

化石制約がベイズ年代推定に及ぼす影響

Jun Inoue, Philip C.J. Donoghue
and Ziheng Yang

Introduction

ベイズ法による年代推定:

DNA などの配列データに化石記録などの情報を盛り込んだ解析ができる.

進化速度

制約年代 ← 化石

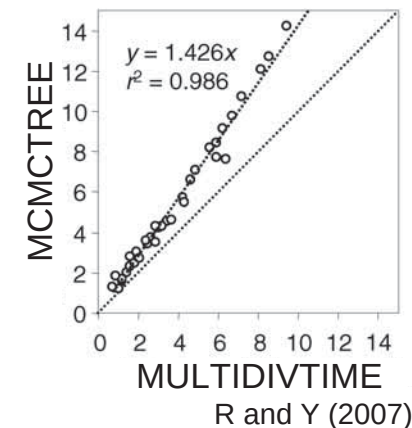
尤度 ← 配列データ

→ MCMC → 事後分布

解析ソフトウェア: MULTIDIVTIME, MCMCTREE, BEAST など

ベイズ年代推定はパラメーターとなる多くの要素が含まれている. これら要素のどれが事前分布 (推定年代など) に大きな影響を及ぼすのか明らかになっていない.

例 . Rannala and Yang(2007) が Jonson et al. (2006) のネコ科のデータ (Jonson et al. 2006) を解析したところ, MCMCTREE では MULTIDIVTIME より 1.4 倍古い推定値が得られた. なぜか?



目的

分岐年代のベイズ推定に，事前分布とその他の要因が及ぼす影響を評価する

MCMCTREE (PAML) と MULTIDIVTIME の違い

	MULTIDIVTIME	MCMCTREE
根幹の年代に対する制約	ガンマ事前分布が必要.	根幹の年代に上限がない場合, ラフな上限が必要
外群	外群が必要.	有根樹であれば内群だけでよい.
年代に対する事前分布の設定方法		次のスライド
化石制約の上限と下限		次のスライド

MULTIDIVTIME と MCMCTREE における分岐年代・事前分布の仕様

MULTIDIVTIME

ガンマ事前分布に基づいて根幹の分岐年代が与えられる。



他の分岐年代の分布が割り当てられる。



MCMC アルゴリズムで上限と下限制約を破った推定年代が取り除かれる。

MCMCTREE

制約が施された分岐 (複数) の年代に対する同時分布が作成される。



制約を施した分岐の年代を仮定して、制約なしの分岐年代の分布が割り当てられる。



MCMC アルゴリズムで切り打 (truncation) が行われる。

MCMCTREE4.2 の下限制約に導入された新しい概念

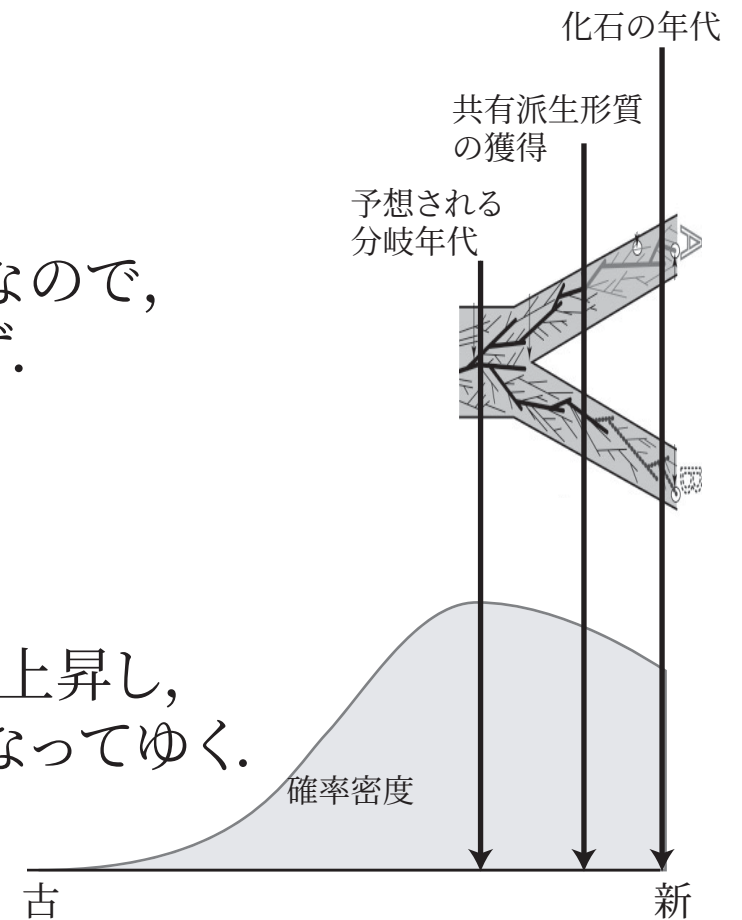
共有派生形質は種の分岐の後で獲得されるはずなので、
実際の分岐年代は化石が出た年代よりも古いはず。
(e.g., Steiper and Young, 2008).



