

Stochastic dynamic modeling of evolving gene expression programs

Lucas Goiriz

Doctoral Thesis Defense

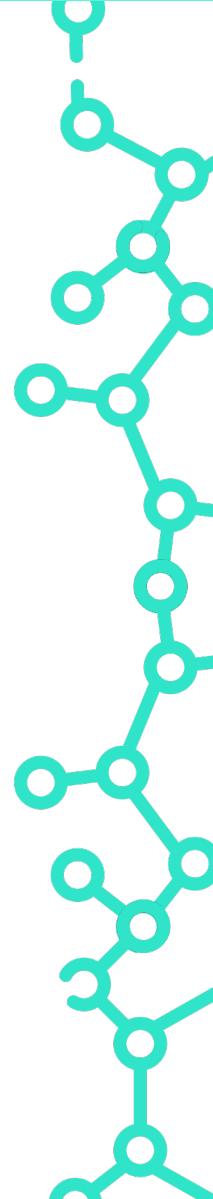
Thesis Advisors

Guillermo Rodrigo Tárrega (CSIC – UV)

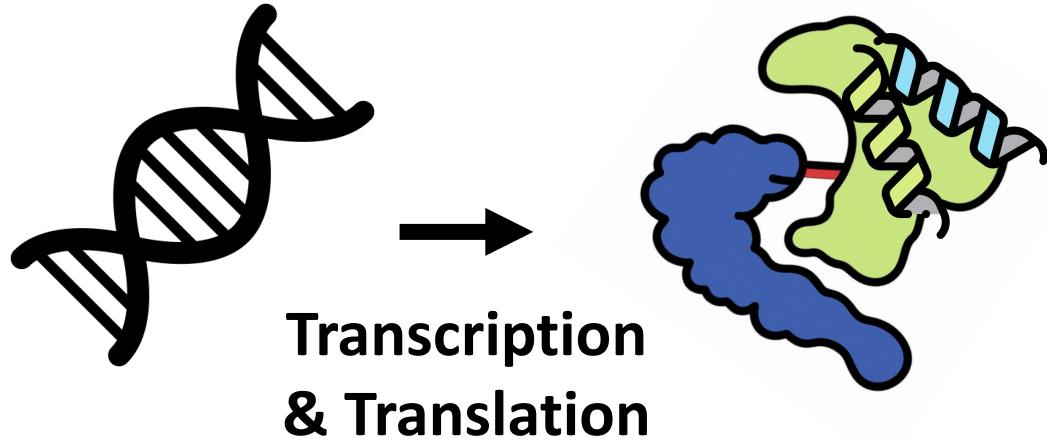
J. Alberto Conejero Casares (IUMPA – UPV)

Index

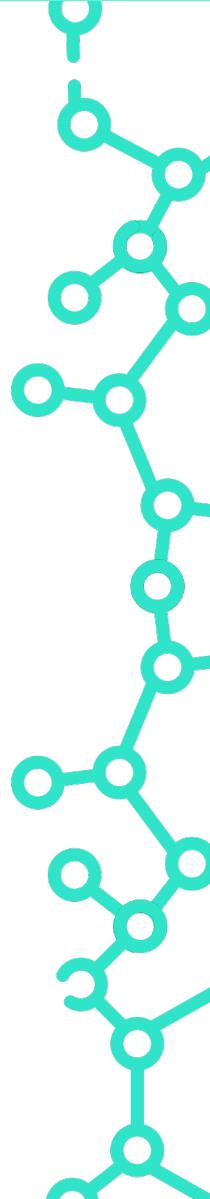
- ❖ **Introduction** (Chapter 1)
- ❖ **Objectives** (Chapter 2)
- ❖ **Chapter 3:** A variant-dependent molecular clock with anomalous diffusion models SARS-CoV-2 evolution in humans
- ❖ **Chapter 4:** PyEvoMotion: a software to perform the temporal statistical analysis of genome evolution
- ❖ **Chapter 5:** Deciphering microscopic drivers of viral genome-scale molecular clock dynamics
- ❖ **General Discussion**
- ❖ **Conclusions**



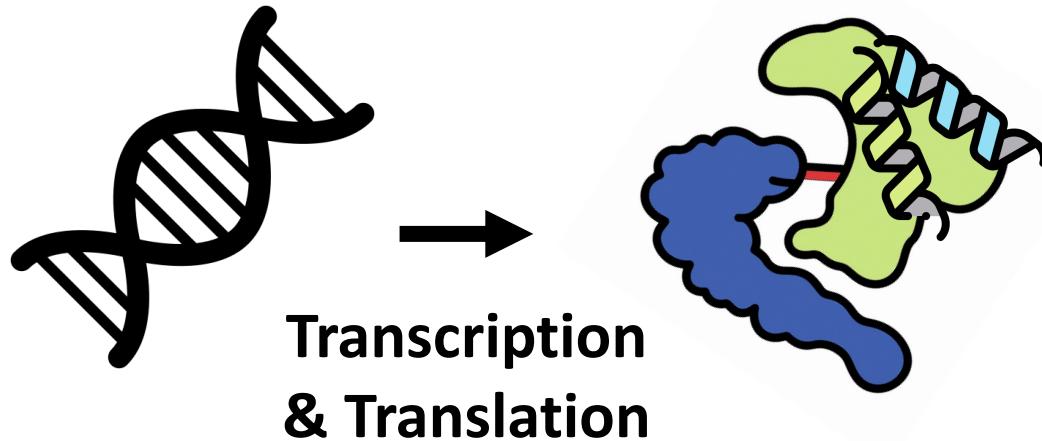
Introduction: Genotype-to-Phenotype - The Program Behind Life



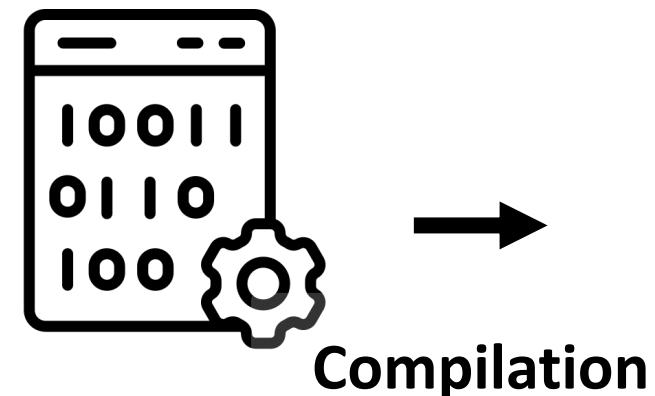
Genes are **not passive**, they “execute code in response to biological inputs”



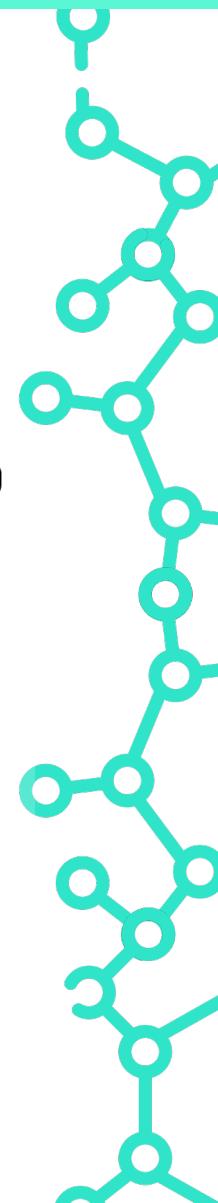
Introduction: Genotype-to-Phenotype - The Program Behind Life



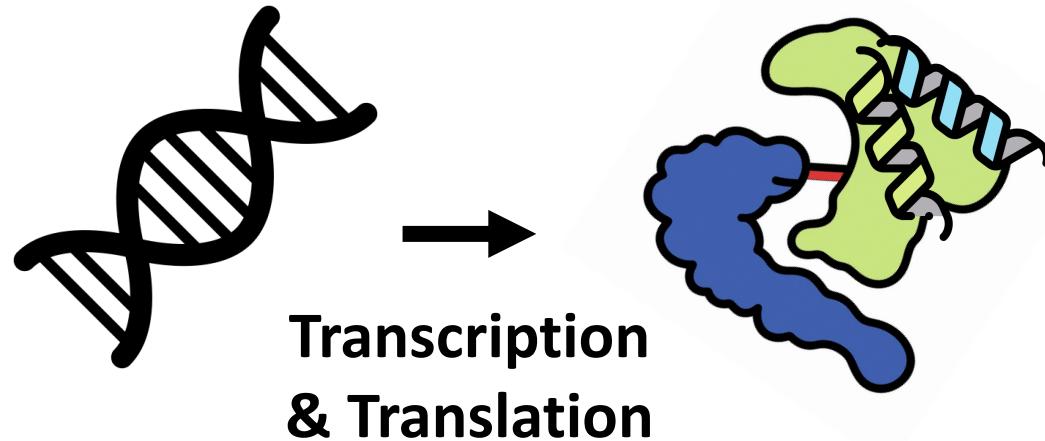
Genes are **not passive**, they “execute code in response to biological inputs”



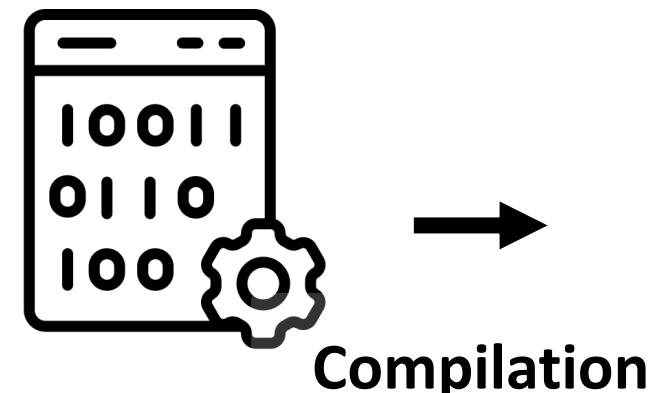
Cells interpret DNA much like **compilers** transform source code into **functional programs**



Introduction: Genotype-to-Phenotype - The Program Behind Life

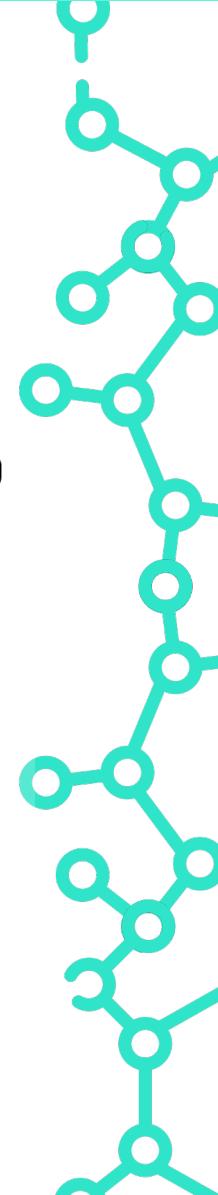


Genes are **not passive**, they “execute code in response to biological inputs”

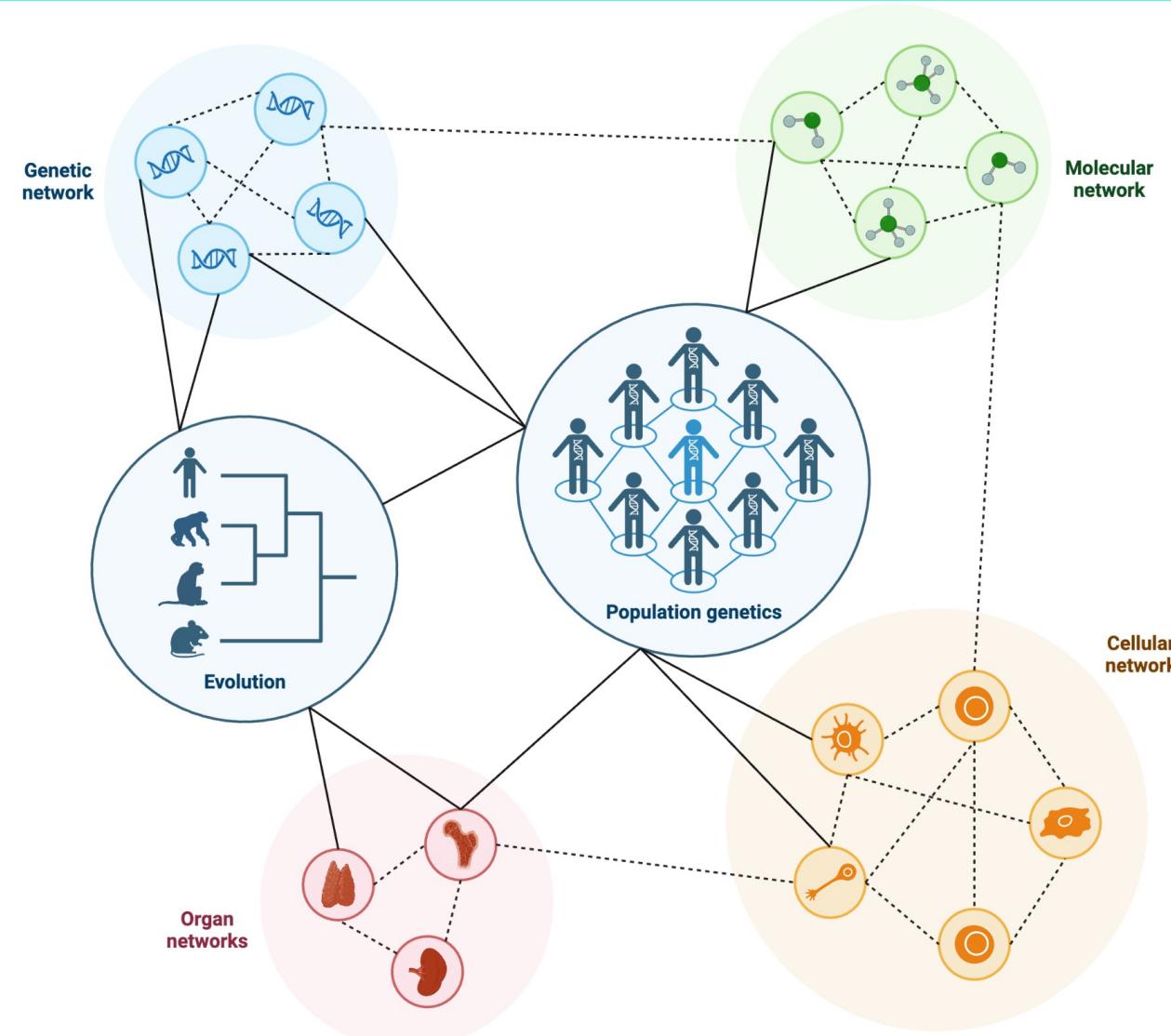


Cells interpret DNA much like **compilers** transform source code into **functional programs**

