

2020年12月4日(金)第29回バイオメディカル研究会「Withコロナ時代の創薬」

バイオインフォマティクスを用いた 感染症研究

北海道大学人獣共通感染症リサーチセンター
バイオインフォマティクス部門
伊藤公人

疫学とゲノムのビッグデータ

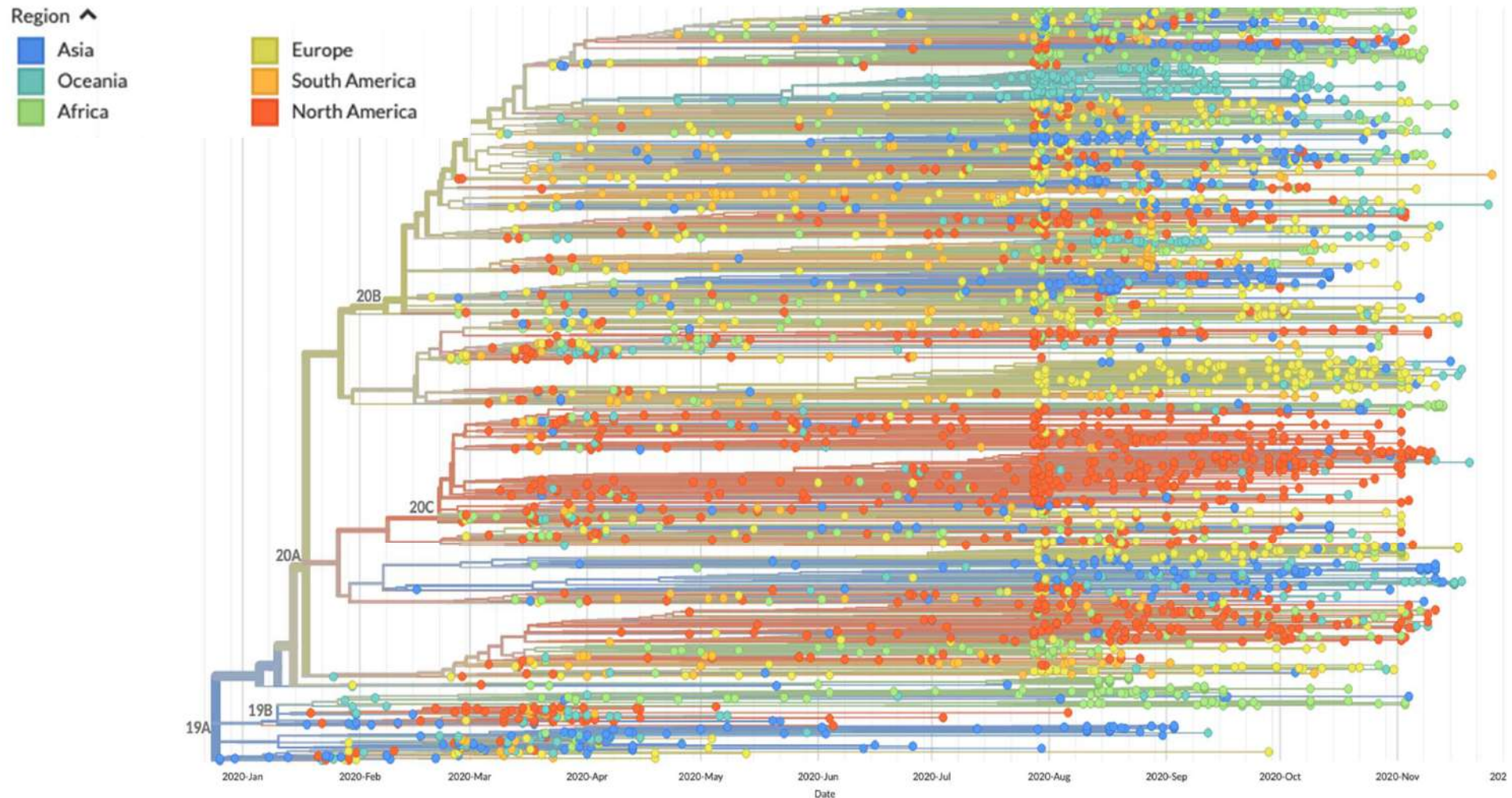
■ インフルエンザ

- 毎週の陽性者数を発表 (WHO, 国)
- 333,056株の塩基配列を公開 (GISAID)
(内 H1N1pdm09ウイルス 60,558株)

■ COVID-19

- 毎日の陽性者数と死者数を発表 (WHO, 国)
- 230,069株の塩基配列を公開 (GISAID)

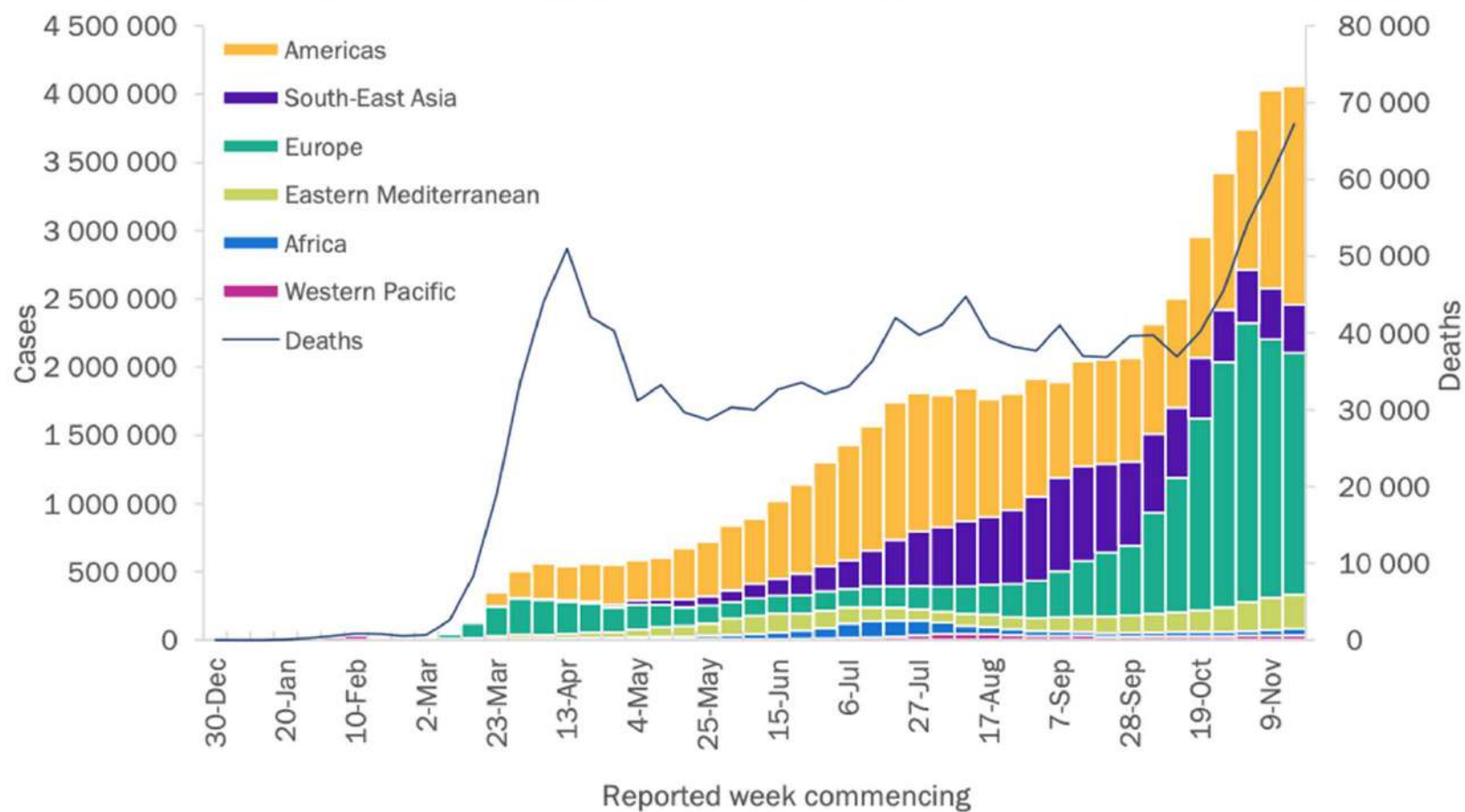
SARS-CoV-2塩基配列の進化



SARS-CoV-2進化のリアルタイム可視化 (<http://nextstrain.org>)

