	<i>carnosus</i>	subsp. carnosus TM300	ACTCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	33.716	29.725	0.283	3.991	1039830	34.633
NC_010999	Firmicutes	Lactobacillales	<i>Lactobacillus</i>	<i>casei</i>	BL23	ACTCCGATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	3.950	5.992	0.192	2.042	1576602	46.336
NC_008526	Firmicutes	Lactobacillales	<i>Lactobacillus</i>	<i>casei</i>	ATCC 334	ACTTCGATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	5.078	2.906	0.185	2.172	1406791	46.619
NC_011999	Firmicutes	Bacillales	<i>Macrocococcus</i>	<i>caseolyticus</i>	JCS5402	ACTCCGATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	17.040	19.468	0.316	2.428	1150672	36.877
NC_011772	Firmicutes	Bacillales	<i>Bacillus</i>	<i>cereus</i>	G9842	ACTACCTATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	6.843	7.038	0.410	0.196	2591269	35.264
NC_004722	Firmicutes	Bacillales	<i>Bacillus</i>	<i>cereus</i>	ATCC 14579	ACTACCTATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	1.634	1.660	0.413	0.026	2681342	35.281
NC_011725	Firmicutes	Bacillales	<i>Bacillus</i>	<i>cereus</i>	B4264	ACTGCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	5.925	6.120	0.408	0.195	2620337	35.304
NC_006274	Firmicutes	Bacillales	<i>Bacillus</i>	<i>cereus</i>	E33L	ACTGCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	5.396	5.468	0.410	0.072	2570998	35.352
NC_011773	Firmicutes	Bacillales	<i>Bacillus</i>	<i>cereus</i>	AH820	ACTGCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	5.167	5.238	0.411	0.070	2575228	35.399
NC_012472	Firmicutes	Bacillales	<i>Bacillus</i>	<i>cereus</i>	Q3BB102	ACTGCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	2.874	2.905	0.407	0.031	2592744	35.441
NC_011969	Firmicutes	Bacillales	<i>Bacillus</i>	<i>cereus</i>	01	ACTGCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	6.661	6.727	0.410	0.065	2510616	35.562
NC_003909	Firmicutes	Bacillales	<i>Bacillus</i>	<i>cereus</i>	ATCC 10987	ACTGCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	1.503	1.576	0.408	0.072	2590323	35.576
NC_011658	Firmicutes	Bacillales	<i>Bacillus</i>	<i>cereus</i>	AH187	ACTGCCATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	4.792	4.818	0.405	0.027	2564385	35.591
NC_010471	Firmicutes	Lactobacillales	<i>Leuconostoc</i>	<i>citreum</i>	KM20	ACTTCGATAAATATATATGTAAGT	(-(-(- (CCCC (---) ---) ---) ---)	2.970	2.996	0.284	0.026	915631	38.989
NC_006582	Firmicutes	Bacillales	<i>Bacillus</i>	<i>clausii</i>	KSM-K16	GGTTCGTGAAGTATATATGTAACAT	(-(-(- (-(- (---) ---) ---) ---)	5.119	4.561	0.220	0.558	2090731	44.753
NC_009674	Firmicutes	Bacillales	<i>Bacillus</i>	<i>cytotoxicus</i>	NVH 391-98	ACTACCTATAAATATATATGTAACAT	(-(-(- (CCCC (---) ---) ---) ---)	6.415	6.467	0.396	0.052	1970686	35.877
NC_008529	Firmicutes	Lactobacillales	<i>Lactobacillus</i>	<i>delbrueckii</i>	subsp. bulgaricus ATCC BAA-365	AGTTCCGATAAATATATATGTAAGC	(-(-(- (CCCC (-) ---) ---) ---)	3.017	2.115	0.141	0.902	930665	49.686

Accession	Phylum	Class	Genus	Species	Strain	dif	Palindrome	dif angle	GC angle	GCSI	A	dif diff	GC%
NC_008054	Firmicutes	Lactobacillales	<i>Lactobacillus</i>	<i>delbrueckii</i>	subsp. bulgaricus ATCC 11842	AGTTCGAATAATACATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	4.820	5.370	0.142	0.550	943973	49.719
NC_004461	Firmicutes	Bacillales	<i>Staphylococcus</i>	<i>epidermidis</i>	ATCC 12228	ACTTCCTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	25.772	22.138	0.280	3.634	1070717	32.095
NC_002976	Firmicutes	Bacillales	<i>Staphylococcus</i>	<i>epidermidis</i>	RP62A	ACTTCCTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	47.259	35.709	0.281	11.550	964781	32.151
NC_004668	Firmicutes	Lactobacillales	<i>Enterococcus</i>	<i>faecalis</i>	V583	ACTTTGTATAATGATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	6.544	6.494	0.287	0.049	1550522	37.527
NC_011567	Firmicutes	Bacillales	<i>Anoxybacillus</i>	<i>flavithermus</i>	WK1	ACTTCCTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	1.818	1.354	0.282	0.464	1437753	41.777
NC_008530	Firmicutes	Lactobacillales	<i>Lactobacillus</i>	<i>gasseri</i>	ATCC 33323	ACTTCGTGAAATATATATATGTAAGT	CCCC(-(-(-(-(-(-(-))))))))))	10.956	10.968	0.335	0.012	1004832	35.260
NC_007168	Firmicutes	Bacillales	<i>Staphylococcus</i>	<i>haemolyticus</i>	JCS1435	ACTTCCTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	30.252	27.445	0.254	2.806	1570847	32.793
NC_007907	Firmicutes	Clostridia	<i>Desulfotobacterium</i>	<i>hafniense</i>	Y51	GGTTCCTATAAATAGATATATGTAAGT	---(-(-(-(-(-(-(-))))))))))	65.107	65.111	0.262	0.004	1827924	47.360
NC_011830	Firmicutes	Clostridia	<i>Desulfotobacterium</i>	<i>hafniense</i>	DCB-2	GGTTCCTATAAATAGATATATGTAAGT	---(-(-(-(-(-(-(-))))))))))	18.675	18.673	0.242	0.002	2913416	47.537
NC_002570	Firmicutes	Bacillales	<i>Bacillus</i>	<i>halodurans</i>	C-125	GGTTCCTATAAATATATATATGTAAGT	---(-(-(-(-(-(-(-))))))))))	12.170	12.105	0.267	0.064	2243234	43.686
NC_010080	Firmicutes	Lactobacillales	<i>Lactobacillus</i>	<i>helveticus</i>	DPC 4571								37.077
NC_007503	Firmicutes	Clostridia	<i>Carboxydotherrnus</i>	<i>hydrogenoformans</i>	Z-2901	ACTTCCTATAAATAGATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	2.751	2.760	0.382	0.009	1219110	42.050
NC_004193	Firmicutes	Bacillales	<i>Oceanobacillus</i>	<i>ihyensis</i>	HTE831	ACTTCCTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	4.268	2.901	0.306	1.367	1772223	35.681
NC_003212	Firmicutes	Bacillales	<i>Listeria</i>	<i>innocua</i>	Clip11262	ACTTCCTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	6.755	6.727	0.262	0.028	1449101	37.439
NC_005362	Firmicutes	Lactobacillales	<i>Lactobacillus</i>	<i>johnsonii</i>	NCC 533	AATTCGTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	18.188	18.306	0.329	0.118	1099729	34.611
NC_006510	Firmicutes	Bacillales	<i>Geobacillus</i>	<i>kaustophilus</i>	HTA426	ACTTCCTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	1.532	1.525	0.243	0.007	1757303	52.086
NC_006270	Firmicutes	Bacillales	<i>Bacillus</i>	<i>licheniformis</i>	ATCC 14580	ACTTCGGAGAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	6.782	6.777	0.181	0.005	2031748	46.194
NC_006322	Firmicutes	Bacillales	<i>Bacillus</i>	<i>licheniformis</i>	ATCC 14580	ACTTCGGAGAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	6.797	6.792	0.181	0.005	2031600	46.195
NC_008531	Firmicutes	Lactobacillales	<i>Leuconostoc</i>	<i>mesenteroides</i>	subsp. mesenteroides ATCC 8293	ACTTCGATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	19.388	18.876	0.305	0.512	1131295	37.710
NC_010337	Firmicutes	Clostridia	<i>Heliobacterium</i>	<i>modesticaldum</i>	Ice1	ATTCGGATAATCGATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	11.995	54.705	0.094	42.710	2529545	56.975
NC_003210	Firmicutes	Bacillales	<i>Listeria</i>	<i>monocytogenes</i>	EGD-e	ACTTCCTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	3.679	3.495	0.260	0.184	1442173	37.981
NC_002973	Firmicutes	Bacillales	<i>Listeria</i>	<i>monocytogenes</i>	str. 4b F2365	ACTTCCTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	3.805	3.620	0.261	0.185	1421891	38.041
NC_012488	Firmicutes	Bacillales	<i>Listeria</i>	<i>monocytogenes</i>	Clip80459	ACTTCCTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	3.765	3.581	0.262	0.184	1425883	38.062
NC_011660	Firmicutes	Bacillales	<i>Listeria</i>	<i>monocytogenes</i>	HCC23	ACTTCCTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	3.994	3.990	0.248	0.004	1197132	38.194
NC_008528	Firmicutes	Lactobacillales	<i>Oenococcus</i>	<i>oeni</i>	PSU-1	GCTTTGAATAATATATATATGTAAGT	--(-(-(-(-(-(-(-))))))))))	16.035	15.655	0.193	0.380	971657	37.889
NC_008525	Firmicutes	Lactobacillales	<i>Pediococcus</i>	<i>pentosaceus</i>	ATCC 25745	ACTTTGTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	6.981	6.995	0.310	0.015	951725	37.358
NC_003366	Firmicutes	Clostridia	<i>Clostridium</i>	<i>perfringens</i>	str. 13								28.566
NC_004567	Firmicutes	Lactobacillales	<i>Lactobacillus</i>	<i>plantarum</i>	WCFS1	ACTTTGTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	0.107	0.664	0.282	0.557	1655120	44.467
NC_012984	Firmicutes	Lactobacillales	<i>Lactobacillus</i>	<i>plantarum</i>	JDM1	ACTTTGTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	4.087	3.277	0.282	0.809	1635182	44.661
NC_009848	Firmicutes	Bacillales	<i>Bacillus</i>	<i>pumilus</i>	SAFR-032	ACTTCGATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	8.368	8.369	0.218	0.001	1766124	41.287
NC_009253	Firmicutes	Clostridia	<i>Desulfotomaculum</i>	<i>reducens</i>	MI-1	GTTCCGATAAATACATATATGTAAGT	---(-(-(-(-(-(-(-))))))))))	22.540	21.333	0.260	1.208	2029964	42.275
NC_013198	Firmicutes	Lactobacillales	<i>Lactobacillus</i>	<i>rhamnosus</i>	GG	ACTTCGTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	8.860	6.154	0.210	2.706	1430973	46.690
NC_013199	Firmicutes	Lactobacillales	<i>Lactobacillus</i>	<i>rhamnosus</i>	Lc 705	ACTTCGTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	4.934	2.112	0.218	2.823	1443610	46.744
NC_007576	Firmicutes	Lactobacillales	<i>Lactobacillus</i>	<i>sakei</i>	subsp. sakei 23K	ACTTCCTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	2.543	4.829	0.261	2.286	955645	41.258
NC_007929	Firmicutes	Lactobacillales	<i>Lactobacillus</i>	<i>salivarius</i>	UCC118	ACTTTGTATAATGATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	2.800	2.106	0.432	0.694	901966	32.941
NC_007350	Firmicutes	Bacillales	<i>Staphylococcus</i>	<i>saprophyticus</i>	subsp. saprophyticus ATCC 15305	ACTTCCTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	22.187	18.612	0.310	3.575	1413387	33.240
NC_010556	Firmicutes	Bacillales	<i>Exiguobacterium</i>	<i>sibiricum</i>	255-15	ACTTCGATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	4.585	2.560	0.186	2.024	1559113	47.718
NC_012673	Firmicutes	Bacillales	<i>Exiguobacterium</i>	sp.	AT1b	TTACTCGATAAATATATATATGTAAGT	---(-(-(-(-(-(-(-))))))))))	4.602	3.371	0.183	1.231	239188	48.463
NC_012793	Firmicutes	Bacillales	<i>Geobacillus</i>	sp.	WCH70	ACTGCCTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	1.887	1.894	0.266	0.007	1714148	42.840
NC_010382	Firmicutes	Bacillales	<i>Lysinibacillus</i>	<i>sphaericus</i>	C3-41	ACTTCCTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	44.107	44.131	0.282	0.025	2888374	37.289
NC_000964	Firmicutes	Bacillales	<i>Bacillus</i>	<i>subtilis</i>	subsp. subtilis str. 168	ACTTCCTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	14.113	14.107	0.215	0.006	1942542	43.514
NC_007644	Firmicutes	Clostridia	<i>Moorella</i>	<i>thermoacetica</i>	ATCC 39073	ACTTCCTATAAATAGATATATGTAAGT	---(-(-(-(-(-(-(-))))))))))	22.993	5.035	0.132	17.958	1149218	55.792
NC_009328	Firmicutes	Bacillales	<i>Geobacillus</i>	<i>thermodenitrificans</i>	NG80-2	ACTTCCTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	6.242	6.192	0.268	0.050	1713597	49.010
NC_012034	Firmicutes	Clostridia	<i>Anaerocellum</i>	<i>thermophilum</i>	DSM 6725	ACTTCCTATAAATATATATATGTAAGT	---(-(-(-(-(-(-(-))))))))))	67.269	1.150	0.401	66.119	914282	35.172
NC_006177	Firmicutes	Clostridia	<i>Symbiobacterium</i>	<i>thermophilum</i>	IAM 14863	AGTCGGATAAATATATATATGTAAGT	---(-(-(-(-(-(-(-))))))))))	9.350	9.347	0.117	0.003	1875685	68.668
NC_010718	Firmicutes	Clostridia	<i>Natranaerobius</i>	<i>thermophilus</i>	JI/WNM-WN-LF	AGTCCTATAAATAACTATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	0.333	0.272	0.321	0.061	1579849	36.322
NC_009454	Firmicutes	Clostridia	<i>Pelotomaculum</i>	<i>thermopropionicum</i>	SW	AGTCGGATAAATAGATATATGTAAGT	---(-(-(-(-(-(-(-))))))))))	11.131	11.193	0.215	0.062	1419141	52.961
NC_005957	Firmicutes	Bacillales	<i>Bacillus</i>	<i>thuringiensis</i>	serovar konkukian str. 97-27	ACTGCCTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	4.027	4.098	0.409	0.071	2560255	35.405
NC_008600	Firmicutes	Bacillales	<i>Bacillus</i>	<i>thuringiensis</i>	str. Al Hakam	ACTGCCTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	2.435	2.650	0.405	0.215	2592992	35.426
NC_010184	Firmicutes	Bacillales	<i>Bacillus</i>	<i>weihenstephanensis</i>	KBAB4	ACTTCGTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	6.912	6.680	0.418	0.232	2530339	35.562
NC_008555	Firmicutes	Bacillales	<i>Listeria</i>	<i>welshimeri</i>	serovar 6b str. SLCC5334	ACTTCCTATAAATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	1.743	1.741	0.276	0.002	1420692	36.352
NC_008346	Firmicutes	Clostridia	<i>Syntrophomonas</i>	<i>wolfiei</i>	subsp. wolfiei str. Goettingen	ACTTCCTATAAATAGATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	5.127	12.545	0.253	7.418	1509914	44.866
NC_012489	Gemmatimonadetes	Gemmatimonadales	<i>Gemmatimonas</i>	<i>aurantiaca</i>	T-27	AATCGATAAATACATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	30.139	30.128	0.071	0.011	1930284	64.274
NC_011296	Nitrospirae	Nitrospirales	<i>Thermodesulfobivrio</i>	<i>yellowstonii</i>	DSM 11347	ATGTCGATAAATGATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	2.814	15.403	0.104	12.588	1017567	34.129
NC_005027	Planctomycetes	Planctomycetacia	<i>Rhodopirellula</i>	<i>baltica</i>	SH 1								55.403
NC_007759	Proteobacteria	Deltaproteobacteria	<i>Syntrophus</i>	<i>aciditrophicus</i>	SB	TTGCTATAAAGATATATATATGTAAGT	---(-(-(-(-(-(-(-))))))))))	8.628	8.632	0.125	0.005	1665844	51.460
NC_010002	Proteobacteria	Betaproteobacteria	<i>Delftia</i>	<i>acidovorans</i>	SPH-1	GTTGCATGACGATATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	40.975	40.797	0.069	0.178	4181915	66.476
NC_008463	Proteobacteria	Gammaproteobacteria	<i>Pseudomonas</i>	<i>aeruginosa</i>	UCBPP-PA14	GTTCCGATAAATGATATATATGTAAGT	C(-(-(-(-(-(-(-))))))))))	3.920	1.472	0.093	2.448	3175681	66.292
NC_011770	Proteobacteria	Gammaproteobacteria	<i>Pseudomonas</i>	<i>aeruginosa</i>	LESB58	GATTCGATAAATGATATATATGTAAGT	---(-(-(-(-(-(-(-))))))))))	5.159	5.817	0.094	0.658	3395491	66.297
NC_009656	Proteobacteria	Gammaproteobacteria	<i>Pseudomonas</i>	<i>aeruginosa</i>	PA7	GATTCGATAAATGATATATATGTAAGT	---(-(-(-(-(-(-(-))))))))))	4.802	6.063	0.094	1.262	3180098	66.449