Phenotypes arise from this dynamic gene expression process

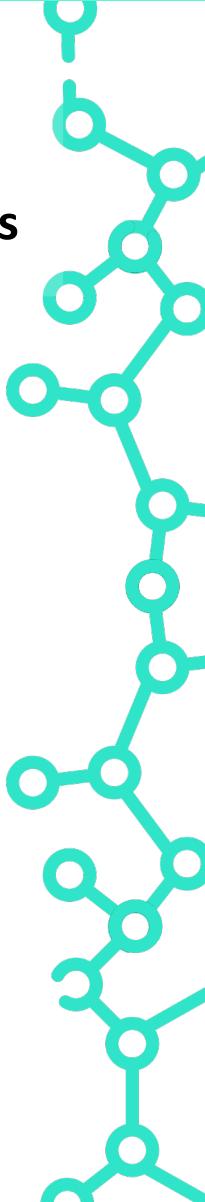


Introduction: Gene Expression as a Dynamic, Modelable System

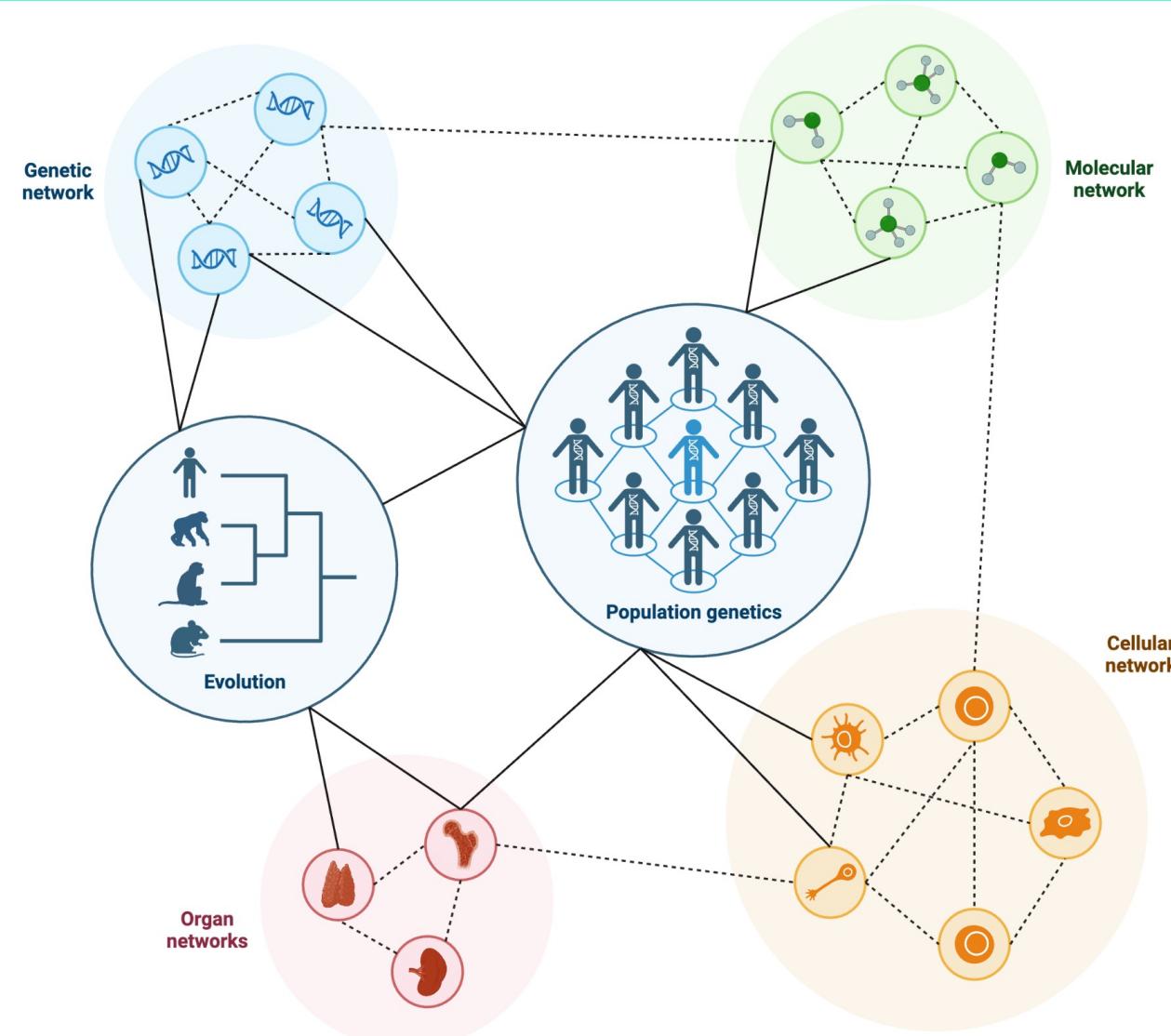


(Adapted from Villa & Sonis (2020))

Gene expression is **dynamic**: it involves RNA synthesis, translation, and molecular **interactions**



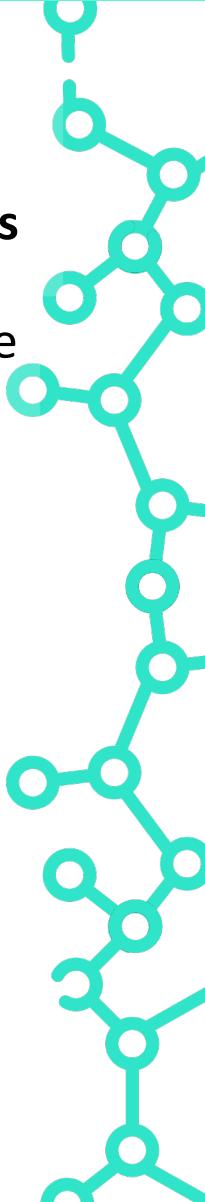
Introduction: Gene Expression as a Dynamic, Modelable System



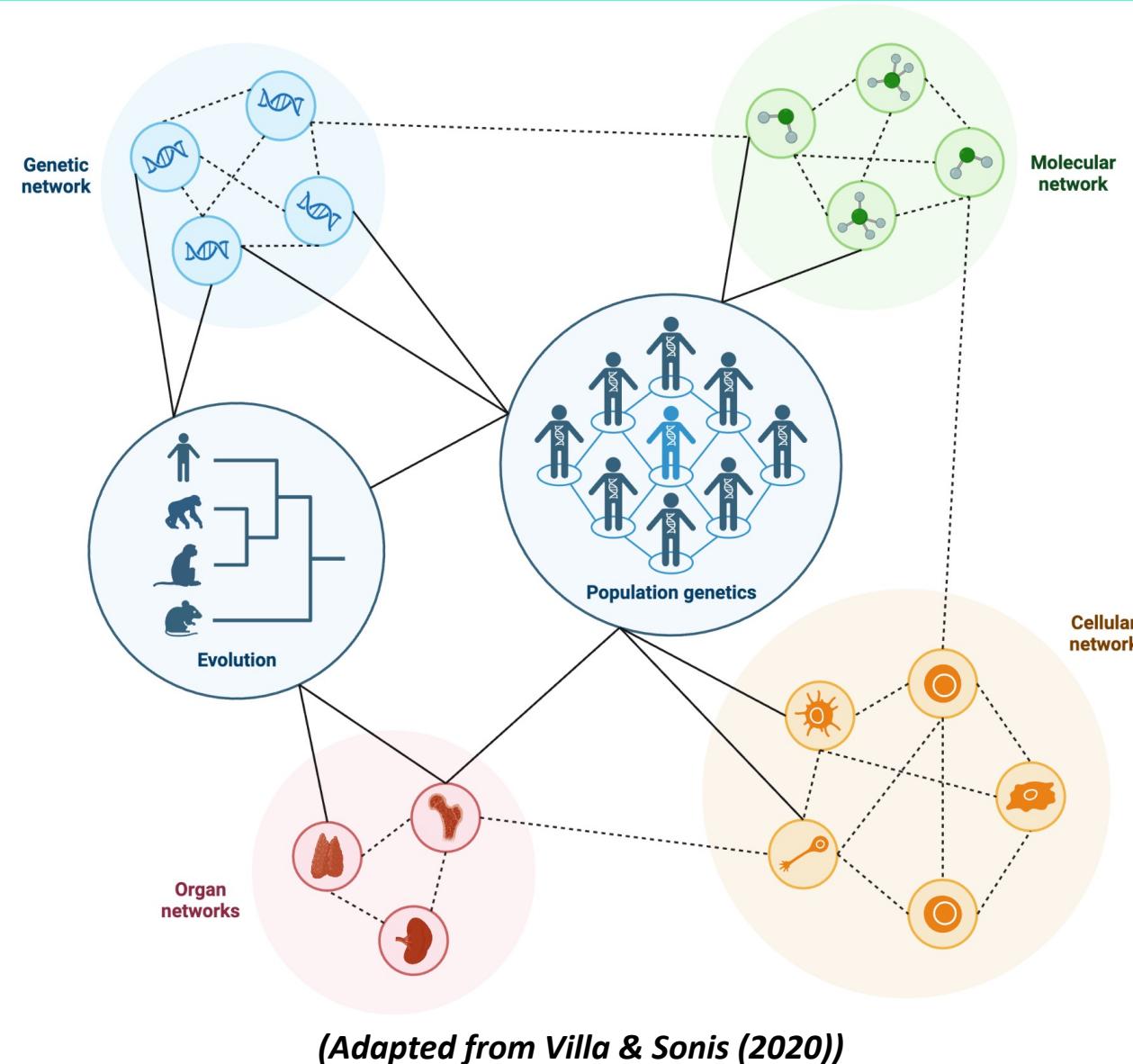
(Adapted from Villa & Sonis (2020))

Gene expression is **dynamic**: it involves RNA synthesis, translation, and molecular **interactions**

Regulatory networks process signals and produce outputs



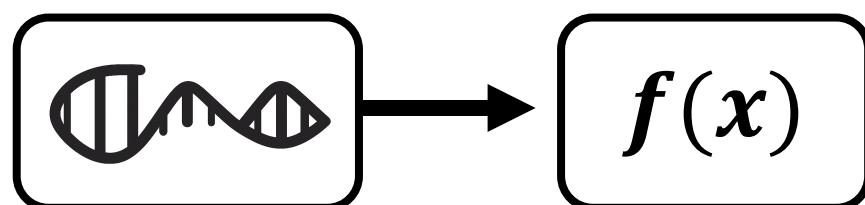
Introduction: Gene Expression as a Dynamic, Modelable System



Gene expression is **dynamic**: it involves RNA synthesis, translation, and molecular **interactions**

Regulatory networks process signals and produce outputs

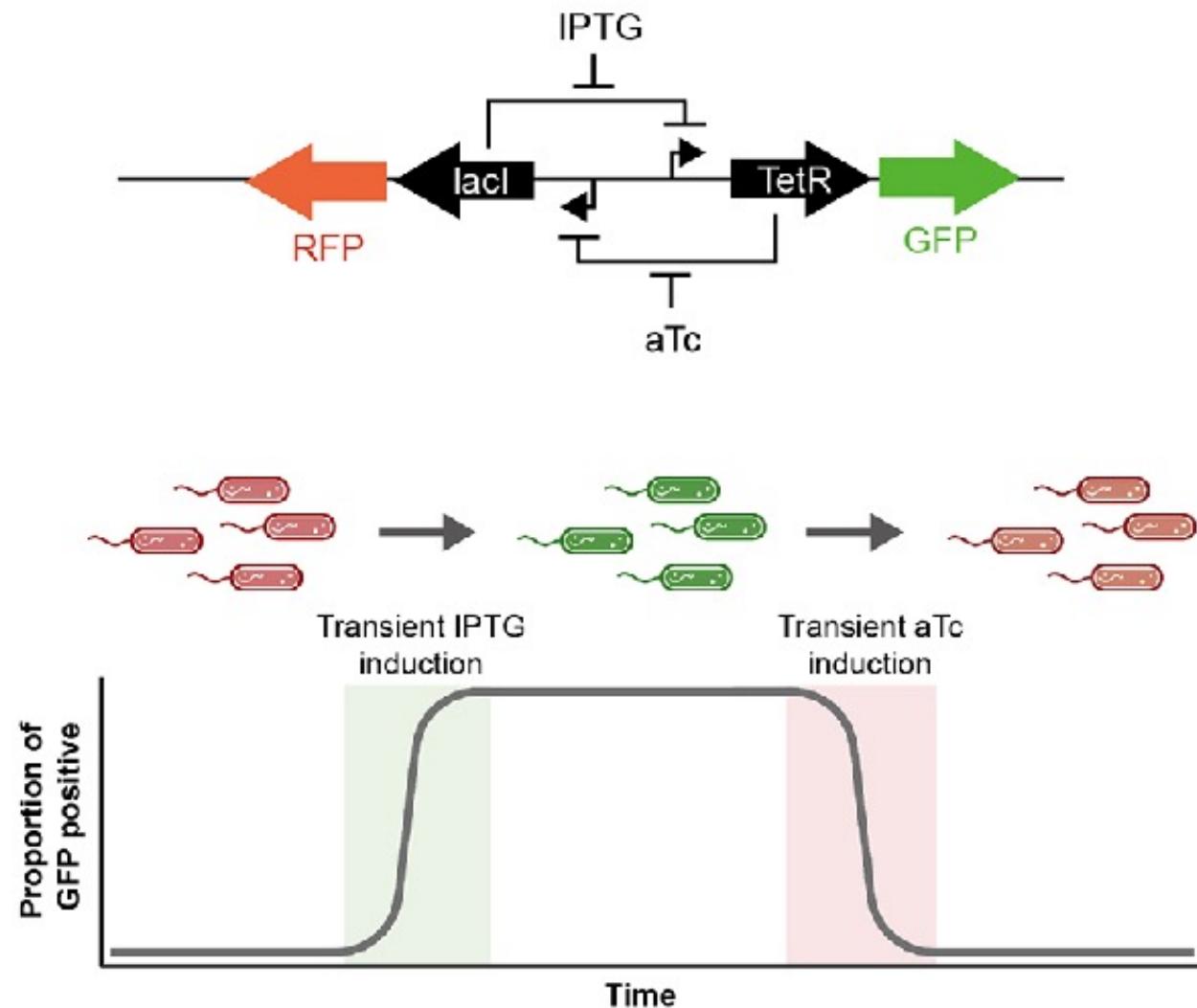
Mathematical models clarify how genes behave, *and misbehave*



$$\frac{dx}{dt} = f(x)$$

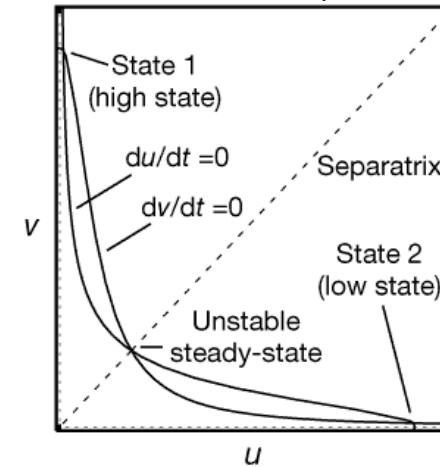


Introduction: Gene Expression as a Dynamic, Modelable System

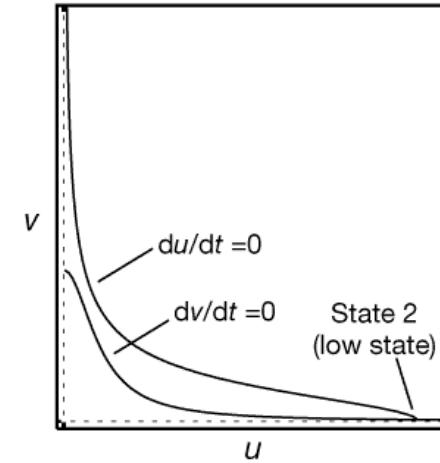


(Adapted from Aggarwal et al. (2022))

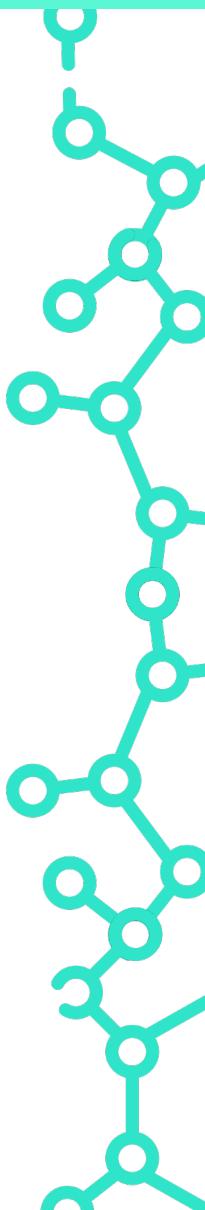
Parameterization 1 (bistability)



