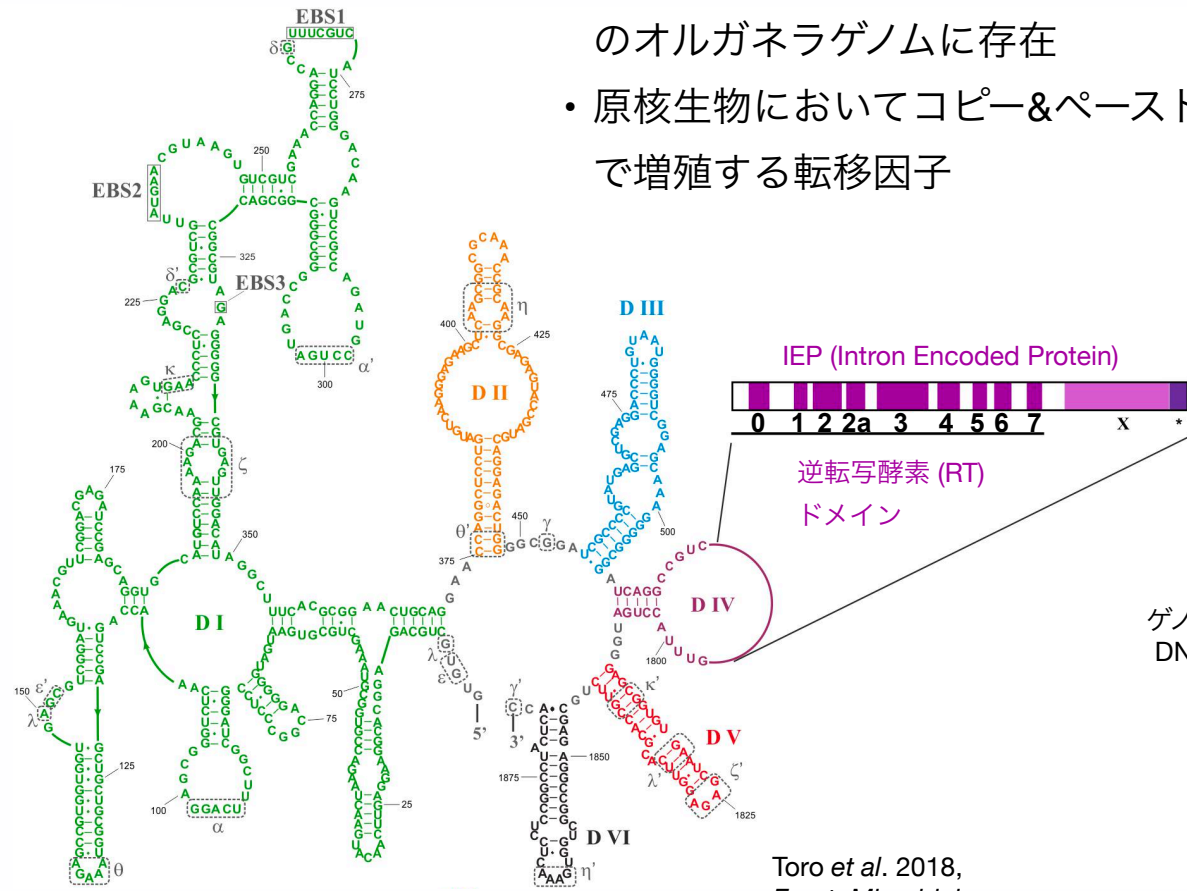


# 原核生物ゲノムにおけるグループIIイントロンの情報科学的解析

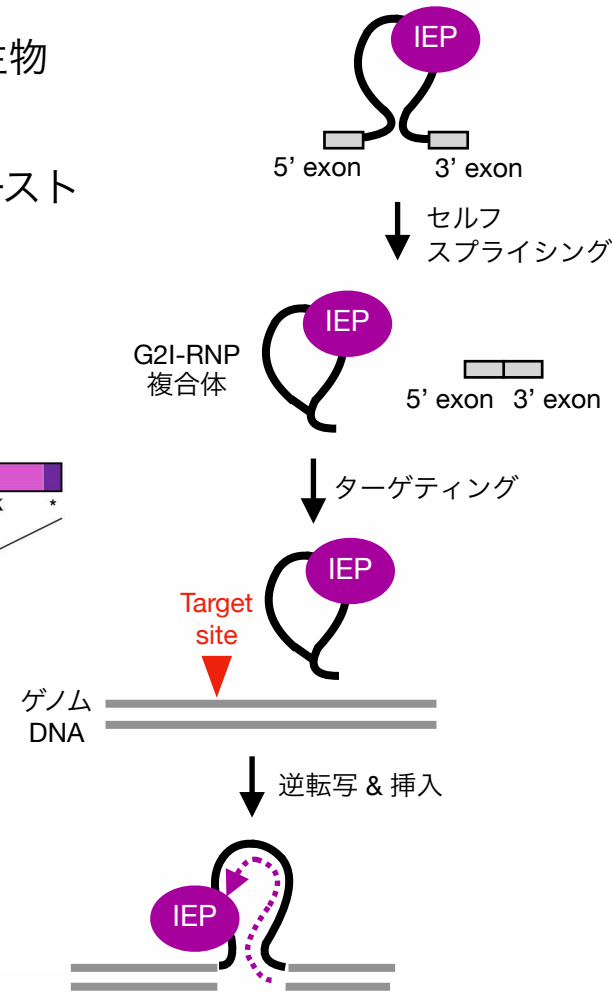
三浦 昌浩 / 学籍番号82025058 / BIプログラム

## グループIIイントロン (G2I) とは

- バクテリア, アーキア, 真核生物のオルガネラゲノムに存在
- 原核生物においてコピー&ペーストで増殖する転移因子

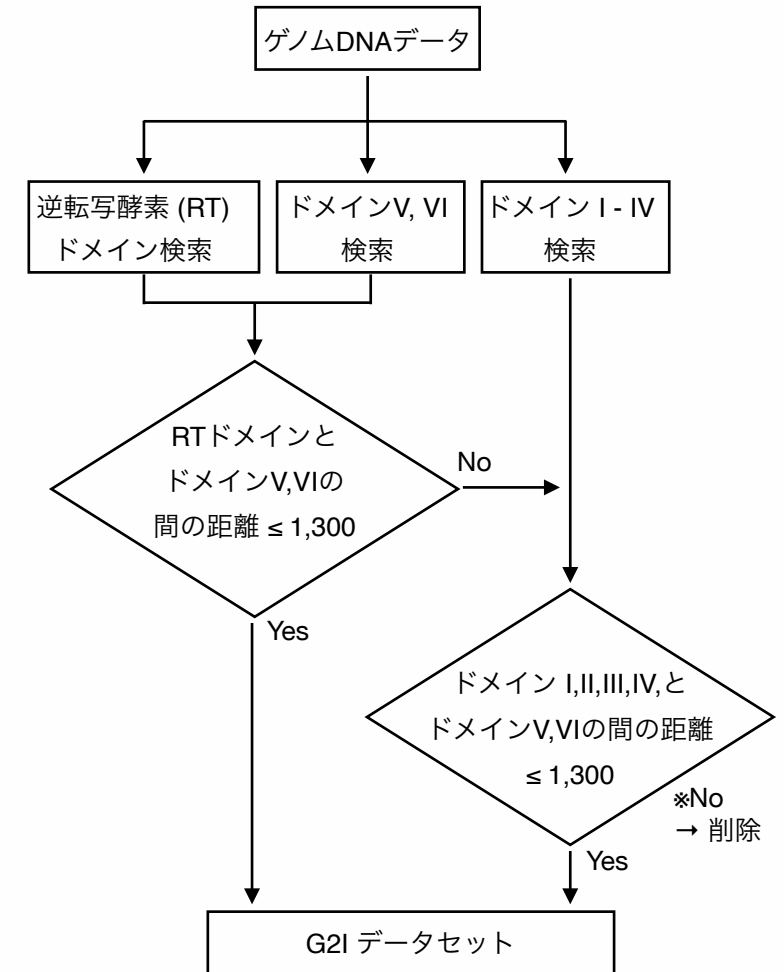


Toro et al. 2018, *Front. Microbiol.*



## 手法

ゲノムデータから G2I を見つけ出す  
バイオインフォマティクスパイプライン



### 逆転写酵素ドメイン検索

BLAST+ ver 2.4.0  
tblastn 425 queries  
(Toro et al. 2014, *PLoS ONE*)  