確率密度は化石の出た年代から古くなるに従って上昇し、
予想される分岐年代で最大に達し、その後は低くなってゆく。

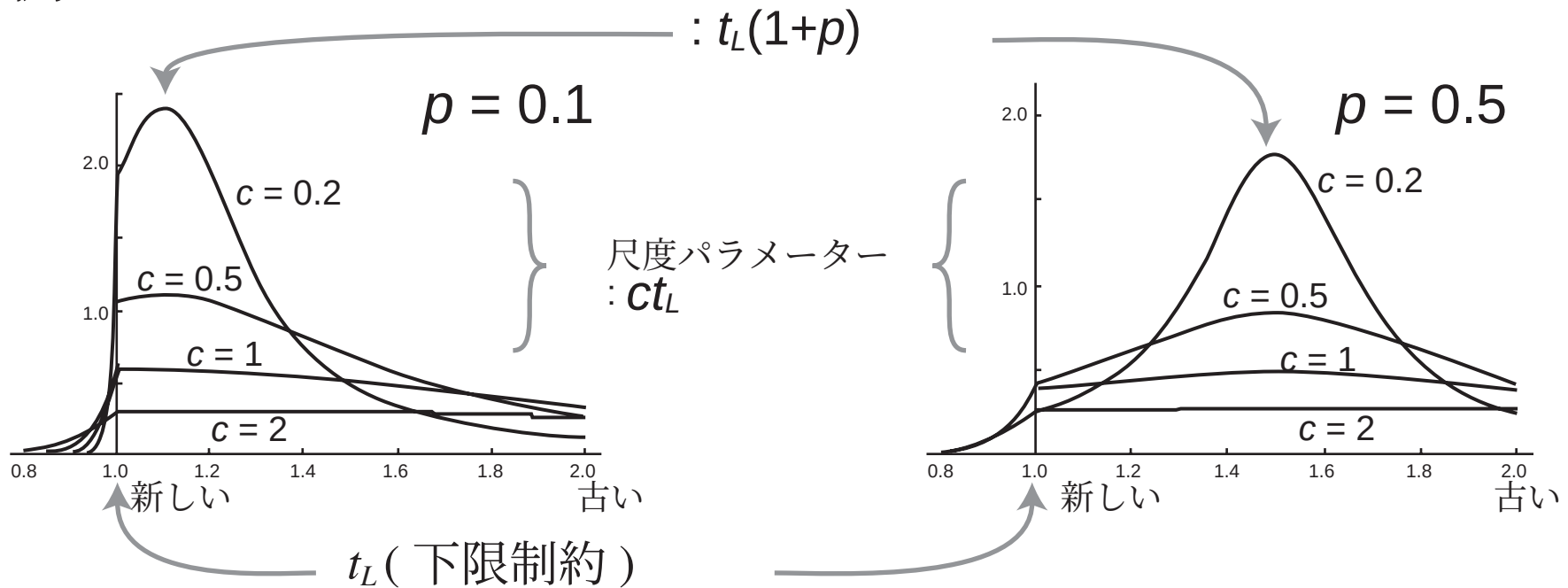


切断コーシー分布 (truncated Cauchy distribution)
を下限制約に導入。



MCMCTREE の下限制約に導入された コーシー分布 ' $L(t_L, p, c)$ '