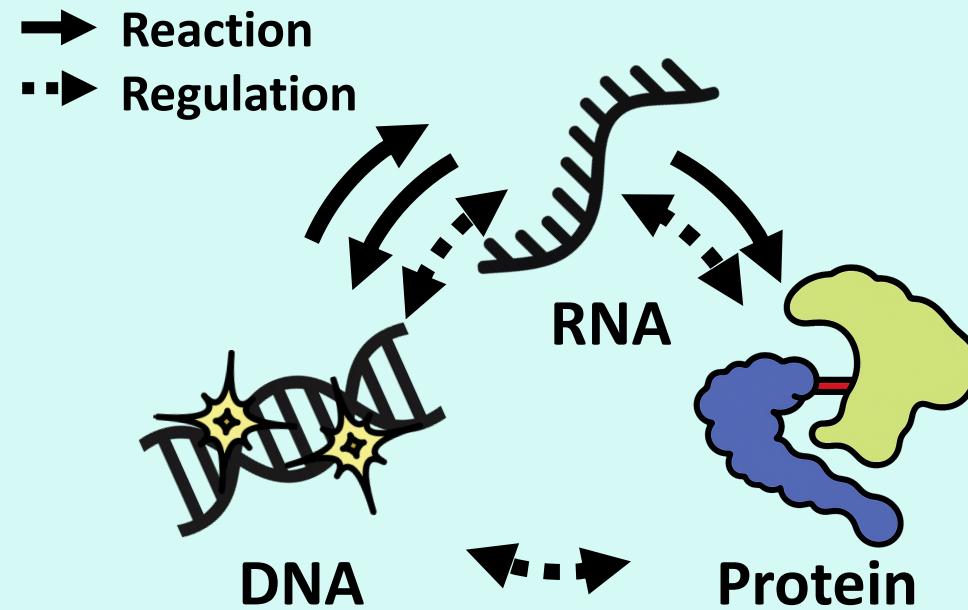
Parameterization 2 (monostability)



(Adapted from Gardner et al. (2000))



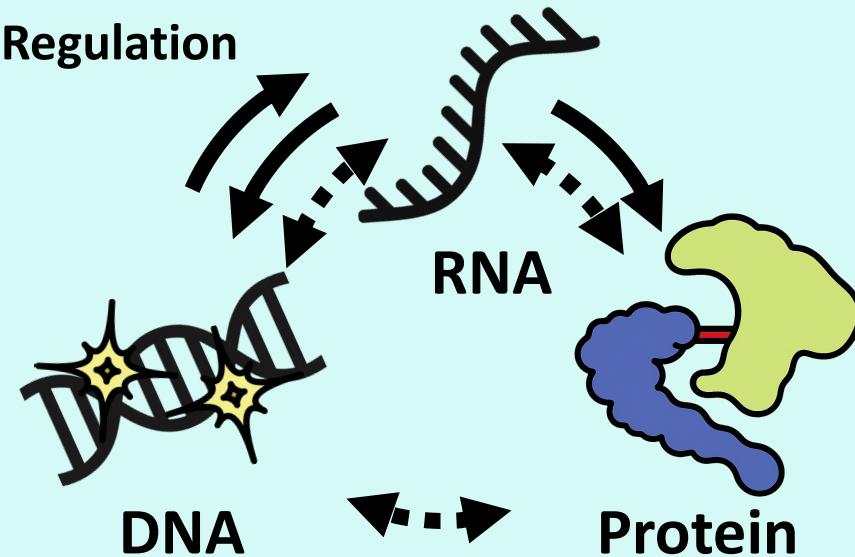
Introduction: Gene Expression as a Dynamic, Modelable System



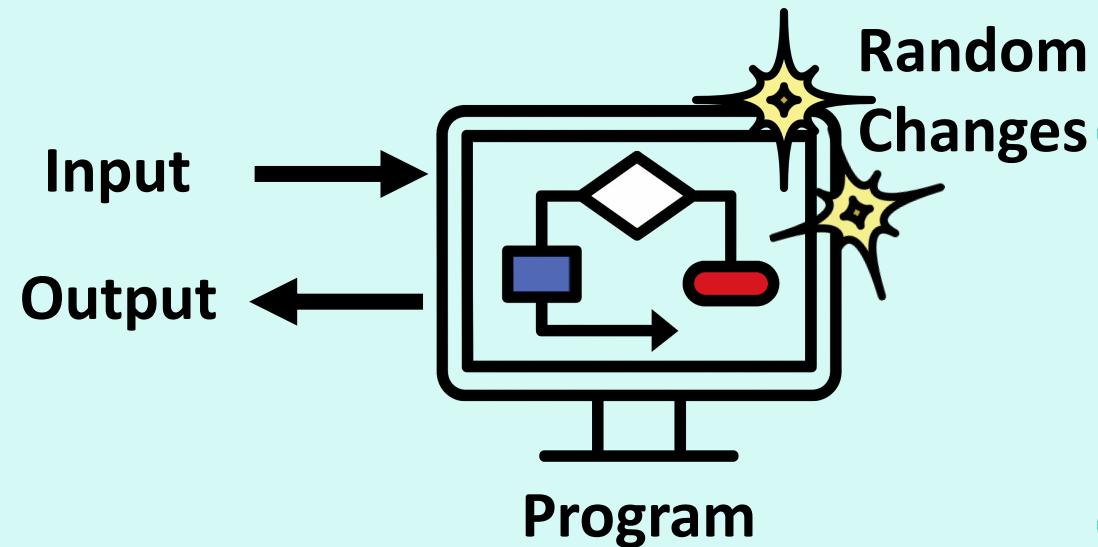
Gene expression *programs* are **not** static: they evolve

Introduction: Gene Expression as a Dynamic, Modelable System

→ Reaction
↔ Regulation



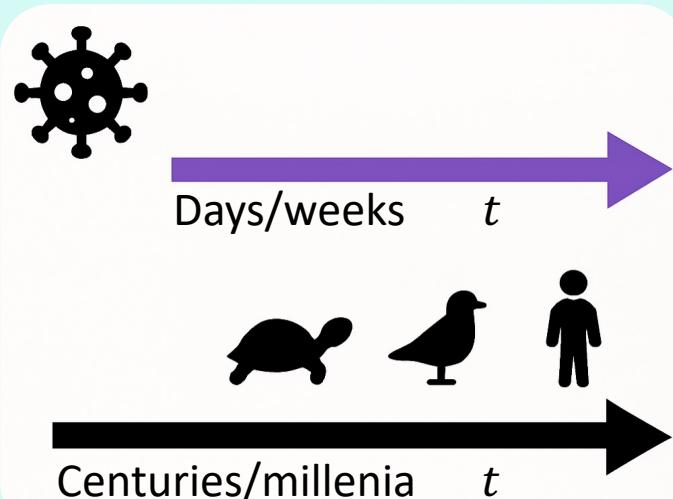
Gene expression *programs* are **not** static: they evolve



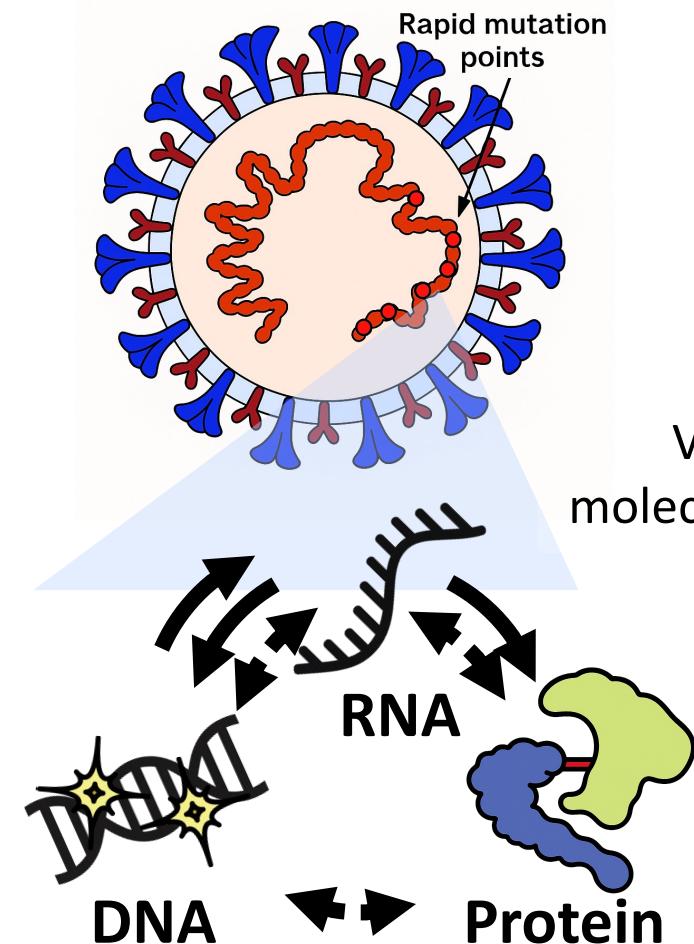
Over generations, random edits shape entirely new regulatory architectures

Introduction: Viruses as Models for Studying Evolution

Compressed Timescales in Viruses



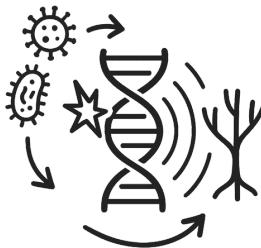
- Short generation times
- Large population sizes
- High evolution rates



(Adapted from Zhang et al. (2021))



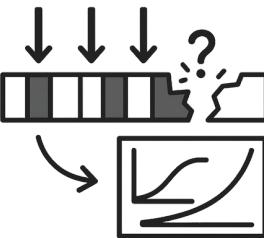
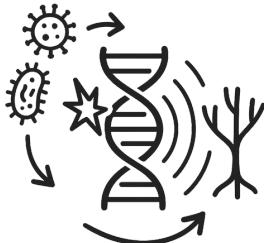
Objectives



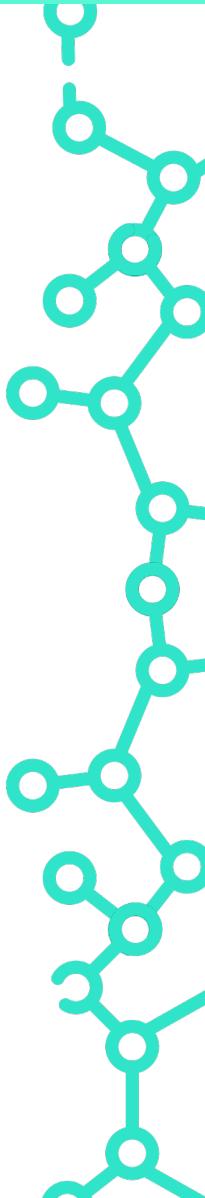
1. Investigate the underlying stochastic processes and localized mutation dynamics in rapidly mutating biological systems, assessing how these micro-level events collectively influence macroscale evolutionary outcomes.



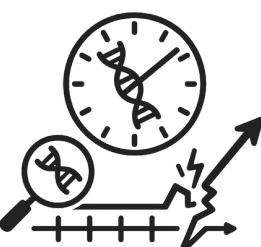
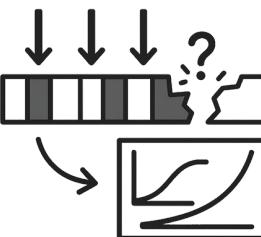
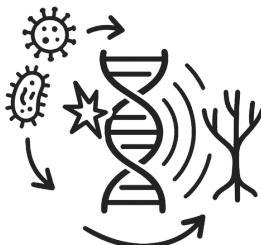
Objectives



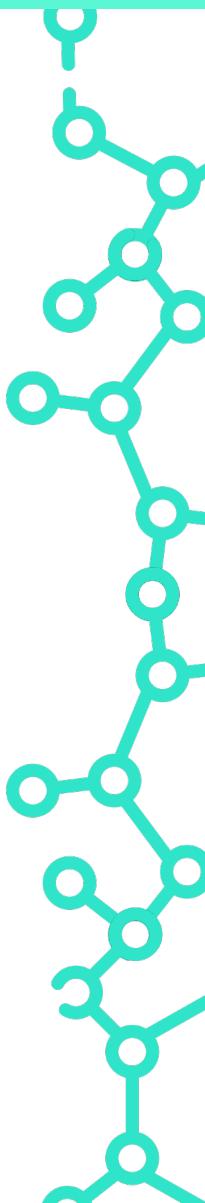
1. Investigate the underlying stochastic processes and localized mutation dynamics in rapidly mutating biological systems, assessing how these micro-level events collectively influence macroscale evolutionary outcomes.
2. Evaluate selective pressures across diverse genomic segments and identify deviations from classical evolutionary models by employing alternative quantitative frameworks to elucidate patterns of adaptation and constraint.



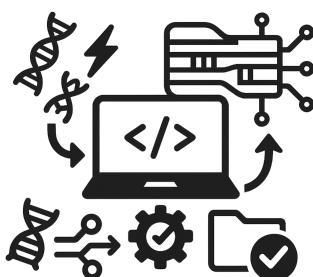
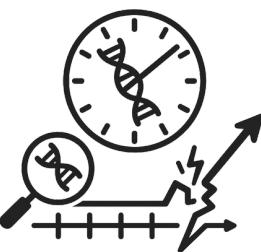
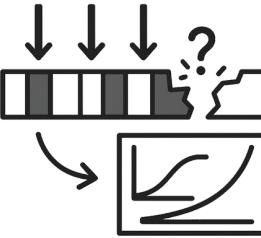
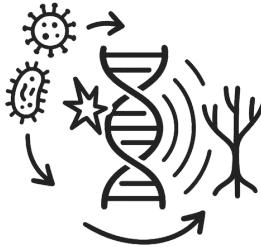
Objectives



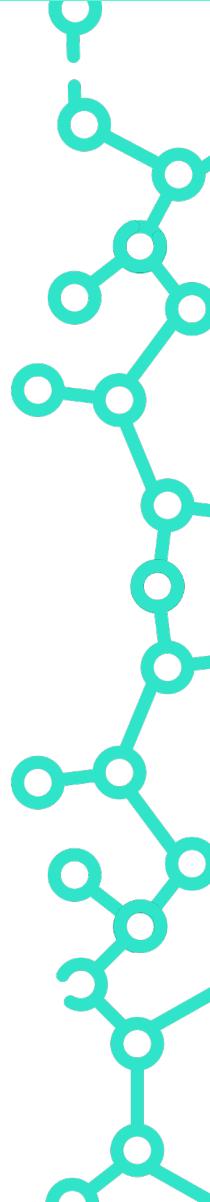
1. Investigate the underlying stochastic processes and localized mutation dynamics in rapidly mutating biological systems, assessing how these micro-level events collectively influence macroscale evolutionary outcomes.
2. Evaluate selective pressures across diverse genomic segments and identify deviations from classical evolutionary models by employing alternative quantitative frameworks to elucidate patterns of adaptation and constraint.
3. Verify whether the molecular clock hypothesis holds in rapidly mutating biological systems or if significant deviations arise due to localized genetic variability.



Objectives



1. Investigate the underlying stochastic processes and localized mutation dynamics in rapidly mutating biological systems, assessing how these micro-level events collectively influence macroscale evolutionary outcomes.
2. Evaluate selective pressures across diverse genomic segments and identify deviations from classical evolutionary models by employing alternative quantitative frameworks to elucidate patterns of adaptation and constraint.
3. Verify whether the molecular clock hypothesis holds in rapidly mutating biological systems or if significant deviations arise due to localized genetic variability.
4. Develop and implement integrated computational and statistical methodologies, coupled with novel large-scale open-source data treatment approaches, to bridge localized evolutionary events with comprehensive genomic-scale dynamics while ensuring reproducibility and scalability.