E-value  $\leq 1E-10$   
coverage  $\geq 40\%$

### ドメイン I-IV 検索

Infernal ver 1.1.2  
7 models (from Rfam)  
cmsearch -rfam  
E-value  $\leq 1E-4$   
BLAST+ ver 2.4.0  
blastn Zbase 346 queries  
(Candales et al. 2012)  
E-value  $< 1E-10$   
coverage  $\geq 60\%$

### ドメインV,VI検索

(セルフスプライシングに重要)  
Infernal ver 1.1.2  
RF00029.cm (from Rfam)  
cmsearch -nohmm  
bit\_score  $\geq 24$

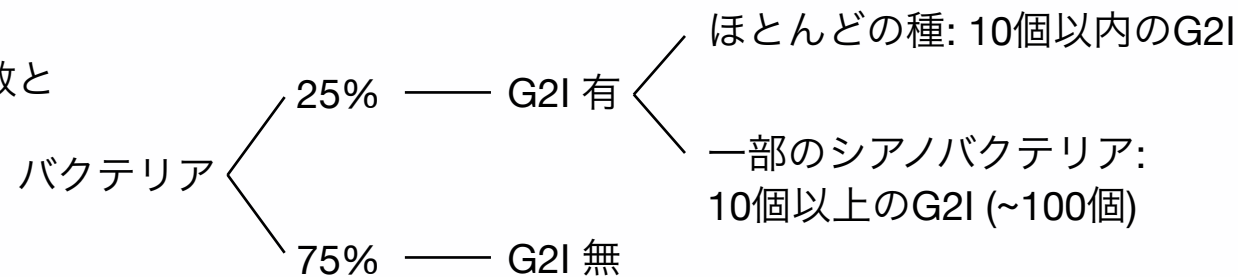
### 先行研究のデータベースに対して検証

(Candales et al. 2012, *Mob DNA*)