世界のCOVID-19の流行状況

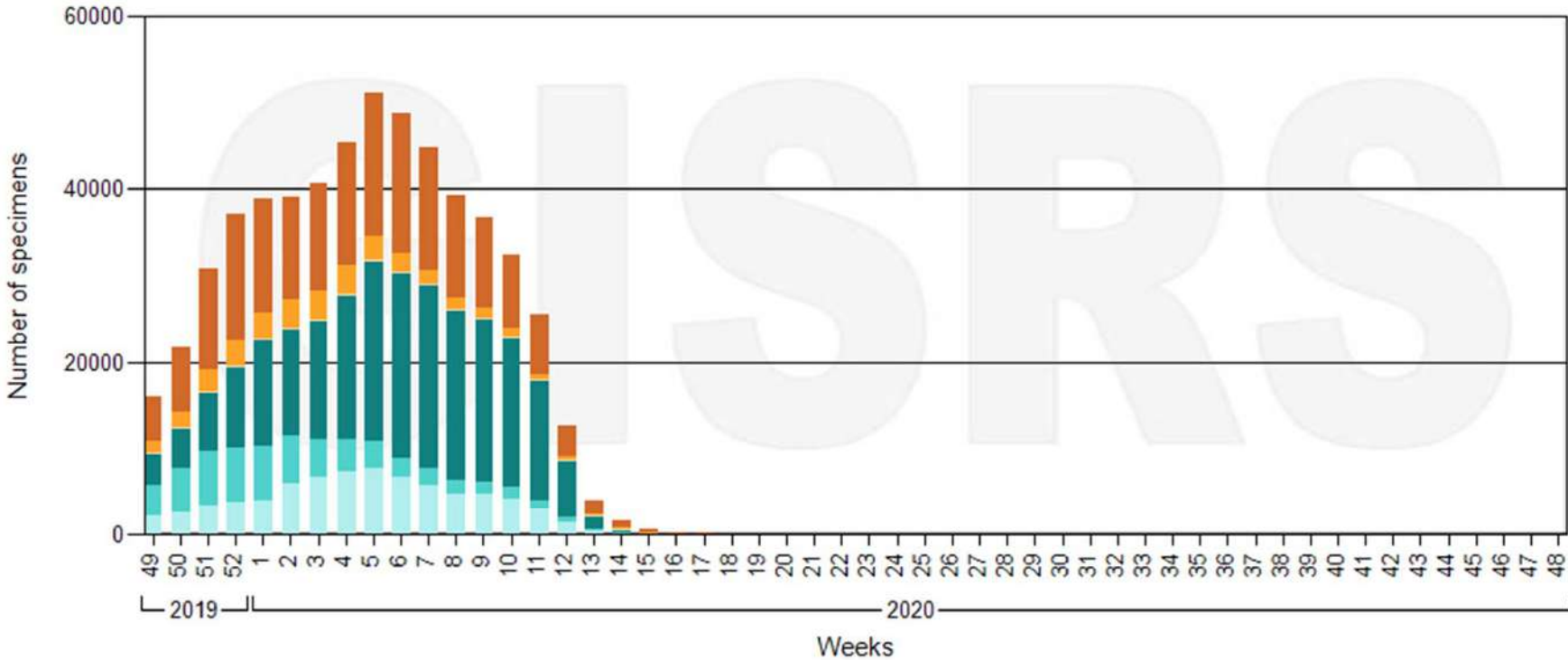
Figure 1: COVID-19 cases reported weekly by WHO Region, and global deaths, as of 22 November 2020**



COVID-19 Weekly Epidemiological Update, WHO

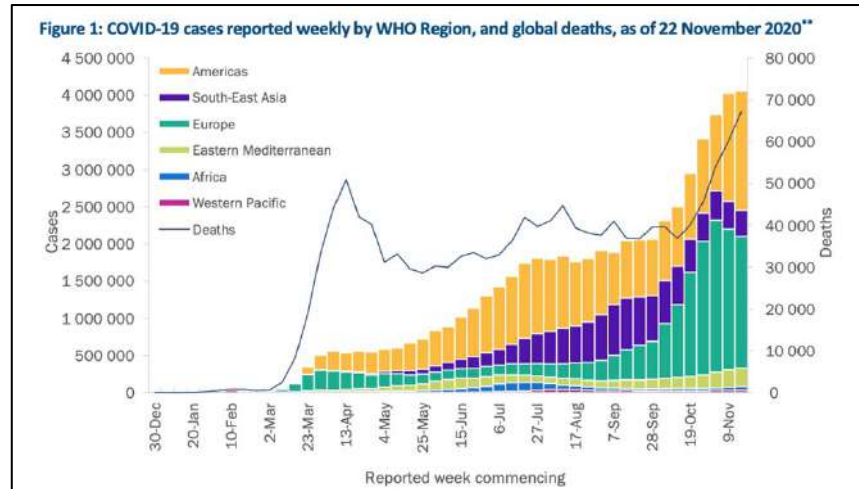
世界のインフルエンザの流行状況

Number of specimens positive for influenza by subtype



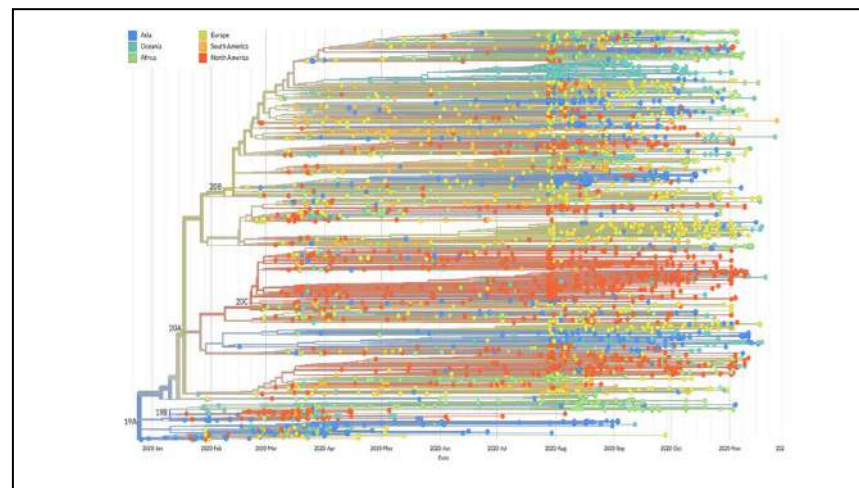
Data from: All sites

疫学解析と進化生物学解析



■ 疫学解析

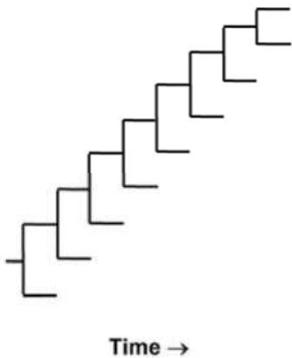
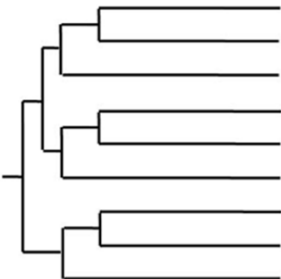
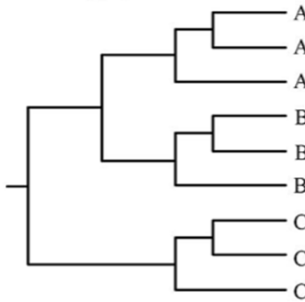
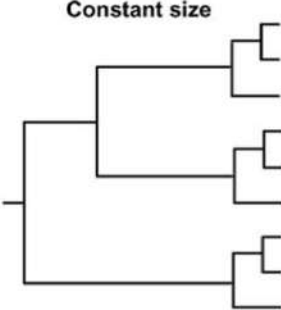
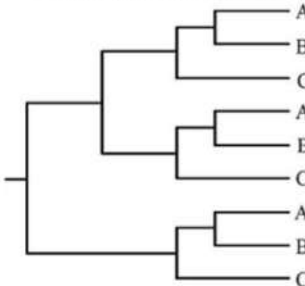
- 基本再生産数(R_0)および実効再生産数(R_t)の推定
- SIRモデルによる R_0 の推定
- 再生方程式による R_t の推定