Chapter 3: A variant-dependent molecular clock with anomalous diffusion models SARS-CoV-2 evolution in humans

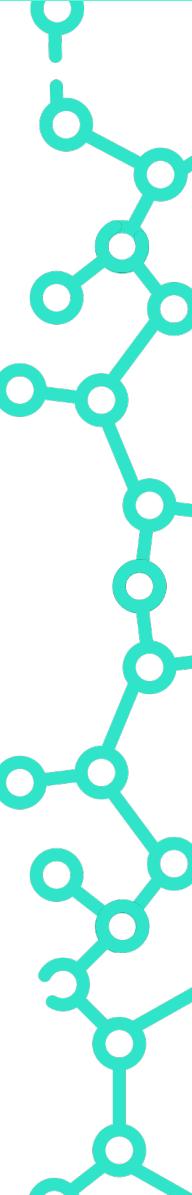
Goiriz L, Ruiz R, Garibo-i-Orts O, Conejero J.A, Rodrigo G. (2023)
A variant-dependent molecular clock with anomalous diffusion models
SARS-CoV-2 evolution in humans. *Proc. Natl. Acad. Sci. U.S.A.*, 120(30): e2303578120



The Molecular Clock Hypothesis

Definition: Molecular Clock (Zuckerkandl & Pauling, 1965)

The rate of evolution of a gene is constant over time



The Molecular Clock Hypothesis

Definition: Molecular Clock (Zuckerkandl & Pauling, 1965)

The rate of evolution of a gene is constant over time

Definition: Neutral Theory of Molecular Evolution (Kimura, 1968)

The number of substitutions over time $N(t)$ follows a **Poisson process** with constant rate μ , such that the probability of observing k mutations in time t is given by

$$P(N(t) = k) = \frac{(\mu t)^k e^{-\mu t}}{k!}, \quad k = 0, 1, 2, \dots$$



The Molecular Clock Hypothesis

Definition: Molecular Clock (Zuckerkandl & Pauling, 1965)

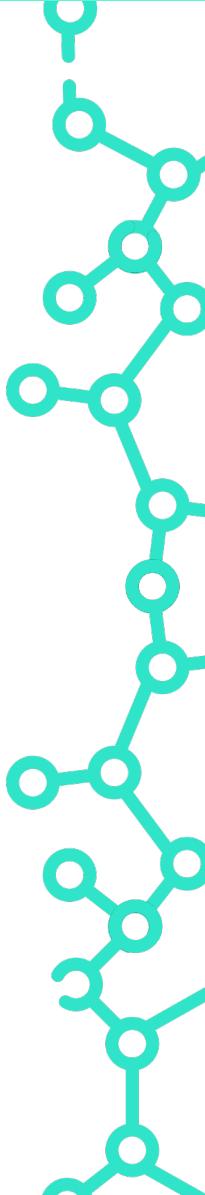
The rate of evolution of a gene is constant over time

Definition: Neutral Theory of Molecular Evolution (Kimura, 1968)

The number of substitutions over time $N(t)$ follows a **Poisson process** with constant rate μ , such that the probability of observing k mutations in time t is given by

$$P(N(t) = k) = \frac{(\mu t)^k e^{-\mu t}}{k!}, \quad k = 0, 1, 2, \dots$$

Substitutions accumulate at a constant rate: $\mathbb{E}[N(t)] = \mu t, \quad \mathbb{V}[N(t)] = \mu t$



The Molecular Clock Hypothesis

Definition: Molecular Clock (Zuckerkandl & Pauling, 1965)

The rate of evolution of a gene is constant over time

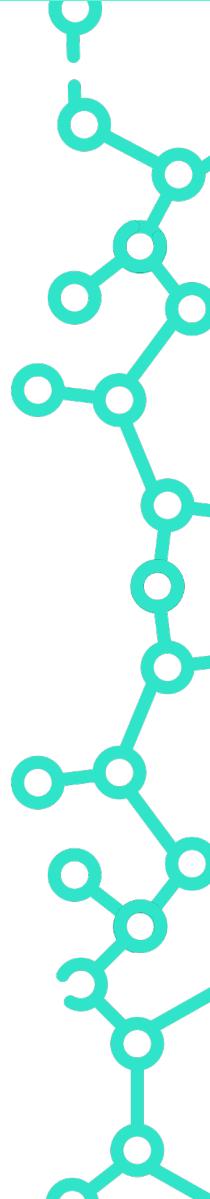
Definition: Neutral Theory of Molecular Evolution (Kimura, 1968)

The number of substitutions over time $N(t)$ follows a **Poisson process** with constant rate μ , such that the probability of observing k mutations in time t is given by

$$P(N(t) = k) = \frac{(\mu t)^k e^{-\mu t}}{k!}, \quad k = 0, 1, 2, \dots$$

Substitutions accumulate at a constant rate: $\mathbb{E}[N(t)] = \mu t, \quad \mathbb{V}[N(t)] = \mu t$

Observed overdispersion challenges this assumption



The Molecular Clock Hypothesis

Definition: Molecular Clock (Zuckerkandl & Pauling, 1965)

The rate of evolution of a gene is constant over time

Definition: Neutral Theory of Molecular Evolution (Kimura, 1968)

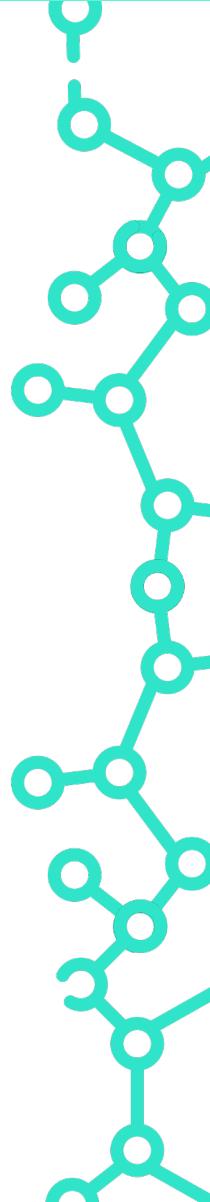
The number of substitutions over time $N(t)$ follows a **Poisson process** with constant rate μ , such that the probability of observing k mutations in time t is given by

$$P(N(t) = k) = \frac{(\mu t)^k e^{-\mu t}}{k!}, \quad k = 0, 1, 2, \dots$$

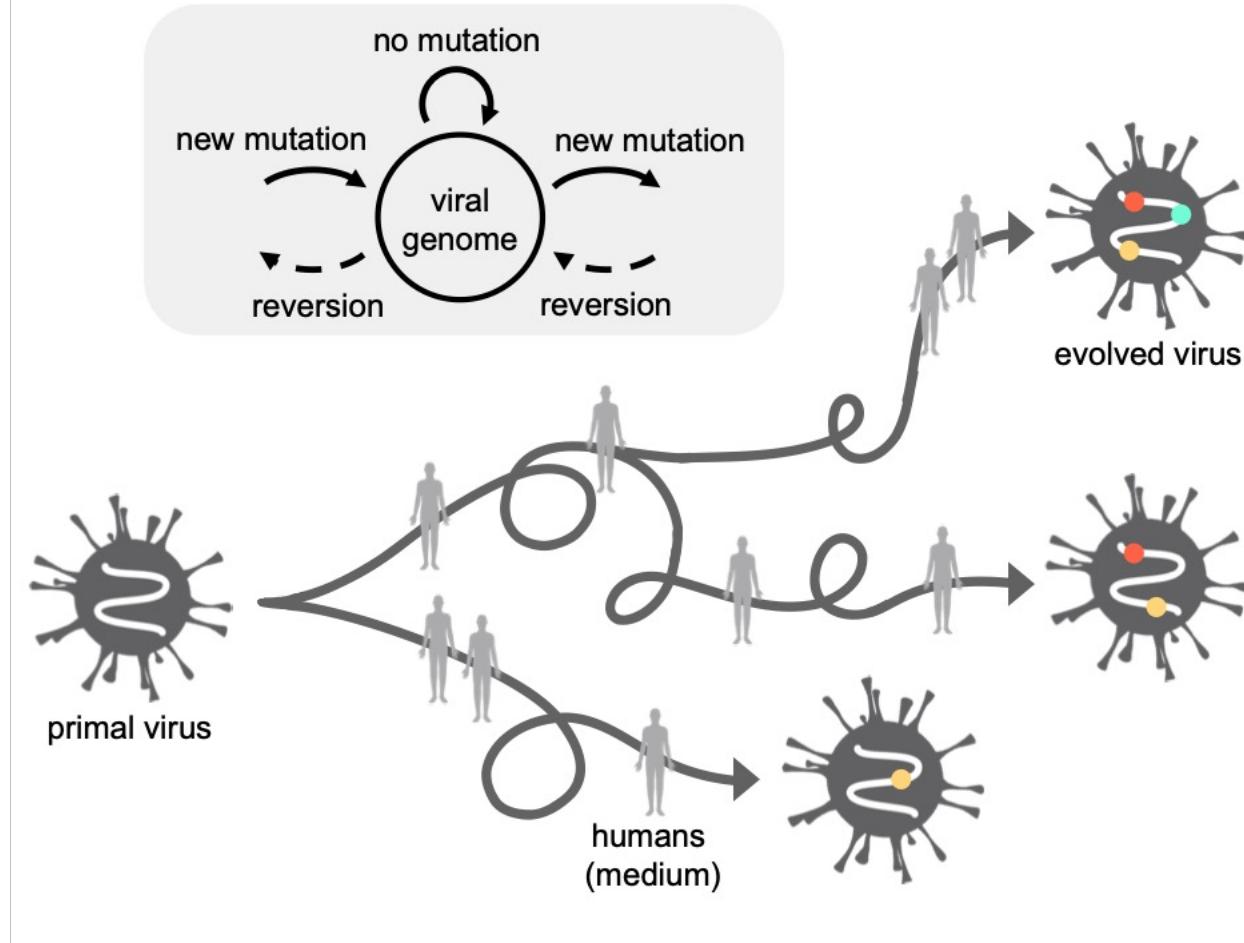
Substitutions accumulate at a constant rate: $\mathbb{E}[N(t)] = \mu t, \quad \mathbb{V}[N(t)] = \mu t$

Observed overdispersion challenges this assumption

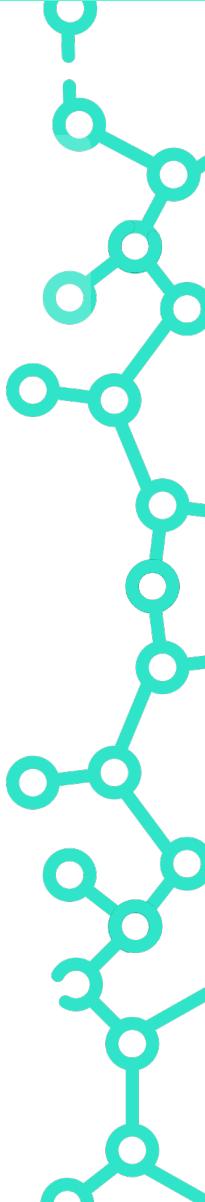
SARS-CoV-2: ideal model to test these deviations



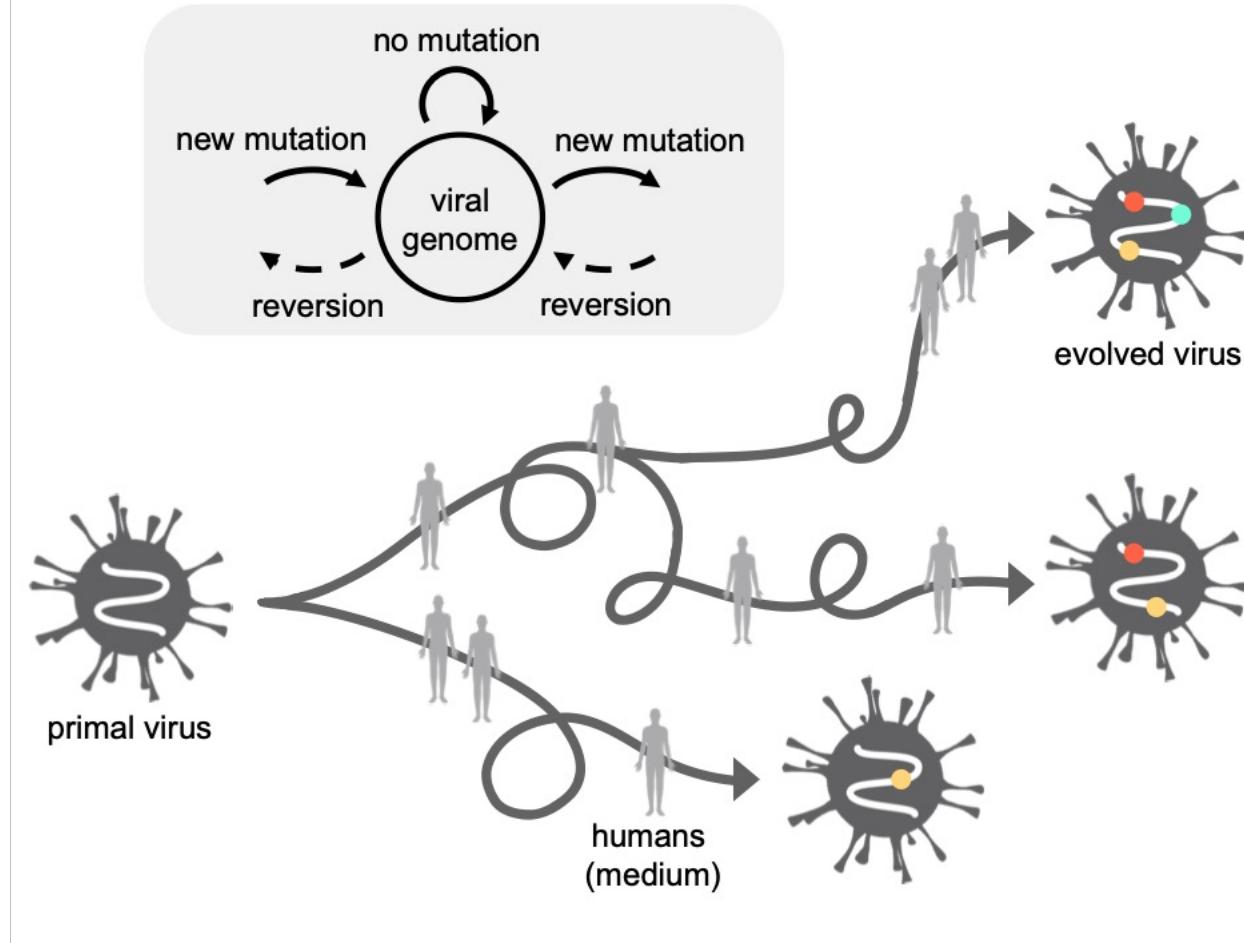
SARS-CoV-2 as a “Natural” Evolutionary Experiment



10 million SARS-CoV-2 genomes sequenced globally (May 2022)



SARS-CoV-2 as a “Natural” Evolutionary Experiment

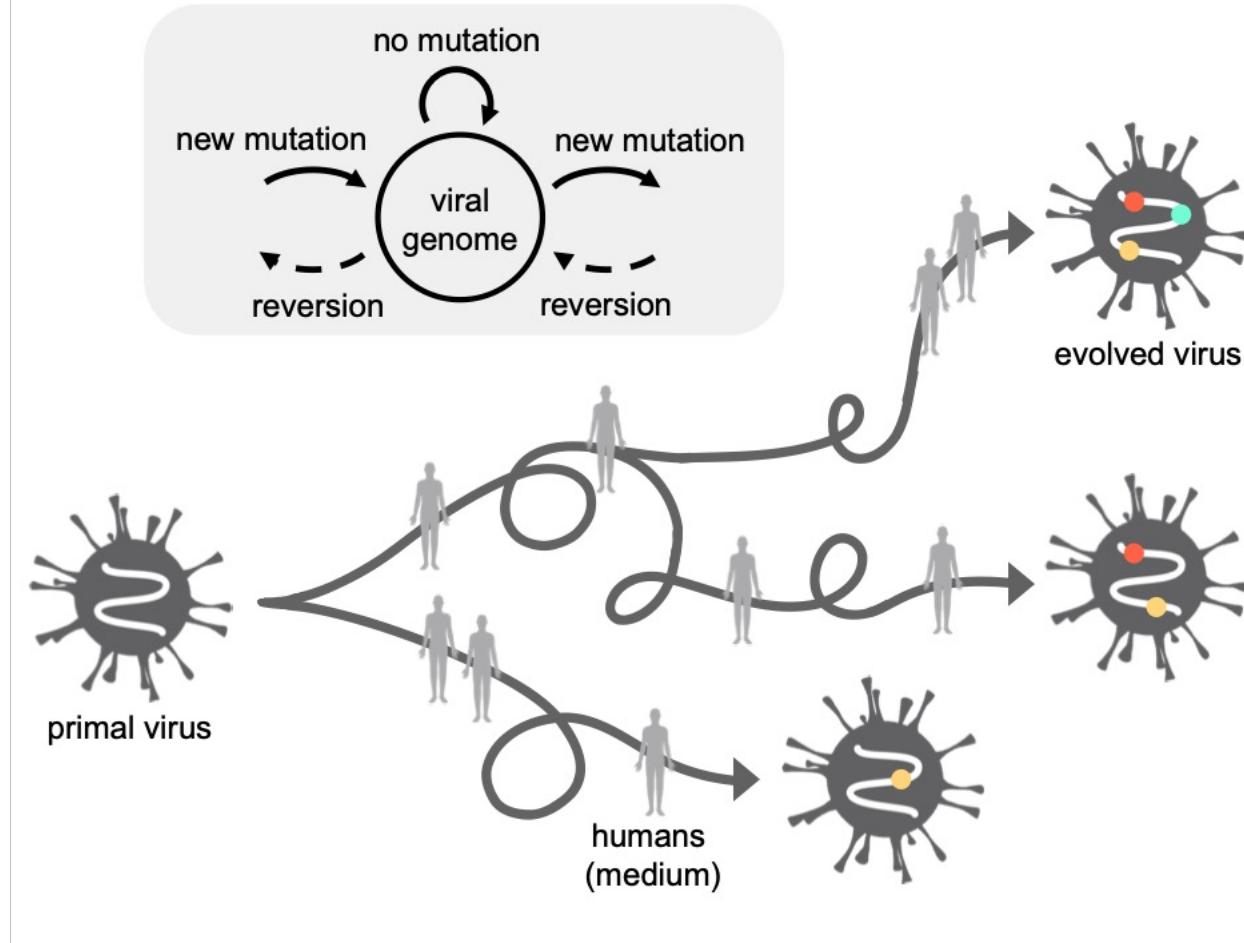


10 million SARS-CoV-2 genomes sequenced globally (May 2022)