ドメインV, VI 検索では400のG2Iの内, 378を同定 (94.5%)  
標準的なG2Iは上記のパイプラインで同定できる

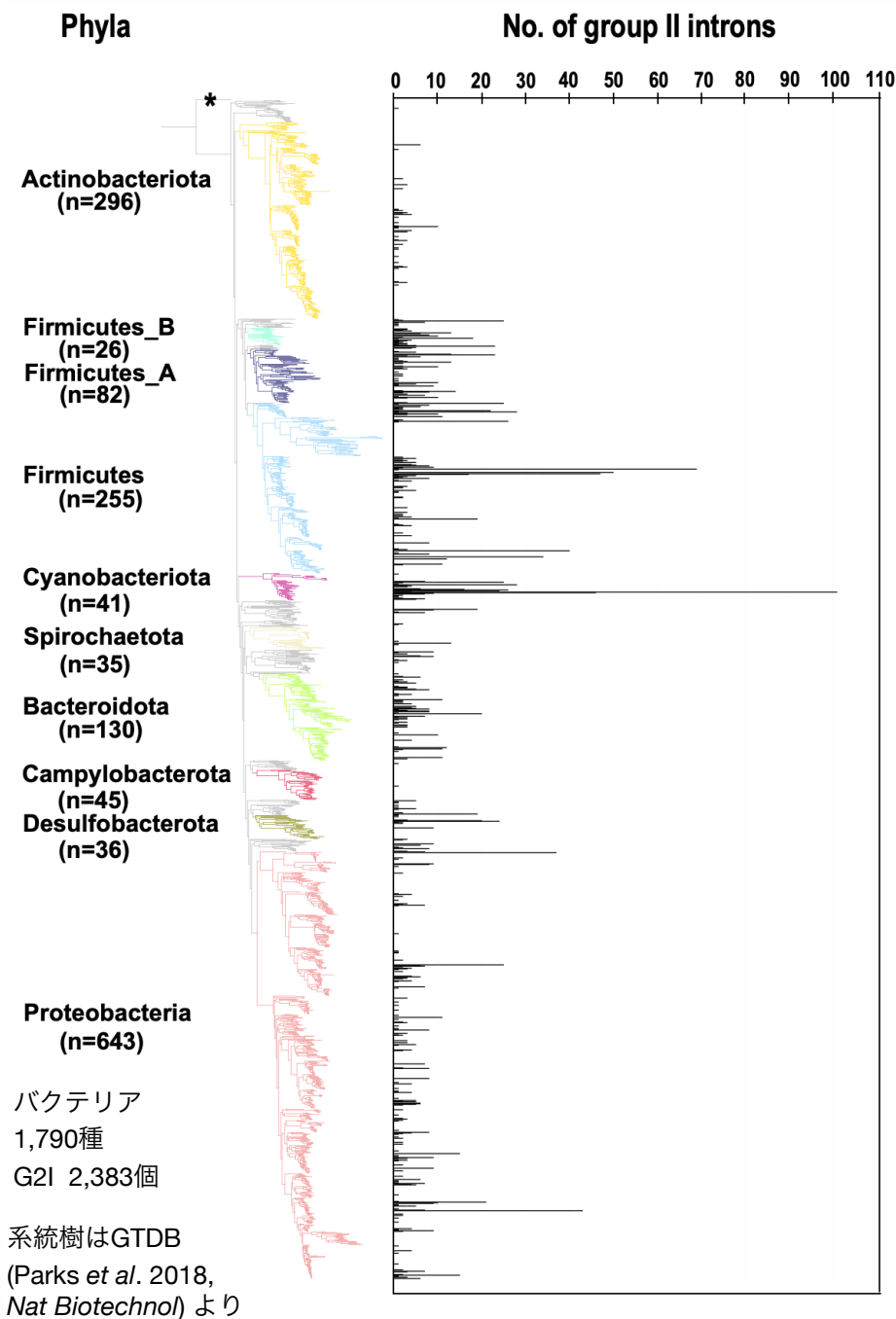
## 目的

グループIIイントロンの個数と  
原核生物の分類群との  
関連性はあるか?



Waldern et al. 2020, *Mol Biol Evol.*  
Candales et al. 2012, *Mob DNA.*

# バクテリア G2I の個数と分布

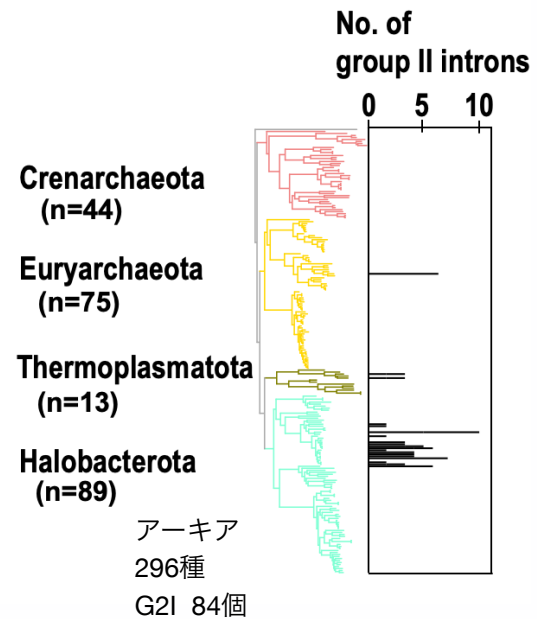


バクテリアの分類群ごとに、ばらつきを見せる

- Firmicutes\_B  
→ G2I の個数が10個以上の種が11%
- Cyanobacteriota  
→ G2I の個数が10個以上の種が16%