■ 進化生物学解析

- 進化系統解析
- Coalescent 理論
- Allele頻度の変化の解析
- 自然選択の解析

Phylodynamics (Grenfell 2004)

	Continual Immune Selection	Weak or Absent Immune Selection Tree shape controlled by non-selective population dynamic processes	
		Population size dynamics	Spatial dynamics
Idealized Phylogeny Shapes		<p>Exponential growth</p> 	<p>Strong spatial structure</p> 
		<p>Constant size</p> 	<p>Weak spatial structure</p> 
Examples	Human influenza A virus intra-host HIV	inter-host HIV inter-host HCV	Measles, rabies inter-host HIV
Tree Inferences	Detection of antigenic escape mutations	Estimation of population growth rates	Estimation of population migration rates

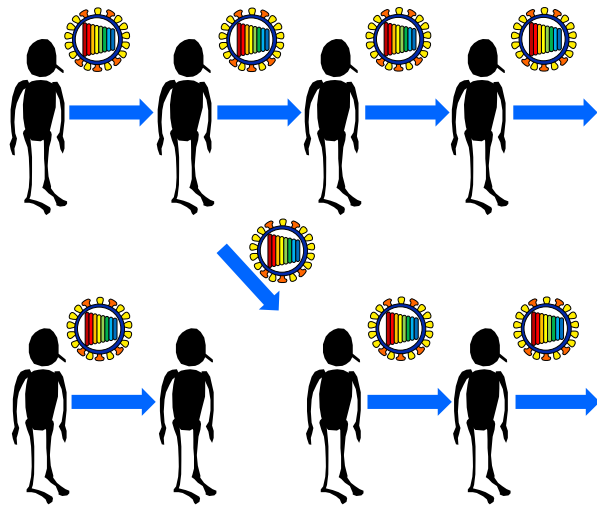
病原体の進化系統樹の形は感染症の流行動態によって異なる

進化系統学と感染症流行動態を融合した解析を“Phylodynamics”と呼ぶ。

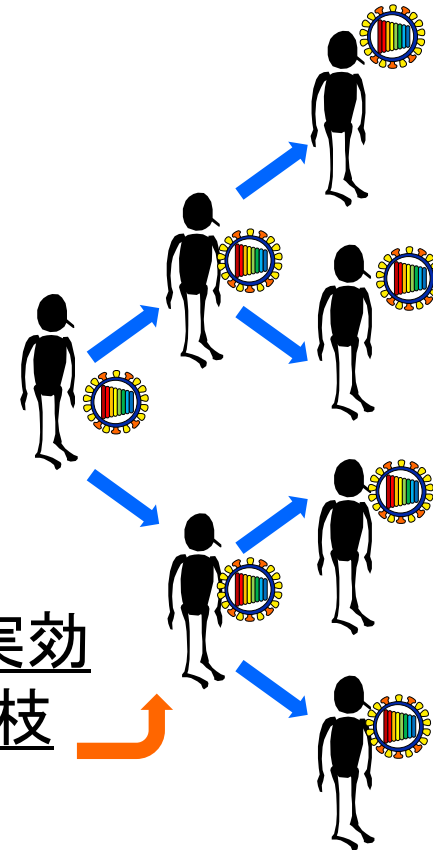
Grenfell et al.
Science (2004)

Phylodynamicsの基本的アイデア

実効再生産数が1の場合



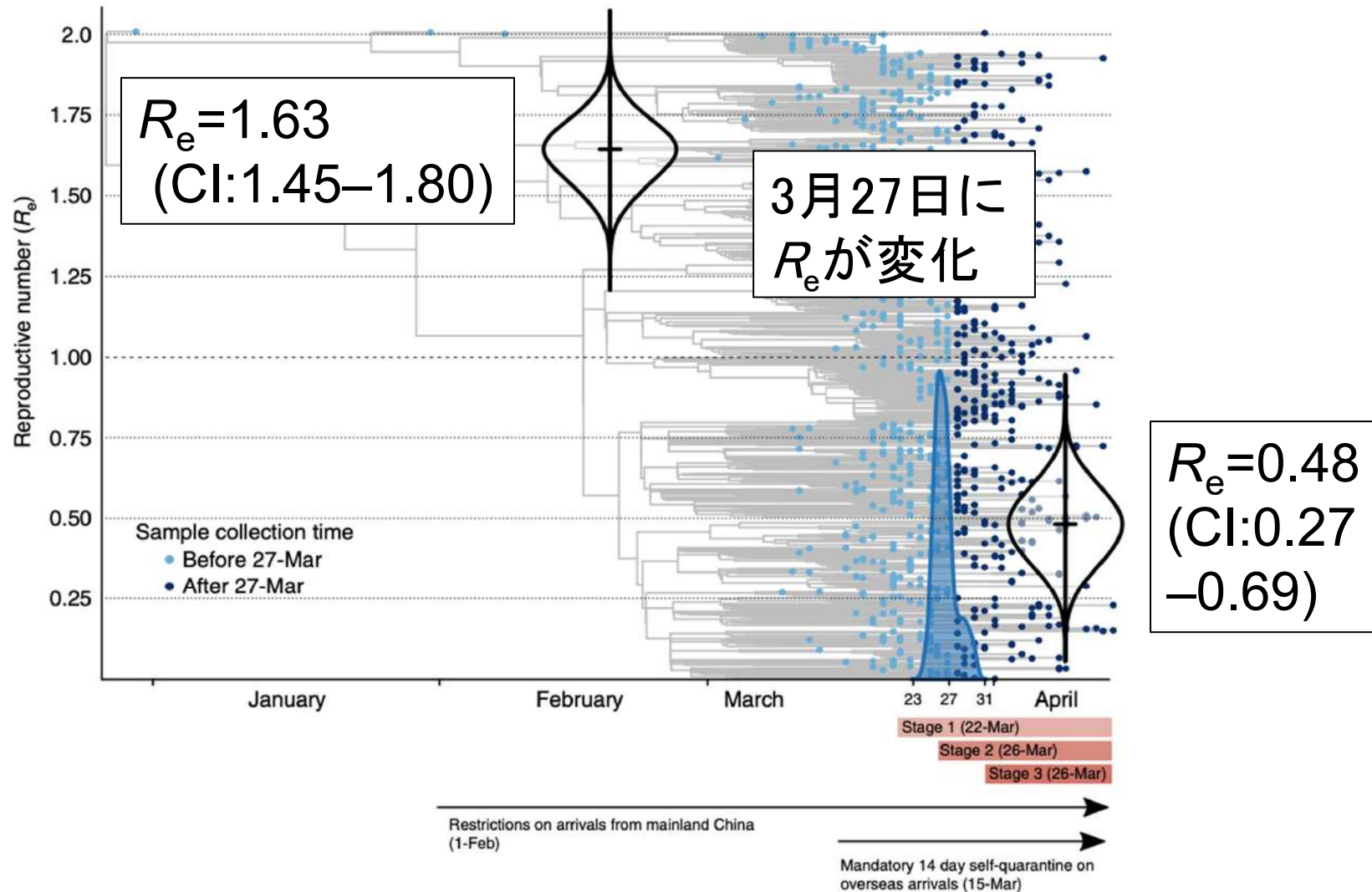
実効再生産数が2の場合



塩基がランダムに変異する時, 実効再生産数が, 進化系統樹の形と枝の長さの分布に影響する

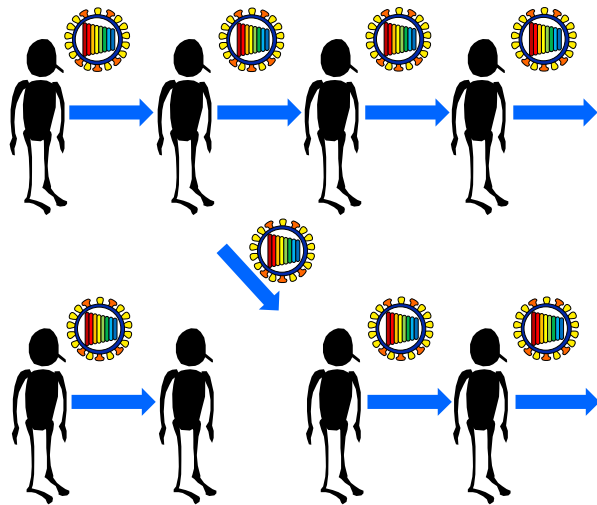
Seemann, et al. *Nature Communications* (2020)