例：



- c は p よりも推定年代の事後分布に大きな影響を与える。
- 厳密には、それぞれの下限制約に異なる c と p を与える必要がある。

本研究に用いた 3 つのプログラムの違い

事前年代分布の割当方法

下限制約の分布

MCMCTREE4.1

制約を施した分岐年代
の同時分布



MCMCTREE4.2

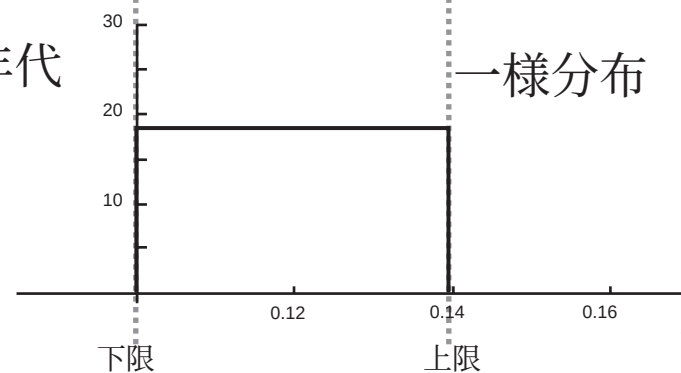
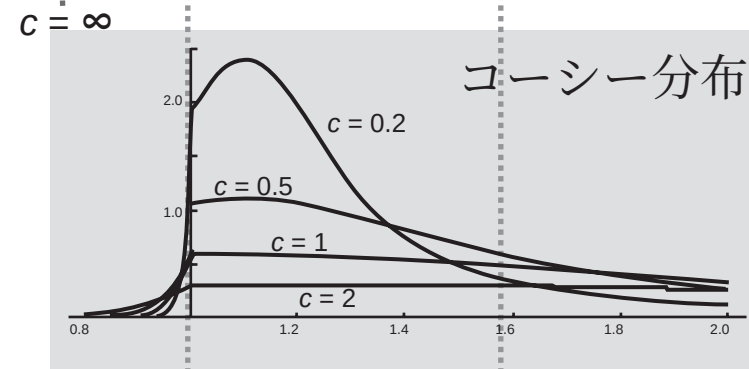
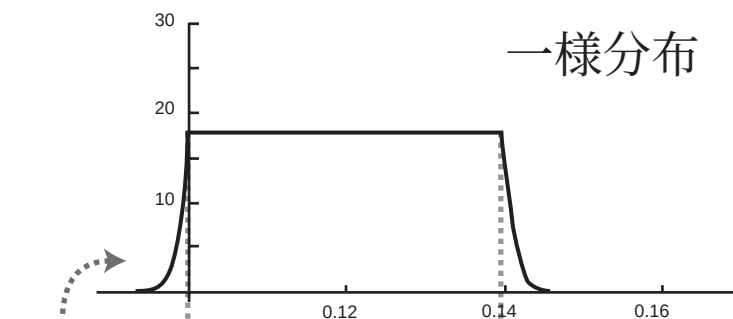
MCMC 解析

MULTIDIVTIME

ガンマ分布による根幹の分岐年代



制約付き年代を用いて
MCMC 解析



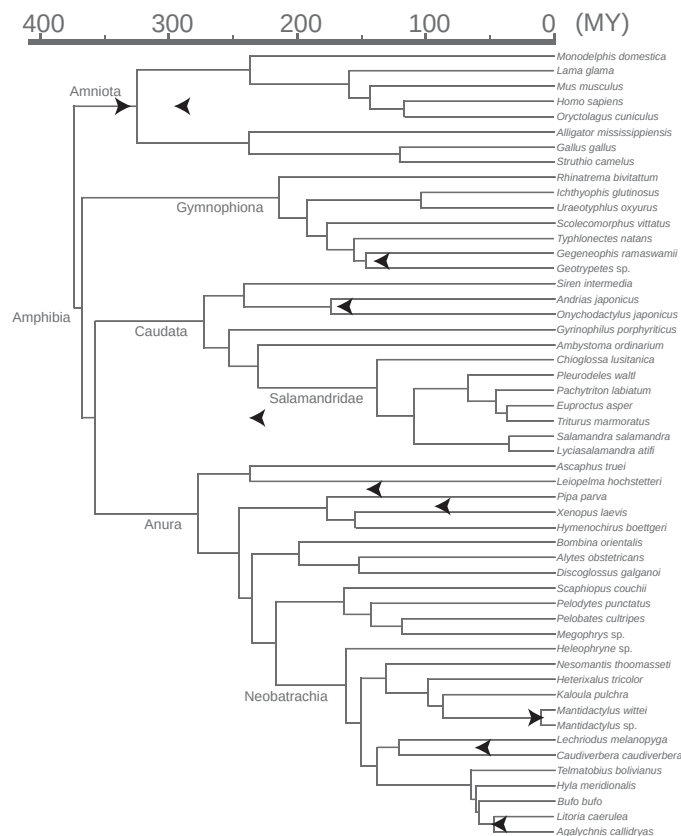
3 プログラムの他のセッティングは できるだけ同じにした

	MULTIDIVTIME		MCMCTREE4.1, 4.2
尤度	MLEs の正規近似	=	MLEs の正規近似
進化モデル	F84+ G ₅	=	F84+ G ₅
速度変動パラメーター のガンマ事前分布	平均, 標準偏差	=	形状パラメーター (α) と 尺度パラメーター (β)
速度相関モデル	Auto-correlated	=	Auto-correlated
根幹年代の ガンマ事前分布	すべてのデータセット	≈	Frog データセットのみ