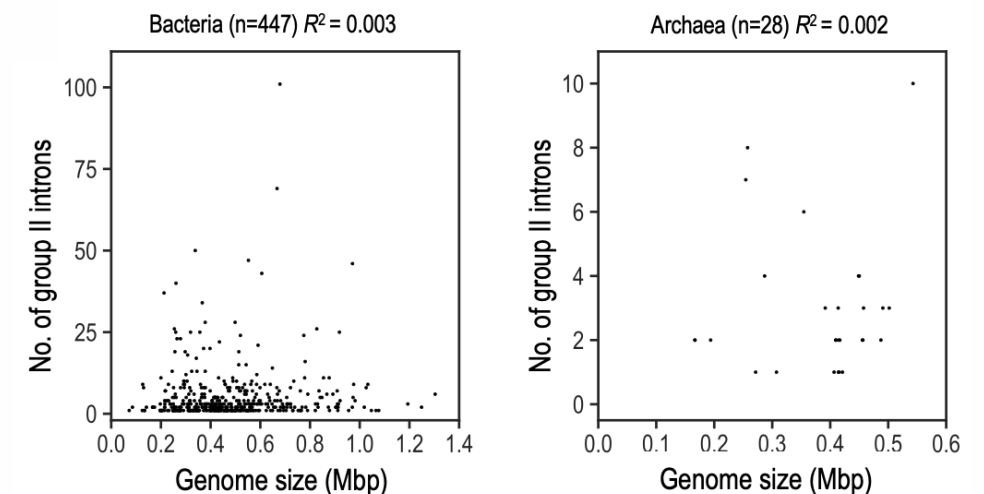
次は、特定の分類群で増えている理由を知りたい

# アーキアの場合



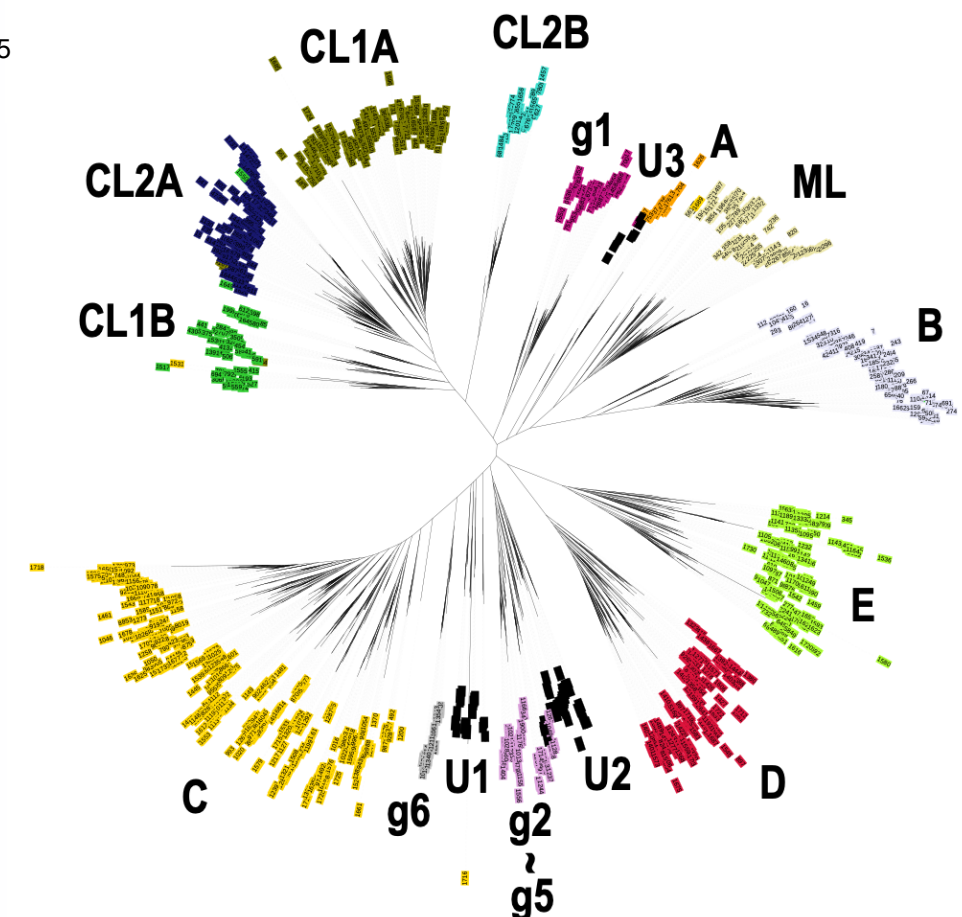
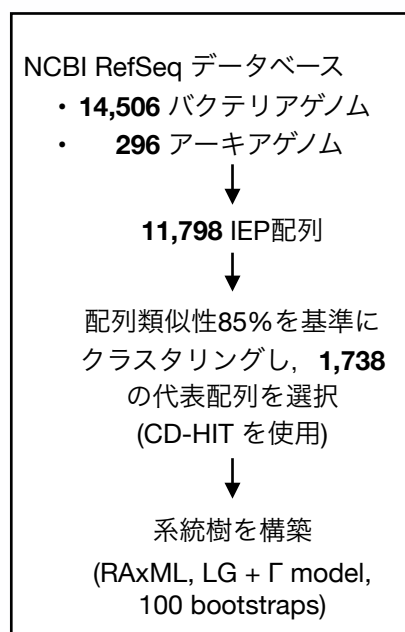
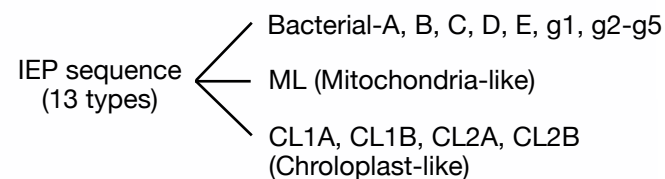
アーキアではG2Iの分布は限定的  
G2Iの個数もせいぜい10個