COVID-19のPhylodynamics研究例

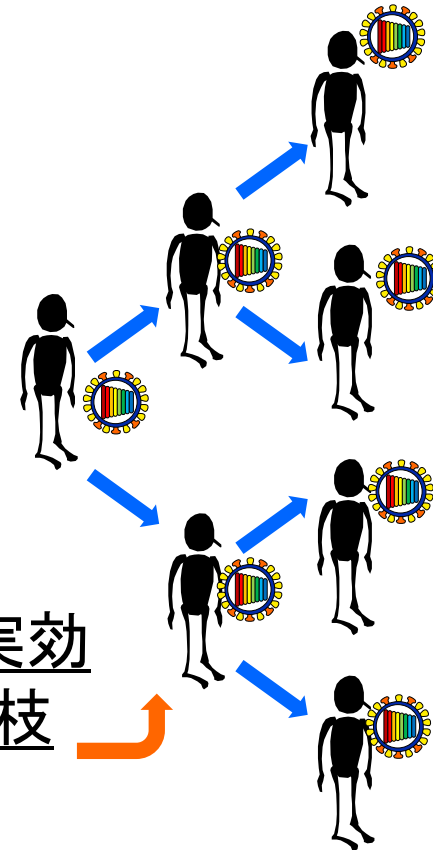


Phylodynamicsの基本的アイデア

実効再生産数が1の場合

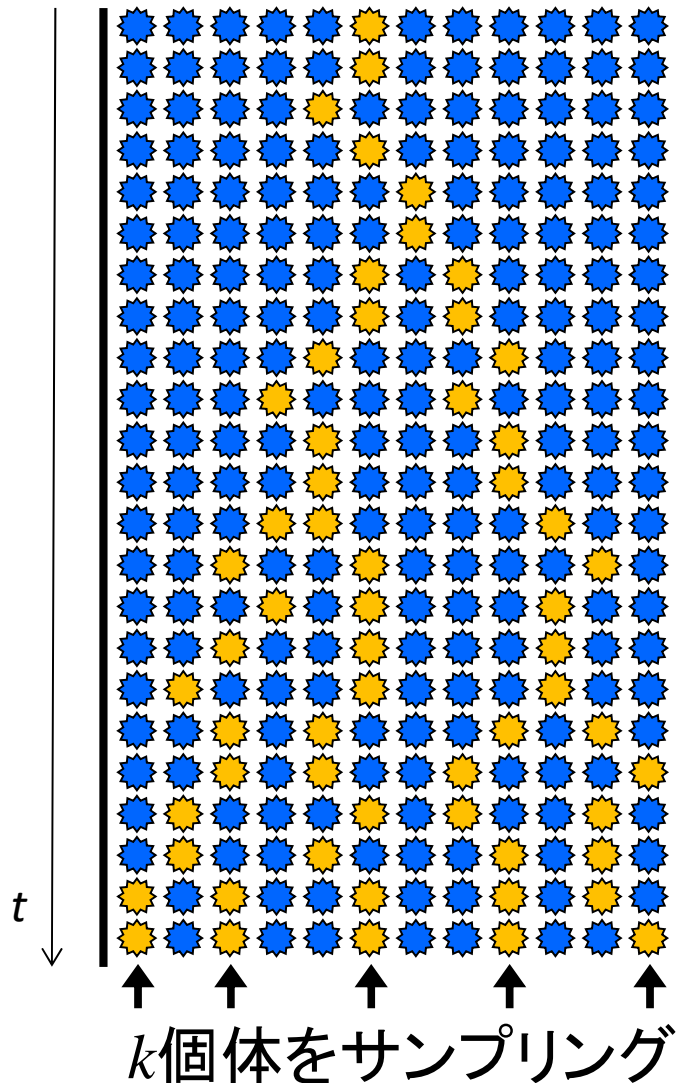


実効再生産数が2の場合



塩基がランダムに変異する時, 実効再生産数が, 進化系統樹の形と枝の長さの分布に影響する

集団遺伝学におけるCoalescent理論



Wright-Fisher モデル:

- 個体数 N は一定で, 集団は一つ.
- 集団は中立進化 (個体の子孫を残す確率が等しい)

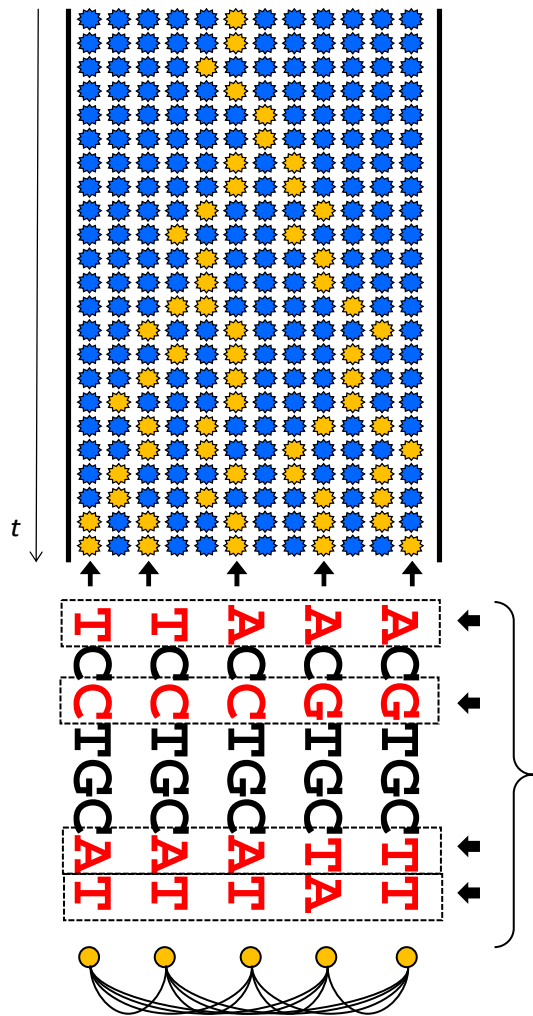
n 世代前に, k 個体のどの二つも共通祖先を持たない確率は,

$$\left(1 - \frac{k(k-1)}{2N}\right)^n \approx \exp\left\{-\frac{k(k-1)}{2N}n\right\}$$

Phylodynamicsでは, 共通祖先の分岐時間から逆に集団サイズ N を推定

Pybus, et al. **Genetics** (2000)

集団遺伝学の指標: Tajima's D



π_{ij} : 塩基置換数

$$D = \frac{\hat{\theta}_T - \hat{\theta}_W}{Std(\hat{\theta}_T - \hat{\theta}_W)},$$

$$\hat{\theta}_T = \frac{2}{n(n-1)} \sum_{i < j} \pi_{ij}$$

塩基置換数(π_{ij})の平均から推定される変異率

S_n
: 変異を持つ位置の数

$$\hat{\theta}_W = \frac{S_n}{\sum_{k=2}^n \frac{1}{(k-1)}}$$

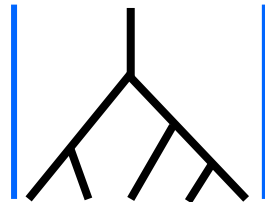
変異を持つ位置の数(S_n)から推定される変異率

Tajima's D の特徴

Demography and selection

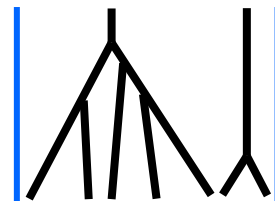
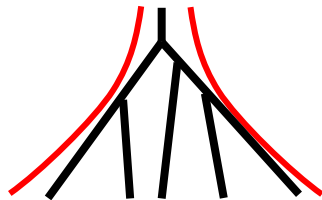
Tajima's D