本研究に用いた3つのデータセット

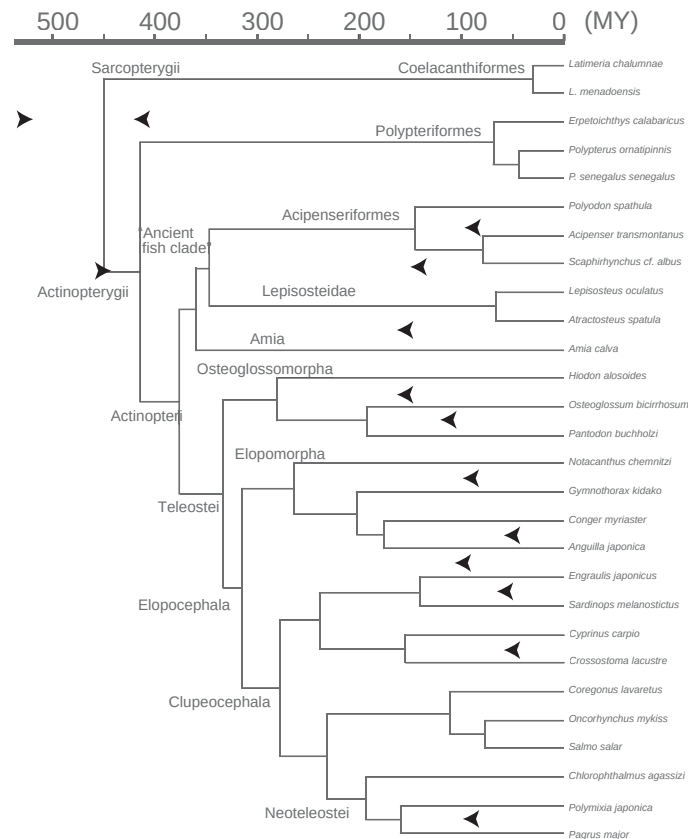
Frog (両生類) San Mauro et al. (2005)

RAG1 遺伝子: 1 遺伝子
52 種
1368 bp.
2 上限, 8 下限 制約



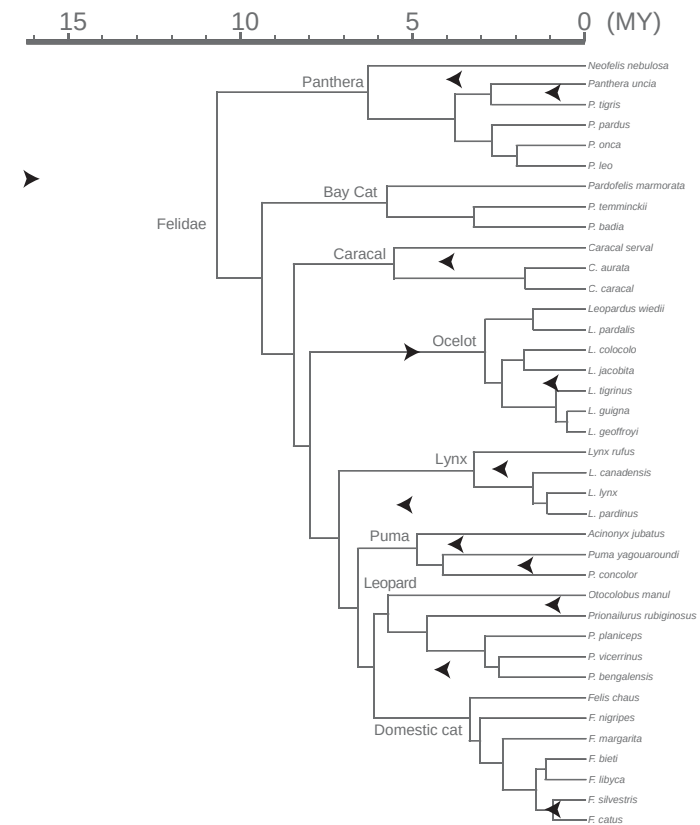
Fish (条鰭類) Inoue et al. (2005)