# G2I の個数とゲノムサイズの関係

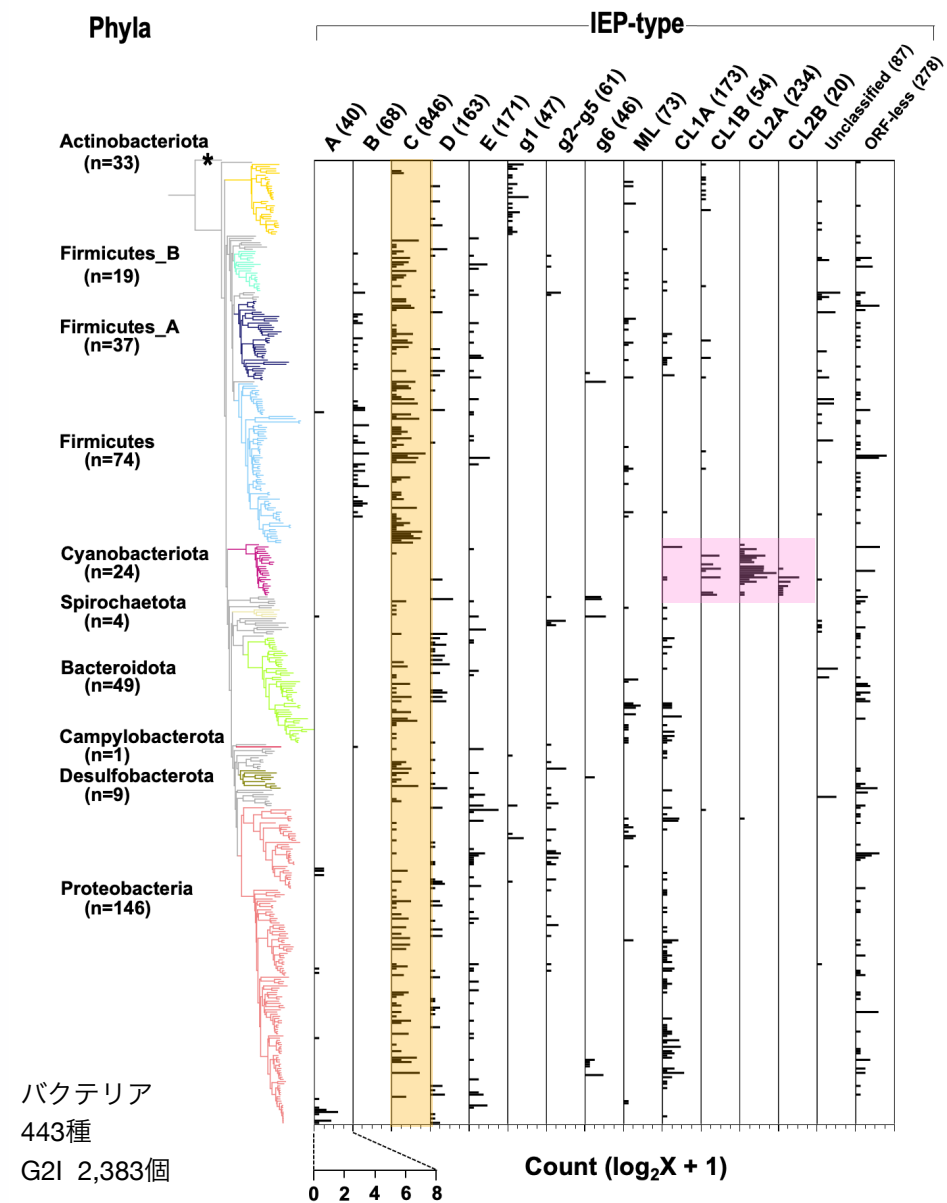


G2I の個数とゲノムサイズの相関はない

# G2I の分類

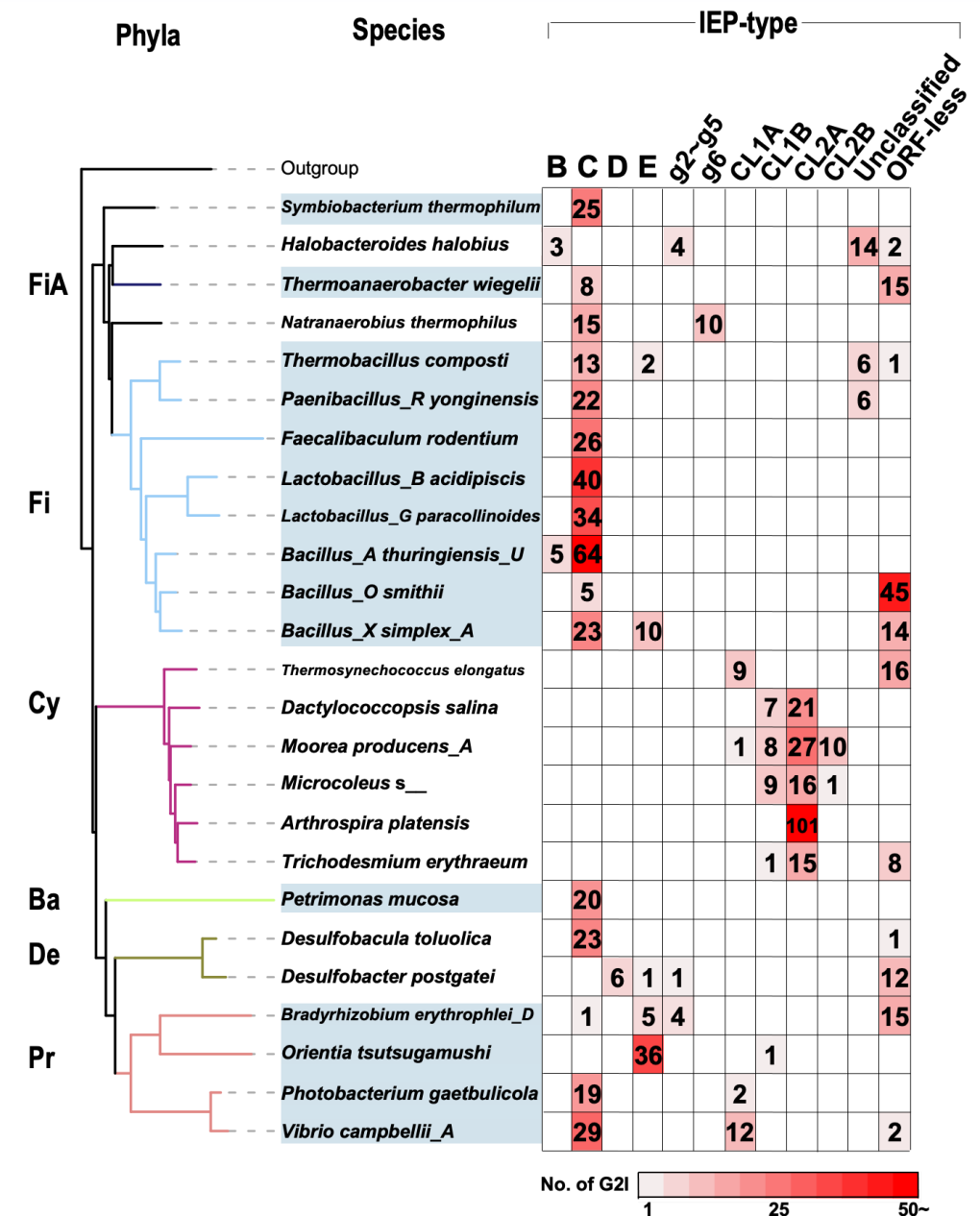


## G2I の分類と個数の増加

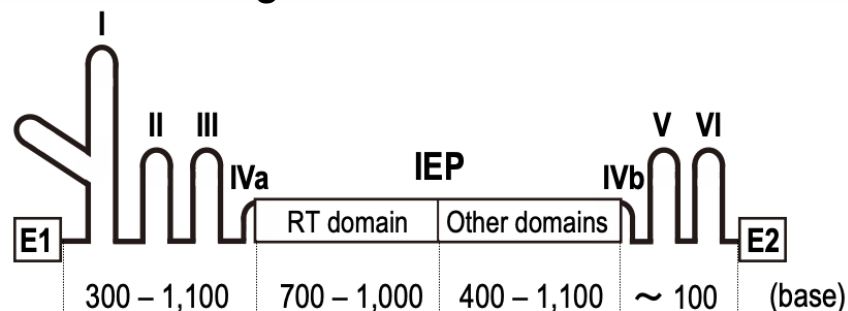


Bacterial-Cタイプ：幅広い門 (Firmicutes等)で増える  
CLタイプ：シアノバクテリアで増える

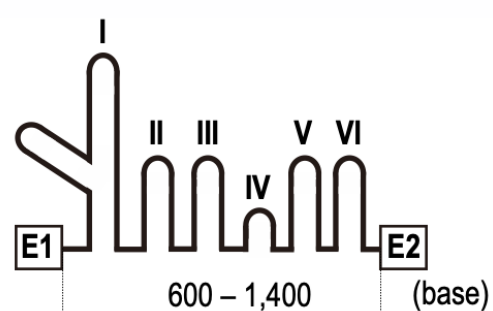
## 特に G2I を多く持つ種



### ORF-containing 型



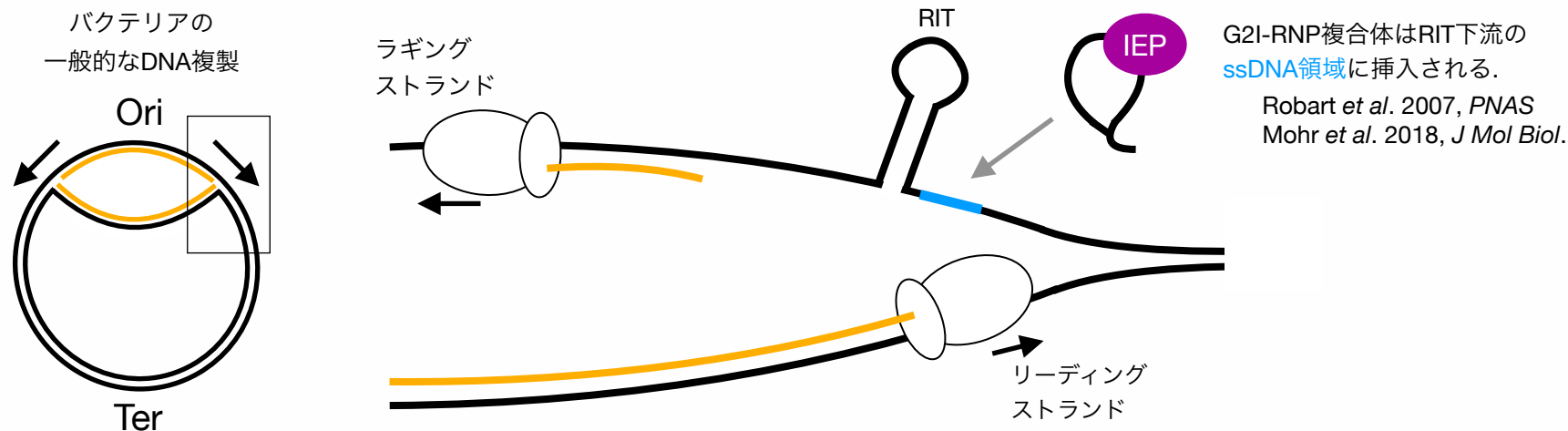