Accession	Phylum	Class	Genus	Species	Strain	dif	Palindrome	dif angle	GC angle	GCSI	A	dif Position	GC%	
NC_002516	Proteobacteria	Gammaproteobacteria	<i>Pseudomonas</i>	<i>aeruginosa</i>	PAO1	GATTCGCATAA	TCATATATGTAAAT	--(- - (CCCC -O -))--	38.341	39.161	0.100	0.820	2443067	66.556
NC_012633	Proteobacteria	Alphaproteobacteria	<i>Rickettsia</i>	<i>africae</i>	ESF-5	GTTCTAATAAC	TATATATGAAAATAA	-(- - (--- -O) ---) -)-	21.033	21.050	0.188	0.017	713968	32.399
NC_009881	Proteobacteria	Alphaproteobacteria	<i>Rickettsia</i>	<i>akari</i>	str. Hartford	TTCTATAACT	TATATGAAAAACAA	(---(- -(- -O) ---) -)-)	5.596	5.518	0.173	0.079	685616	32.335
NC_008048	Proteobacteria	Alphaproteobacteria	<i>Sphingopyxis</i>	<i>alaskensis</i>	RB2256	ATTTGTAATA	TATATGTTAAATAC	-(- - (- --- CO) ---) -)-)	41.750	42.560	0.057	0.809	1314809	65.502
NC_011768	Proteobacteria	Deltaproteobacteria	<i>Desulfatibacillum</i>	<i>alkenivivans</i>	AK-01	ATGCCGCAAA	TATATATGTAAC	-(- - (--- -O))--)	79.652	72.663	0.031	6.989	1128887	54.485
NC_008700	Proteobacteria	Gammaproteobacteria	<i>Shewanella</i>	<i>amazonensis</i>	SB2B	AATGCCAAG	TATATGTTAAAT	-(- - (--- -O))--)	2.326	2.063	0.121	0.263	2180892	53.586
NC_012913	Proteobacteria	Gammaproteobacteria	<i>Aggregatibacter</i>	<i>aphrophilus</i>	NJ8700	ACTTCGCAAT	TCATATGTTAAAT	-(- - (--- -O))--)	5.393	4.679	0.151	0.714	779982	42.231
NC_008740	Proteobacteria	Mariniproteobacteria	<i>Marinobacter</i>	<i>aquaeolei</i>	VT8	ACTGGCAAT	GCATATATGTTAAAT	-(- - (--- -O))--)	46.758	46.753	0.152	0.006	1601433	57.272
NC_007204	Proteobacteria	Gammaproteobacteria	<i>Psychrobacter</i>	<i>arcticus</i>	273-4	TTTTGGCAAT	GCATATATGTTAAAT	-(- - (--- -O))--)	3.133	1.540	0.116	1.593	1302283	42.800
NC_007298	Proteobacteria	Betaproteobacteria	<i>Dechloromonas</i>	<i>aromatica</i>	RCB	AACGGCATAA	TTCATATGTTAAAT	(---(- -(- -O))--)	5.037	5.511	0.140	0.474	2192493	59.247
NC_007794	Proteobacteria	Alphaproteobacteria	<i>Novosphingobium</i>	<i>aromaticivivans</i>	DSM 12444	TATTTACATAA	TGATATGTTAAAT	-(- - (--- -O))--)	28.086	26.478	0.056	1.609	2048158	65.153
NC_006513	Proteobacteria	Betaproteobacteria	<i>Aromatoleum</i>	<i>aromaticum</i>	EbN1	ATATCTCACA	TATATGTTGGCAAT	(-(- - (--- -O))--)	28.351	3.177	0.086	25.174	2053683	65.121
NC_009138	Proteobacteria	Betaproteobacteria	<i>Hermiiniomonas</i>	<i>arsenicioxydans</i>		ATTTGCTAA	TATATGTTAAAT	(---(- -(- -O))--)	14.600	14.650	0.134	0.050	1523751	54.295
NC_012985	Proteobacteria	Alphaproteobacteria	<i>Liberibacter</i>	<i>asiaticus</i>	asiaticus str. psy62	GTCGCAATAA	TAGTATGGAACATTT	(---(- -(- -O))--)	9.829	8.256	0.318	1.573	1083369	36.479
NC_012962	Proteobacteria	Gammaproteobacteria	<i>Photorhabdus</i>	<i>asymbiotica</i>		AGTGGCGTAA	TGATATGTTAAAT	-(- - (--- -O))--)	5.575	5.582	0.236	0.007	2453972	42.153
NC_008228	Proteobacteria	Gammaproteobacteria	<i>Pseudoalteromonas</i>	<i>atlantica</i>	T6c	TATTGCTAA	TATGATATGTTAAAT	-(- - (--- -O))--)	2.796	2.940	0.119	0.144	2553223	44.616
NC_004547	Proteobacteria	Gammaproteobacteria	<i>Pectobacterium</i>	<i>atrosepticum</i>	SCR11043	GGTTCGCAAT	GCATATGTTAAAT	-(- - (--- -O))--)	0.008	1.521	0.163	1.514	2532119	50.967
NC_012691	Proteobacteria	Gammaproteobacteria	<i>Tolumonas</i>	<i>auensis</i>	DSM 9187	ATTTGCGTAA	TGATATGTTAAAT	(---(- -(- -O))--)	2.367	3.109	0.137	0.742	1683258	49.027
NC_012108	Proteobacteria	Deltaproteobacteria	<i>Desulfobacterium</i>	<i>autotrophicum</i>	HRM2	TTGCGATAAT	TATGTTAAATC	(---(- -(- -O))--)	8.503	6.994	0.122	1.509	2926553	48.847
NC_009720	Proteobacteria	Alphaproteobacteria	<i>Xanthobacter</i>	<i>autotrophicus</i>	Py2	GAATGCGAGA	GAGATATGGAAC	(-(- - (--- -O))--)	148.918	135.106	0.038	13.812	4867206	67.465
NC_010645	Proteobacteria	Betaproteobacteria	<i>Bordetella</i>	<i>avium</i>	197N	AATACGCATA	TGATATGTTAAAG	(---(- -(- -O))--)	0.500	1.287	0.142	0.787	1855830	61.583
NC_003919	Proteobacteria	Gammaproteobacteria	<i>Xanthomonas</i>	<i>aromopidis</i>	pv. citri str. 306	ATTTGCGTAA	TGATATGTTAAAT	(---(- -(- -O))--)	0.150	0.154	0.074	0.004	2486038	64.772
NC_008783	Proteobacteria	Alphaproteobacteria	<i>Bartonella</i>	<i>bacilliformis</i>	KC583	ATTTGCAATA	TATGTTGGCAATTT	(---(- -(- -O))--)	26.348	26.513	0.367	0.165	837738	38.239
NC_005363	Proteobacteria	Deltaproteobacteria	<i>Bdellovibrio</i>	<i>bacteriovorus</i>	HD100	TCTGATAAGT	TATATGTAACGGTAA	(---(- -(- -O))--)	5.269	4.709	0.167	0.560	1946840	50.646
NC_013173	Proteobacteria	Deltaproteobacteria	<i>Desulfomicrobium</i>	<i>baclarium</i>	DSM 4028	TTGACACATA	TAGTAAATGTTAACT	(---(- -(- -O))--)	7.953	6.604	0.138	1.349	1644132	58.647
NC_012982	Proteobacteria	Alphaproteobacteria	<i>Hirschia</i>	<i>baltica</i>	ATCC 49814	AACGCAATACT	ACATATGTTGCAATTT	(---(- -(- -O))--)	1.223	1.202	0.266	0.021	1733476	45.230
NC_009997	Proteobacteria	Gammaproteobacteria	<i>Shewanella</i>	<i>baltica</i>	OS195	GCTGGCACA	TGATATGTTAAAT	-(- - (--- -O))--)	2.552	4.195	0.161	1.623	2711546	46.283
NC_009052	Proteobacteria	Gammaproteobacteria	<i>Shewanella</i>	<i>baltica</i>	OS155	GCTGGCACA	TGATATGTTAAAT	-(- - (--- -O))--)	1.475	0.924	0.157	0.551	2542678	46.283
NC_009665	Proteobacteria	Gammaproteobacteria	<i>Shewanella</i>	<i>baltica</i>	OS185	GCTGGCACA	TGATATGTTAAAT	-(- - (--- -O))--)	0.692	1.346	0.162	0.655	2624890	46.312
NC_011663	Proteobacteria	Gammaproteobacteria	<i>Shewanella</i>	<i>baltica</i>	OS223	GCTGGCACA	TGATATGTTAAAT	-(- - (--- -O))--)	3.660	4.552	0.160	0.891	2625272	46.314
NC_009085	Proteobacteria	Gammaproteobacteria	<i>Acinetobacter</i>	<i>baumannii</i>	ATCC 17978	TGTTGCTATA	TGATATGTTAAAT	-(- - (--- -O))--)	4.094	5.080	0.165	0.987	2033595	38.941
NC_010611	Proteobacteria	Gammaproteobacteria	<i>Acinetobacter</i>	<i>baumannii</i>	ACICU	TGTTGCTATA	TGATATGTTAAAT	-(- - (--- -O))--)	2.106	2.612	0.169	0.506	1936244	39.031
NC_011595	Proteobacteria	Acinetobacter	<i>baumannii</i>		AB307-0294	TGTTGCTATA	TGATATGTTAAAT	-(- - (--- -O))--)	3.660	2.897	0.172	0.763	1842254	39.044
NC_010400	Proteobacteria	Gammaproteobacteria	<i>Acinetobacter</i>	<i>baumannii</i>	SDF	TGTTGCTATA	TGATATGTTAAAT	-(- - (--- -O))--)	13.510	14.707	0.156	1.197	1839393	39.192
NC_011586	Proteobacteria	Gammaproteobacteria	<i>Acinetobacter</i>	<i>baumannii</i>	AB0057	GCTTCGATAA	TGATATGTTAAAT	-(- - (--- -O))--)	7.795	7.789	0.170	0.006	2150158	39.205
NC_010410	Proteobacteria	Gammaproteobacteria	<i>Acinetobacter</i>	<i>baumannii</i>	AYE	ACTTCGATAA	TGATATGTTAAAT	-(- - (--- -O))--)	4.664	4.658	0.166	0.006	1898699	39.380
NC_009883	Proteobacteria	Alphaproteobacteria	<i>Rickettsia</i>	<i>bellii</i>	OSU 85-389	GTTCTAATAAC	TATATGAAAATAA	-(- - (--- -O))--)	29.990	29.580	0.123	0.410	772678	31.628
NC_007940	Proteobacteria	Alphaproteobacteria	<i>Rickettsia</i>	<i>bellii</i>	RML369-C	GTTCTAATAA	TATATGAAAATAA	-(- - (--- -O))--)	2.719	2.617	0.134	0.102	755654	31.646
NC_011146	Proteobacteria	Deltaproteobacteria	<i>Geobacter</i>	<i>bemidjiensis</i>	Bem	AGTCGCCATA	AAGTATATATGTTAAAT	-(- - (--- -O))--)	14.399	34.497	0.048	20.099	2492163	60.267
NC_008343	Proteobacteria	Alphaproteobacteria	<i>Granulibacter</i>	<i>bethesdensis</i>	CGDNIH1	GTCGTAAAG	TATATGGAACATAA	-(- - (--- -O))--)	3.139	1.774	0.156	1.365	1330560	59.072
NC_008260	Proteobacteria	Gammaproteobacteria	<i>Alcanivorax</i>	<i>borkumensis</i>	SK2	ACTAGCATAA	TGATATGTTAAAT	-(- - (--- -O))--)	2.906	6.708	0.252	3.802	1534889	54.728
NC_007613	Proteobacteria	Gammaproteobacteria	<i>Shigella</i>	<i>boydii</i>	Sb227	GGTGGCATAA	TGATATGTTAAAT	-(- - (--- -O))--)	10.930	15.722	0.049	4.792	1539611	51.213
NC_010658	Proteobacteria	Gammaproteobacteria	<i>Shigella</i>	<i>boydii</i>	CDC 3083-94	GGTGGCATAA	TGATATGTTAAAT	-(- - (--- -O))--)	54.531	57.001	0.056	2.470	1608794	51.333
NC_002927	Proteobacteria	Betaproteobacteria	<i>Bordetella</i>	<i>bronchiseptica</i>	RB50	AATTCGATAA	TGATATGTTAAAG	-(- - (--- -O))--)	18.035	19.614	0.103	1.578	2955841	68.077
NC_007508	Proteobacteria	Gammaproteobacteria	<i>Xanthomonas</i>	<i>campestris</i>	pv. vesicatoria str. 85-10	ATTTGCGATA	TGATATGTTAAAT	(---(- -(- -O))--)	12.514	11.251	0.074	1.263	2782410	64.748
NC_007086	Proteobacteria	Gammaproteobacteria	<i>Xanthomonas</i>	<i>campestris</i>	pv. campestris str. 8004	ATTTGCGATA	TGATATGTTAAAT	(---(- -(- -O))--)	3.194	3.164	0.076	0.030	2537462	64.957
NC_010688	Proteobacteria	Gammaproteobacteria	<i>Xanthomonas</i>	<i>campestris</i>	pv. campestris	ATTTGCGATA	TGATATGTTAAAT	-(- - (--- -O))--)	1.292	1.293	0.078	0.001	2769795	65.042
NC_003902	Proteobacteria	Gammaproteobacteria	<i>Xanthomonas</i>	<i>campestris</i>	pv. campestris str. ATCC 33913	ATTTGCGATA	TGATATGTTAAAT	-(- - (--- -O))--)	25.363	25.331	0.072	0.032	2441748	65.068
NC_009879	Proteobacteria	Alphaproteobacteria	<i>Rickettsia</i>	<i>canadensis</i>	str. McKiel	TTCTATAA	TCCATATATGAAAATAA	-(- - (- -(- -O))--)	17.424	17.454	0.191	0.030	516602	31.054
NC_007354	Proteobacteria	Alphaproteobacteria	<i>Ehrlichia</i>	<i>canis</i>	str. Jake	ATTACATAAT	TATATGAAAATAC	-(- - (--- -O))--)	0.029	0.693	0.533	0.664	657409	28.958
NC_002977	Proteobacteria	Gammaproteobacteria	<i>Methylococcus</i>	<i>capsulatus</i>	str. Bath	TATGGCATAA	TGATATGTTAAAT	-(- - (--- -O))--)	17.405	13.019	0.124	4.386	1492511	63.585
NC_007498	Proteobacteria	Deltaproteobacteria	<i>Pelobacter</i>	<i>carbinolicus</i>	DSM 2380	GGTTCGATAA	TGATATGTTAAAC	(-(- - (--- -O))--)	4.584	5.444	0.184	0.860	1879625	55.114
NC_011386	Proteobacteria	Alphaproteobacteria	<i>Oligotropha</i>	<i>carboxidoovorans</i>	OMS	GATTTCGATA	AAGTATATGTTAAAT	-(- - (--- -O))--)	16.726	16.292	0.096	0.434	2070413	62.399
NC_012917	Proteobacteria	Gammaproteobacteria	<i>Pectobacterium</i>	<i>carotovorum</i>	subsp. carotovorum PC1	GGTTCGATAA	TGATATGTTAAAT	-(- - (--- -O))--)	0.309	1.598	0.164	1.289	2371447	51.932
NC_009937	Proteobacteria	Alphaproteobacteria	<i>Azorhizobium</i>	<i>caulinodans</i>	ORS 571	GATCGGTAGA	GAAGATATATGGAAC	(---(- -(- -O))--)	4.470	0.132	0.058	4.338	2586370	67.319
NC_010162	Proteobacteria	Deltaproteobacteria	<i>Sorangium</i>	<i>cellulosum</i>	So ce 56'	GGATCGCATA	AAGCAACATATGTTAACT	(---(- -(- -O))--)	4.983	1.004	0.037	3.978	5024591	71.376
NC_011420	Proteobacteria	Alphaproteobacteria	<i>Rhodospirillum</i>	<i>centenum</i>	SW	AAGTCGATAA	TGATATGTTAAAT	-(- - (--- -O))--)	135.597	51.810	0.011	83.788	130296	70.461
NC_007799	Proteobacteria	Alphaproteobacteria	<i>Ehrlichia</i>	<i>chaffensis</i>	str. Arkansas	ATTACATAAT	TATATGAAAATAA	-(- - (--- -O))--)	97.765	4.807	0.513	92.958	926614	30.101
NC_007645	Proteobacteria	Gammaproteobacteria	<i>Hahella</i>	<i>chejuensis</i>	KCTC 2396	AGTGGCATAA	TATATGTTAAAT	-(- - (--- -O))--)	8.511	8.450	0.179	0.062	3437046	53.870
NC_011757	Proteobacteria	Alphaproteobacteria	<i>Methylobacterium</i>	<i>chloromethanicum</i>	CM4	TTGCGATAA	TATATGGAACATGA	(---(- -(- -O))--)	141.573	74.511	0.013	67.061	4823122	68.210