集団サイズ一定 & 中立進化



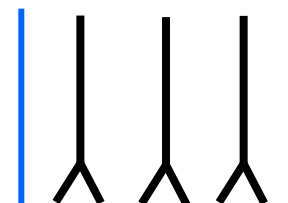
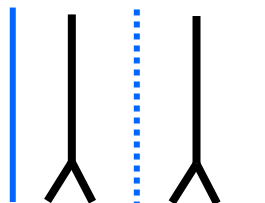
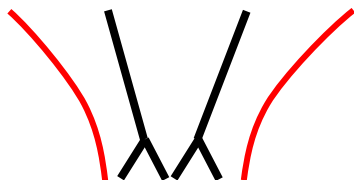
$$D = 0$$

集団サイズが増加 or 特定の株が選択 (Selective sweep)



$$D < 0$$

集団サイズの減少 or 集団の分離 or 少数株の選択 (Balancing selection)



$$D > 0$$

Tajima's Dを用いた 基本再生産数の推定

ウイルスの塩基配列からTajima' Dを計算し, 感染症の基本再生産数(R_0)を推定する。

Input:

- ウイルスの塩基配列とそれらから計算される Tajima's D の時系列

Output:

- 基本再生産数 (R_0)

Kim, K., Omori, R., & Ito, K. (2017). Inferring epidemiological dynamics of infectious diseases using Tajima's D statistic on nucleotide sequences of pathogens. *Epidemics*.

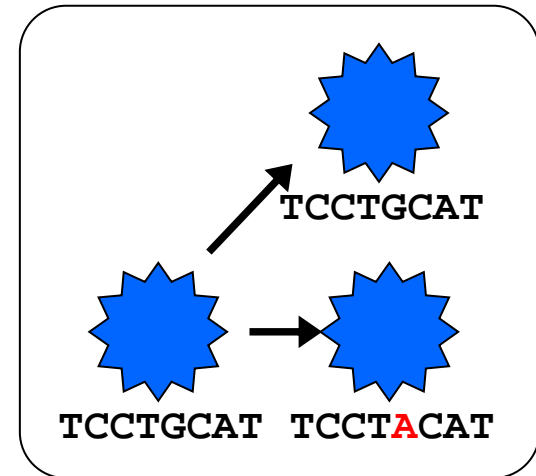
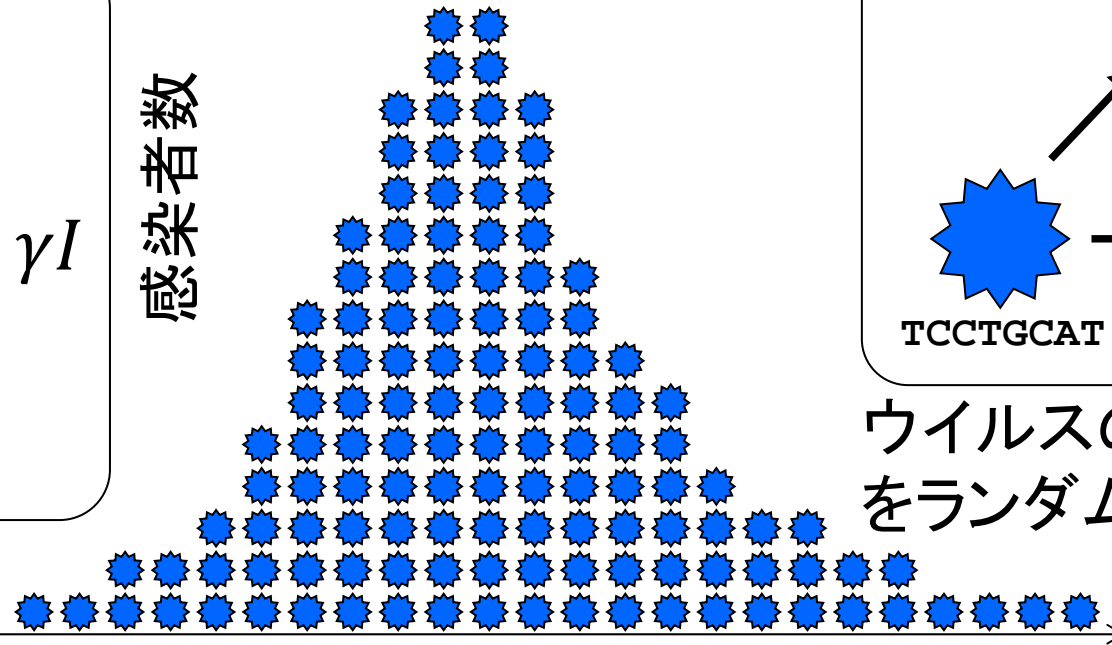
Tajima's D を用いた 感染症流行モデルのパラメータ推定