Variant emergence and transitions:



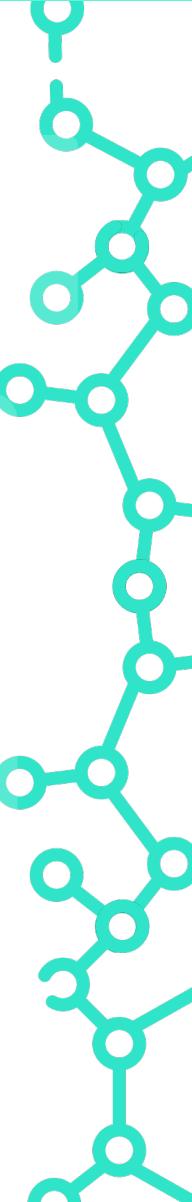
SARS-CoV-2 as a “Natural” Evolutionary Experiment



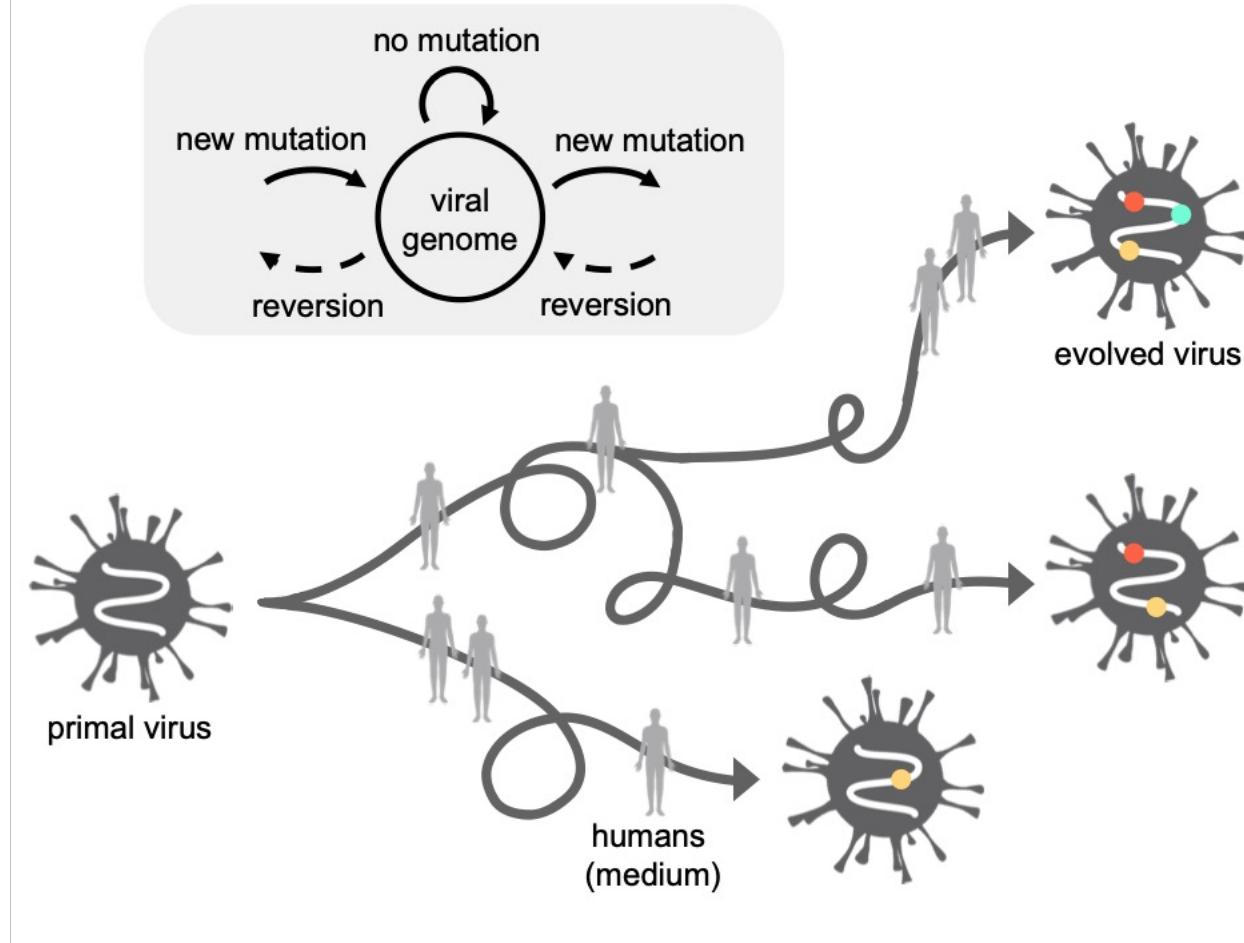
10 million SARS-CoV-2 genomes sequenced globally (May 2022)

Variant emergence and transitions:

- ❖ Mutation accumulation rates



SARS-CoV-2 as a “Natural” Evolutionary Experiment



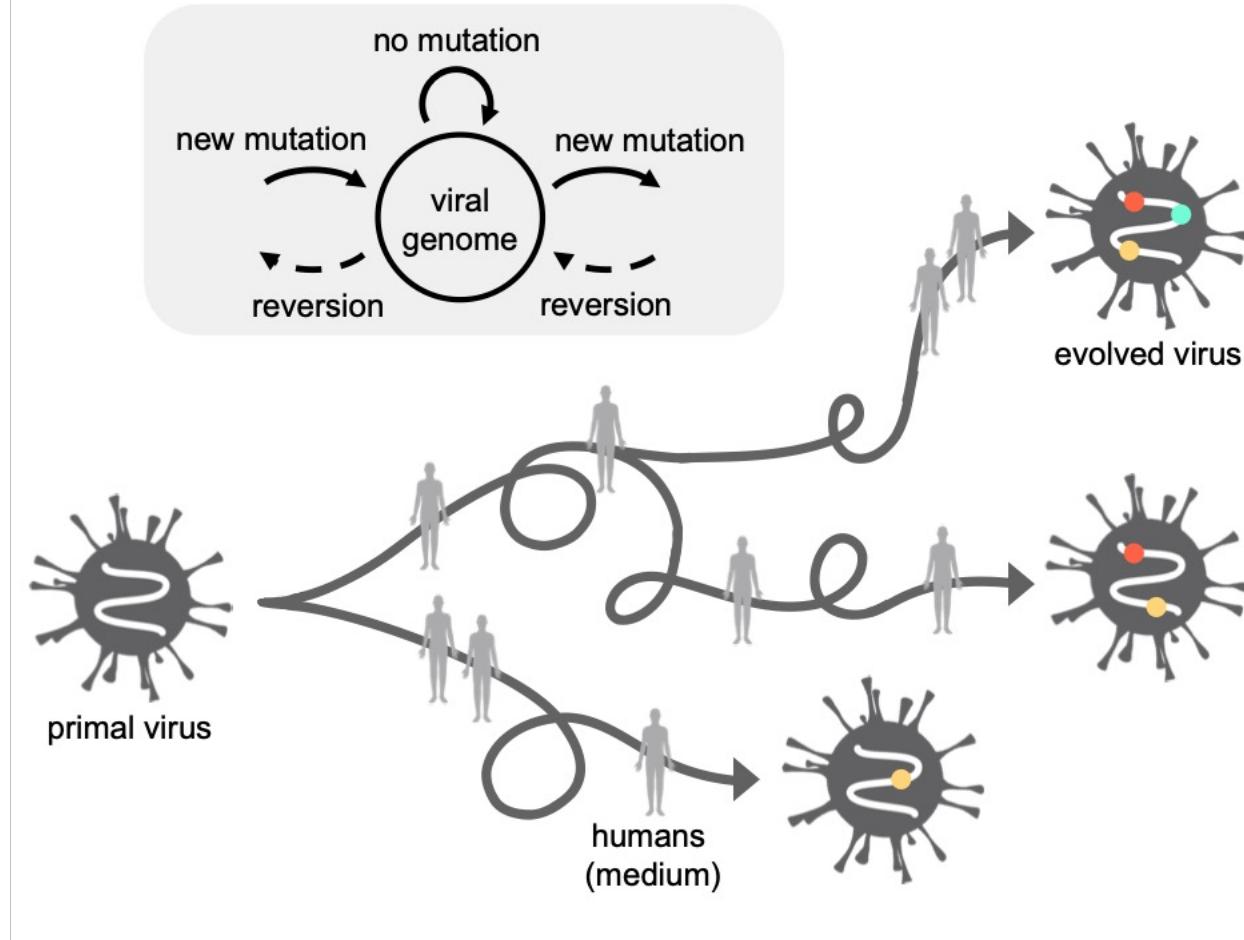
10 million SARS-CoV-2 genomes sequenced globally (May 2022)

Variant emergence and transitions:

- ❖ Mutation accumulation rates
- ❖ Transmission dynamics and selective pressures



SARS-CoV-2 as a “Natural” Evolutionary Experiment



10 million SARS-CoV-2 genomes sequenced globally (May 2022)

Variant emergence and transitions:

- ❖ Mutation accumulation rates
- ❖ Transmission dynamics and selective pressures
- ❖ Ideal system to probe evolutionary motion statistically



Modeling Viral Evolution as Stochastic Motion

Evolution modeled as displacement \Rightarrow mutations as motion from original genome.

$$\frac{dm(t)}{dt} = \kappa + \xi(t)$$

where $m(t)$ is the number of mutations over time, κ evolution rate and $\xi(t)$ is a noise source

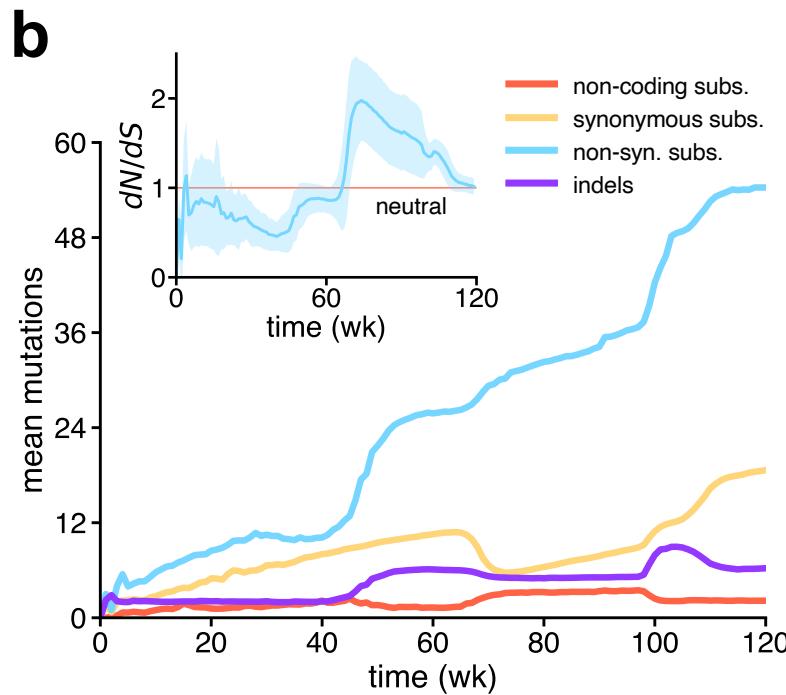
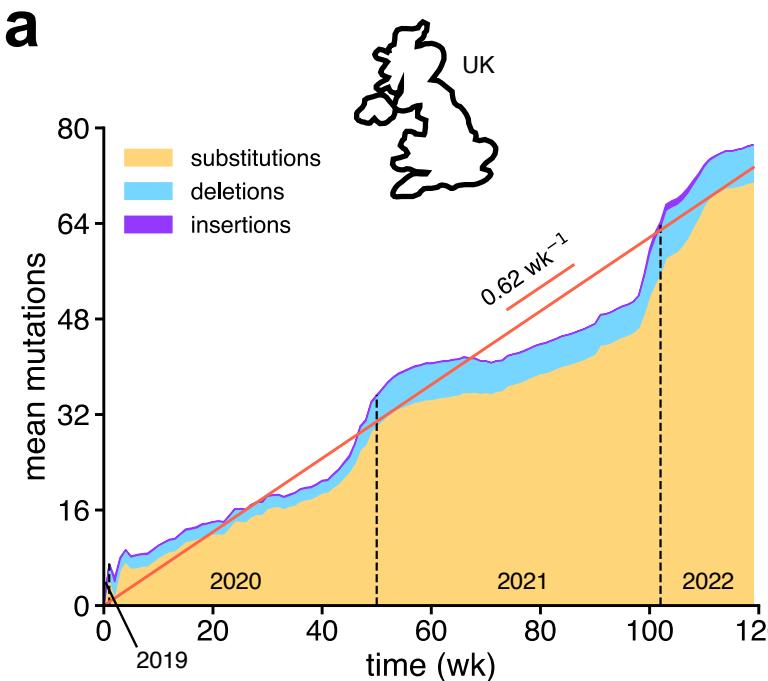


Modeling Viral Evolution as Stochastic Motion

Evolution modeled as displacement \Rightarrow mutations as motion from original genome.

$$\frac{dm(t)}{dt} = \kappa + \xi(t)$$

where $m(t)$ is the number of mutations over time, κ evolution rate and $\xi(t)$ is a noise source



Time-binned weekly analysis
reveals $\kappa = 0.62 \text{ wk}^{-1}$

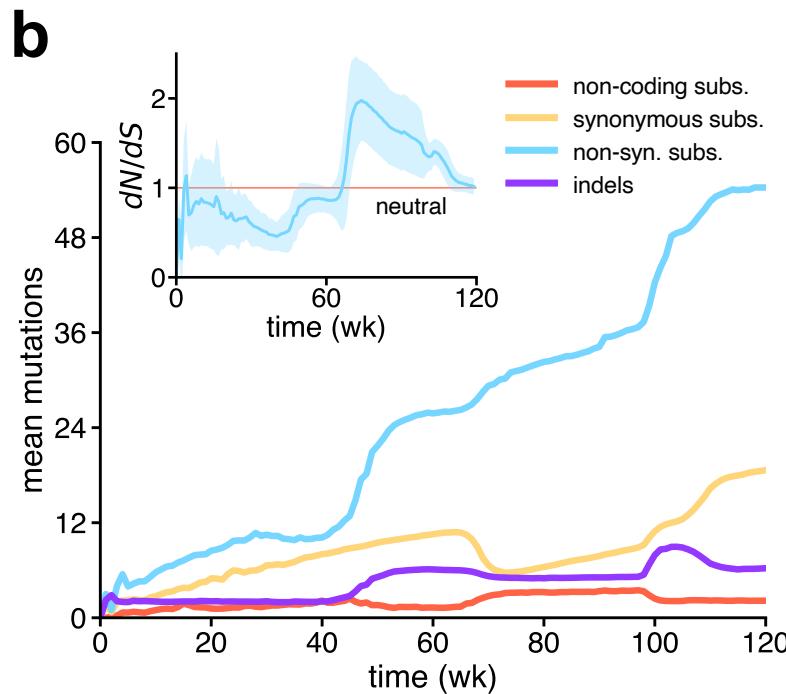
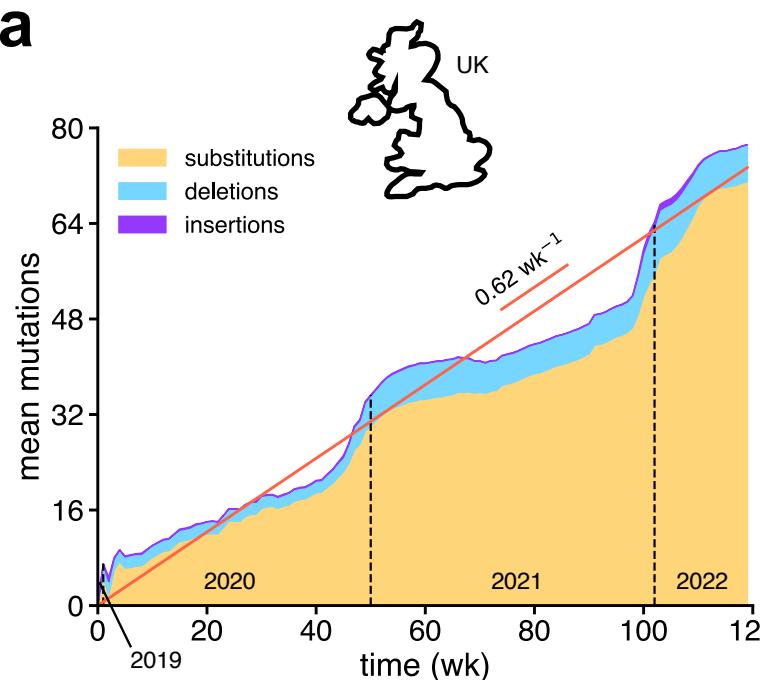


Modeling Viral Evolution as Stochastic Motion

Evolution modeled as displacement \Rightarrow mutations as motion from original genome.