Accession	Phylum	Class	Genus	Species	Strain	dif	Palindrome	dif angle	GC angle	GCSI	A	dif Position	GC%
NC_010524	Proteobacteria	Betaproteobacteria	<i>Leptothrix</i>	<i>cholodnii</i>	SP-6	ATGCGTGACGATGAATATATGCAACT	-(--- (-(- (---))--)-)-	27.011	29.745	0.039	2.734	2066218	68.900
NC_008752	Proteobacteria	Betaproteobacteria	<i>Acidovorax</i>	<i>citrulli</i>	AA000-1	AGTGAACGACATGATATATTGTCAAT	-(-(-(-(-(-(-))--)-)-)-)	23.646	24.986	0.062	1.340	3051233	68.532
NC_002655	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	O157:H7 EDL933	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	3.298	0.584	0.114	2.714	1973154	50.383
NC_004431	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	CFT073	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	2.006	2.810	0.107	0.804	1781761	50.475
NC_010498	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	SECEC SMS-3-5	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	1.028	2.028	0.101	1.001	1665735	50.497
NC_013008	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	O157:H7 str. TW14359	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	0.961	1.752	0.117	0.791	2028866	50.507
NC_008253	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	536	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	6.464	8.611	0.098	2.146	1551242	50.518
NC_011353	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	O157:H7 str. EC4115	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	0.401	3.093	0.120	2.691	2029890	50.522
NC_002695	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	O157:H7 str. Sakai	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	9.524	6.797	0.111	2.727	2115289	50.537
NC_008563	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	APEC O1	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	2.440	4.292	0.108	1.852	1643800	50.547
NC_011601	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	O127:H6 str. E2348/69	GGTGGCGATAAGTGATATTATGTTAAAG	-(--- (CCCC (-))))) ---	2.673	2.676	0.098	0.003	1702574	50.566
NC_007946	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	UT189	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	0.043	2.460	0.101	2.417	1652986	50.604
NC_009801	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	E24377A	AGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	4.087	7.387	0.101	3.300	1701154	50.622
NC_011750	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	IA139	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	6.573	9.387	0.098	2.814	1869330	50.631
NC_011748	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	55989	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	1.541	4.447	0.106	2.906	1706398	50.656
NC_011742	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	S88	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	1.920	4.067	0.106	2.147	1591005	50.681
NC_011993	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	LF82	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	2.477	3.664	0.106	1.186	1552227	50.696
NC_011751	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	UMN026	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	0.804	4.257	0.104	3.453	1799797	50.716
NC_011745	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	ED1a	GATGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	8.444	8.732	0.121	0.288	1654531	50.733
NC_012967	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	B str. REL606	AGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	0.291	3.478	0.098	3.187	1567316	50.769
NC_010473	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	str. K-12 substr. DH10B	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	0.020	2.852	0.095	2.832	1678398	50.781
NC_000913	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	MG1655	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	1.165	4.141	0.097	2.976	1588773	50.790
NC_012759	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	BW2952	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	3.306	6.323	0.098	3.016	1480832	50.791
NC_011741	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	IA1I	GATGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	3.886	7.591	0.107	3.705	1600998	50.794
NC_011415	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	SE11	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	3.306	6.371	0.107	3.065	1691106	50.797
AC_000091	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	str. K-12 substr. W3110	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	15.848	12.876	0.094	2.972	1592463	50.800
NC_009800	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	HS	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	1.817	4.801	0.102	2.984	1607572	50.820
NC_012892	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	BL21	AGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	1.130	2.109	0.097	0.979	1548256	50.830
NC_012947	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	BL21(DE3)	AGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	4.092	0.862	0.095	3.230	2233508	50.841
NC_010468	Proteobacteria	Gammaproteobacteria	<i>Escherichia</i>	<i>coli</i>	ATCC 8739	AGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	1.519	1.394	0.098	0.125	2353088	50.867
NC_003103	Proteobacteria	Alphaproteobacteria	<i>Rickettsia</i>	<i>conorii</i>	str. Malish 7	GTTCATAACTCTAATTATGGAAAATAA	-(-(-(-(-(-(-))--)-)-)-)	17.672	13.507	0.186	4.165	696659	32.439
NC_011916	Proteobacteria	Alphaproteobacteria	<i>Caulobacter</i>	<i>crenscentus</i>	NA1000	AGTTGCTCAAGATATATTACGAGAAG	-(--- (-(-(-) (-))--)-)-)	4.090	5.335	0.053	1.245	1970161	67.172
NC_002696	Proteobacteria	Alphaproteobacteria	<i>Caulobacter</i>	<i>crenscentus</i>	CB15	AGTTCGTCAAGATATATTACGAGAAG	-(--- (-(-(-) (-))--)-)-)	5.280	6.533	0.053	1.253	1944185	67.214
NC_007520	Proteobacteria	Gammaproteobacteria	<i>Thiomicrospira</i>	<i>crunigena</i>	XCL-2	TTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	6.876	5.882	0.086	0.994	1320200	43.130
NC_007969	Proteobacteria	Gammaproteobacteria	<i>Psychrobacter</i>	<i>cryohalolentis</i>	K5	TTTTCCGATAAGTGATATTATGTTAAAC	-(--- (CCCC (-))))) ---	17.500	16.685	0.114	0.814	1678680	42.291
NC_009484	Proteobacteria	Alphaproteobacteria	<i>Acidiphilium</i>	<i>cryptum</i>	JF-5	ATGTCGCAAGAAAATATTGGAACT	-(-(-(-(-(-) (-))--)-)-)	10.298	90.745	0.043	80.447	1081116	67.993
NC_012880	Proteobacteria	Gammaproteobacteria	<i>Dickeya</i>	<i>dadantii</i>	Ech703	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	2.427	1.909	0.184	0.518	2250817	55.015
NC_011891	Proteobacteria	Deltaproteobacteria	<i>Anaeromyxobacter</i>	<i>dehalogenans</i>	2CP-1	AGTCGGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	3.019	2.996	0.042	0.022	2536077	74.716
NC_007760	Proteobacteria	Deltaproteobacteria	<i>Anaeromyxobacter</i>	<i>dehalogenans</i>	2CP-C	AGTCGGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	39.773	39.747	0.047	0.026	1906463	74.905
NC_008209	Proteobacteria	Alphaproteobacteria	<i>Roseobacter</i>	<i>denitrificans</i>	OCh 114	TATTTATAACTCAACTATTGTAAATAA	-(-(-(-(-(-) (-))--)-)-)	5.041	5.110	0.098	0.068	2529441	58.969
NC_007954	Proteobacteria	Gammaproteobacteria	<i>Shewanella</i>	<i>denitrificans</i>	OS217	GATGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	6.013	6.634	0.154	0.621	2182540	45.149
NC_007575	Proteobacteria	Epsilonproteobacteria	<i>Sulfurimonas</i>	<i>denitrificans</i>	DSM 1251								34.462
NC_007404	Proteobacteria	Betaproteobacteria	<i>Thiobacillus</i>	<i>denitrificans</i>	ATCC 25259	ACTTCGATAAGTGATATTATGTTAAAT	-(-(-(-(-(-) (-))--)-)-)	3.609	2.426	0.053	1.183	1430769	66.067
NC_007519	Proteobacteria	Deltaproteobacteria	<i>Desulfovibrio</i>	<i>desulfuricans</i>	subsp. desulfuricans str. G20	GTTCATAAGTGATATTATGTTAAATTT	-(--- (-(-(-) (-))--)-)-)	14.190	11.345	0.154	2.844	2012146	57.836
NC_011883	Proteobacteria	Deltaproteobacteria	<i>Desulfovibrio</i>	<i>desulfuricans</i>	subsp. desulfuricans str. ATCC 27774	GTCCATAAGTGATATTATGTTAAATTTT	-(--- (-(-(-) (-))--)-)-)	1.355	0.240	0.126	1.115	1425901	58.072
NC_011365	Proteobacteria	Alphaproteobacteria	<i>Gluconacetobacter</i>	<i>diazotrophicus</i>	PAI 5	ATCGTATAATATATTATGCCAACAG	-(-(-(-(-(-) (-))--)-)-)	3.820	23.536	0.081	19.716	212909	66.378
NC_010125	Proteobacteria	Alphaproteobacteria	<i>Gluconacetobacter</i>	<i>diazotrophicus</i>	PAI 5	ATCGTATAATATATTATGCCAACAG	-(-(-(-(-(-) (-))--)-)-)	29.911	35.726	0.072	5.815	2006457	66.431
NC_012988	Proteobacteria	Alphaproteobacteria	<i>Methylobacterium</i>	<i>dichloromethanicum</i>		TTGGCATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	147.373	92.466	0.016	54.907	5104353	68.088
NC_002940	Proteobacteria	Gammaproteobacteria	<i>Haemophilus</i>	<i>ducreyi</i>	35000HP	ACTTCGATAAGTGATATTATGTTAAAT	-(-(-(-(-(-) (-))--)-)-)	69.500	73.142	0.219	3.642	1177470	38.220
NC_007606	Proteobacteria	Gammaproteobacteria	<i>Shigella</i>	<i>dysenteriae</i>	Sd197	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	52.970	52.589	0.060	0.381	1482367	51.250
NC_008340	Proteobacteria	Gammaproteobacteria	<i>Alkalilimnicola</i>	<i>ehrlichii</i>	MLHE-1	ACTGGCGATAAGTGATATTATGTTAAAT	-(-(-(-(-(-) (-))--)-)-)	4.321	3.899	0.110	0.422	1617316	67.531
NC_008786	Proteobacteria	Betaproteobacteria	<i>Verminephrobacter</i>	<i>eiseniae</i>	EF01-2	GAGACATACGATGATATTATGTTAAAT	-(--- (-(-(-) (-))--)-)-)	65.735	54.445	0.021	11.290	3780130	65.279
NC_006833	Proteobacteria	Alphaproteobacteria	<i>Wolbachia</i>	<i>endosymbiont</i>	strain TRS of Brugia malayi	TTACATAATATATTATGTTAAATAAA	-(-(-(-(-(-) (-))--)-)-)	49.496	46.460	0.172	3.036	391541	34.179
NC_010981	Proteobacteria	Alphaproteobacteria	<i>Wolbachia</i>	<i>endosymbiont</i>	of Culex quinquefasciatus Pel	ATTCATAAGTGATATTATGTTAAATA	-(--- (-(-(-) (-))--)-)-)	3.138	21.124	0.114	17.986	16977	34.190
NC_002978	Proteobacteria	Alphaproteobacteria	<i>Wolbachia</i>	<i>endosymbiont</i>	of Drosophila melanogaster	TTCCATAAGTGATATTATGTTAAATTTCC	-(--- (-(-(-) (-))--)-)-)	6.409	79.377	0.088	72.969	438676	35.235
NC_010067	Proteobacteria	Gammaproteobacteria	<i>Salmonella</i>	<i>enterica</i>	subsp. arizonae serovar 62:z4,z23:-	AGTGGCGATAAGTGATATTATGTTAAAT	-(-(-(-(-(-) (-))--)-)-)	2.690	1.373	0.130	1.317	1445404	51.377
NC_004631	Proteobacteria	Gammaproteobacteria	<i>Salmonella</i>	<i>enterica</i>	subsp. enterica serovar Typhi str. Ty2	GAGCGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	11.939	12.190	0.118	0.250	1513293	52.052
NC_011149	Proteobacteria	Gammaproteobacteria	<i>Salmonella</i>	<i>enterica</i>	subsp. enterica serovar Agona str. SL483	AGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	1.260	1.040	0.128	0.220	1585903	52.084
NC_011083	Proteobacteria	Gammaproteobacteria	<i>Salmonella</i>	<i>enterica</i>	subsp. enterica serovar Heidelberg str. SL476	GGTGGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	2.581	2.581	0.125	0.598	1676807	52.087
NC_003198	Proteobacteria	Gammaproteobacteria	<i>Salmonella</i>	<i>enterica</i>	subsp. enterica serovar Typhi str. CT18	GGAGCGATAAGTGATATTATGTTAAAT	-(--- (CCCC (-))))) ---	8.105	7.859	0.121	0.246	1468603	52.093