mt ゲノム: 36 遺伝子
28 種
10,327 bp.
2 上限, 12 下限 制約

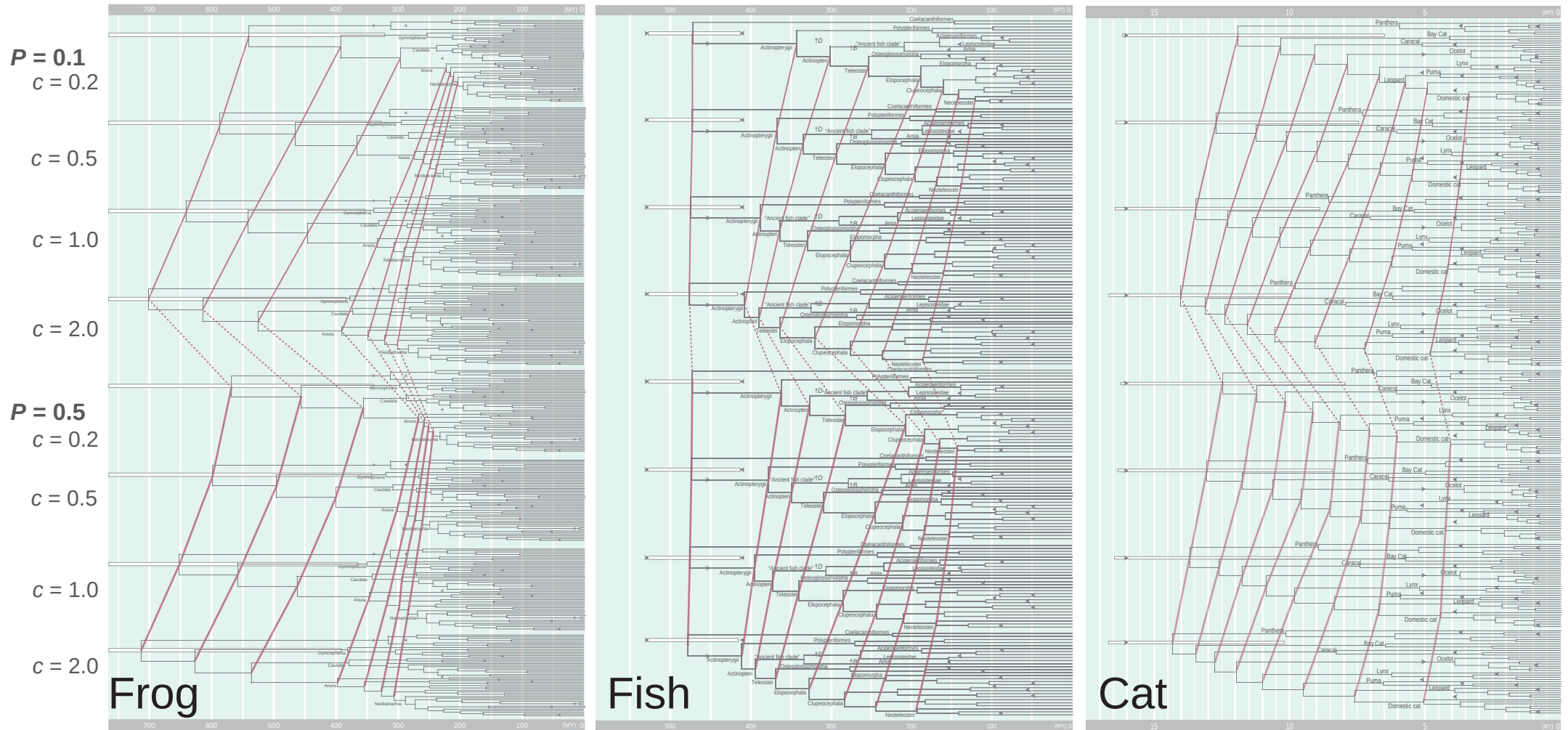


Cat (ネコ科) Johnson et al. (2006)