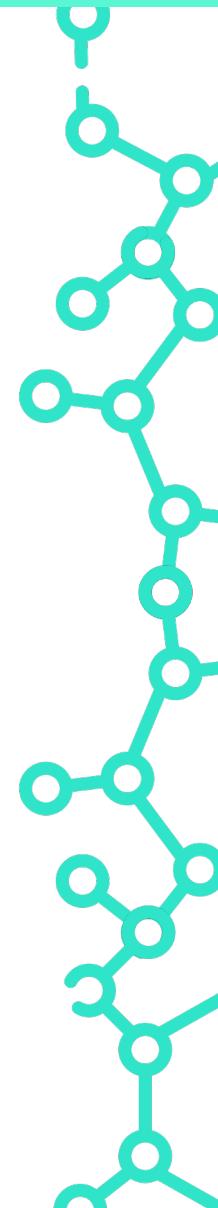
$$\frac{dm(t)}{dt} = \kappa + \xi(t)$$

where $m(t)$ is the number of mutations over time, κ evolution rate and $\xi(t)$ is a noise source



Time-binned weekly analysis reveals $\kappa = 0.62 \text{ wk}^{-1}$

$\frac{dN}{dS}$ reveals transitions

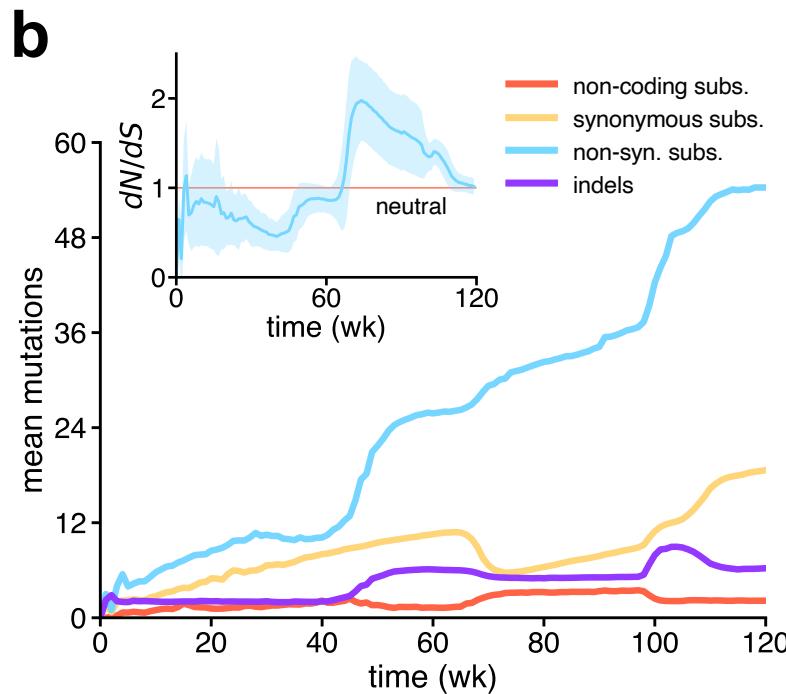
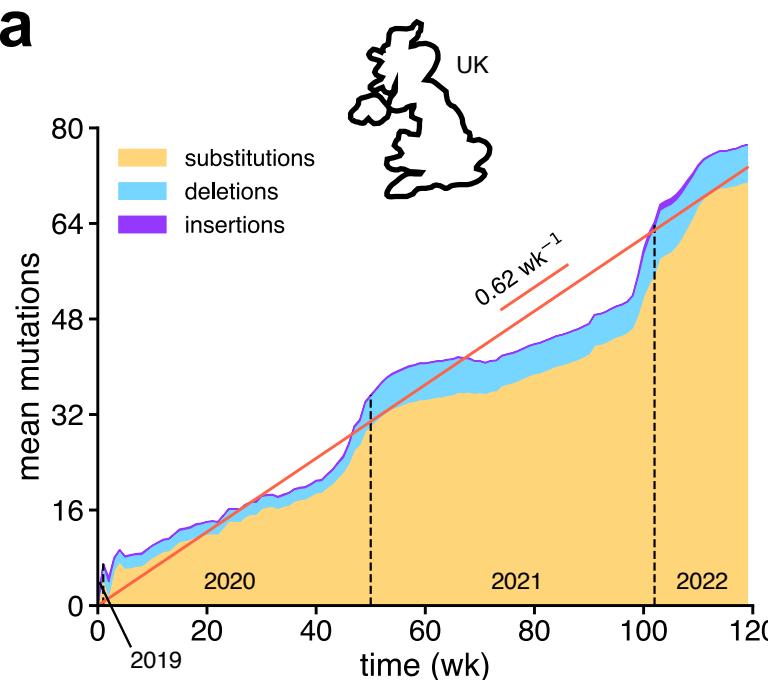


Modeling Viral Evolution as Stochastic Motion

Evolution modeled as displacement \Rightarrow mutations as motion from original genome.

$$\frac{dm(t)}{dt} = \kappa + \xi(t)$$

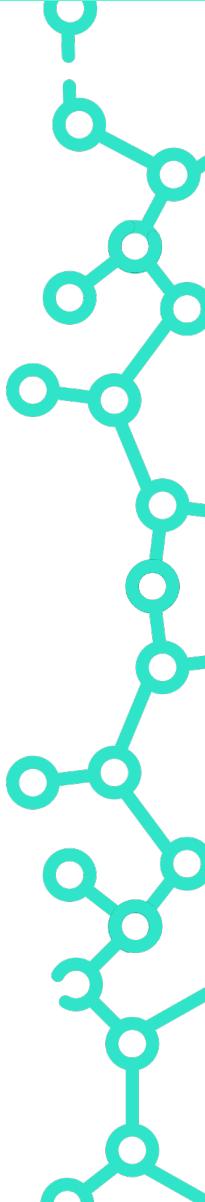
where $m(t)$ is the number of mutations over time, κ evolution rate and $\xi(t)$ is a noise source



Time-binned weekly analysis reveals $\kappa = 0.62 \text{ wk}^{-1}$

$\frac{dN}{dS}$ reveals transitions

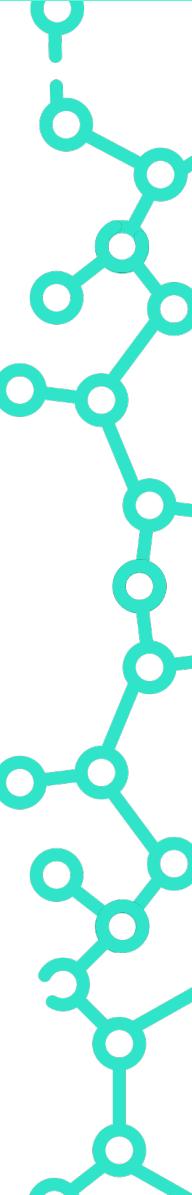
Apparent acceleration in the evolution rate was observed \Rightarrow **deviation** from classical molecular clock



Mutation Variance and Dispersion Index

The dispersion index $\rho(t)$ is a measure of the relative variability of a stochastic process $X(t)$:

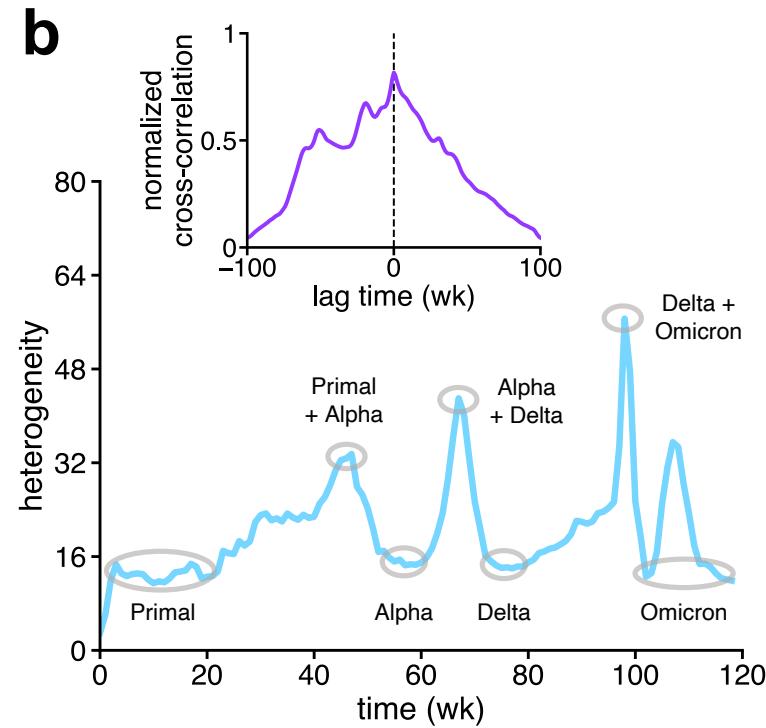
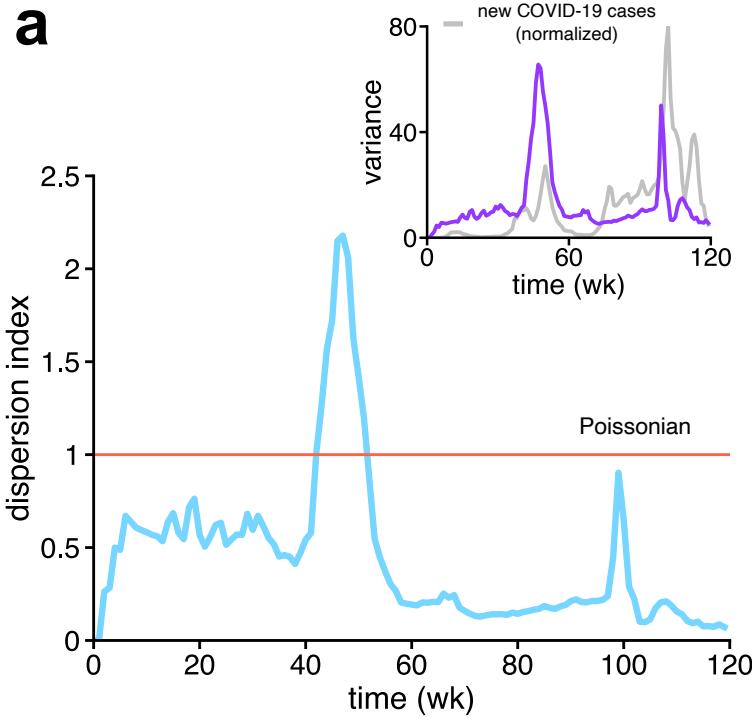
$$\rho(t) = \frac{\mathbb{V}[X(t)]}{\mathbb{E}[X(t)]}$$



Mutation Variance and Dispersion Index

The dispersion index $\rho(t)$ is a measure of the relative variability of a stochastic process $X(t)$:

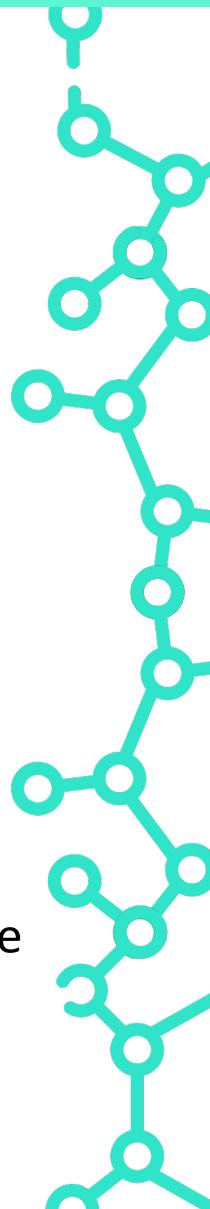
$$\rho(t) = \frac{\mathbb{V}[X(t)]}{\mathbb{E}[X(t)]}$$



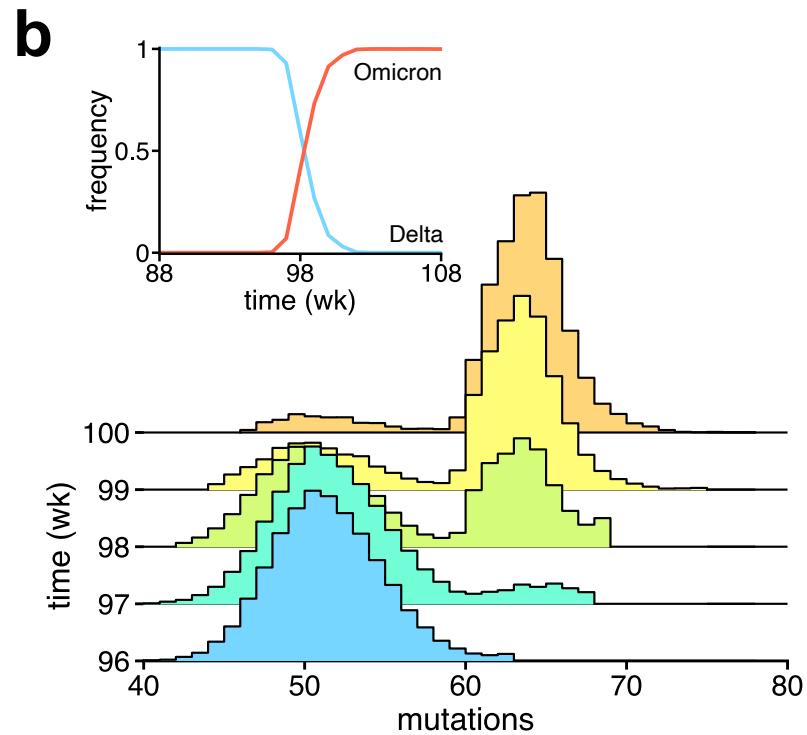
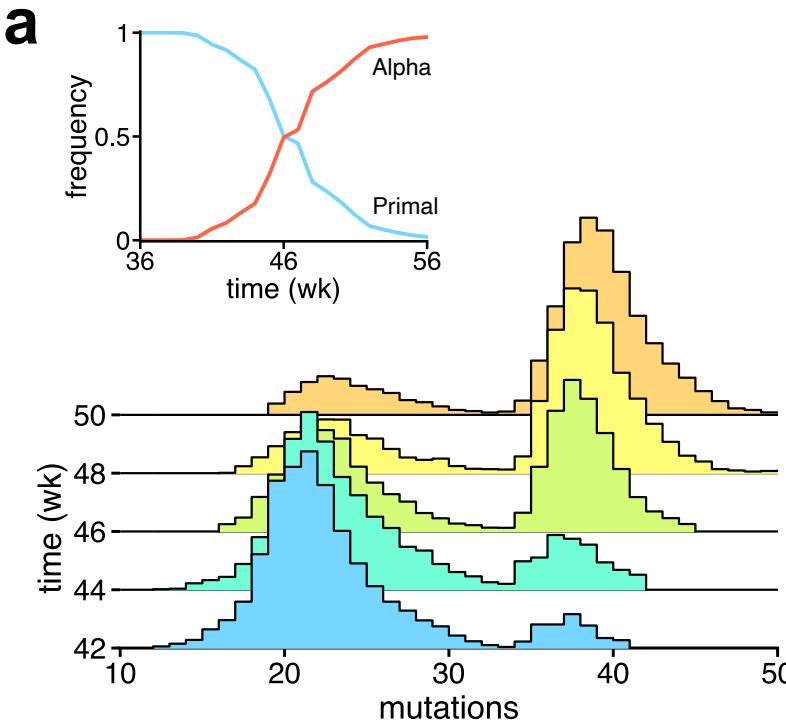
Overall, an apparent **sub-Poissonian** behavior is observed:

$$\rho(t) < 1$$

Existence of a constraint in the evolutionary motion



Dispersion Bursts

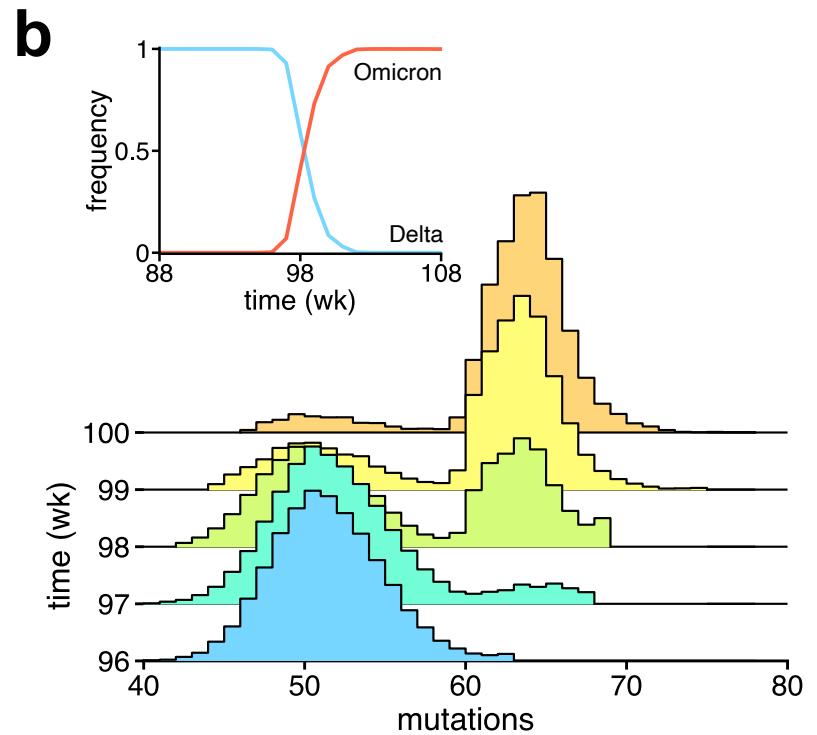
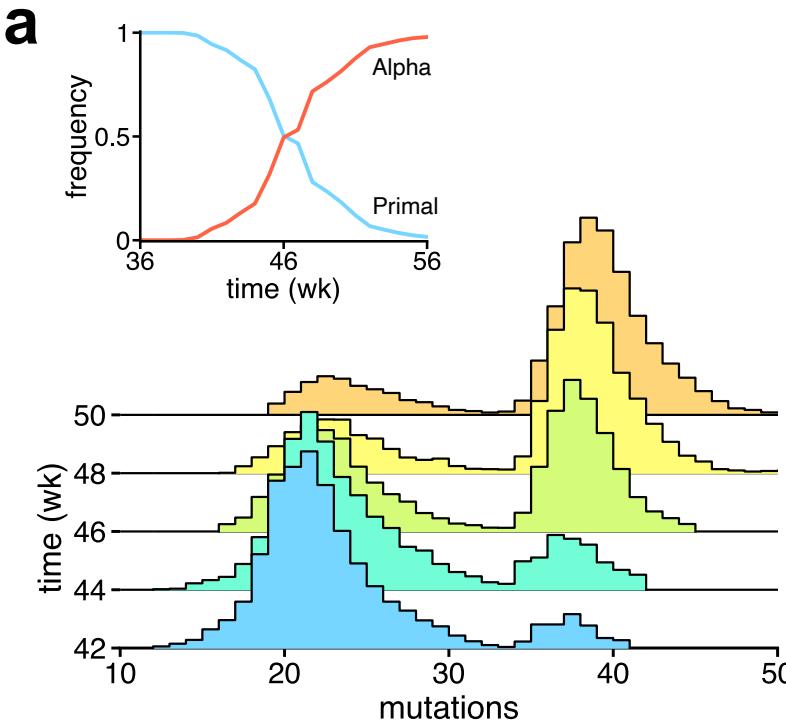


Two major dispersion burst events:

- ❖ Primal \Rightarrow Alpha
- ❖ Delta \Rightarrow Omicron



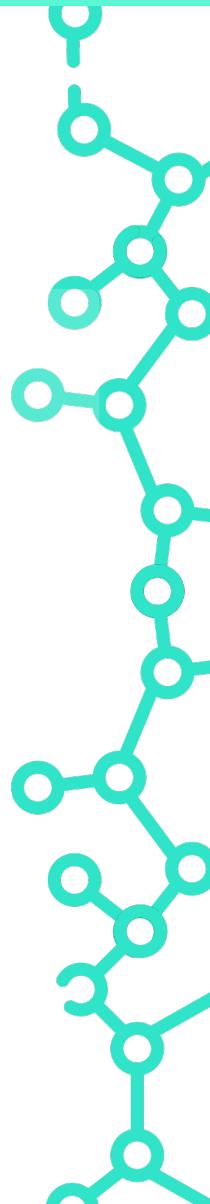
Dispersion Bursts



Two major dispersion burst events:

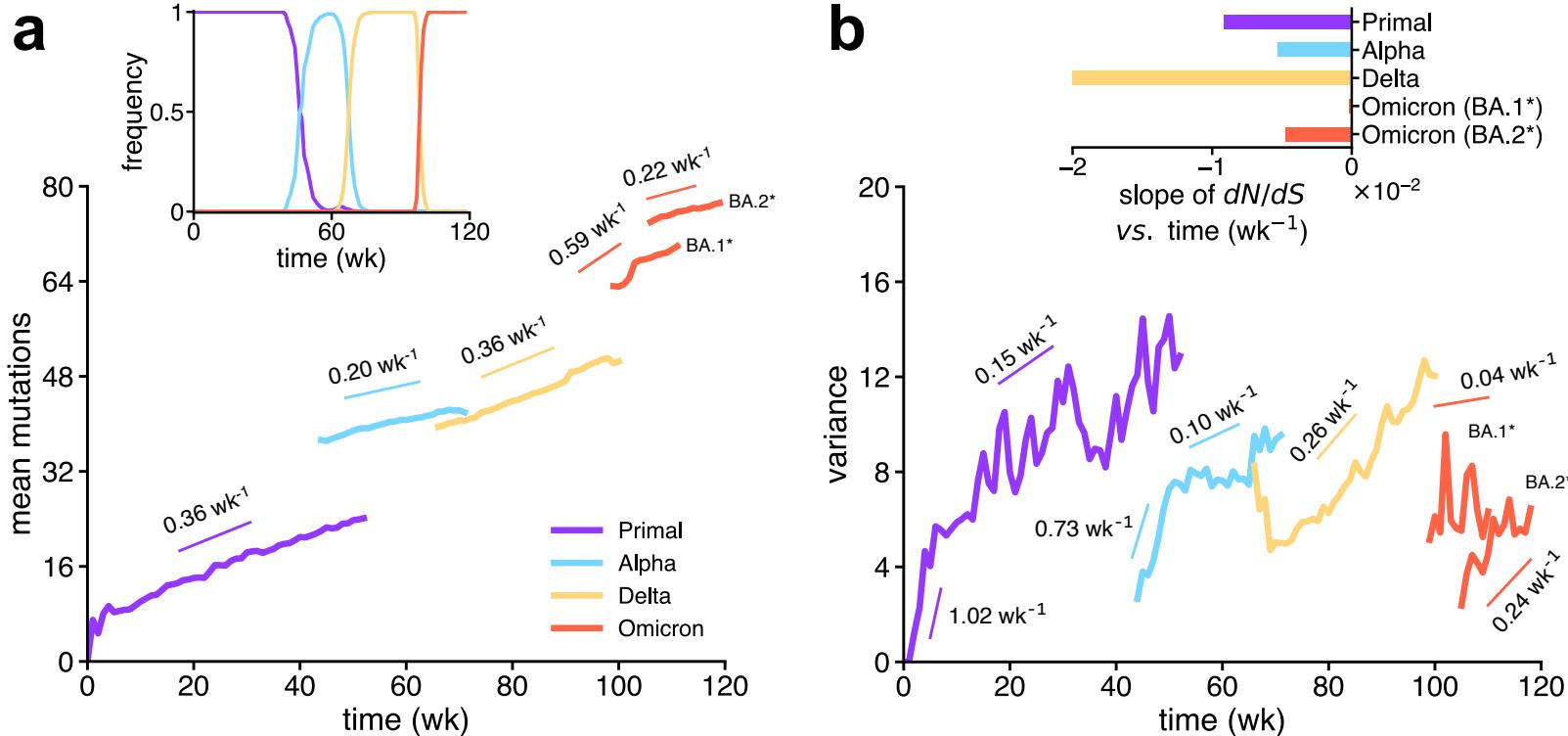
- ❖ Primal \Rightarrow Alpha
- ❖ Delta \Rightarrow Omicron

Bimodal mutation distributions \Rightarrow coexistence of divergent genotypes



Variant-Specific Mutation Dynamics

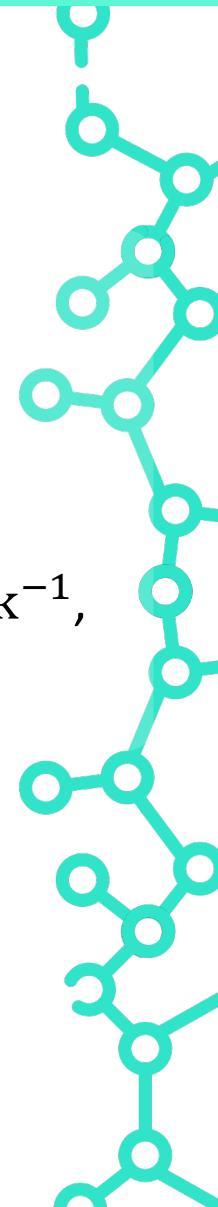
Per-variant analysis: Primal, Alpha, Delta & Omicron



Variant-specific rates:

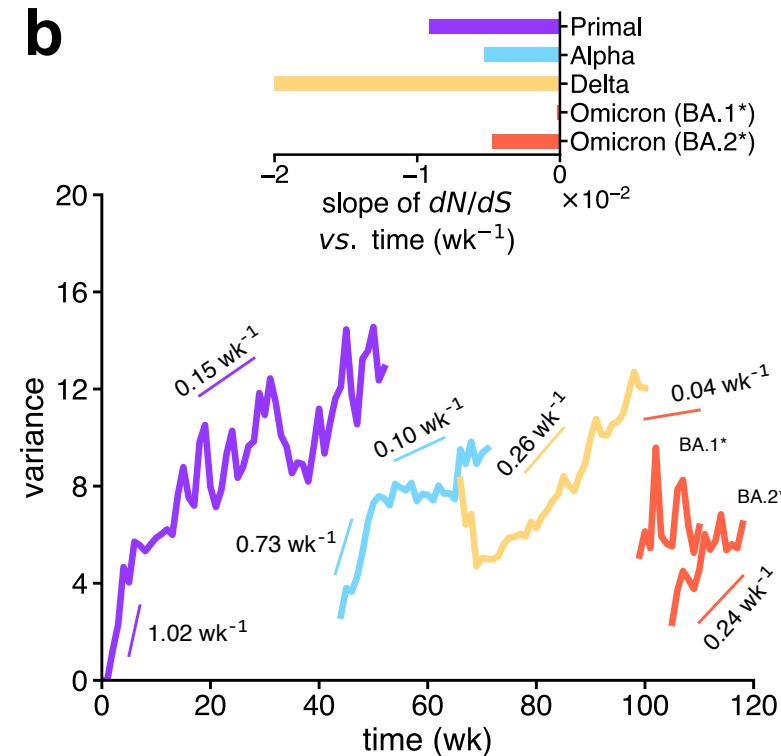
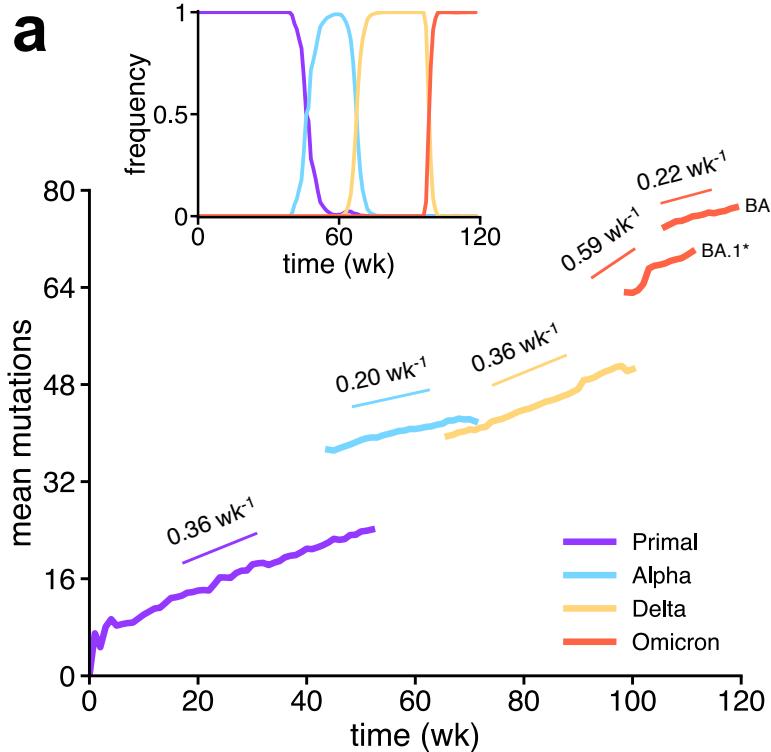
- ❖ Primal: $\sim 0.36 \text{ wk}^{-1}$
- ❖ Alpha: $\sim 0.20 \text{ wk}^{-1}$
- ❖ Delta: $\sim 0.36 \text{ wk}^{-1}$
- ❖ Omicron BA.1: $\sim 0.24 \text{ wk}^{-1}$,
BA.2: $\sim 0.22 \text{ wk}^{-1}$

Global rate **overestimates**
variant-level dynamics.



Variant-Specific Mutation Dynamics

Per-variant analysis: Primal, Alpha, Delta & Omicron

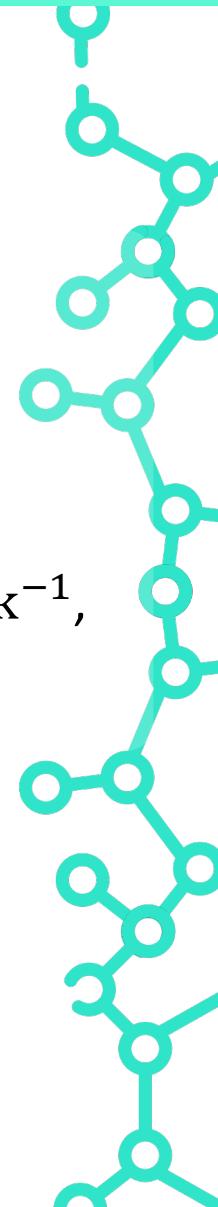


Variant-specific rates:

- ❖ Primal: $\sim 0.36 \text{ wk}^{-1}$
- ❖ Alpha: $\sim 0.20 \text{ wk}^{-1}$
- ❖ Delta: $\sim 0.36 \text{ wk}^{-1}$
- ❖ Omicron BA.1: $\sim 0.24 \text{ wk}^{-1}$, BA.2: $\sim 0.22 \text{ wk}^{-1}$

Global rate **overestimates** variant-level dynamics.