Accession	Phylum	Class	Genus	Species	Strain	dif	Palindrome	dif angle	GC angle	GCSI	Δ	dif Position	GC%
NC_010102	Proteobacteria	Gammaproteobacteria	Salmonella	enterica	subsp. enterica serovar Paratyphi B str. SPB7	GGTGCCACATAGTATATTATGTAAAT	--(- (((((- -)))))) --	7.817	8.298	0.126	0.480	1493985	52.110
NC_006905	Proteobacteria	Gammaproteobacteria	Salmonella	enterica	subsp. enterica serovar Choleraesuis str. SC-B67	GGTGCCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	0.848	0.385	0.124	0.463	1652777	52.157
NC_012125	Proteobacteria	Gammaproteobacteria	Salmonella	enterica	subsp. enterica serovar Paratyphi C strain RKS4594	GGTGCCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	12.617	11.863	0.120	0.753	2247159	52.163
NC_006511	Proteobacteria	Gammaproteobacteria	Salmonella	enterica	subsp. enterica serovar Paratyphi A str. ATCC 9150	GGTGCCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	13.070	12.347	0.131	0.723	1400306	52.164
NC_011294	Proteobacteria	Gammaproteobacteria	Salmonella	enterica	subsp. enterica serovar Enteritidis str. P125109	GGTGCCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	0.721	1.080	0.128	0.359	1602396	52.167
NC_011094	Proteobacteria	Gammaproteobacteria	Salmonella	enterica	subsp. enterica serovar Schwarzengrund str. CVM19633	GGTGCCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	0.651	1.056	0.126	0.405	1597329	52.174
NC_011205	Proteobacteria	Gammaproteobacteria	Salmonella	enterica	subsp. enterica serovar Dublin str. CT_02021853	GGTGCCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	3.394	4.409	0.127	1.015	1737129	52.180
NC_011147	Proteobacteria	Gammaproteobacteria	Salmonella	enterica	subsp. enterica serovar Paratyphi A str. AKU_12601	GGTGCCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	13.253	12.530	0.131	0.723	1395487	52.184
NC_011274	Proteobacteria	Gammaproteobacteria	Salmonella	enterica	subsp. enterica serovar Gallinarum str. 287/91	GGTGCCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	18.263	19.120	0.123	0.857	1651259	52.205
NC_011080	Proteobacteria	Gammaproteobacteria	Salmonella	enterica	subsp. enterica serovar Newport str. SL254	GGTGCCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	0.758	0.327	0.127	0.430	1626651	52.217
NC_003197	Proteobacteria	Gammaproteobacteria	Salmonella	enterica	subsp. enterica serovar Typhimurium str. LT2	GGTGCCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	1.925	1.451	0.126	0.474	1629675	52.222
NC_008800	Proteobacteria	Gammaproteobacteria	Yersinia	enterocolitica	subsp. enterocolitica 8081	GGTGCCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	6.652	6.890	0.131	0.237	2222656	47.271
NC_008027	Proteobacteria	Gammaproteobacteria	Pseudomonas	entomophila	subsp. entomophila 8081	TGTGGCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	2.166	2.705	0.099	0.539	2919252	64.160
NC_007761	Proteobacteria	Alphaproteobacteria	Rhizobium	etli	CFN 42	AGATCACATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	5.272	3.653	0.065	1.618	2254965	61.272
NC_010994	Proteobacteria	Alphaproteobacteria	Rhizobium	etli	CIAT 652	AGATCACATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	1.836	17.373	0.060	15.537	2233649	61.665
NC_004757	Proteobacteria	Betaproteobacteria	Nitrosomonas	europaea	ATCC 19718	ATTGGTATAAAGTATATTATGTAAAT	(((--- (((((- -)))))) ---))	54.878	56.004	0.182	1.126	974204	50.718
NC_008344	Proteobacteria	Betaproteobacteria	Nitrosomonas	europhaa	C91	ATTGGTATAAAGTATATTATGTAAAT	(((--- (((((- -)))))) ---))	20.672	20.288	0.164	0.384	1279899	48.486
NC_010172	Proteobacteria	Alphaproteobacteria	Methylobacterium	extorquens	PA1	TTCGCATAAAGTATATTATGGAACATGA	(----- (----- (C) -----) -----)	142.905	77.967	0.014	64.938	4654342	68.184
NC_012808	Proteobacteria	Alphaproteobacteria	Methylobacterium	extorquens	AM1	TTCGCATAAAGTATATTATGGAACATGA	(----- (----- (C) -----) -----)	144.495	86.389	0.016	58.106	4690847	68.710
NC_010577	Proteobacteria	Gammaproteobacteria	Xylella	fastidiosa	M23	ACTTGGCATAAAGTATATTATGTCCGCG	(----- (((((- -)))))) -----	17.606	8.925	0.314	8.681	1143839	51.757
NC_004556	Proteobacteria	Gammaproteobacteria	Xylella	fastidiosa	Temecula1	ACTTGGCATAAAGTATATTATGTCCGCT	(----- (((((- -)))))) -----	17.907	9.209	0.319	8.698	1134561	51.776
NC_010513	Proteobacteria	Gammaproteobacteria	Xylella	fastidiosa	M12	ACTTGGCATAAAGTATATTATGTCCGCG	(----- (((((- -)))))) -----	8.915	1.096	0.320	7.820	1176268	51.917
NC_002488	Proteobacteria	Gammaproteobacteria	Xylella	fastidiosa	9a5c	ATTGGCATAAAGTATATTATGTCCGCT	(----- (((((- -)))))) -----	59.819	59.813	0.318	0.006	1784857	52.674
NC_007109	Proteobacteria	Alphaproteobacteria	Rickettsia	felis	URRWXCa2	GTTCTATAAAGTATATTATGGAACATGA	(----- (((((- -)))))) -----	35.131	35.103	0.153	0.028	887503	32.450
NC_011740	Proteobacteria	Gammaproteobacteria	Escherichia	fergusonii	ATCC 35469	AGTGGCATAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	19.725	20.767	0.098	1.043	1600724	49.936
NC_007908	Proteobacteria	Betaproteobacteria	Rhodospirillum	ferrireducens	T118	ACTGTAGCAGATGATATTATGTAAAT	(((--- (((((- -)))))) ---))	8.370	9.627	0.066	1.257	2472536	59.879
NC_011761	Proteobacteria	Gammaproteobacteria	Acidithiobacillus	ferrooxidans	ATCC 23270	GGTGCCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	26.792	25.789	0.116	1.003	1713151	58.772
NC_011206	Proteobacteria	Gammaproteobacteria	Acidithiobacillus	ferrooxidans	ATCC 53993	GTTGGCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	18.557	17.521	0.111	1.037	1594258	58.850
NC_007947	Proteobacteria	Betaproteobacteria	Methylobacillus	flagellatus	KT	ACTTGGCATAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	9.556	10.775	0.171	1.219	1564639	55.719
NC_004337	Proteobacteria	Gammaproteobacteria	Shigella	flexneri	2a str. 301	ACTAGCATAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	7.302	3.453	0.088	3.849	1611575	50.889
NC_004741	Proteobacteria	Gammaproteobacteria	Shigella	flexneri	2a str. 2457T	ACTAGCATAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	0.207	4.161	0.084	3.955	1651213	50.906
NC_008258	Proteobacteria	Gammaproteobacteria	Shigella	flexneri	5 str. 8401	ACTGGCATAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	2.530	3.474	0.071	0.944	1603268	50.918
NC_012660	Proteobacteria	Gammaproteobacteria	Pseudomonas	fluorescens	SBW25	ATTGGCATAAAGTATATTATGTAAAT	(((--- (((((- -)))))) ---))	1.834	2.196	0.081	0.363	3327027	60.505
NC_007492	Proteobacteria	Gammaproteobacteria	Pseudomonas	fluorescens	PF0-1	GTTAGCATAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	3.248	3.249	0.090	0.001	3150828	60.521
NC_004129	Proteobacteria	Gammaproteobacteria	Pseudomonas	fluorescens	PF-5	CATTGGCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	2.327	1.528	0.114	0.798	3461815	63.305
NC_008345	Proteobacteria	Gammaproteobacteria	Shewanella	frigidimarina	NCIMB 400	CATGGCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	4.519	3.600	0.148	0.919	2380403	41.584
NC_008554	Proteobacteria	Deltaproteobacteria	Syntrophobacter	fumaroxidans	MPOB	ACGTCTGAATGATATTATGTAAAT	(((--- (((((- -)))))) ---))	16.128	12.236	0.070	3.892	1053758	59.949
NC_007712	Proteobacteria	Gammaproteobacteria	Sodalis	glossiniidius	str. 'morsitans'	AGTAGCATAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	33.277	32.683	0.107	0.593	2471133	54.699
NC_011035	Proteobacteria	Betaproteobacteria	Neisseria	gonorrhoeae	NCCP11945	AGTTGGCATAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	45.168	40.384	0.146	4.784	835969	52.366
NC_002946	Proteobacteria	Betaproteobacteria	Neisseria	gonorrhoeae	FA 1090	AGTTGGCATAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	6.074	6.527	0.160	0.452	778161	52.690
NC_012846	Proteobacteria	Alphaproteobacteria	Bartonella	grahamii	as4aup	ATCGCATAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	7.747	7.736	0.297	0.011	1221051	38.057
NC_010334	Proteobacteria	Gammaproteobacteria	Shewanella	halifaxensis	HAW-EB4	TATGGCATAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	2.955	3.894	0.172	0.939	2656359	44.591
NC_008789	Proteobacteria	Gammaproteobacteria	Halorhodospira	halophila	SL1	GCTGGCATAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	173.054	173.653	0.091	0.599	2598940	67.982
NC_007964	Proteobacteria	Alphaproteobacteria	Nitrospira	hamburgensis	X14	AATTGGCATAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	8.634	8.017	0.061	0.617	2229955	61.714
NC_005956	Proteobacteria	Alphaproteobacteria	Bartonella	henselae	str. Houston-1	ATCGCATAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	1.493	1.896	0.327	0.403	957515	38.230
NC_004917	Proteobacteria	Epsilonproteobacteria	Helicobacter	hepaticus	ATCC 51449	TCTAAGCAGATGATATTATGTAAAT	(----- (((((- -)))))) -----	17.509	41.193	0.223	23.684	119688	35.931
NC_008570	Proteobacteria	Gammaproteobacteria	Aeromonas	hydrophila	subsp. hydrophila ATCC 7966	ACCGGCCATAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	10.828	9.352	0.113	1.476	2514922	61.548
NC_012779	Proteobacteria	Gammaproteobacteria	Edwardsiella	ictaluri	93-146	AATTGGCATAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	0.462	1.453	0.180	0.991	1901266	57.437
NC_010581	Proteobacteria	Alphaproteobacteria	Beijerinckia	indica	subsp. indica ATCC 9039	GATTGGCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	19.933	10.803	0.101	9.130	1730914	57.060
NC_009566	Proteobacteria	Gammaproteobacteria	Haemophilus	influenzae	PittEE	TTCGCATAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	4.261	4.583	0.112	0.322	932931	38.400
NC_000907	Proteobacteria	Gammaproteobacteria	Haemophilus	influenzae	Rd KW20	ATTGGCATAAAGTATATTATGTAAAT	(((--- (((((- -)))))) ---))	9.866	9.517	0.104	0.349	1473961	38.148
NC_007146	Proteobacteria	Gammaproteobacteria	Haemophilus	influenzae	86-028NP	ATTGGCATAAAGTATATTATGTAAAT	(((--- (((((- -)))))) ---))	21.098	21.399	0.110	0.301	1623101	38.157
NC_008709	Proteobacteria	Gammaproteobacteria	Psychromonas	ingrahamii	37	AATGGCATAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	3.160	3.086	0.143	0.074	2335814	40.086
NC_008011	Proteobacteria	Deltaproteobacteria	Lawsonia	intracellularis	PHE/MN1-00	GTCGCAAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	15.715	8.224	0.259	7.491	314416	33.284
NC_004463	Proteobacteria	Alphaproteobacteria	Bradyrhizobium	japonicum	USDA 110	GATTGGCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	10.204	14.254	0.041	4.049	4996160	64.059
NC_010995	Proteobacteria	Gammaproteobacteria	Cellvibrio	japonicus	Ueda107	ACTGGCATAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	11.950	12.094	0.171	0.144	2127091	51.991
NC_013166	Proteobacteria	Gammaproteobacteria	Kangiella	koreensis	DSM 16069	TATTGGCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	0.423	0.863	0.202	0.440	1422688	43.687
NC_009792	Proteobacteria	Gammaproteobacteria	Citrobacter	koseri	ATCC BAA-895	AGTGGCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	66.263	66.258	0.135	0.006	1491360	53.828
NC_009719	Proteobacteria	Alphaproteobacteria	Parvibaculum	lavamentivorans	DS-1	ATAGCCATAAAGTATATTATGTAAAT	(----- (((((- -)))))) -----	73.943	92.174	0.086	18.232	3329662	62.328
NC_012850	Proteobacteria	Alphaproteobacteria	Rhizobium	leguminosarum	bv. trifolii WSM1325	GTTGGCATAAAGTATATTATGTAAAT	--(- (((((- -)))))) --	1.627	15.088	0.063	13.461	1964456	61.087

Accession	Phylum	Class	Genus	Species	Strain	dif	Palindrome	dif angle	GC angle	GCSI	A	dif Position	GC%
NC_008380	Proteobacteria	Alphaproteobacteria	<i>Rhizobium</i>	<i>leguminosarum</i>	bv. viciae 3841	CTATCGCATAAGATAGATTATGGAACATA	---(- (CCCC(-O-))))) ---))	4.426	8.986	0.060	4.559	2590747	61.090
NC_011369	Proteobacteria	Alphaproteobacteria	<i>Rhizobium</i>	<i>leguminosarum</i>	bv. trifolii WSM2304	AAATCGCATAAGACAGATTATGGAACAT	-(- (CCCC(-)) ---)) ---))	0.464	138.952	0.062	138.488	1837698	61.531
NC_007722	Proteobacteria	Alphaproteobacteria	<i>Erythrobacter</i>	<i>litoralis</i>	HTCC2594	GAATGATAATATATATATGTAAMATAA	-(- (-(--- (CO) ---)) ---))	30.912	31.145	0.037	0.233	1297704	63.069
NC_009092	Proteobacteria	Gammaproteobacteria	<i>Shewanella</i>	<i>loihica</i>	PV-4	CATCGGCACATGTATATATGTTAAAT	-(-(- (CCC(-O-))))) ---))	0.631	0.636	0.156	0.005	2292256	53.667
NC_006512	Proteobacteria	Gammaproteobacteria	<i>Idiomarina</i>	<i>lohiensis</i>	L2TR	ATTGCGTATAAGTATATATATGTTAAAT	(((--- (CCCC(-O-))))) ---))	2.912	91.323	0.169	88.411	1387609	47.040
NC_002678	Proteobacteria	Alphaproteobacteria	<i>Mesorhizobium</i>	<i>loti</i>	MAFF303099	AAGTCGCATAAGATAGATTATGGAACAT	(((--- (CCCC(-O-))))) ---))	28.235	32.566	0.050	4.331	299607	62.747
NC_010814	Proteobacteria	Deltaproteobacteria	<i>Geobacter</i>	<i>lovleyi</i>	SZ	TADGCCATAAAGCTATATATGTTAAAGT	----- (CCC(-O-))))) -----	7.292	7.782	0.118	0.491	2038233	54.770
NC_005126	Proteobacteria	Gammaproteobacteria	<i>Photorhabdus</i>	<i>luminescens</i>	subsp. laumondii TTO1	AGTAGCATAAAGTATATATGTTAAAT	-(-(- (CCCC(-O-))))) ---))	11.452	9.554	0.231	1.899	2646810	42.825
NC_011138	Proteobacteria	Gammaproteobacteria	<i>Alteromonas</i>	<i>macleodii</i>	Deep ecotype'	ATTACGCATAAAGTATATATGTTAAAT	(((--- (CCCC(-O-))))) ---))	1.588	2.534	0.103	0.945	2186675	44.879
NC_007626	Proteobacteria	Alphaproteobacteria	<i>Magnetospirillum</i>	<i>magneticum</i>	AMB-1	CGGCATAATGTATATATGGCGAGCA	---(- (--- (CO) ---) ---))	9.772	10.883	0.067	1.111	2610323	65.094
NC_012796	Proteobacteria	Deltaproteobacteria	<i>Desulfovibrio</i>	<i>magneticus</i>	RS-1	TGCTCATAAGTAAATATGTTAACT	---(- (CCCC(-O-))))) ---))	14.965	10.380	0.090	4.585	2464808	62.769
NC_011071	Proteobacteria	Gammaproteobacteria	<i>Stenotrophomonas</i>	<i>maltophilia</i>	R551-3	ATTTCGCATAAGTATATATGTTACAG	-(- (CCCC(-O-))))) ---))	5.106	4.679	0.074	0.427	2212261	66.300
NC_010943	Proteobacteria	Gammaproteobacteria	<i>Stenotrophomonas</i>	<i>maltophilia</i>	K279a	ATTTCGCATAAGTATATATGTTCAAG	---(- (CCCC(-O-))))) ---))	2.585	0.281	0.074	2.304	2436410	66.324
NC_004842	Proteobacteria	Alphaproteobacteria	<i>Anaplasma</i>	<i>marginale</i>	str. St. Maries	AATACATAATATATATATGGAAATAA	----- (--- (CO) ---) -----	21.498	21.747	0.168	0.250	527323	49.763
NC_012026	Proteobacteria	Alphaproteobacteria	<i>Anaplasma</i>	<i>marginale</i>	str. Florida	AATACATAATATATATATGGAAATAA	----- (--- (CO) ---) -----	22.467	22.715	0.171	0.249	526177	49.774
NC_008347	Proteobacteria	Alphaproteobacteria	<i>Maricaulis</i>	<i>maris</i>	MCS10	TTCCGCATAAATAATATATGAGAAAT	-(- (CCCC(-O-))))) ---))	1.829	1.283	0.083	0.547	1591541	62.728
NC_009900	Proteobacteria	Alphaproteobacteria	<i>Rickettsia</i>	<i>massiliae</i>	MTU5	GTTCATAAATCTATATATGGAAATAA	-(- (--- (CO) ---) ---))	34.511	34.499	0.180	0.012	810910	32.543
NC_009636	Proteobacteria	Alphaproteobacteria	<i>Sinorhizobium</i>	<i>medicae</i>	WSM419	GATCGCATAAAGATAGATTATGGAACATA	-(- (CCC(-O-))))) ---))	10.080	8.797	0.061	1.283	1384114	61.502
NC_003047	Proteobacteria	Alphaproteobacteria	<i>Sinorhizobium</i>	<i>meliloti</i>	1021	GCATCATAAAGATAGATTATGGAACATA	-(- (CCC(-O-))))) ---))	9.009	8.472	0.059	0.537	1735626	62.729
NC_009439	Proteobacteria	Gammaproteobacteria	<i>Pseudomonas</i>	<i>mendocina</i>	ymp	TGTCGCATAAAGTATATATGTTAAAT	---(- (CCCC(-O-))))) ---))	5.792	5.794	0.099	0.002	2441422	64.676
NC_003112	Proteobacteria	Betaproteobacteria	<i>Neisseria</i>	<i>meningitidis</i>	MCS58	AGTTCCGATAAGTATATATGTTAAAT	-(- (CCCC(-O-))))) ---))	12.581	13.022	0.162	0.441	1229332	51.528
NC_008767	Proteobacteria	Betaproteobacteria	<i>Neisseria</i>	<i>meningitidis</i>	FAM18	AGTTCCGATAAGTATATATGTTAAAT	-(- (CCCC(-O-))))) ---))	0.634	1.090	0.167	0.456	1101344	51.624
NC_010120	Proteobacteria	Betaproteobacteria	<i>Neisseria</i>	<i>meningitidis</i>	53442	AGTTCCGATAAGTATATATGTTAAAT	-(- (CCCC(-O-))))) ---))	0.061	0.404	0.163	0.344	1090911	51.700
NC_003116	Proteobacteria	Betaproteobacteria	<i>Neisseria</i>	<i>meningitidis</i>	Z2491	AGTTCCGATAAGTATATATGTTAAAT	-(- (CCCC(-O-))))) ---))	8.137	7.685	0.158	0.452	1290393	51.809
NC_013016	Proteobacteria	Betaproteobacteria	<i>Neisseria</i>	<i>meningitidis</i>	alpha14	AGTTCCGATAAGTATATATGTTAAAT	-(- (CCCC(-O-))))) ---))	0.720	0.026	0.171	0.694	1068358	51.947
NC_007517	Proteobacteria	Deltaproteobacteria	<i>Geobacter</i>	<i>metallireducens</i>	GS-15	ACGTCGCATAAAGATATATATGTTCAAGT	-(- (CCCC(-O-))))) ---))	2.245	4.211	0.080	1.966	2023640	59.510
NC_010554	Proteobacteria	Gammaproteobacteria	<i>Proteus</i>	<i>mirabilis</i>	HI4320	GGTTCCGATAAAGTATATATGTTAAAT	-(- (CCCC(-O-))))) ---))	2.938	0.935	0.173	2.003	1358031	38.900
NC_012968	Proteobacteria	Betaproteobacteria	<i>Methylotenera</i>	<i>mobilis</i>	JLW8	ATTTGCTATAAGTATATATGTTAAAT	(((--- (CCCC(-O-))))) ---))	10.337	7.348	0.224	2.990	1200631	45.510
NC_006526	Proteobacteria	Alphaproteobacteria	<i>Zymomonas</i>	<i>mobilis</i>	subsp. mobilis ZM4	GTCTGATAATATATATGTTAAATAA	-(- (--- (CO) ---) ---))	13.520	124.908	0.138	111.388	1105438	46.328
NC_002663	Proteobacteria	Gammaproteobacteria	<i>Pasteurella</i>	<i>multocida</i>	subsp. multocida str. Pm70	ACCTTCGCATAAGTATATATGTTAAAT	-(- (CCCC(-O-))))) ---))	26.312	26.537	0.150	0.225	713836	40.405
NC_008781	Proteobacteria	Betaproteobacteria	<i>Polaromonas</i>	<i>naphthalenivorans</i>	CJ2	ACATTCGACAGTGATATATGTTAAAT	-(- (--- (CO) ---) ---))	4.541	5.560	0.050	1.018	2113787	62.525
NC_010531	Proteobacteria	Betaproteobacteria	<i>Polynucleobacter</i>	<i>necessarius</i>	subsp. necessarius STIR1	TTCGGTATAAATCATATATGTTAAAT	-(-(- (CCCC(-O-))))) ---))	5.135	6.418	0.199	1.283	757975	45.557
NC_008358	Proteobacteria	Alphaproteobacteria	<i>Hyphomonas</i>	<i>neptunium</i>	ATCC 15444	GTCGCATAAATCATATATGTTAAAT	----- (--- (CO) ---) -----	8.344	6.163	0.081	2.181	1856932	61.930
NC_009446	Proteobacteria	Gammaproteobacteria	<i>Dichelobacter</i>	<i>nodosus</i>	VRS1703A	TATTCGCAGAAATGATATATGTTAAAT	-(- (--- (CO) ---) ---))	17.886	18.191	0.170	0.305	763703	44.396
NC_011894	Proteobacteria	Alphaproteobacteria	<i>Methylobacterium</i>	<i>nodulans</i>	OCS 2060	GATTCGCATAAAGGAGATTATGGAACAT	-(- (--- (CO) ---) ---))	17.024	128.756	0.008	111.732	2476039	68.857
NC_008601	Proteobacteria	Gammaproteobacteria	<i>Francisella</i>	<i>novicida</i>	U112	CATTCCGATAAATATATATGTTAAAT	-(- (CCCC(-O-))))) ---))	6.831	14.030	0.281	7.200	918774	32.475
NC_007484	Proteobacteria	Gammaproteobacteria	<i>Nitrosococcus</i>	<i>oceanii</i>	ATCC 19707	TGTTCCGATAAAGATATATGTTAAAT	-(- (CCCC(-O-))))) ---))	11.327	11.267	0.202	0.060	1850395	50.318
NC_009943	Proteobacteria	Deltaproteobacteria	<i>Desulfococcus</i>	<i>oleovorans</i>	Hxd3	ACGCTGATAAAGATATATGTTAAAT	-(- (--- (CO) ---) ---))	0.781	1.109	0.087	0.328	2066614	56.169
NC_004347	Proteobacteria	Gammaproteobacteria	<i>Shewanella</i>	<i>oneidensis</i>	MR-1	ACTGGCACAATGTATATGTTAAAT	-(- (--- (CO) ---) ---))	0.579	0.101	0.160	0.478	2476914	45.961
NC_010717	Proteobacteria	Gammaproteobacteria	<i>Xanthomonas</i>	<i>oryzae</i>	pv. oryzae PXO99A	ACTTCGCATAAAGTATATGTTAAACAT	-(- (CCCC(-O-))))) ---))	6.609	8.267	0.069	1.659	2372929	63.634
NC_006834	Proteobacteria	Gammaproteobacteria	<i>Xanthomonas</i>	<i>oryzae</i>	pv. oryzae KACC10331	ATTTCGCATAAAGTATATGTTAAACAT	-(- (CCCC(-O-))))) ---))	8.410	10.655	0.067	2.245	2390366	63.693
NC_007705	Proteobacteria	Gammaproteobacteria	<i>Xanthomonas</i>	<i>oryzae</i>	pv. oryzae MAFF 311018	ACTTCGCATAAAGTATATGTTAAACAT	-(- (CCCC(-O-))))) ---))	7.170	9.527	0.068	2.357	2369935	63.701
NC_006677	Proteobacteria	Alphaproteobacteria	<i>Gluconobacter</i>	<i>oxydans</i>	621H	GTCTGATAAAGTATATGTTAAACAT	-(- (--- (CO) ---) ---))	35.687	15.701	0.182	19.986	18710	61.072
NC_008435	Proteobacteria	Alphaproteobacteria	<i>Rhodospseudomonas</i>	<i>palustris</i>	BisA53	GATTCCGATAAAGGATATATGTTAAAT	---(- (CCCC(-O-))))) ---))	4.441	3.565	0.045	0.877	3224547	64.443
NC_007958	Proteobacteria	Alphaproteobacteria	<i>Rhodospseudomonas</i>	<i>palustris</i>	BisB5	GATTCCGATAAAGGATATATGTTAAAT	-(- (CCCC(-O-))))) ---))	9.340	7.670	0.057	1.670	3057023	64.808
NC_011004	Proteobacteria	Alphaproteobacteria	<i>Rhodospseudomonas</i>	<i>palustris</i>	TIE-1	TATTCGCATAAAGGATATATGTTAAAT	---(- (CCCC(-O-))))) ---))	11.018	10.483	0.063	0.534	3368038	64.860
NC_007925	Proteobacteria	Alphaproteobacteria	<i>Rhodospseudomonas</i>	<i>palustris</i>	BisB18	AATTCCGATAAAGGATATATGTTAAAT	-(- (CCCC(-O-))))) ---))	12.554	12.451	0.045	0.103	2941089	64.957
NC_005296	Proteobacteria	Alphaproteobacteria	<i>Rhodospseudomonas</i>	<i>palustris</i>	CGA009	GATTCCGATAAAGGATATATGTTAAAT	---(- (CCCC(-O-))))) ---))	5.937	5.375	0.063	0.562	3156836	65.038
NC_007778	Proteobacteria	Alphaproteobacteria	<i>Rhodospseudomonas</i>	<i>palustris</i>	HaA2	AGTTCCGATAAAGGATATATGTTAAAT	-(- (CCCC(-O-))))) ---))	29.461	28.896	0.047	0.565	3071275	65.038
NC_002928	Proteobacteria	Betaproteobacteria	<i>Bordetella</i>	<i>parapertussis</i>	12822	AATTCCGATAAAGTATATGTTAAAGT	-(- (CCCC(-O-))))) ---))	37.668	37.266	0.106	0.402	2905082	68.100
NC_011852	Proteobacteria	Gammaproteobacteria	<i>Haemophilus</i>	<i>parvus</i>	SH0165	TGTTCCGATAAAGTATATGTTAAAT	-(- (CCCC(-O-))))) ---))	118.850	118.369	0.049	0.481	433934	39.993
NC_013209	Proteobacteria	Alphaproteobacteria	<i>Acetobacter</i>	<i>pasteurianus</i>	IFO 3283-01	ATTGTATAAGATATATATGTTAAACAA	---(- (--- (CO) ---) ---))	29.158	6.225	0.182	22.933	1218259	53.037
NC_012730	Proteobacteria	Alphaproteobacteria	<i>Rickettsia</i>	<i>peacockii</i>	str. Rustic	GTTCATAAATCTATATATGTTAAATAA	-(- (--- (CO) ---) ---))	47.412	86.831	0.097	39.419	792989	32.561
NC_009901	Proteobacteria	Gammaproteobacteria	<i>Shewanella</i>	<i>pealeana</i>	ATCC 700345	TADGCCCAATGTATATATGTTAAAT	-(- (CCCC(-O-))))) ---))	6.380	6.235	0.179	0.145	2678992	44.662
NC_002929	Proteobacteria	Betaproteobacteria	<i>Bordetella</i>	<i>pertussis</i>	Tohama I	AATTCCGATAAAGTATATGTTAAAGT	-(- (CCCC(-O-))))) ---))	10.206	10.112	0.059	0.094	2229054	67.721
NC_008149	Proteobacteria	Gammaproteobacteria	<i>Yersinia</i>	<i>pestis</i>	Nepal516	GGTGACATAAATGTATATGTTAAAT	-(- (CCCC(-O-))))) ---))	24.084	24.186	0.153	0.102	1963931	47.584
NC_010159	Proteobacteria	Gammaproteobacteria	<i>Yersinia</i>	<i>pestis</i>	Angola	GATGACATAAATGTATATGTTAAAT	-(- (CCCC(-O-))))) ---))	46.984	132.047	0.027	85.062	2382781	47.598
NC_003143	Proteobacteria	Gammaproteobacteria	<i>Yersinia</i>	<i>pestis</i>	Con92	GGTGACATAAATGTATATGTTAAAT	-(- (CCCC(-O-))))) ---))	18.260	18.292	0.110	0.032	2562905	47.636
NC_004088	Proteobacteria	Gammaproteobacteria	<i>Yersinia</i>	<i>pestis</i>	KIM	GGTGACATAAATGTATATGTTAAAT	-(- (CCCC(-O-))))) ---))	21.698	21.798	0.133	0.100	2330525	47.641
NC_009381	Proteobacteria	Gammaproteobacteria	<i>Yersinia</i>	<i>pestis</i>	Pestoides F	GGTGACATAAATGTATATGTTAAAT	-(- (CCCC(-O-))))) ---))	99.167	99.004	0.141	0.163	981520	47.644