核遺伝子: 30 遺伝子
38 種
19,984 bp.
1 上限, 11 下限 制約

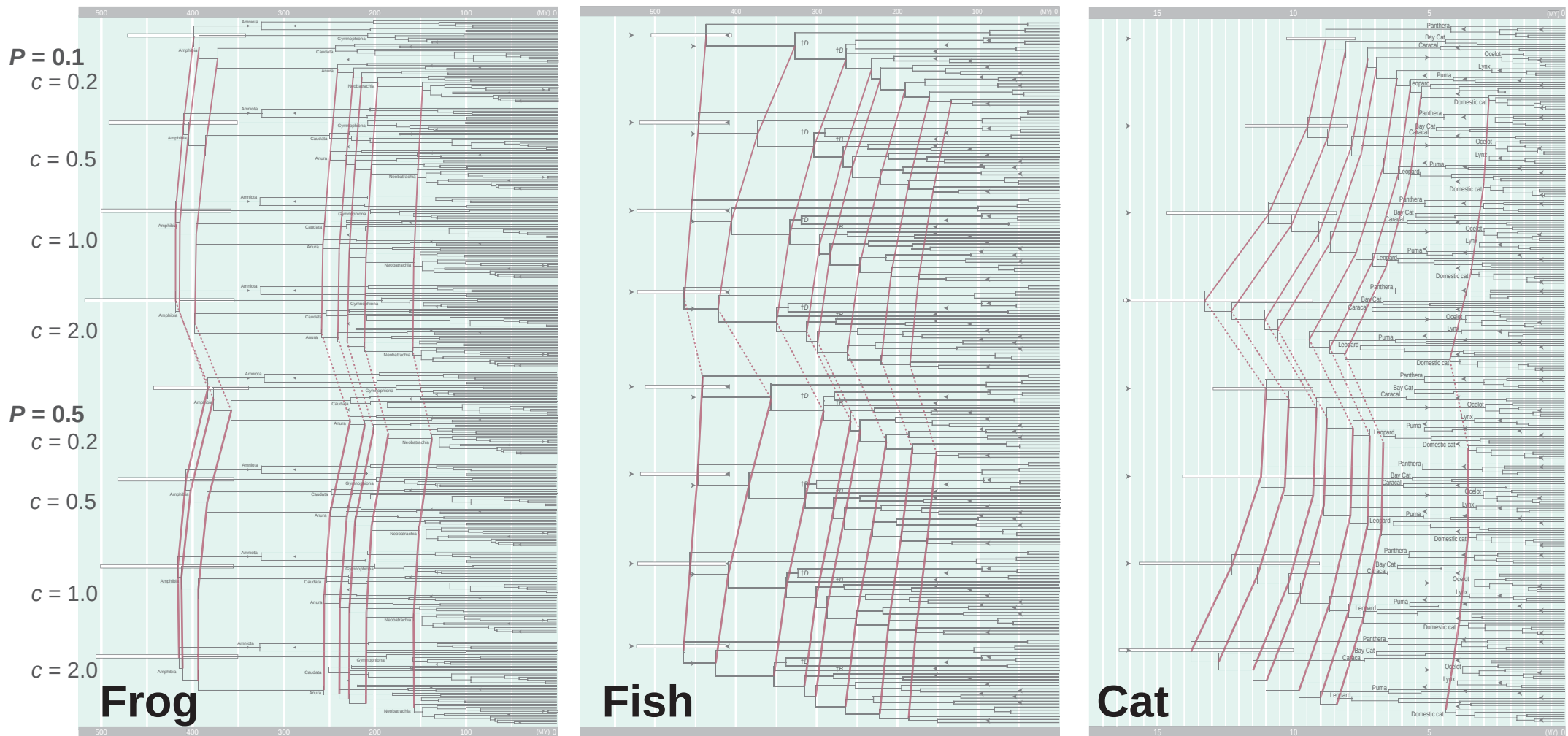


MCMCTREE4.2 の下限制約に用いるパラメーター p と c が推定年代 (平均値) の事前分布に及ぼす影響



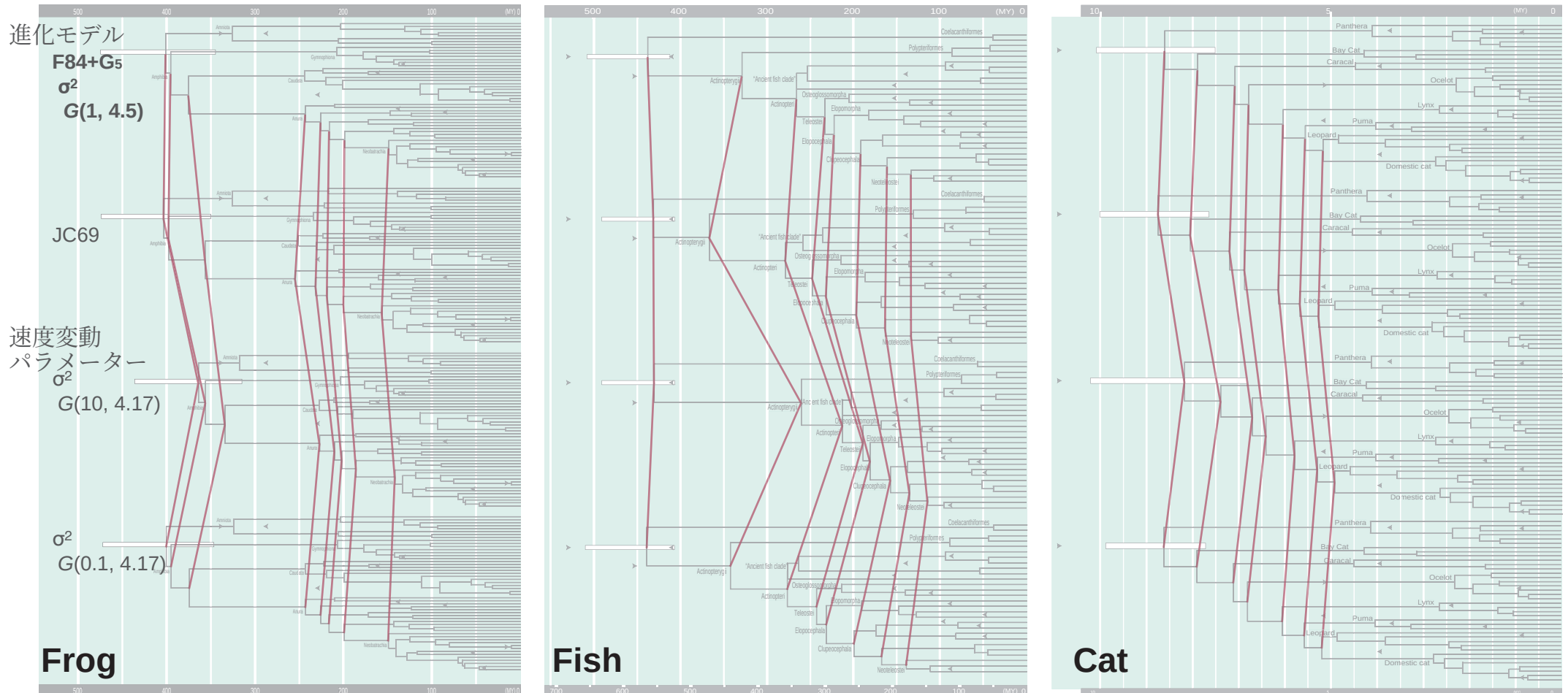
p と c のどちらを増加させても、推定年代の事前分布は古くなった。
 c の方がより大きな影響を及ぼした。