### ORF-less 型



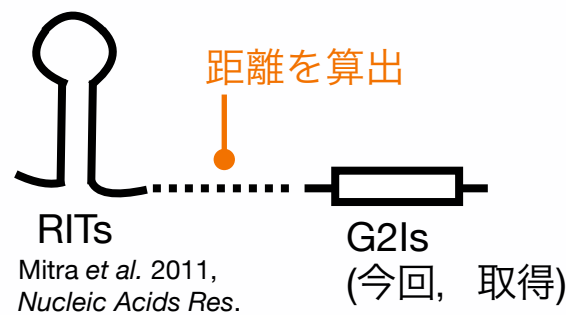
G2I の増加に寄与する因子として2つを検討

- Rho-independent transcription terminator
- GC-skew

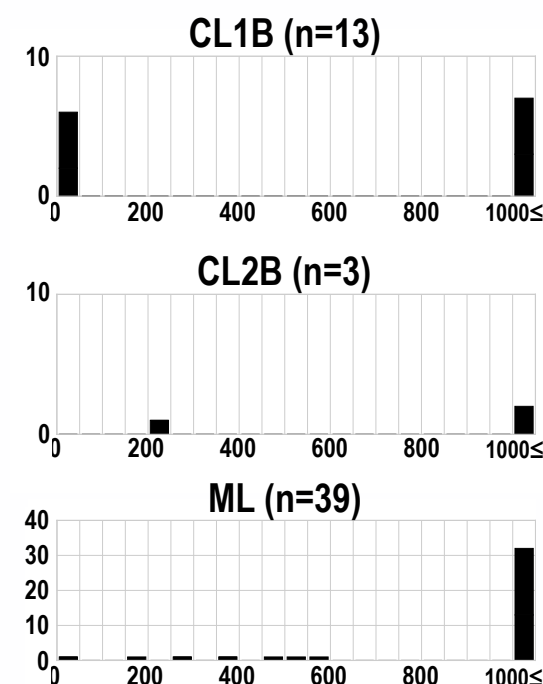
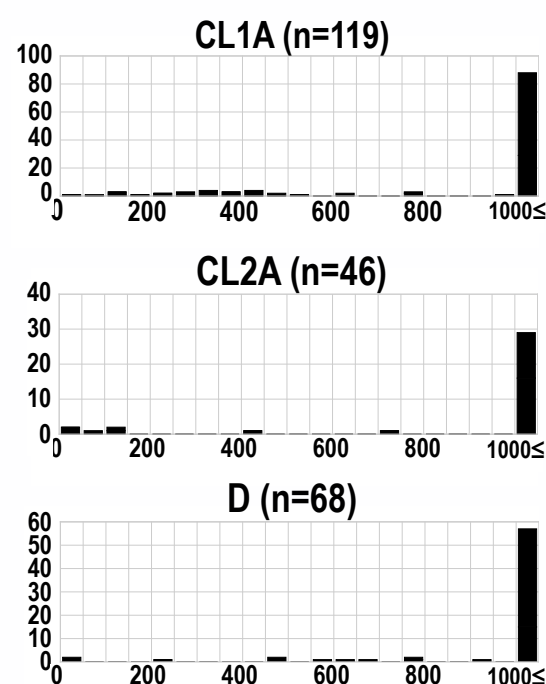
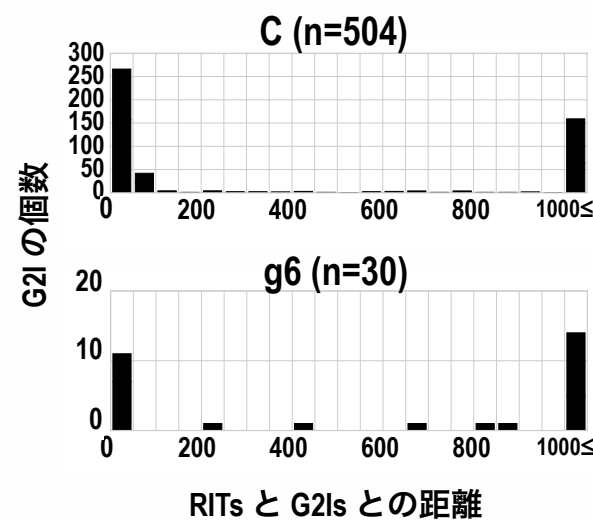
# Bacterial-Cタイプの転移機構は RIT (Rho-independent transcription terminator) と関連する