Accession	Phylum	Class	Genus	Species	Strain	dif	Palindrome	dif angle	GC angle	GCSI	A	dif Position	GC%
NC_005810	Proteobacteria	Gammaproteobacteria	<i>Yersinia</i>	<i>pestis</i>	biovar Microtus str. 91001	GGTGACATAAATGATATATGTTAAAT	--(C (CCCC (O))))) --	0.078	0.022	0.049	0.056	2298533	47.650
NC_008150	Proteobacteria	Gammaproteobacteria	<i>Yersinia</i>	<i>pestis</i>	Antiqua	GGTGACATAAATGATATATGTTAAAT	--(C (CCCC (O))))) --	40.293	39.899	0.106	0.394	1824845	47.701
NC_010170	Proteobacteria	Betaproteobacteria	<i>Bordetella</i>	<i>petrii</i>	DSM 12804	ATTGCCGATAATGATATATGTTAAAT	-(C (CCCC (O))))) --	4.512	3.076	0.063	1.436	2556070	65.483
NC_008825	Proteobacteria	Betaproteobacteria	<i>Methylobium</i>	<i>petroleiphilum</i>	PM1	AACCGCCCAATGATATATGTTCAACT	-(C (CCCC (O))))) --	3.395	3.020	0.050	0.374	1983963	69.198
NC_007797	Proteobacteria	Alphaproteobacteria	<i>Anaplasma</i>	<i>phagocytophilum</i>	HZ	AACACATAAATCTATTATGGAAATAAT	----- (----- O) -----	2.709	1.706	0.277	1.004	724568	41.636
NC_010336	Proteobacteria	Gammaproteobacteria	<i>Francisella</i>	<i>philomiragia</i>	subsp. philomiragia ATCC 25017	AATTCCGATAAATATATATGTTAAAT	-(C (CCCC (O))))) --	11.925	6.214	0.298	5.711	1884294	32.573
NC_013194	Proteobacteria	Betaproteobacteria	<i>Accumulibacter</i>	<i>phosphatis</i>	phosphatis clade IIA str. UW-1	AGTAGCCATAAATGTTGATATGTTAAAT	-(C (CCCC (O))))) --	21.707	21.367	0.068	0.341	2834280	64.111
NC_011566	Proteobacteria	Gammaproteobacteria	<i>Shewanella</i>	<i>piezotolerans</i>	WP3	AATGCCGCAATGATATATGTTAAAT	-(C (CCCC (O))))) --	0.219	0.607	0.175	0.388	2694957	43.292
NC_010939	Proteobacteria	Gammaproteobacteria	<i>Actinobacillus</i>	<i>pleuropneumoniae</i>	serovar 7 str. AP76	ACTTCCGATAAAGATATATGTTAAAT	-(C (CCCC (O))))) --	2.635	1.676	0.143	0.960	768440	41.212
NC_010278	Proteobacteria	Gammaproteobacteria	<i>Actinobacillus</i>	<i>pleuropneumoniae</i>	serovar 3 str. JL03	ACTTCCGATAAAGATATATGTTAAAT	-(C (CCCC (O))))) --	5.970	6.968	0.154	0.998	695433	41.233
NC_009053	Proteobacteria	Gammaproteobacteria	<i>Actinobacillus</i>	<i>pleuropneumoniae</i>	L20	ACTTCCGATAAAGATATATGTTAAAT	-(C (CCCC (O))))) --	1.987	2.972	0.152	0.984	742267	41.300
NC_011283	Proteobacteria	Gammaproteobacteria	<i>Klebsiella</i>	<i>pneumoniae</i>	342	AGTGCCGATAATGATATATGTTAAAT	-(C (CCCC (O))))) --	4.026	4.037	0.102	0.011	2724469	57.286
NC_009648	Proteobacteria	Gammaproteobacteria	<i>Klebsiella</i>	<i>pneumoniae</i>	subsp. pneumoniae MGH 78578	AGTGCCGATAATGATATATGTTAAAT	-(C (CCCC (O))))) --	0.546	1.089	0.099	0.543	1892068	57.479
NC_012731	Proteobacteria	Gammaproteobacteria	<i>Klebsiella</i>	<i>pneumoniae</i>	NTUH-K2044	AGTGCCGATAATGATATATGTTAAAT	-(C (CCCC (O))))) --	0.658	1.127	0.101	0.469	2633851	57.677
NC_003911	Proteobacteria	Alphaproteobacteria	<i>Ruegeria</i>	<i>pomeroyi</i>	DSS-3	TTCCTATAATCGGATATGATATGTTAAAT	-(C (CCCC (O))))) --	2.855	3.009	0.082	0.154	2087306	64.223
NC_010725	Proteobacteria	Alphaproteobacteria	<i>Methylobacterium</i>	<i>populi</i>	BJ001	TGTTCCGATAAAGATATATGTTAAAT	-(C (CCCC (O))))) --	145.015	82.402	0.015	62.613	5016360	69.390
NC_012115	Proteobacteria	Epsilonproteobacteria	<i>Nautilia</i>	<i>profundicola</i>	AmH	TTAGAGATAACATATGTTAACTAA	-(C (CCCC (O))))) --	4.382	4.349	0.289	0.033	858626	33.508
NC_008609	Proteobacteria	Deltaproteobacteria	<i>Pelobacter</i>	<i>propionicus</i>	DSM 2379	ATGCCGATAAAGATATATGTTAAAT	-(C (CCCC (O))))) --	11.159	4.989	0.083	6.171	1887808	59.024
NC_009832	Proteobacteria	Gammaproteobacteria	<i>Serratia</i>	<i>proteamaculans</i>	568	GGTGCCGATAATGATATATGTTAAAT	-(C (CCCC (O))))) --	2.887	2.881	0.127	0.007	2680725	55.074
NC_000963	Proteobacteria	Alphaproteobacteria	<i>Rickettsia</i>	<i>proვაქიი</i>	str. Madrid E	TTCCTATAATGATATGTTAAATAA	-(C (CCCC (O))))) --	13.063	23.455	0.214	10.392	596094	29.000
NC_010634	Proteobacteria	Gammaproteobacteria	<i>Yersinia</i>	<i>pseudotuberculosis</i>	PB1/+	GGTGCCGATAATGATATATGTTAAAT	-(C (CCCC (O))))) --	12.543	11.507	0.149	1.035	2511410	47.529
NC_010465	Proteobacteria	Gammaproteobacteria	<i>Yersinia</i>	<i>pseudotuberculosis</i>	YPIII	GGTGCCGATAATGATATATGTTAAAT	-(C (CCCC (O))))) --	9.513	8.573	0.145	0.941	2187957	47.530
NC_009708	Proteobacteria	Gammaproteobacteria	<i>Yersinia</i>	<i>pseudotuberculosis</i>	IP 31758	GGTGCCGATAATGATATATGTTAAAT	-(C (CCCC (O))))) --	15.854	15.066	0.151	0.788	2153648	47.538
NC_006155	Proteobacteria	Gammaproteobacteria	<i>Yersinia</i>	<i>pseudotuberculosis</i>	IP 32953	GGTGCCGATAATGATATATGTTAAAT	-(C (CCCC (O))))) --	16.694	15.866	0.149	0.828	2592362	47.610
NC_010501	Proteobacteria	Gammaproteobacteria	<i>Pseudomonas</i>	<i>putida</i>	W619	GGTGCCGATAAATGATATATGTTAAAT	-(C (CCCC (O))))) --	1.274	0.019	0.093	1.255	2907604	61.436
NC_002947	Proteobacteria	Gammaproteobacteria	<i>Pseudomonas</i>	<i>putida</i>	KT2440	GGTGCCGATAAATGATATATGTTAAAT	-(C (CCCC (O))))) --	37.261	37.222	0.105	0.040	3730777	61.524
NC_009512	Proteobacteria	Gammaproteobacteria	<i>Pseudomonas</i>	<i>putida</i>	F1	GTTATGCGATAATGCGATATGTTAAAT	-(C (CCCC (O))))) --	8.940	8.942	0.100	0.002	2794947	61.862
NC_010322	Proteobacteria	Gammaproteobacteria	<i>Pseudomonas</i>	<i>putida</i>	GB-1	GTTACTATAATGATATATGTTAAAT	-(C (CCCC (O))))) --	5.124	5.177	0.098	0.053	2927155	61.942
NC_009438	Proteobacteria	Gammaproteobacteria	<i>Shewanella</i>	<i>putrefaciens</i>	CN-32	ACTGCCGACAAATGATATATGTTAAAT	-(C (CCCC (O))))) --	0.136	3.839	0.176	3.702	2284415	44.452
NC_005955	Proteobacteria	Alphaproteobacteria	<i>Bartonella</i>	<i>quintana</i>	str. Toulouse	TTCGATAAATATATATGTTAAATTTT	-(C (CCCC (O))))) --	15.890	15.339	0.330	0.552	720890	38.797
NC_010505	Proteobacteria	Alphaproteobacteria	<i>Methylobacterium</i>	<i>radiotolerans</i>	JCM 2831	TGTTCCGATAAAGATATATGTTAAAT	-(C (CCCC (O))))) --	84.055	178.139	0.013	94.084	1619825	71.502
NC_013223	Proteobacteria	Deltaproteobacteria	<i>Desulfobalobium</i>	<i>retbaense</i>	DSM 5692	TGTTCTATAATGATATATGTTAAAT	-(C (CCCC (O))))) --	22.032	6.432	0.171	15.600	1579797	57.540
NC_010263	Proteobacteria	Alphaproteobacteria	<i>Rickettsia</i>	<i>rickettsii</i>	str. Iowa	GTTCTATAAATCTATAATGTTAAATAA	-(C (CCCC (O))))) --	20.027	15.881	0.185	4.146	704636	32.447
NC_009882	Proteobacteria	Alphaproteobacteria	<i>Rickettsia</i>	<i>rickettsii</i>	str. 'Sheila Smith'	GTTCTATAAATCTATAATGTTAAATAA	-(C (CCCC (O))))) --	21.450	17.266	0.185	4.184	703793	32.467
NC_013009	Proteobacteria	Alphaproteobacteria	<i>Neorickettsia</i>	<i>rusticii</i>	str. Illinois	ACACATAATTTGCTTATGTTAAATAA	-(C (CCCC (O))))) --	18.202	1.207	0.241	16.995	479647	41.271
NC_007643	Proteobacteria	Alphaproteobacteria	<i>Rhodospirillum</i>	<i>rubrum</i>	ATCC 11170	TGTCCGATAAATGATATATGTTAAAT	-(C (CCCC (O))))) --	0.664	4.130	0.089	3.466	1986368	65.448
NC_005295	Proteobacteria	Alphaproteobacteria	<i>Ehrlichia</i>	<i>ruminantium</i>	str. Welgevonden	TTCATAATGATATATGTTAAATAA	-(C (CCCC (O))))) --	2.425	3.995	0.517	1.570	747964	27.479
NC_006832	Proteobacteria	Alphaproteobacteria	<i>Ehrlichia</i>	<i>ruminantium</i>	str. Welgevonden	TTCATAATGATATATGTTAAATAA	-(C (CCCC (O))))) --	5.894	7.468	0.519	1.574	727812	27.479
NC_006831	Proteobacteria	Alphaproteobacteria	<i>Ehrlichia</i>	<i>ruminantium</i>	str. Gardel	TTCATAATGATATATGTTAAATAA	-(C (CCCC (O))))) --	5.424	7.008	0.516	1.584	723483	27.513
NC_009778	Proteobacteria	Gammaproteobacteria	<i>Cronobacter</i>	<i>sakazakii</i>	ATCC BAA-894	GGTGCCGATAATGATATATGTTAAAT	-(C (CCCC (O))))) --	0.431	2.057	0.073	1.626	1787147	56.767
NC_007963	Proteobacteria	Gammaproteobacteria	<i>Chromohalobacter</i>	<i>salexigens</i>	DSM 3043	GTTTCCGATAAATGATATATGTTAAAT	-(C (CCCC (O))))) --	28.102	27.628	0.137	0.474	2136885	63.912
NC_012881	Proteobacteria	Deltaproteobacteria	<i>Desulfobalobium</i>	<i>salexigens</i>	DSM 2638	TGTTCTATAATGATATATGTTAAAT	-(C (CCCC (O))))) --	3.664	3.084	0.208	0.580	2130676	47.093
NC_009348	Proteobacteria	Gammaproteobacteria	<i>Aeromonas</i>	<i>salmonicida</i>	subsp. salmonicida A449	ACTGCCGATAATGATATATGTTAAAT	-(C (CCCC (O))))) --	16.564	16.486	0.104	0.079	2134832	58.511
NC_009831	Proteobacteria	Gammaproteobacteria	<i>Shewanella</i>	<i>sediminis</i>	HAW-EB3	TGTCCGACAAATGATATATGTTAAAT	-(C (CCCC (O))))) --	5.152	4.119	0.163	1.033	2663706	46.078
NC_007798	Proteobacteria	Alphaproteobacteria	<i>Neorickettsia</i>	<i>senetsu</i>	str. Miyayama	ACACATAATTTGCTTATGTTAAATAA	-(C (CCCC (O))))) --	16.044	3.705	0.244	12.339	467786	41.084
NC_009952	Proteobacteria	Alphaproteobacteria	<i>Dinoroseobacter</i>	<i>shibae</i>	DFL 12	ATTTCTATAATGATATATGTTAAATAA	-(C (CCCC (O))))) --	26.373	21.346	0.070	5.027	1617172	66.015
NC_011666	Proteobacteria	Alphaproteobacteria	<i>Methylocella</i>	<i>silvestris</i>	BL2	GATTCCGATAAATGATATATGTTAAAT	-(C (CCCC (O))))) --	0.223	2.909	0.035	2.686	598971	63.083
NC_003295	Proteobacteria	Betaproteobacteria	<i>Ralstonia</i>	<i>solonacearum</i>	GMI1000	CCATCCGATAAATGATATATGTTAAAT	-(C (CCCC (O))))) --	16.336	14.783	0.082	1.552	2031205	67.037
NC_008309	Proteobacteria	Gammaproteobacteria	<i>Haemophilus</i>	<i>somnus</i>	129PT	ATTTCTATAATGATATATGTTAAAT	-(C (CCCC (O))))) --	4.628	4.884	0.160	0.256	953368	37.197
NC_010519	Proteobacteria	Gammaproteobacteria	<i>Haemophilus</i>	<i>somnus</i>	2336	ATTTCTATAATGATATATGTTAAAT	-(C (CCCC (O))))) --	49.963	50.203	0.153	0.240	1535015	37.378
NC_007384	Proteobacteria	Gammaproteobacteria	<i>Shigella</i>	<i>sonnei</i>	Ss046	TGTCCGATAAATGATATATGTTAAAT	-(C (CCCC (O))))) --	14.236	17.305	0.085	3.069	1698224	51.010
NC_008782	Proteobacteria	Betaproteobacteria	<i>Acidovorax</i>	sp.	JS42	ACTAGATACGATATATGTTAAAT	-(C (CCCC (O))))) --	5.957	9.132	0.076	3.175	2156645	66.172
NC_005966	Proteobacteria	Gammaproteobacteria	<i>Acinetobacter</i>	sp.	ADP1	GATTCTGATAAATGATATATGTTAAAT	-(C (CCCC (O))))) --	4.942	4.747	0.169	0.195	1848716	40.430
NC_009675	Proteobacteria	Deltaproteobacteria	<i>Anaeromyxobacter</i>	sp.	Fw109-5	GAGTCCGATAAATGATATATGTTAAAT	-(C (CCCC (O))))) --	12.768	12.765	0.042	0.003	2429132	73.531
NC_011145	Proteobacteria	Deltaproteobacteria	<i>Anaeromyxobacter</i>	sp.	K	ACTGCCGATAAATGATATATGTTAAAT	-(C (CCCC (O))))) --	1.370	1.353	0.042	0.017	2468334	74.841
NC_008702	Proteobacteria	Betaproteobacteria	<i>Azoarcus</i>	sp.	BH72	TACTGCCATAATTTGATATGTTAAACT	-(C (CCCC (O))))) --	10.509	0.190	0.096	10.319	2045092	67.916
NC_009485	Proteobacteria	Alphaproteobacteria	<i>Bradyrhizobium</i>	sp.	BTAi1	AATTCCGATAAAGTATATGTTAAAT	-(C (CCCC (O))))) --	8.693	9.866	0.063	1.173	4292334	64.920
NC_009445	Proteobacteria	Alphaproteobacteria	<i>Bradyrhizobium</i>	sp.	ORS278	AAITCCGATAAAGTATATGTTAAAT	-(C (CCCC (O))))) --	0.502	0.470	0.043	0.032	4009009	65.510
NC_010338	Proteobacteria	Alphaproteobacteria	<i>Caulobacter</i>	sp.	K31	TATTCCGATAAAGTATATGTTAAACT	-(C (CCCC (O))))) --	13.765	13.543	0.041	0.223	2884570	67.463
NC_008254	Proteobacteria	Alphaproteobacteria	<i>Chelatvibrans</i>	sp.	BNC1	GGATTGATAAAGATATGTTAAACT	-(C (CCCC (O))))) --	21.782	22.793	0.047	1.011	1734815	61.067