$$\begin{aligned}\frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

SIRモデル

S: 感受性,
I: 感染性,
R: 回復個体

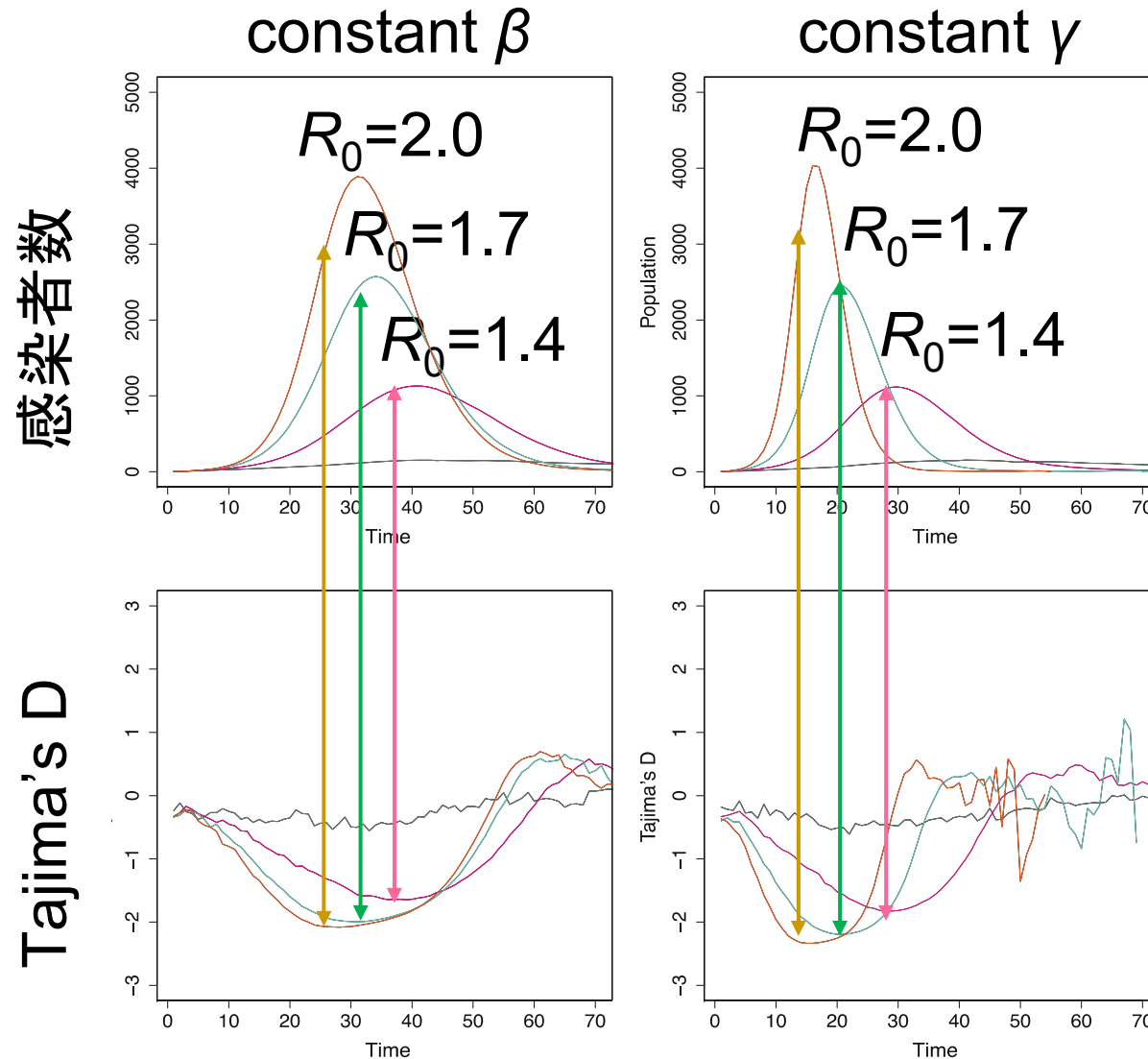
流行のシミュレーション



ウイルスの塩基配列
をランダムに変異



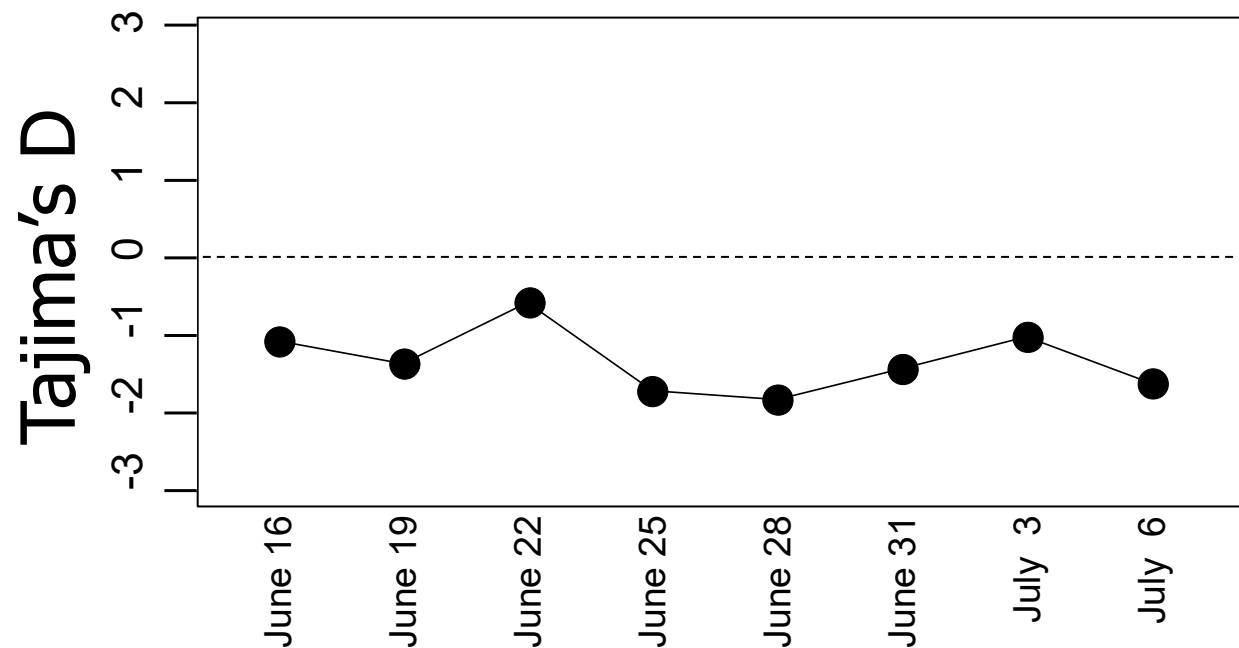
Tajima's D と基本再生産数の関係



感染者数が増加する時、Tajima's D は減少する。

2009年のパンデミック H1N1の Tajima's D

- ブエノスアイレスで2009年8月に集められたNA遺伝子の塩基配列 (Barrero *et al.* 2011).
 - 塩基配列数 = 212
 - 塩基配列の長さ = 357

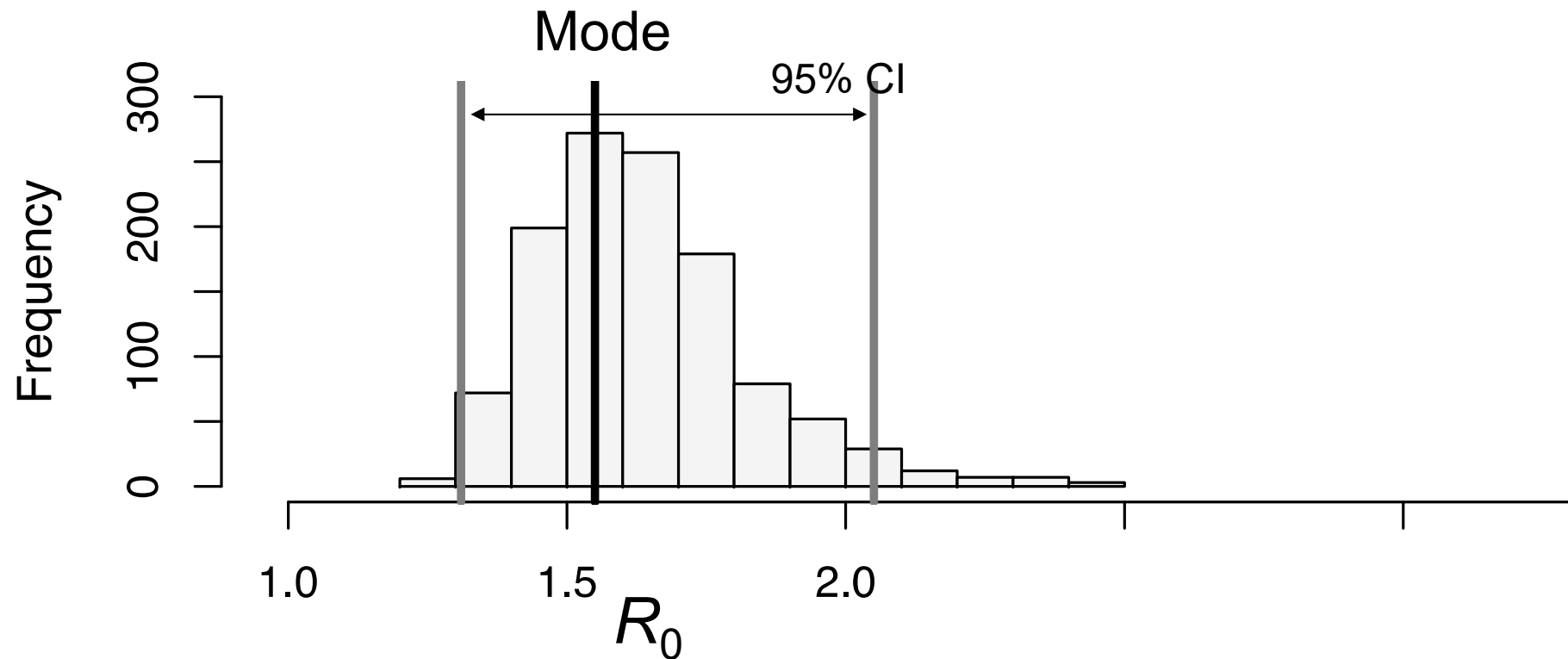


パラメータ (R_0) の推定

Approximate Bayesian Computation (ABC)

1. **Withdraw parameters from prior distributions** of basic reproduction number, recovery rate, and mutation rate
2. **Conduct the compartment simulation 100,000 times** using the parameters and draw sequences from simulation and calculate Tajima's D periodically.
3. **Compare Tajima's D values** of simulated sequences, after matching the minimum D value position.
4. **Calculate distance δ from the observed Tajima's D**
5. **Accept simulations that have small δ .**
6. **Estimating posterior distribution** of parameters from accepted simulations were used for.