## IEPタイプごとの G2I と RIT との距離



Bacterial-Cタイプ: RITs の直後にある  
CLタイプ: ほとんどは RITs の直後にはない



RITs と G2Is との距離

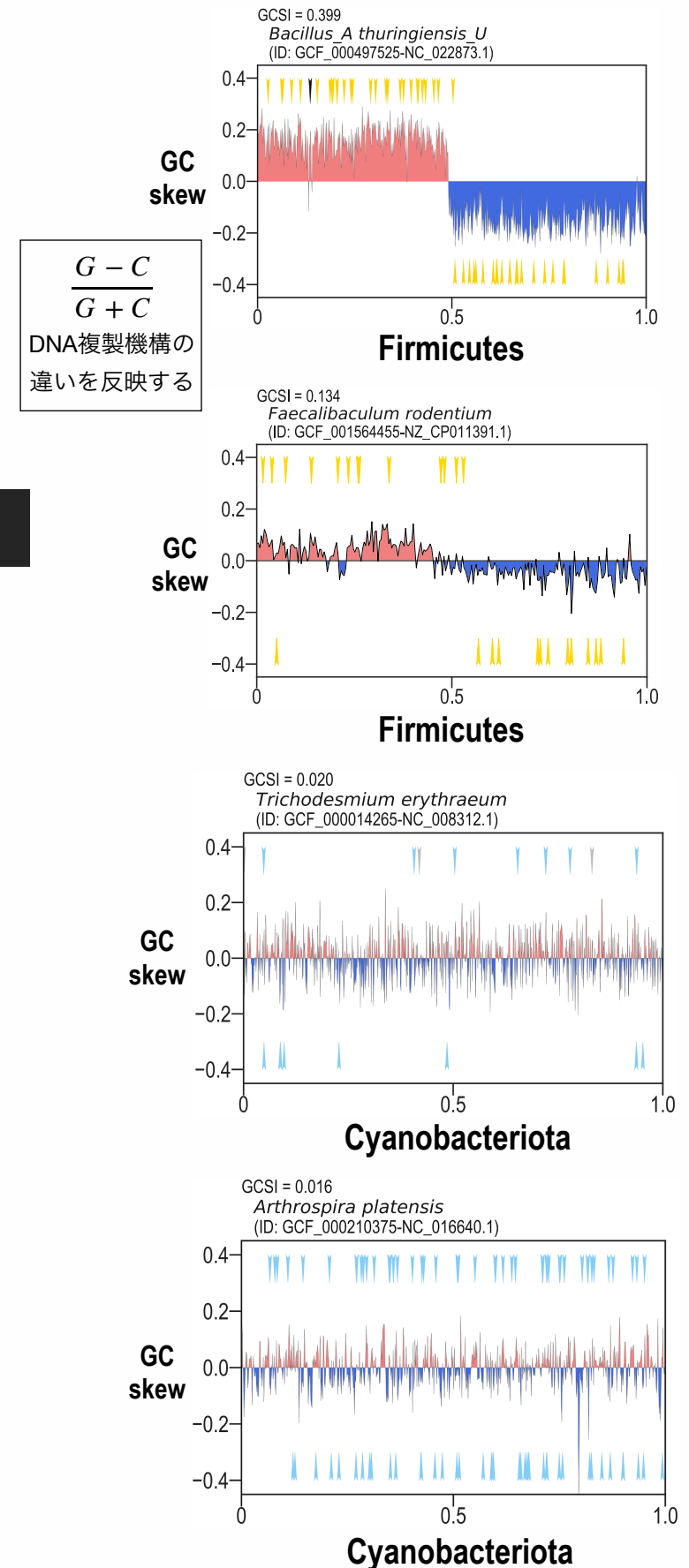
## GC skew



G2Is C-type

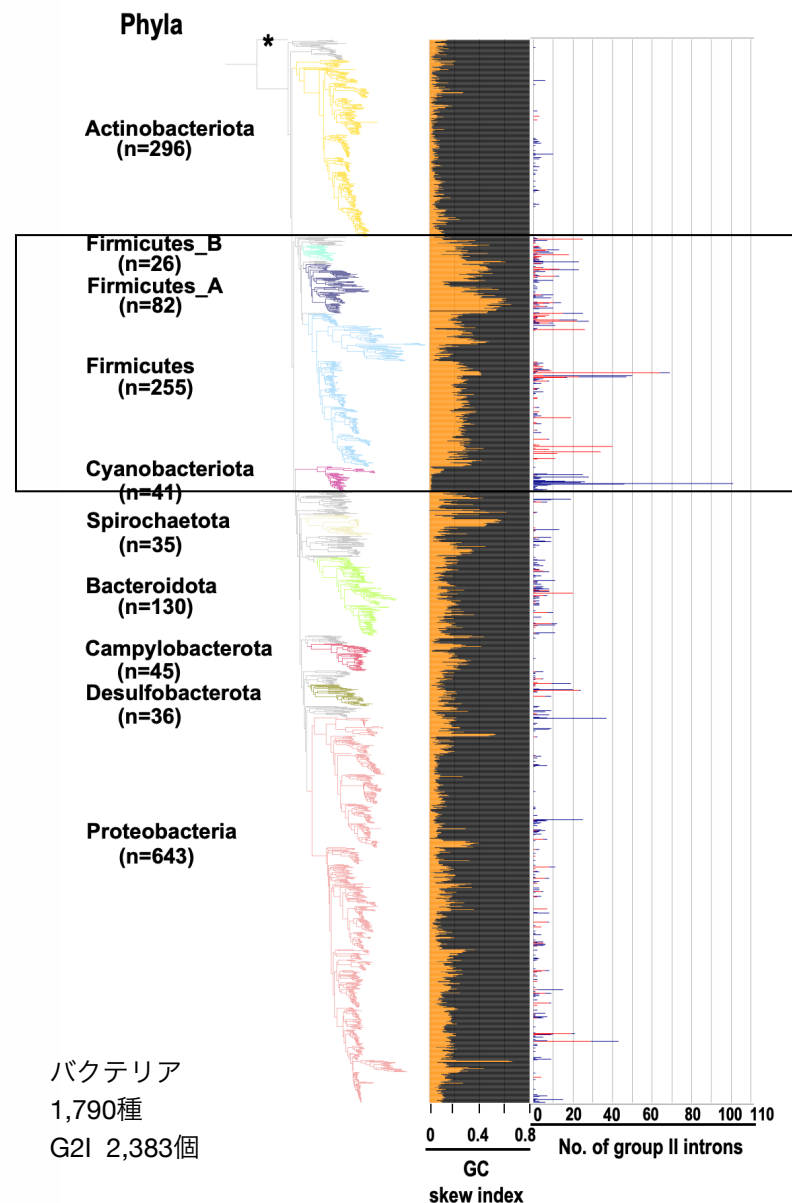


G2Is CL-type

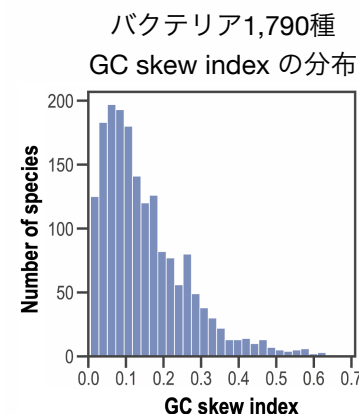




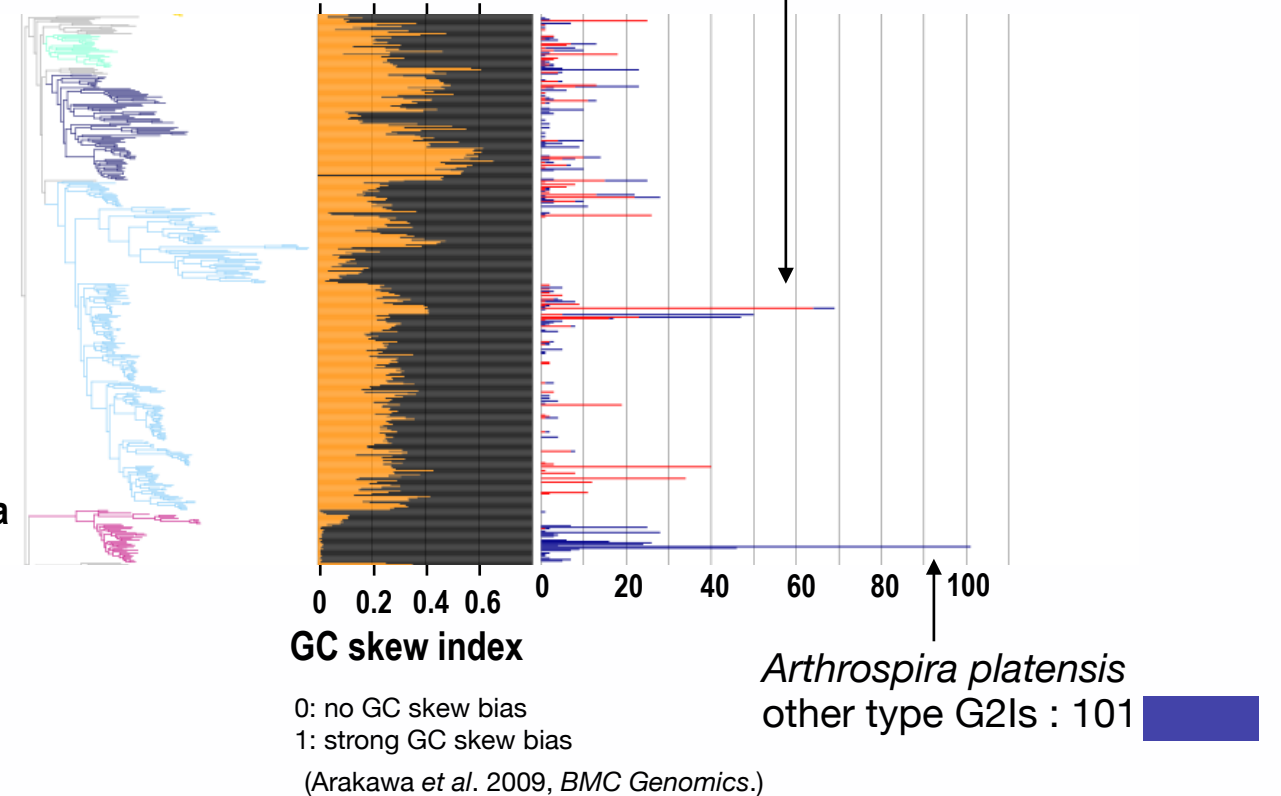
# GC skew の偏りと G2I の個数との関連性



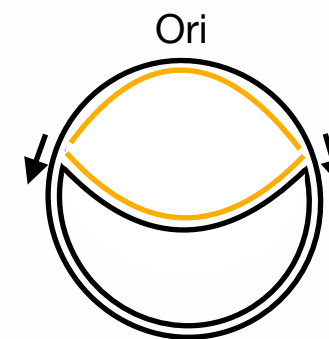
バクテリア  
1,790種  
G2I 2,383個



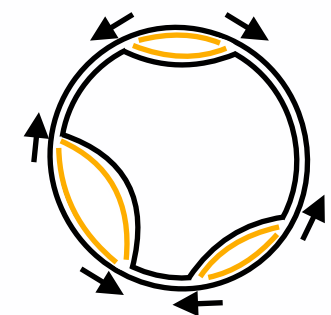
Firmicutes\_B (n=26)  
Firmicutes\_A (n=82)  
Firmicutes (n=255)  
Cyanobacteriota (n=41)



- C-type G2I は比較的高い GC skew index のゲノムで増幅
- CL-type G2I は比較的低い GC skew index のゲノムで増幅



ほとんどのバクテリアは  
単一の複製開始点  
→ 強い GC skew bias



シアノバクテリアの一部は  
特定の複製開始点を持たない  
→ 弱い GC skew bias

## まとめ

G2I が多い生物種と少ない生物種との違いは何か？

G2I を見つけるパイプラインを構築 → 多量のゲノムデータに適用

特定のIEPタイプの G2I が増幅している

- C-type → 幅広い系統 (特に Firmicutes)
- CL-type → シアノバクテリア

G2I 自体の分類、生物種ごとのゲノム構造の違いの関与を示唆

## 謝辞

以上の方々にお世話になりました。  
永田祥平さん、玉木聡志さん、富田勝さん、金井昭夫さん、RNA機能解析グループの皆様