Accession	Phylum	Class	Genus	Species	Strain	dif	Palindrome	dif angle	GC angle	GCSI	A	dif Position	GC%
NC_011992	Proteobacteria	Betaproteobacteria	<i>Diaphorobacter</i>	sp.	TPSY	ACTAGATACGATGATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	6.032	1.271	0.082	4.761	1988108	66.834
NC_009436	Proteobacteria	Gammaproteobacteria	<i>Enterobacter</i>	sp.	638	AGTGGCGATAAGTATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	2.008	2.014	0.090	0.006	2202843	52.982
NC_011979	Proteobacteria	Deltaproteobacteria	<i>Geobacter</i>	sp.	FRC-32	GGCTCCGAATGAATATATGTTAAAGT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	77.105	77.672	0.084	0.568	3074186	53.472
NC_012918	Proteobacteria	Deltaproteobacteria	<i>Geobacter</i>	sp.	M21	ACGCCCAATGAATATATGTTCAAGT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	2.755	1.766	0.045	0.989	2409227	60.470
NC_007802	Proteobacteria	Alphaproteobacteria	<i>Jamnaschia</i>	sp.	CCS1	GGCCAGTATCATGATATGTTAAAGT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	0.658	3.880	0.098	3.222	2302393	62.332
NC_009659	Proteobacteria	Betaproteobacteria	<i>Janthinobacterium</i>	sp.	Marseille	AGTTCGCAATTTATATGTTCAAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	0.955	0.196	0.129	0.759	1977534	54.234
NC_008576	Proteobacteria	Magnetococcus	<i>Magnetococcus</i>	sp.	MC-1	ATGTCGCAATGAATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	4.259	3.025	0.215	1.234	2303953	54.167
NC_010511	Proteobacteria	Alphaproteobacteria	<i>Methylobacterium</i>	sp.	4-46	GATTTCGCAATGAATATGTTCAAGT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	101.023	47.817	0.007	53.206	7108795	71.583
NC_012969	Proteobacteria	Betaproteobacteria	<i>Methylovorus</i>	sp.	SIP3-4	ACTCGGCAATGATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	5.122	7.884	0.113	2.762	1455133	54.860
NC_009662	Proteobacteria	Epsilonproteobacteria	<i>Nitratiruptor</i>	sp.	SB155-2								39.689
NC_007948	Proteobacteria	Betaproteobacteria	<i>Polaromonas</i>	sp.	JS666	ATATTCCAGCTGATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	0.061	0.895	0.064	0.834	2606263	62.472
NC_009379	Proteobacteria	Betaproteobacteria	<i>Polynucleobacter</i>	sp.	QLW-P1DMWA-1	TTTGGCATAAGTATATATGTTCAAC	(-(-(-(-(-(-(-)-)-)-)-)-)-)	19.424	18.465	0.168	0.959	963229	44.834
NC_009524	Proteobacteria	Gammaproteobacteria	<i>Psychrobacter</i>	sp.	PRwf-1	TTTTCGATAAGTATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	7.982	8.170	0.129	0.189	1423441	44.875
NC_012587	Proteobacteria	Alphaproteobacteria	<i>Rhizobium</i>	sp.	NGR234	AAATCGCAATGAATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	4.305	2.035	0.054	2.271	1653335	63.029
NC_008044	Proteobacteria	Alphaproteobacteria	<i>Ruegeria</i>	sp.	TM1040	ATCCATAATCAGTATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	9.658	7.716	0.105	1.942	1344015	60.413
NC_008750	Proteobacteria	Gammaproteobacteria	<i>Shewanella</i>	sp.	W3-18-1	ACTGGCACAATGATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	1.928	0.826	0.172	1.103	2328968	44.633
NC_008322	Proteobacteria	Gammaproteobacteria	<i>Shewanella</i>	sp.	MR-7	ACTGGCACAATGATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	4.581	5.146	0.157	0.565	2327458	47.866
NC_008321	Proteobacteria	Gammaproteobacteria	<i>Shewanella</i>	sp.	MR-4	ACTGGCACAATGATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	2.616	3.198	0.158	0.582	2387341	47.894
NC_009663	Proteobacteria	Epsilonproteobacteria	<i>Sulfurovum</i>	sp.	NBC37-1	TCCGACAATCTATATGTTAAAGT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	173.479	9.867	0.189	163.612	2354140	43.871
NC_011662	Proteobacteria	Betaproteobacteria	<i>Thauera</i>	sp.	MZ1T	TCTACATACAAAAGTATGTTAAAGT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	3.395	3.020	0.050	0.375	1983962	68.404
NC_011901	Proteobacteria	Betaproteobacteria	<i>Thioalkalivibrio</i>	sp.	HL-EbGR7	AGTTCGCAATGATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	11.540	0.386	0.082	11.155	1827430	65.063
NC_012416	Proteobacteria	Alphaproteobacteria	<i>Wolbachia</i>	sp.	wR1	TTACATAATACATATGTTAAAGT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	87.619	132.535	0.113	44.916	285002	35.165
NC_009428	Proteobacteria	Alphaproteobacteria	<i>sphaeroides</i>		ATCC 17025	GATTTCGATAATCAAGCTGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	8.854	8.858	0.054	0.004	1551218	68.478
NC_009434	Proteobacteria	Gammaproteobacteria	<i>Pseudomonas</i>		A1501	ATTTGGCATAAGTATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	0.264	0.327	0.097	0.063	2295802	63.878
NC_006300	Proteobacteria	Gammaproteobacteria	<i>Mannheimia</i>		MBEL55E	ATTTGGCATAAGTATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	12.694	14.154	0.168	1.460	1059670	42.538
NC_009655	Proteobacteria	Gammaproteobacteria	<i>Actinobacillus</i>		130Z	ATTTGGCATAAGTATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	1.258	0.711	0.196	0.547	1547971	44.919
NC_002939	Proteobacteria	Deltaproteobacteria	<i>Geobacter</i>		PCA	ACGCCCAATGAATATATGTTAAAGT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	2.178	0.336	0.066	1.842	1891863	60.937
NC_005773	Proteobacteria	Gammaproteobacteria	<i>Pseudomonas</i>		<i>pv. phaseolicola</i> 1448A	GGTAGCATAAGTATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	9.245	10.162	0.088	0.916	3225909	58.015
NC_004578	Proteobacteria	Gammaproteobacteria	<i>Pseudomonas</i>		<i>pv. tomat</i> str. DC3000	GTACGCAATGATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	0.743	0.407	0.077	0.336	3211772	58.399
NC_007005	Proteobacteria	Gammaproteobacteria	<i>Pseudomonas</i>		<i>pv. syringae</i> B728a	GTTACGCAATGATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	11.537	11.743	0.085	0.206	3235233	59.233
NC_010694	Proteobacteria	Gammaproteobacteria	<i>Erwinia</i>		Et1/99	GGTGGCATAAGTATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	1.870	3.214	0.101	1.344	1943872	53.732
NC_010161	Proteobacteria	Alphaproteobacteria	<i>Bartonella</i>		<i>tribocorum</i> CIP 105476	ATCGCATAAGTATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	8.845	8.798	0.350	0.047	1245185	38.856
NC_010793	Proteobacteria	Alphaproteobacteria	<i>Orientia</i>		<i>tsutsugamushi</i> str. Ikeda	GTTATATAATTAATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	66.863	64.828	0.098	2.035	1385283	30.514
NC_009488	Proteobacteria	Alphaproteobacteria	<i>Orientia</i>		<i>tsutsugamushi</i> str. Boryong	GTTATATAATTAATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	38.575	51.784	0.095	13.209	156501	30.531
NC_007880	Proteobacteria	Gammaproteobacteria	<i>Francisella</i>		<i>subsp. holarctica</i>	CATTGGATAAATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	22.622	31.139	0.257	8.518	1166746	32.153
NC_008369	Proteobacteria	Gammaproteobacteria	<i>Francisella</i>		<i>subsp. holarctica</i> OSU18	CATTGGATAAATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	23.462	31.892	0.256	8.430	1170697	32.158
NC_009749	Proteobacteria	Gammaproteobacteria	<i>Francisella</i>		<i>subsp. holarctica</i> FTNF002-00	CATTGGATAAATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	23.293	31.740	0.258	8.447	1167148	32.161
NC_010677	Proteobacteria	Gammaproteobacteria	<i>Francisella</i>		<i>subsp. mediasiatica</i> FSC147	CATTGGATAAATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	12.761	11.138	0.256	1.623	1049124	32.254
NC_006570	Proteobacteria	Gammaproteobacteria	<i>Francisella</i>		<i>subsp. tularensis</i> SCHU 54	CATTGGATAAATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	9.184	0.802	0.262	3.882	994676	32.259
NC_008245	Proteobacteria	Gammaproteobacteria	<i>Francisella</i>		<i>subsp. tularensis</i> FSC198	CATTGGATAAATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	9.191	0.809	0.261	3.883	994629	32.260
NC_009257	Proteobacteria	Gammaproteobacteria	<i>Francisella</i>		<i>subsp. tularensis</i> WY96-3418	CATTGGATAAATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	17.646	26.129	0.246	8.483	856179	32.269
NC_006142	Proteobacteria	Alphaproteobacteria	<i>Rickettsia</i>		<i>str. Wilmington</i>	GTTCTATAATGATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	16.228	16.251	0.198	0.023	599240	28.919
NC_009483	Proteobacteria	Deltaproteobacteria	<i>Geobacter</i>		Rf4	ACGCCGATAAGTATATGTTCAAGT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	6.033	7.086	0.104	1.053	2482102	54.238
NC_012560	Proteobacteria	Gammaproteobacteria	<i>Azotobacter</i>		DJ	ACTTCGATAAGTATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	9.564	9.530	0.100	0.034	2545995	65.676
NC_005085	Proteobacteria	Betaproteobacteria	<i>Chromobacterium</i>		ATCC 12472	TATTGGCATAAGTATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	0.416	0.413	0.137	0.003	2619354	64.834
NC_008751	Proteobacteria	Desulfobacteria	<i>Desulfobacterium</i>		DP4	ATGCCATAAGTATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	1.817	0.033	0.154	1.784	1713964	63.012
NC_002937	Proteobacteria	Deltaproteobacteria	<i>Desulfobacterium</i>		<i>str. Hildenborough</i>	ATGCCATAAGTATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	0.567	1.225	0.151	0.658	1754262	63.143
NC_011769	Proteobacteria	Deltaproteobacteria	<i>Desulfobacterium</i>		<i>str. Miyazaki F'</i>	ATGTCATAAGTATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	13.112	6.013	0.118	7.099	429001	67.109
NC_007406	Proteobacteria	Alphaproteobacteria	<i>Nitrospira</i>		Nb-255	AAITGGCATAAGTATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	0.604	0.664	0.078	0.060	1591522	62.049
NC_009511	Proteobacteria	Alphaproteobacteria	<i>Sphingomonas</i>		RW1	GTCGATAAATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	6.938	4.682	0.050	2.255	524091	68.398
NC_010506	Proteobacteria	Gammaproteobacteria	<i>Shewanella</i>		ATCC 51908	TATGGCACAATGATATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	1.645	0.354	0.210	1.291	2980128	43.701
NC_008095	Proteobacteria	Deltaproteobacteria	<i>Myxococcus</i>		DK 1622	AGGCCGATAAGTATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	2.458	0.864	0.108	1.594	4489683	68.891
NC_012912	Proteobacteria	Gammaproteobacteria	<i>Dickeya</i>		Echl591	GTTTCGATAAGTATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	2.716	3.405	0.164	0.689	2443245	54.517
NC_011144	Proteobacteria	Alphaproteobacteria	<i>Phenylobacterium</i>		HLK1	TATTCGCAATGAAGTATGTTCAAGT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	18.009	13.654	0.045	4.355	1806314	71.346
NC_012225	Spirochaetes	Spirochaetales	<i>Brachyspira</i>		<i>hyodysenteriae</i> WA1	TATTGGCATAAGTATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	14.695	3.753	0.166	10.941	497170	27.063
NC_010163	Verrucomicrobia	Mollicutes	<i>Acholeplasma</i>		PG-8A	AGTCCGATAAGTATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	7.680	7.718	0.256	0.038	716560	31.928
NC_010794	Verrucomicrobia	Unclassified Verrucomicrobia	<i>Methylacidiphilium</i>		<i>infernum</i> V4	ACTTCGCAATGATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	2.378	6.860	0.192	4.482	1128462	45.483
NC_010655	Verrucomicrobia	Verrucomicrobiae	<i>Akkermansia</i>		<i>ATCC BAA-835</i>	ACTTCGCAATGATATGTTAAAT	(-(-(-(-(-(-(-)-)-)-)-)-)-)	3.027	3.659	0.223	0.632	2312983	55.762
NC_010571	Verrucomicrobia	Opitutae	<i>Opitutus</i>		<i>terrae</i> PB90-1	GCTCCGATCAATGATGATGCAAGA	(-(-(-(-(-(-(-)-)-)-)-)-)-)	31.373	15.577	0.163	15.796	3459633	65.341