MCMCTREE4.2 の下限制約に用いるパラメーター p と c が推定年代 (平均値) の事後分布に及ぼす影響



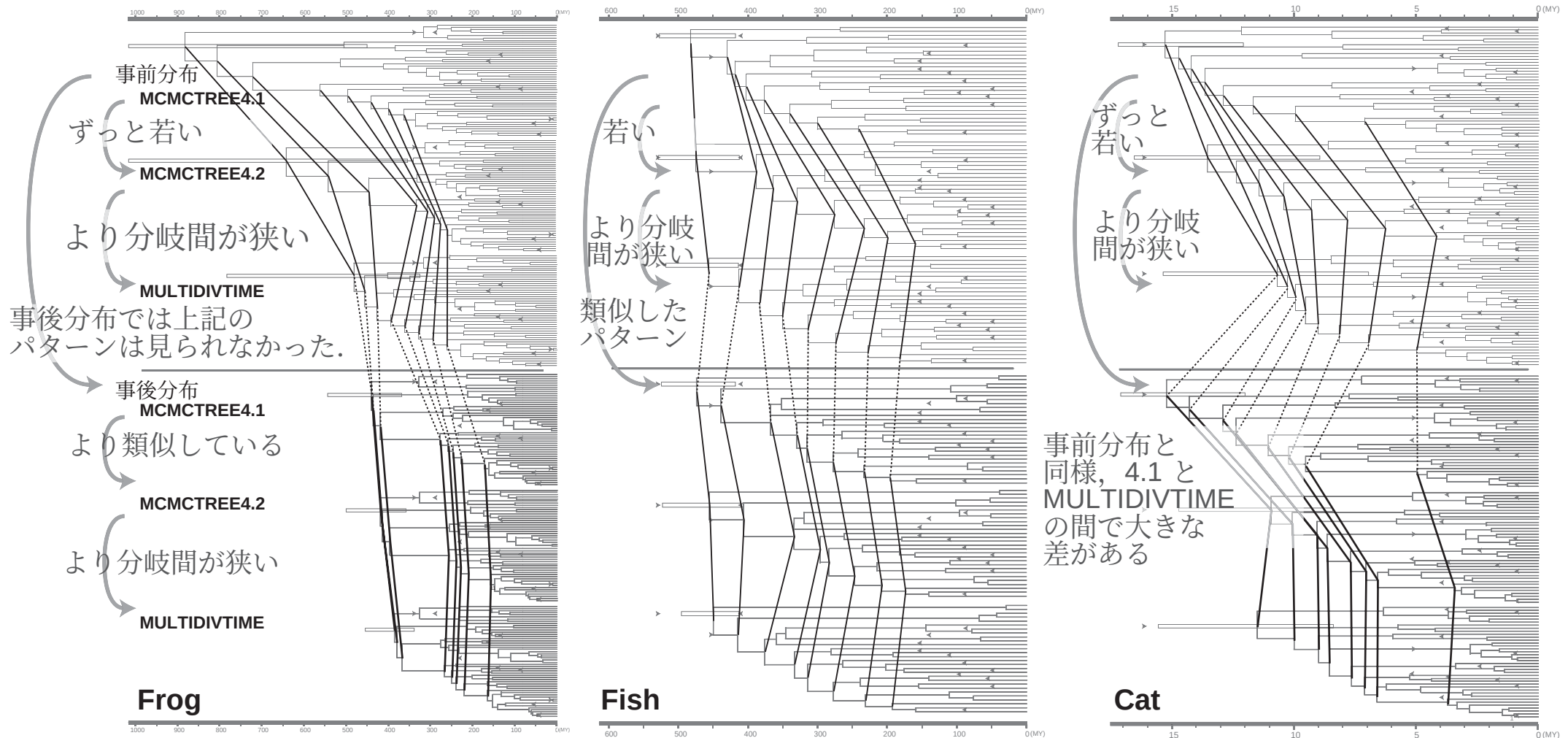
事前分布と同様に、 p と c のどちらを増加させても、推定年代は古くなった。
予備的な解析に基づいて、 $p = 0.1, c = 1.0$ を MCMCTREE のデフォルトにした。
(本来は化石ごとに異なる p と c の値を設定する必要がある)