実効再生産数 R_0 の推定値



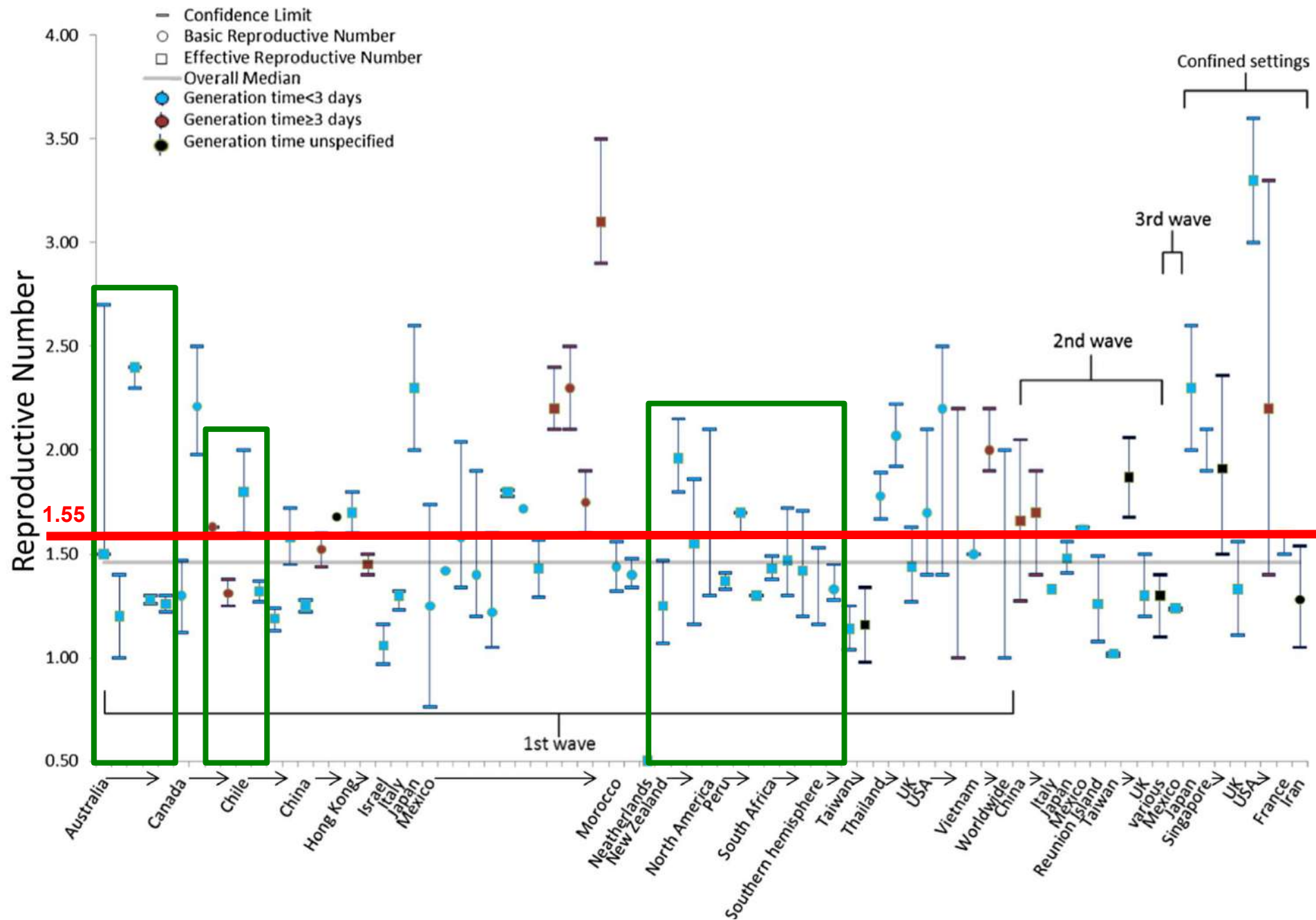
Mode

1.55

95% CI

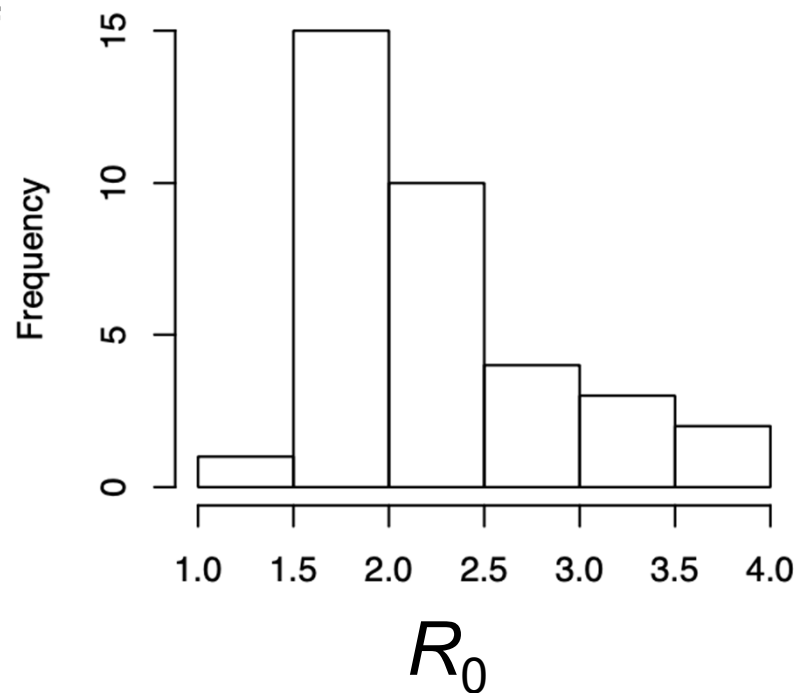
[1.31, 2.05]

疫学データからの推定値 Biggerstaff1 et al. (2014)との比較



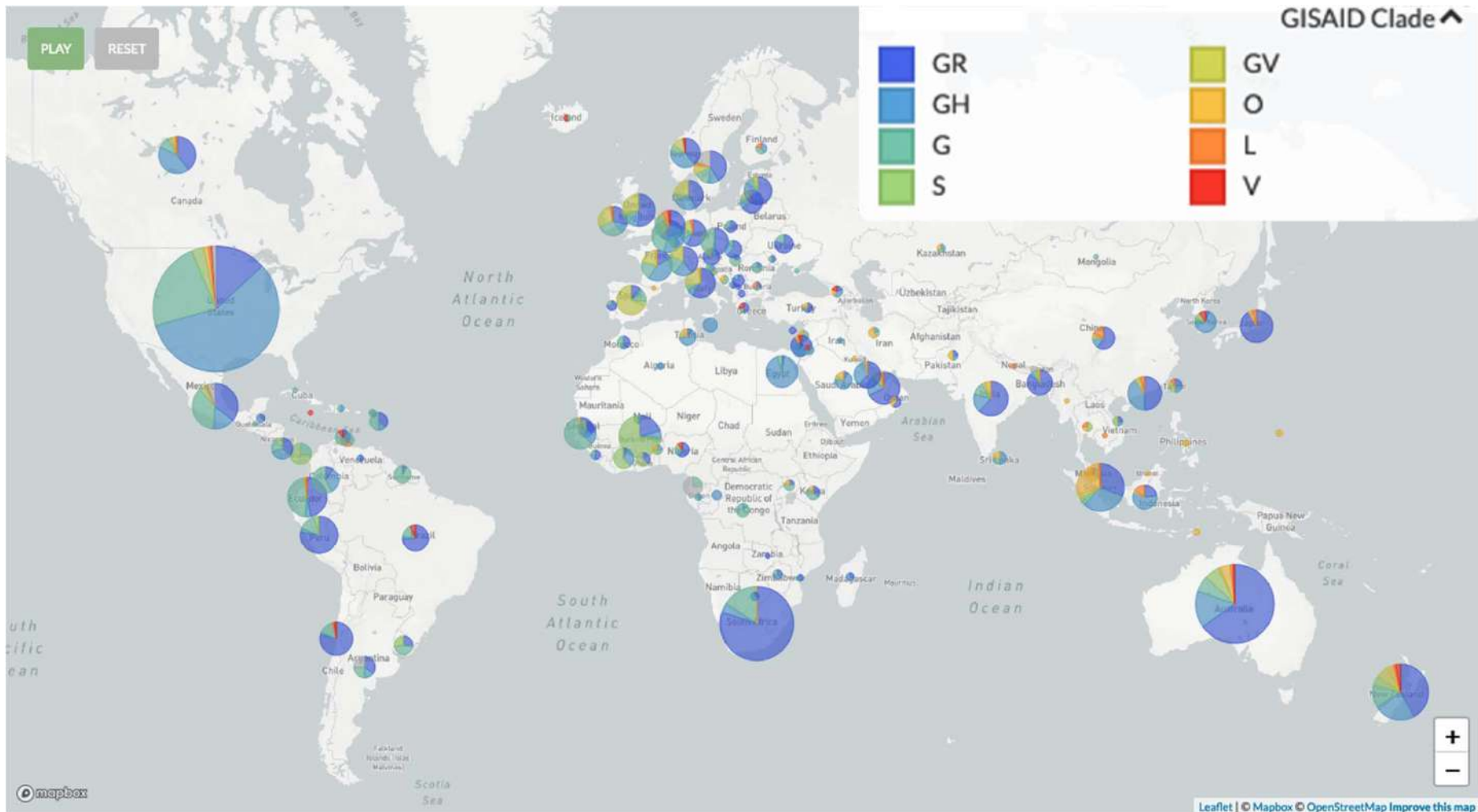
COVID-19の塩基配列への応用

- 2020年2月10日塩基配列を取得
- 湖北省のウイルスに限定
- 塩基配列数 = 81
- 塩基配列の長さ= 2979
- 流行開始日 = 2019-12-01
- GISAIDと同じ日にシミュレーションから塩基配列を取得
- 週毎にTajimaのDを計算
- ABCで R_0 を推定
- 疫学データからの推定とほぼ一致
- 流行初期段階では使える。
 - 集団の分離がない為