Accession	Phylum	Class	Genus	Species	Strain	Chromosome	<i>dif</i>	Palindrome	<i>dif</i> angle	GC angle	GCSI	A	Position	GC%
NC_012441	Proteobacteria	Alphaproteobacteria	<i>Brucella</i>	<i>melitensis</i>	ATCC 23457	I	TAATCGCATAAAGTAGATTATGGAACGTG	-(--(((---(-O)-)))---))--	17.413	13.240	0.112	4.173	1053104	57.152
NC_012442	Proteobacteria	Alphaproteobacteria	<i>Brucella</i>	<i>melitensis</i>	ATCC 23457	II	AAATCAGATAAATGTATTATGGAACAT	-(--(((---(-O)-)))---))--	24.168	24.038	0.100	0.130	511841	57.345
NC_003317	Proteobacteria	Alphaproteobacteria	<i>Brucella</i>	<i>melitensis</i>	bv. 1 str. 16M	I	TAATCGCATAAAGTAGATTATGGAACGTG	-(--(((---(-O)-)))---))--	17.658	13.585	0.112	4.073	954725	57.158
NC_003318	Proteobacteria	Alphaproteobacteria	<i>Brucella</i>	<i>melitensis</i>	bv. 1 str. 16M	II	AAATCAGATAAATGTATTATGGAACAT	-(--(((---(-O)-)))---))--	23.178	23.077	0.108	0.081	758169	57.338
NC_004310	Proteobacteria	Alphaproteobacteria	<i>Brucella</i>	<i>suis</i>	1330	I	TAATCGCATAAAGTAGATTATGGAACGTG	-(--(((---(-O)-)))---))--	15.891	11.671	0.112	4.220	1034378	57.214
NC_004311	Proteobacteria	Alphaproteobacteria	<i>Brucella</i>	<i>suis</i>	1330	II	AAATCAGATAAATGTATTATGGAACAT	-(--(((---(-O)-)))---))--	21.048	20.920	0.107	0.128	531770	57.316
NC_006932	Proteobacteria	Alphaproteobacteria	<i>Brucella</i>	<i>abortus</i>	bv. 1 str. 9-941	I	TAATCGCATAAAGTAGATTATGGAACGTG	-(--(((---(-O)-)))---))--	17.248	17.045	0.111	0.203	1051548	57.156
NC_006933	Proteobacteria	Alphaproteobacteria	<i>Brucella</i>	<i>abortus</i>	bv. 1 str. 9-941	II	AAATCAGATAAATGTATTATGGAACAT	-(--(((---(-O)-)))---))--	31.411	31.092	0.113	0.319	682506	57.336
NC_007493	Proteobacteria	Alphaproteobacteria	<i>Rhodobacter</i>	<i>sphaeroides</i>	2.4.1	I	AGAATACATAACAGATTATCCGGACTC	-----((---(-O)-)))-----	0.226	0.122	0.060	0.104	1436830	69.005
NC_007494	Proteobacteria	Alphaproteobacteria	<i>Rhodobacter</i>	<i>sphaeroides</i>	2.4.1	II	TTATCTGATAAGCAAGATTATGTAATCA	-(--(((---(-O)-)))---))--	8.667	7.662	0.041	1.006	371559	69.011
NC_007618	Proteobacteria	Alphaproteobacteria	<i>Brucella</i>	<i>melitensis</i>	biovar Abortus 2308	I	TAATCGCATAAAGTAGATTATGGAACGTG	-(--(((---(-O)-)))---))--	17.032	13.143	0.112	3.889	1048699	57.157
NC_007624	Proteobacteria	Alphaproteobacteria	<i>Brucella</i>	<i>melitensis</i>	biovar Abortus 2308	II	AAATCAGATAAATGTATTATGGAACAT	-(--(((---(-O)-)))---))--	32.367	32.047	0.113	0.320	682494	57.341
NC_008686	Proteobacteria	Alphaproteobacteria	<i>Paracoccus</i>	<i>denitrificans</i>	PD1222	I								66.706
NC_008687	Proteobacteria	Alphaproteobacteria	<i>Paracoccus</i>	<i>denitrificans</i>	PD1222	II								66.824
NC_009049	Proteobacteria	Alphaproteobacteria	<i>Rhodobacter</i>	<i>sphaeroides</i>	ATCC 17029	I	ATAATACATAACAGATTATCCGGACTC	-----((---(-O)-)))-----	2.150	2.158	0.056	0.008	1483610	69.093
NC_009050	Proteobacteria	Alphaproteobacteria	<i>Rhodobacter</i>	<i>sphaeroides</i>	ATCC 17029	II	TTATCCGATAAGCAAGATTATGTAATCA	-(--(((---(-O)-)))---))--	21.171	24.340	0.041	3.169	1190939	68.628
NC_009505	Proteobacteria	Alphaproteobacteria	<i>Brucella</i>	<i>ovis</i>	ATCC 25840	I	TAATCGCATAAAGTAGATTATGGAACGTG	-(--(((---(-O)-)))---))--	17.063	12.855	0.111	4.208	1043436	57.177
NC_009504	Proteobacteria	Alphaproteobacteria	<i>Brucella</i>	<i>ovis</i>	ATCC 25840	II	AAATCAGATAAATGTATTATGGAACAT	-(--(((---(-O)-)))---))--	23.074	22.942	0.107	0.132	506160	57.226
NC_009667	Proteobacteria	Alphaproteobacteria	<i>Ochrobactrum</i>	<i>anthropi</i>	ATCC 49188	I	CGATCACATAAAGTAGATTATGGAACCA	-(--(((---(-O)-)))---))--	101.180	106.838	0.096	5.658	2255140	56.051
NC_009668	Proteobacteria	Alphaproteobacteria	<i>Ochrobactrum</i>	<i>anthropi</i>	ATCC 49188	II	AAATCAGATAAATGTATTATGGAACAA	-(--(((---(-O)-)))---))--	4.964	0.019	0.087	4.945	971981	56.223
NC_010103	Proteobacteria	Alphaproteobacteria	<i>Brucella</i>	<i>canis</i>	ATCC 23365	I	TAATCGCATAAAGTAGATTATGGAACGTG	-(--(((---(-O)-)))---))--	15.798	11.573	0.112	4.224	1032830	57.208
NC_010104	Proteobacteria	Alphaproteobacteria	<i>Brucella</i>	<i>canis</i>	ATCC 23365	II	AAATCAGATAAATGTATTATGGAACAT	-(--(((---(-O)-)))---))--	20.976	20.848	0.107	0.128	531755	57.304
NC_010169	Proteobacteria	Alphaproteobacteria	<i>Brucella</i>	<i>suis</i>	ATCC 23445	I	TAATCGCATAAAGTAGATTATGGAACGTG	-(--(((---(-O)-)))---))--	38.294	33.700	0.104	4.594	1053967	57.116
NC_010167	Proteobacteria	Alphaproteobacteria	<i>Brucella</i>	<i>suis</i>	ATCC 23445	II	AAATCAGATAAATGTATTATGGAACAT	-(--(((---(-O)-)))---))--	43.188	43.078	0.114	0.110	532368	57.328
NC_010742	Proteobacteria	Alphaproteobacteria	<i>Brucella</i>	<i>abortus</i>	S19	I	TAATCGCATAAAGTAGATTATGGAACGTG	-(--(((---(-O)-)))---))--	17.121	16.918	0.111	0.203	1049878	57.159
NC_010740	Proteobacteria	Alphaproteobacteria	<i>Brucella</i>	<i>abortus</i>	S19	II	AAATCAGATAAATGTATTATGGAACAT	-(--(((---(-O)-)))---))--	31.301	30.982	0.113	0.319	681709	57.336
NC_011963	Proteobacteria	Alphaproteobacteria	<i>Rhodobacter</i>	<i>sphaeroides</i>	KD131	I	ATAATACATAACAGATTATCCGGACTC	-----((---(-O)-)))-----	2.741	2.749	0.059	0.008	1102424	69.176
NC_011958	Proteobacteria	Alphaproteobacteria	<i>Rhodobacter</i>	<i>sphaeroides</i>	KD131	II	TTATCCGATAAGCAAGATTATGTAATAG	-(--(((---(-O)-)))---))--	2.590	2.510	0.036	0.081	402600	68.706
NC_011985	Proteobacteria	Alphaproteobacteria	<i>Agrobacterium</i>	<i>radiobacter</i>	K84	I	AAGTCAAGATAAAGTAGATTATGGAACCA	-(--(((---(-O)-)))---))--	8.209	4.985	0.090	3.224	1911238	60.337
NC_011983	Proteobacteria	Alphaproteobacteria	<i>Agrobacterium</i>	<i>radiobacter</i>	K84	II	AAGTCAAGATAAAGTAGATTATGGAACCA	-(--(((---(-O)-)))---))--	55.938	61.969	0.065	6.031	1759891	59.762
NC_011989	Proteobacteria	Alphaproteobacteria	<i>Agrobacterium</i>	<i>vitis</i>	S4	I	AAGTCAAGATAAAGTAGATTATGGAACCA	-(--(((---(-O)-)))---))--	14.969	15.030	0.094	0.062	2018128	57.658
NC_011988	Proteobacteria	Alphaproteobacteria	<i>Agrobacterium</i>	<i>vitis</i>	S4	II	AAAATGATAAAGTAGATTATGGAACCA	-(--(((---(-O)-)))---))--	177.611	179.990	0.078	2.380	8517	57.519
NC_013119	Proteobacteria	Alphaproteobacteria	<i>Brucella</i>	<i>microti</i>	CCM 4915	I	TAATCGCATAAAGTAGATTATGGAACGTG	-(--(((---(-O)-)))---))--	15.803	15.537	0.111	0.266	1038638	57.225
NC_013118	Proteobacteria	Alphaproteobacteria	<i>Brucella</i>	<i>microti</i>	CCM 4915	II	AAATCAGATAAATGTATTATGGAACAT	-(--(((---(-O)-)))---))--	22.307	26.588	0.107	4.281	533213	57.304
NC_002505	Proteobacteria	Gammaproteobacteria	<i>Vibrio</i>	<i>cholerae</i>	O1 biovar El Tor str. N16961	I	AGTGGCTATTGATGTATTATGTTAAAT	-(--(((---(-O)-)))---))--	10.155	10.219	0.162	0.064	1564103	47.695
NC_002506	Proteobacteria	Gammaproteobacteria	<i>Vibrio</i>	<i>cholerae</i>	O1 biovar El Tor str. N16961	II	AATGGCCATTACGTCGGCTATGTTAAAT	-(--(((---(-O)-)))---))--	9.459	9.766	0.173	0.307	507982	46.914
NC_004459	Proteobacteria	Gammaproteobacteria	<i>Vibrio</i>	<i>vulnificus</i>	CMCP6	I	AGTGGCCATTATGATATTATGTTAAAT	-(--(((---(-O)-)))---))--	0.720	1.616	0.177	0.897	2661289	46.450
NC_004460	Proteobacteria	Gammaproteobacteria	<i>Vibrio</i>	<i>vulnificus</i>	CMCP6	II	GGTGGCCATTATGACGTTATGTTAAAT	-(--(((---(-O)-)))---))--	7.706	9.195	0.147	1.489	313959	47.119
NC_004603	Proteobacteria	Gammaproteobacteria	<i>Vibrio</i>	<i>parahaemolyticus</i>	RIMD 2210633	I	AGTGGCCATTATGATATTATGTTAAAT	-(--(((---(-O)-)))---))--	1.584	0.417	0.158	1.167	1658749	45.386
NC_004605	Proteobacteria	Gammaproteobacteria	<i>Vibrio</i>	<i>parahaemolyticus</i>	RIMD 2210633	II	AGTGGCTATTATGATATTATGTTAAAT	-(--(((---(-O)-)))---))--	0.965	2.067	0.154	1.101	933572	45.352
NC_005139	Proteobacteria	Gammaproteobacteria	<i>Vibrio</i>	<i>vulnificus</i>	YJ016	I	AGTGGCCATTATGATATTATGTTAAAT	-(--(((---(-O)-)))---))--	4.802	7.090	0.177	2.288	1721995	46.411
NC_005140	Proteobacteria	Gammaproteobacteria	<i>Vibrio</i>	<i>vulnificus</i>	YJ016	II	AGTGGCCATTATGACGTTATGTTAAAT	-(--(((---(-O)-)))---))--	5.836	5.723	0.151	0.113	898432	47.214
NC_006370	Proteobacteria	Gammaproteobacteria	<i>Photobacterium</i>	<i>profundum</i>	SS9	I	AGTGGCCATTATGATATTATGTTAAAT	-(--(((---(-O)-)))---))--	2.001	2.075	0.188	0.074	2065356	41.969
NC_006371	Proteobacteria	Gammaproteobacteria	<i>Photobacterium</i>	<i>profundum</i>	SS9	II	AGTGGCCATTATGACCTTATGTTAAAT	-(--(((---(-O)-)))---))--	7.163	6.873	0.167	0.290	1074443	41.223
NC_006840	Proteobacteria	Gammaproteobacteria	<i>Vibrio</i>	<i>fischeri</i>	ES114	I	AGTGGCCATTATGATATTATGTTAAAT	-(--(((---(-O)-)))---))--	0.281	1.199	0.211	0.918	1446504	38.953
NC_006841	Proteobacteria	Gammaproteobacteria	<i>Vibrio</i>	<i>fischeri</i>	ES114	II	AGTGGCCATTATATCTTATGTTAAAT	-(--(((---(-O)-)))---))--	1.844	1.793	0.239	0.051	658353	37.023
NC_007481	Proteobacteria	Gammaproteobacteria	<i>Pseudoalteromonas</i>	<i>haloplanktis</i>	TAC125	I	ACAACCTATAAATTTATGTTAAAT	-(--(((---(-O)-)))---))--	3.020	1.709	0.151	1.311	1580505	40.235
NC_007482	Proteobacteria	Gammaproteobacteria	<i>Pseudoalteromonas</i>	<i>haloplanktis</i>	TAC125	II	ACAGCCCAATAAATAGCTATGTTAAAT	-(--(((---(-O)-)))---))--	156.699	159.136	0.085	2.438	619751	39.367
NC_009456	Proteobacteria	Gammaproteobacteria	<i>Vibrio</i>	<i>cholerae</i>	O395	I	AGTGGCTATTACGTCGGCTATGTTAAAT	-(--(((---(-O)-)))---))--	4.387	4.402	0.142	0.015	567629	46.884
NC_009457	Proteobacteria	Gammaproteobacteria	<i>Vibrio</i>	<i>cholerae</i>	O395	II	AGTGGCTATTACGTCGGCTATGTTAAAT	-(--(((---(-O)-)))---))--	5.204	5.206	0.152	0.002	1114637	47.780
NC_009783	Proteobacteria	Gammaproteobacteria	<i>Vibrio</i>	<i>harveyi</i>	ATCC BAA-1116	I	AGTGGCCATTATGTTATGTTAAAT	-(--(((---(-O)-)))---))--	0.764	0.212	0.128	0.552	2305578	45.547
NC_009784	Proteobacteria	Gammaproteobacteria	<i>Vibrio</i>	<i>harveyi</i>	ATCC BAA-1116	II	AGTGGCCATTATGATATTATGTTAAAT	-(--(((---(-O)-)))---))--	7.841	9.695	0.127	1.854	968428	45.271
NC_011184	Proteobacteria	Gammaproteobacteria	<i>Vibrio</i>	<i>fischeri</i>	MJ11	I	AGTGGCCATTATGATATTATGTTAAAT	-(--(((---(-O)-)))---))--	1.113	1.120	0.211	0.007	1461498	38.873
NC_011186	Proteobacteria	Gammaproteobacteria	<i>Vibrio</i>	<i>fischeri</i>	MJ11	II	AGTGGCCATTATATCTTATGTTAAAT	-(--(((---(-O)-)))---))--	7.675	7.098	0.239	0.576	679176	37.250
NC_011312	Proteobacteria	Gammaproteobacteria	<i>Aliivibrio</i>	<i>salmonicida</i>	LF11238	I	AGTGGCTATTATATCTTATGTTAAAT	-(--(((---(-O)-)))---))--	0.907	83.827	0.158	82.920	1670958	39.228
NC_011313	Proteobacteria	Gammaproteobacteria	<i>Aliivibrio</i>	<i>salmonicida</i>	LF11238	II	AGTGGCCATTATGATATTATGTTAAAT	-(--(((---(-O)-)))---))--	7.223	7.208	0.184	0.015	685835	38.238
NC_011744	Proteobacteria	Gammaproteobacteria	<i>Vibrio</i>	<i>splendidus</i>	LGP32	I	ACTGGCTATTATGTCGGCTATGTTAAAT	-(--(((---(-O)-)))---))--	5.903	4.387	0.164	1.516	808228	43.642
NC_011753	Proteobacteria	Gammaproteobacteria	<i>Vibrio</i>	<i>splendidus</i>	LGP32	II	AGTGGCCATTATGATATTATGTTAAAT	-(--(((---(-O)-)))---))--	6.647	4.126	0.163	2.521	1588735	44.044
NC_012578	Proteobacteria	Gammaproteobacteria	<i>Vibrio</i>	<i>cholerae</i>	M66-2	I	AGTGGCCATTATGATATTATGTTAAAT	-(--(((---(-O)-)))---))--	9.370	11.276	0.161	1.906	1521546	47.832