進化モデルや rate-drift parameters が MCMCTREE4.2 の解析結果に及ぼす影響



進化モデルによる影響は小さかった。
速度変動パラメーター σ^2 は、より大きな影響を及ぼした。

MCMCTREE と MULTIDIVTIME 解析によって 得られた分岐年代 (平均) の事前分布と事後分布



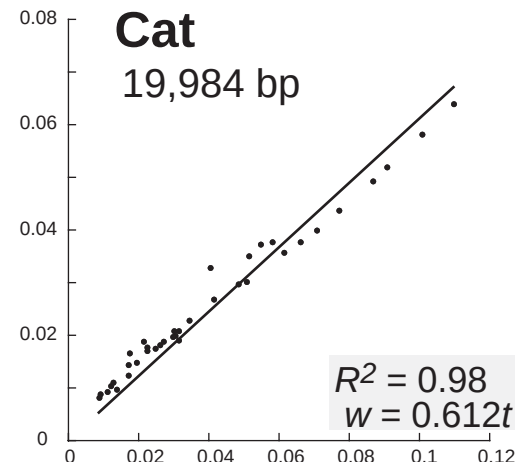
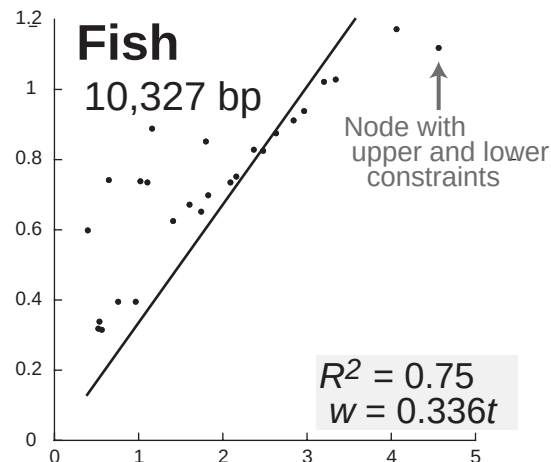
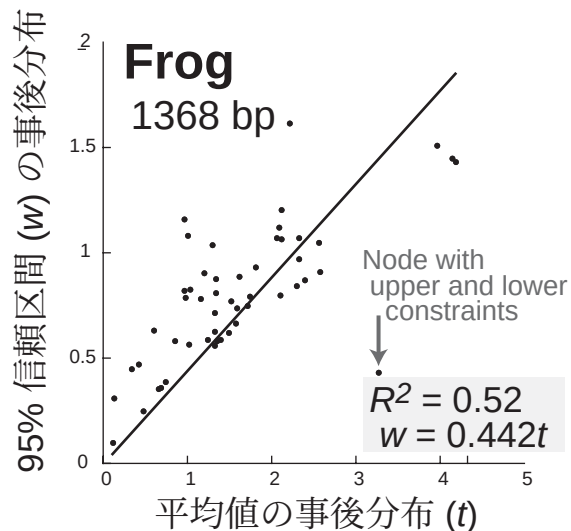
事後分布であっても，上限制約と下限制約にどのような分布を用いるかによって，得られる結果が異なっていた。

Infinite-sites plot

配列データが無限に長くなる



それぞれの分岐年代の平均と 95% 信頼区間
が直線上に並ぶ



相関 (R^2)



配列の長さ

弱い



不十分

強くなる



より良い

ほぼ完璧



ほぼ飽和に達している

傾き (w)



年代制約

平均が 1 Myr 古くなると信頼
区間 0.44 MY 広がる

より小さい



より情報が多い

より大きい



より情報が少ない



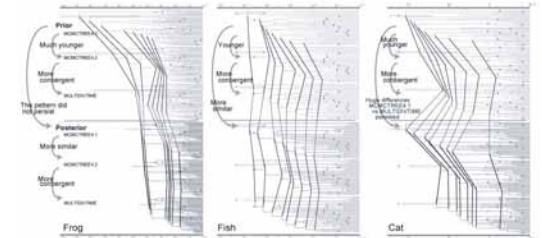
配列を長くしても、事後確率の曖昧さや事前確率から受ける影響がなくなることは無い。

まとめ

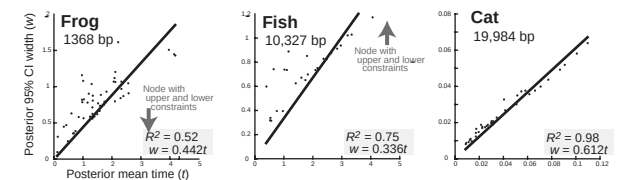
各要因が結果に与える影響：化石制約による影響が最も大きかった。



化石制約の利用：分岐年代の事後分布は上限・下限制約に施す事前分布の違いに影響される



Infinite-sites theory: 配列を長くしても，事後確率の曖昧さや事前確率から受ける影響がなくなることは無い。



- 分子データに基づく年代推定では化石制約が極めて重要。
- 年代推定に利用可能な化石記録をまとめる必要がある。