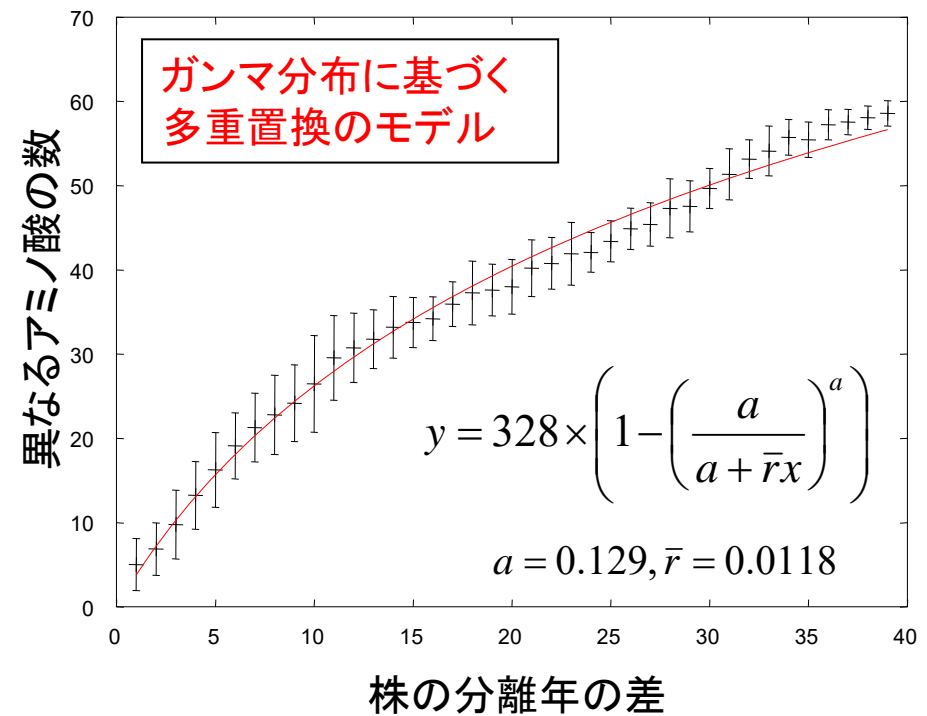
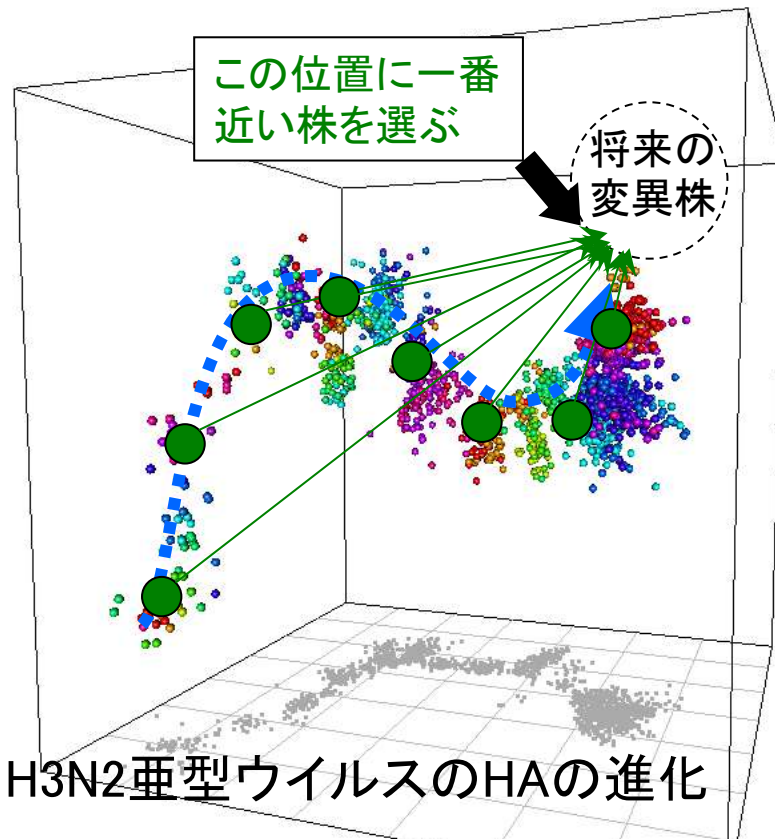
Median	95% CI
2.01	[1.53, 3.63]

各系統の流行状況



SARS-CoV-2系統のリアルタイム可視化 (<http://nextstrain.org>)

H3N2インフルエンザウイルスの HA上のアミノ酸置換の予測



ウイルス株の分離年の差とアミノ酸置換数の関係から、翌年起こるアミノ酸置換をSensitivity=70%, Precision=45%予測できる [Ito K, *PLoS one*, 2010]。

まとめ

- 感染症のリアルタイムビッグデータ
 - 感染者数の疫学データ(感染症流行モデリング)
 - 病原体の塩基配列データ(進化系統解析, 集団遺伝学解析)
- Phylodynamics
 - 進化系統解析による感染症流行動態の推定
- 集団遺伝学 と 感染症流行モデリング の融合
 - Tajima's D による実効再生産数の推定
 - ウイルスの変異予測
- 統計的機械学習による流行予測

共同研究・共同指導の募集

- <http://www.czc.hokudai.ac.jp/bioinform>

The screenshot shows the 'Members' page of the czc-bioinformatics website. The page is titled 'Members' and is part of the 'Division of Bioinformatics'. It lists 'Current Members' and 'Alumni and Past Members'. The 'Current Members' section includes profiles for Kimihito Ito (Professor, Ph.D. in Engineering), Ryosuke Omori (Associate Professor, Ph.D. in Biology), Kiyoon Kim (Post-Doctoral Fellow, Ph.D. in Veterinary Medicine; DVM), Chayada Piantham (Ph.D Student, DVM), Sayaka Iida (Technical Assistant, Master in Fisheries Science), and Wessam Mohamed (Visiting Scholar, Ph.D in Veterinary Medicine). The 'Alumni and Past Members' section lists Gabriel Gonzalez (University College Dublin, Ireland), Jednipit Borthong (Rajamangala University of Technology Thanyaburi, Thailand), Heidi Gurung (University of Wisconsin-Eau Claire, USA), Nipawit Karnbunchob, Mayumbo Nyirenda (The University of Zambia, Zambia), Wei Shi (Kyushu University, Japan), Keisuke Ueno, Kouki Yonezawa (Meijo University, Japan), Makoto Ohigashi, and Manabu Igarashi (Research Center for Zoonosis Control, Hokkaido University). The page also includes contact information for the Division of Bioinformatics, such as the address in Sapporo, Hokkaido, and the contact email itok@czc.hokudai.ac.jp.

- 感染症バイオインフォマティクスの共同研究のご提案を随時募集しています。
- 卒論や修論の共同指導のご提案も歓迎します。
- 研究打ち合わせはオンラインで行います。
- 詳しくはホームページをご覧ください。
- 連絡先:
itok@czc.hokudai.ac.jp