表S2 : *Escherichia* の *dif* 配列

Accession	genus	species	strain	<i>dif</i> sequence
AC_000091	<i>Escherichia</i>	<i>coli</i>	K-12 substr. W3110	GGTGCGCATAATGTATATTATGTTAAAT
NC_002655	<i>Escherichia</i>	<i>coli</i>	O157:H7 EDL933	GGTGCGCATAATGTATATTATGTTAAAT
NC_002695	<i>Escherichia</i>	<i>coli</i>	O157:H7 str. Sakai	GGTGCGCATAATGTATATTATGTTAAAT
NC_004431	<i>Escherichia</i>	<i>coli</i>	CFT073	GGTGCGCATAATGTATATTATGTTAAAT
NC_007946	<i>Escherichia</i>	<i>coli</i>	UTI89	GGTGCGCATAATGTATATTATGTTAAAT
NC_008253	<i>Escherichia</i>	<i>coli</i>	536	GGTGCGCATAATGTATATTATGTTAAAT
NC_008563	<i>Escherichia</i>	<i>coli</i>	APEC O1	GGTGCGCATAATGTATATTATGTTAAAT
NC_009800	<i>Escherichia</i>	<i>coli</i>	HS	GGTGCGCATAATGTATATTATGTTAAAT
NC_009801	<i>Escherichia</i>	<i>coli</i>	E24377A	AGTGCGCATAATGTATATTATGTTAAAT
NC_010468	<i>Escherichia</i>	<i>coli</i>	ATCC 8739	AGTGCGCATAATGTATATTATGTTAAAT
NC_010473	<i>Escherichia</i>	<i>coli</i>	K-12 substr. DH10B	GGTGCGCATAATGTATATTATGTTAAAT
NC_010498	<i>Escherichia</i>	<i>coli</i>	SECEC SMS-3-5	GGTGCGCATAATGTATATTATGTTAAAT
NC_011353	<i>Escherichia</i>	<i>coli</i>	O157:H7 str. EC4115	GGTGCGCATAATGTATATTATGTTAAAT
NC_011415	<i>Escherichia</i>	<i>coli</i>	SE11	GGTGCGCATAATGTATATTATGTTAAAT
NC_011601	<i>Escherichia</i>	<i>coli</i>	O127:H6 str. E2348/69	GGTGCGCATAATGTATATTATGTTAAAG
NC_011740	<i>Escherichia</i>	<i>fergusonii</i>	ATCC 35469	AGTGCGCATAATGTATATTATGTTAAAT
NC_011741	<i>Escherichia</i>	<i>coli</i>	IA11	GATGCGCATAATGTATATTATGTTAAAT
NC_011742	<i>Escherichia</i>	<i>coli</i>	S88	GGTGCGCATAATGTATATTATGTTAAAT
NC_011745	<i>Escherichia</i>	<i>coli</i>	ED1a	GATGCGCATAATGTATATTATGTTAAAT
NC_011748	<i>Escherichia</i>	<i>coli</i>	55989	GGTGCGCATAATGTATATTATGTTAAAT
NC_011750	<i>Escherichia</i>	<i>coli</i>	IA139	GGTGCGCATAATGTATATTATGTTAAAT
NC_011751	<i>Escherichia</i>	<i>coli</i>	UMN026	GGTGCGCATAATGTATATTATGTTAAAT
NC_011993	<i>Escherichia</i>	<i>coli</i>	LF82	GGTGCGCATAATGTATATTATGTTAAAT
NC_012759	<i>Escherichia</i>	<i>coli</i>	BW2952	GGTGCGCATAATGTATATTATGTTAAAT
NC_012892	<i>Escherichia</i>	<i>coli</i>	BL21	AGTGCGCATAATGTATATTATGTTAAAT
NC_012947	<i>Escherichia</i>	<i>coli</i>	BL21(DE3)	AGTGCGCATAATGTATATTATGTTAAAT
NC_012967	<i>Escherichia</i>	<i>coli</i>	REL606	AGTGCGCATAATGTATATTATGTTAAAT
NC_013008	<i>Escherichia</i>	<i>coli</i>	O157:H7 str. TW14359	GGTGCGCATAATGTATATTATGTTAAAT
Consensus sequence				RRTGCGCATAATGTATATTATGTTAAAK

表S3 : *Bacillus* の *dif* 配列

Accession	genus	species	strain	<i>dif</i> sequence
NC_002570	<i>Bacillus</i>	<i>halodurans</i>	C-125	GGTTCCTAATATATATTATGTTAACT
NC_003909	<i>Bacillus</i>	<i>cereus</i>	ATCC 10987	ACTGCCTAATATATATTATGTTAACT
NC_003997	<i>Bacillus</i>	<i>anthracis</i>	Ames	ACTGCCTAATATATATTATGTTAACT
NC_004722	<i>Bacillus</i>	<i>cereus</i>	ATCC 14579	ACTACCTAATATATATTATGTTAACT
NC_005945	<i>Bacillus</i>	<i>anthracis</i>	Sterne	ACTGCCTAATATATATTATGTTAACT
NC_005957	<i>Bacillus</i>	<i>thuringiensis</i>	serovar konkukian str. 97-27	ACTGCCTAATATATATTATGTTAACT
NC_006274	<i>Bacillus</i>	<i>cereus</i>	E33L	ACTGCCTAATATATATTATGTTAACT
NC_007530	<i>Bacillus</i>	<i>anthracis</i>	Ames Ancestor	ACTGCCTAATATATATTATGTTAACT
NC_008600	<i>Bacillus</i>	<i>thuringiensis</i>	Al Hakam	ACTGCCTAATATATATTATGTTAACT
NC_009674	<i>Bacillus</i>	<i>cytotoxicus</i>	cytotoxis NVH 391-98	ACTACCTAATATATATTATGTTAACT
NC_010184	<i>Bacillus</i>	<i>weihenstephanensis</i>	KBAB4	ACTTCGTATAATATATATTATGTTAACT
NC_011658	<i>Bacillus</i>	<i>cereus</i>	AH187	ACTGCCTAATATATATTATGTTAACT
NC_011725	<i>Bacillus</i>	<i>cereus</i>	B4264	ACTGCCTAATATATATTATGTTAACT
NC_011969	<i>Bacillus</i>	<i>cereus</i>	Q1	ACTGCCTAATATATATTATGTTAACT
NC_012472	<i>Bacillus</i>	<i>cereus</i>	03BB102	ACTGCCTAATATATATTATGTTAACT
NC_012581	<i>Bacillus</i>	<i>anthracis</i>	CDC 684	ACTGCCTAATATATATTATGTTAACT
NC_012659	<i>Bacillus</i>	<i>anthracis</i>	A0248	ACTGCCTAATATATATTATGTTAACT
Consensus sequence				RCTDCSTATAATATATATTATGTTAACT

表S4 : *N. farcinica* と *M. avium* の *dif* 配列

Accession	genus	species	strain	<i>dif</i> sequence
NC_006361	<i>Nocardia</i>	<i>farcinica</i>	IFM 10152	TACGCCGATAATCTACATTATGTTAAGT
NC_002944	<i>Mycobacterium</i>	<i>avium</i>	paratuberculosis K-10	TCTACCGATAAGCGACATTATGTCAAGT
Consensus sequence				TMYRCCGATAAKCKACATTATGTYAAGT

研究成果一覽

査読付き国際学術誌掲載論文 (筆頭2報, 共著3報)

1. Kono N, Arakawa K*, Tomita M (2006) MEGU: Pathway mapping web-service based on KEGG and SVG. In *Silico Biology* 6:621-655
2. Kono N, Arakawa K*, Ogawa R, Kido N, Oshita K, Ikegami K, Tamaki S, Tomita M (2009) Pathway Projector: Web-Based Zoomable Pathway Browser Using KEGG Atlas and Google Maps API. *PLoS One* 4(11):e7710
3. Arakawa K, Kono N, Yamada Y, Mori H, Tomita M* (2005) KEGG-based pathway visualization tool for complex omics data. In *Silico Biology* 5:0039
4. Tamaki S, Arakawa K*, Kono N, Tomita M (2007) Restauro-G: A Rapid Genome Re-Annotation System for Comparative Genomics. *Genomics Proteomics Bioinformatics* 5:53-58
5. Arakawa K*, Tamaki S, Kono N, Kido N, Ikegami K, Ogawa R, Tomita M (2009) Genome Projector: zoomable genome map with multiple views. *BMC Bioinformatics* 10:31

学会口頭発表 (筆頭2回)

1. Kono N, Arakawa K, Kido N, Ogawa R, Oshita K, Ikegami K, Tamaki S, Tomita M. "Pathway Projector" SIG-MBI 2009 Ishikawa Japan
2. Kono N, Arakawa K, Tamaki S, Kido N, Ikegami K, Ogawa R, Tomita M. "Genome Projector" SIG-MBI 2008 Ishikawa Japan

学会ポスター発表(筆頭14回)

1. 河野暢明, 荒川和晴, 富田勝 “The prediction of bacterial dif sequences and analysis of the evolutionary conservation.” 日本分子生物学会 2009 横浜
2. Kono N, Arakawa K, Ogawa R, Kido N, Oshita K, Ikegami K, Tamaki S, Tomita M. “Pathway Projector: an Interactive Pathway Browser with Zoomable User Interface” GIW 2009 Yokohama Japan
3. 河野暢明, 荒川和晴, 富田勝 “バクテリアゲノムの複製終結部位に存在するdif配列の進化” 日本進化学会 2009 札幌
4. Kono N, Arakawa K, Kido N, Ogawa R, Oshita K, Ikegami K, Tamaki S, Tomita M. “Pathway Projector: Web-based Zoomable Pathway Browser using KEGG Atlas and Google Maps API” ISMB 2009 Stockholm Sweden
5. Kono N, Arakawa K, Kido N, Ogawa R, Oshita K, Ikegami K, Tamaki S, Tomita M. “Pathway Projector: KEGG AtlasとGoogle Maps APIを用いたパスウェイブラウザ” Open-Bio 2009 石川
6. 河野暢明, 荒川和晴, 富田勝 “Growth Efficiency of Bacteria Depends on the Genomic Polarity”, 分子生物学会 2008 神戸
7. Kono N, Arakawa K, Kido N, Ogawa R, Oshita K, Ikegami K, Tamaki S, Tomita M. “Web-based Zoomable Pathway Browser using KEGG Atlas and Google Maps API” GIW 2008 Gold Coast Australia
8. Kono N, Arakawa K, Tomita M. “Growth Efficiency of Bacteria Depends on the Genomic Polarity”, 3R Symposium 2008 Shizuoka Japan
9. Kono N, Arakawa K, Tomita M. “Pathway Visualization Web Application for Omics Data Based on KEGG” 日本分子生物学会年会 2007 横浜
10. Kono N, Arakawa K, Yachie N, Tomita M. “E-Cell Simulation Environment 3D” AGW 2007 Tokyo Japan
11. Kono N, Arakawa K, Tomita M. “MEGU: Pathway mapping web-service on KEGG and SVG” ISMB 2007 Vienna Austria
12. Kono N, Arakawa K, Tomita M. “Pathway Mapping Web-Service Based on KEGG and SVG” GIW 2006 Yokohama Japan
13. Kono N, Arakawa K, Tomita M, “KEGG-based pathway visualization tool for complex omics data” ISMB 2006 Fortaleza Brazil
14. Kono N, Arakawa K, Yamada Y, Mori H, Tomita M. “KEGG-Based Pathway Visualization Tool for Complex Omics Data” GIW 2005 Yokohama Japan

学会ポスター発表(共著10回)

1. Arakawa K, Kono N, Tamaki S, Yoshida H, Tomita M, "GEM System version 2 for the analysis of metabolic pathways", ISMB, 2006, Fortaleza, Brazil
2. Tamaki S, Arakawa K, Kono N, Tomita M “Restauro-G: A Rapid Genome Re-Annotation System” GIW 2006 Yokohama Japan
3. Ikegami K, Kono N, Tamaki S, Arakawa K, Tomita M “Identification of mega-palindromic structures in bacterial chromosomes” 日本バイオインフォマティクス学会年会 2007 東京

4. 玉木聡志, 荒川和晴, 河野暢明, 池上慶太, 木戸信博, 小川隆, 富田勝 “Genome Projector: ズーマブルユーザーインターフェイスを備えたゲノムブラウザ” 日本分子生物学会年会 2007 横浜
5. Kido N, Kono N, Arakawa K, Tomita M “Automatic layout tool for large-scale metabolic pathway models based on KEGG Atlas” GIW 2008 Gold Coast Australia
6. 木戸信博, 河野暢明, 荒川和晴, 富田勝 “KEGG Atlasを用いた大規模代謝モデルの自動レイアウトツール” 日本分子生物学会年会 2008 神戸
7. 木戸信博, 河野暢明, 荒川和晴, 富田勝 “KEGG Atlasを用いた大規模代謝モデルの自動レイアウトツール” オープンバイオ研究会 2009 石川
8. Kido N, Kono N, Arakawa K, Tomita M “Automatic layout tool for large-scale metabolic pathway models using KEGG Atlas, SBML and SBGN” ISMB 2009 Stockholm Sweden
9. 池上慶太, 河野暢明, 荒川和晴, 富田勝 “ゲノムサイズの縮小に伴う遺伝子機能の保存傾向の解析” 日本分子生物学会年会 2009 横浜
10. 中川真菜, 河野暢明, 荒川和晴, 小川隆, 富田勝 “BioUbiGraph: 細胞シミュレーションの3次元可視化” 日本分子生物学会年会 2009 横浜

受賞歴

- GIW 2008 Student Bursary (2008.12)
- Mashup Award 4 (cowin) (2008.10)
- 2007年度日本学生支援機構 優秀学生顕彰 学術分野大賞 (2007.11)
- 2007年度慶應義塾大学 SFC Award (2008.3)
- 2006年度慶應義塾大学 SFC Award (2007.3)

Keio University



Copyright © 2010

Nobuaki Kono

All right reserved

高精度な複製終結点予測アルゴリズムによる
バクテリアゲノム対称性の解析

2010年3月25日 初版発行

著者 河野暢明

監修 富田勝・荒川和晴

発行 慶應義塾大学 湘南藤沢学会
〒252-0816 神奈川県藤沢市遠藤5322
TEL:0466-49-3437

Printed in Japan 印刷・製本 ワキプリントピア

ISBN 978-4-87762-237-4
SFC-MT 2009-005

■ 本論文は修士論文において優秀と認められ、出版されたものです。