Nonlinear patterns in the variance **challenge** the molecular clock

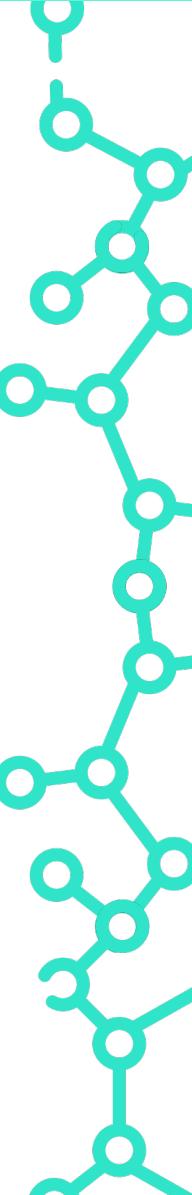


Anomalous Diffusion in *Evolutionary Motion*

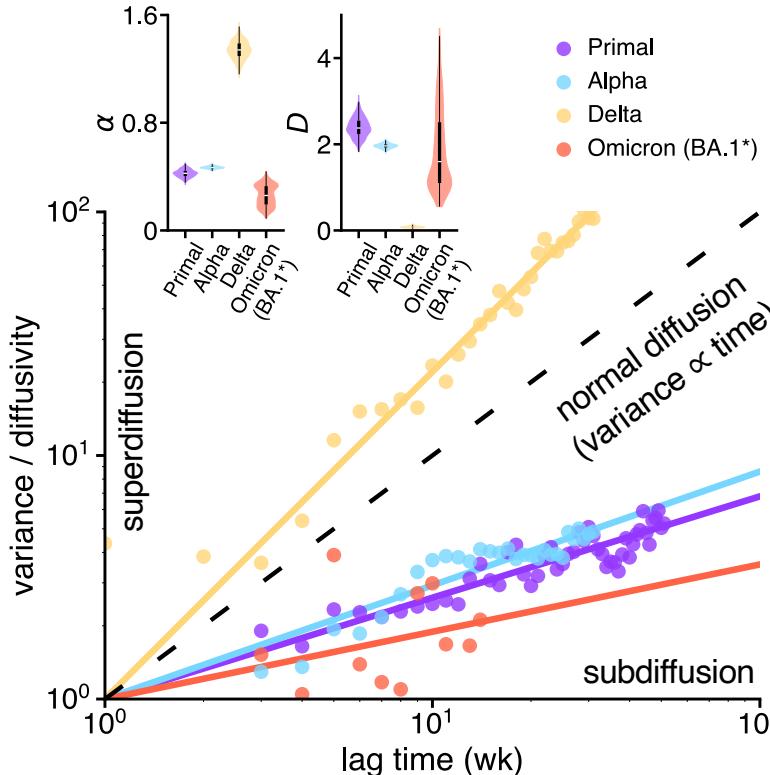
The noise source $\xi(t)$ is not a **white noise** (classical Brownian Motion), and rather obeys

$$\mathbb{E}[\xi(t)] = 0, \quad \text{Cov}[\xi(t), \xi(s)] = \frac{1}{2} D\alpha(\alpha - 1) |t - s|^{\alpha-2}$$

where α is the diffusion exponent.



Anomalous Diffusion in *Evolutionary Motion*



$$\log(\mathbb{V}[m(t)]/D) = \alpha \log t$$

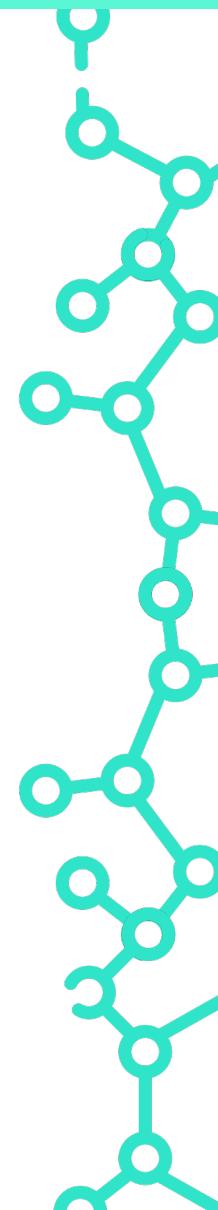
The noise source $\xi(t)$ is not a **white noise** (classical Brownian Motion), and rather obeys

$$\mathbb{E}[\xi(t)] = 0, \quad \text{Cov}[\xi(t), \xi(s)] = \frac{1}{2} D \alpha (\alpha - 1) |t - s|^{\alpha - 2}$$

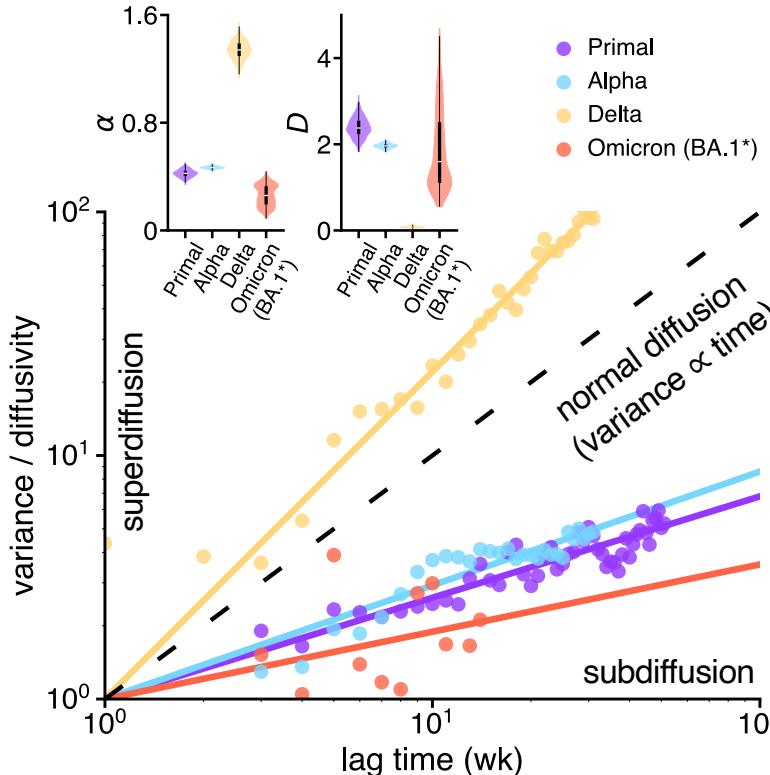
where α is the diffusion exponent. With this formulation

$$\mathbb{E}[m(t)] = kt, \quad \mathbb{V}[m(t)] = Dt^\alpha$$

i.e. variance obeys a power law (given $\alpha \neq 1$). $\frac{dm(t)}{dt}$ describes a **Fractional Brownian Motion**.



Anomalous Diffusion in *Evolutionary Motion*



$$\log(\mathbb{V}[m(t)]/D) = \alpha \log t$$

The noise source $\xi(t)$ is not a **white noise** (classical Brownian Motion), and rather obeys

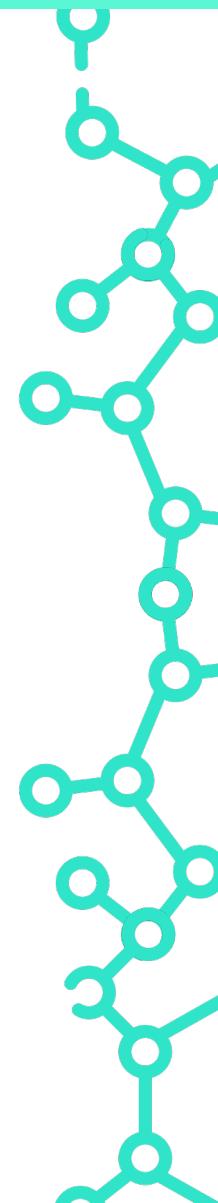
$$\mathbb{E}[\xi(t)] = 0, \quad \text{Cov}[\xi(t), \xi(s)] = \frac{1}{2} D \alpha (\alpha - 1) |t - s|^{\alpha - 2}$$

where α is the diffusion exponent. With this formulation

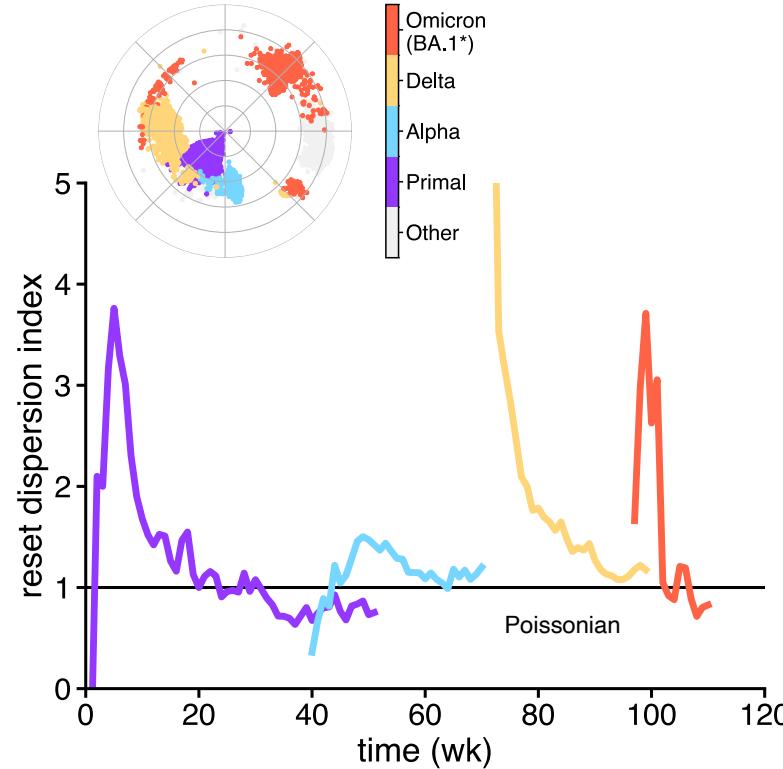
$$\mathbb{E}[m(t)] = kt, \quad \mathbb{V}[m(t)] = Dt^\alpha$$

i.e. variance obeys a power law (given $\alpha \neq 1$). $\frac{dm(t)}{dt}$ describes a **Fractional Brownian Motion**.

$\alpha = 1$ recovers classical Brownian motion, $\alpha < 1$ indicates **subdiffusion** and $\alpha > 1$ indicates **superdiffusion**.



Reset Dispersion and Asymptotic Poisson Behavior

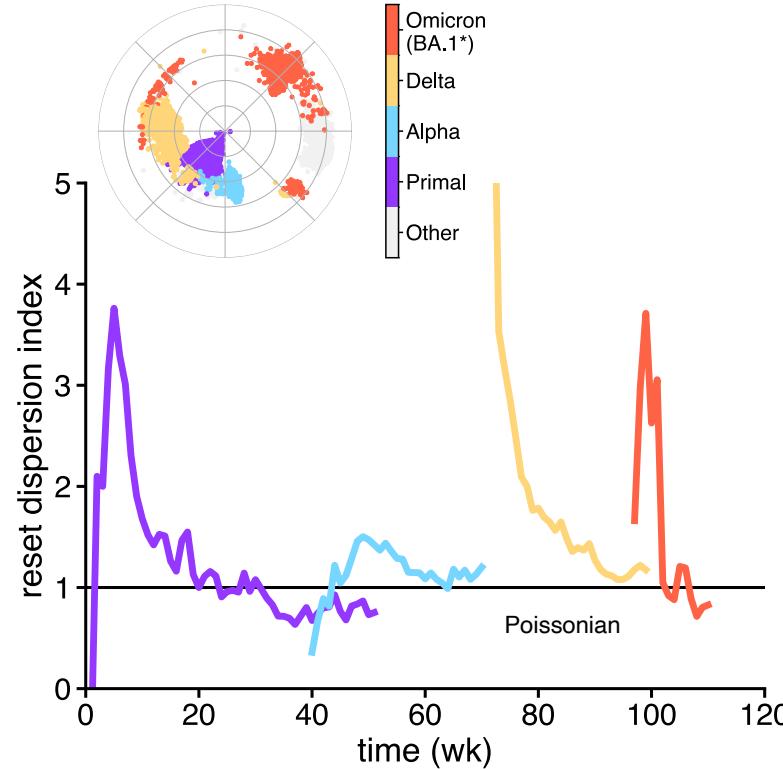


The reset dispersion index ρ_k^{reset} for a particular variant v is defined as

$$\rho_k^{\text{reset}} = \frac{\mathbb{V}[m_{v,k}]}{\mathbb{E}[m_{v,k}] - \mu_v}$$

where $m_{v,k}$ is the number of mutations for variant v on the k^{th} week, and μ_v is the mean number of mutations in the week in which $\mathbb{V}[m_{v,k}]$ is minimal.

Reset Dispersion and Asymptotic Poisson Behavior



The reset dispersion index ρ_k^{reset} for a particular variant v is defined as

$$\rho_k^{\text{reset}} = \frac{\mathbb{V}[m_{v,k}]}{\mathbb{E}[m_{v,k}] - \mu_v}$$

where $m_{v,k}$ is the number of mutations for variant v on the k^{th} week, and μ_v is the mean number of mutations in the week in which $\mathbb{V}[m_{v,k}]$ is minimal.

Resets mutation count at variant emergence

When $k \rightarrow \infty$, $\rho_k^{\text{reset}} \rightarrow 1 \Rightarrow \text{asymptotic Poissonian behavior}$ within variants



Closing Remarks

- 1. Challenges the molecular clock paradigm:** Viral evolution is *non-uniform* and *variant-dependent*, departing from classical Poisson models.
- 2. Reveals anomalous diffusion:** Evolutionary dynamics show *nonlinear mutation scaling* and *overdispersion* tied to selection and population shifts.
- 3. Calls for refined models:** Highlights the need for *stochastic, variant-aware frameworks* to better capture real-time viral evolution.



Chapter 4: PyEvoMotion: a software to perform the temporal statistical analysis of genome evolution

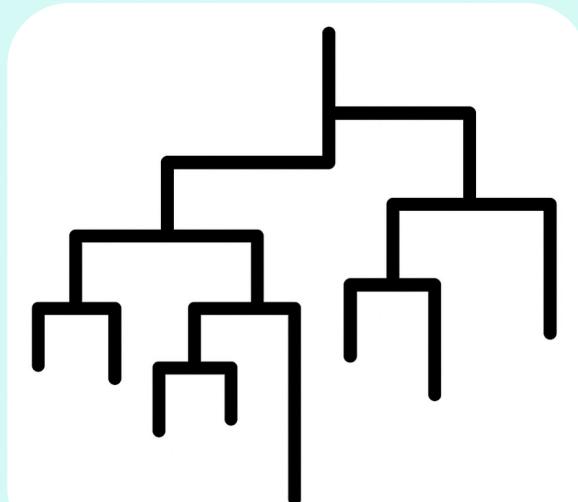
Goiriz L, Rodrigo G. (2025)

PyEvoMotion: a Python tool for population-based time-course analysis of genome evolution
Bioinformatics. (Under Review)

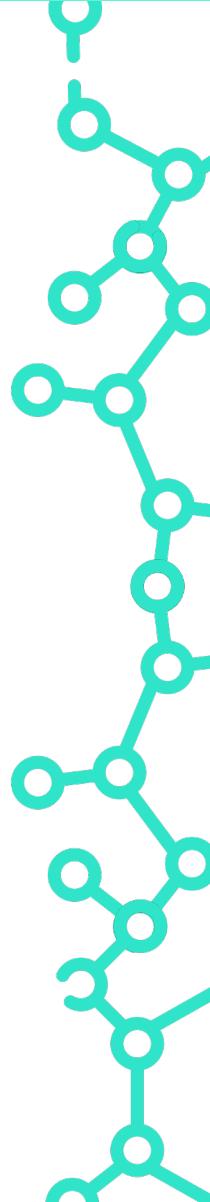


Phylogeny vs. Statistical Analysis

Classical Phylogenetic Methods

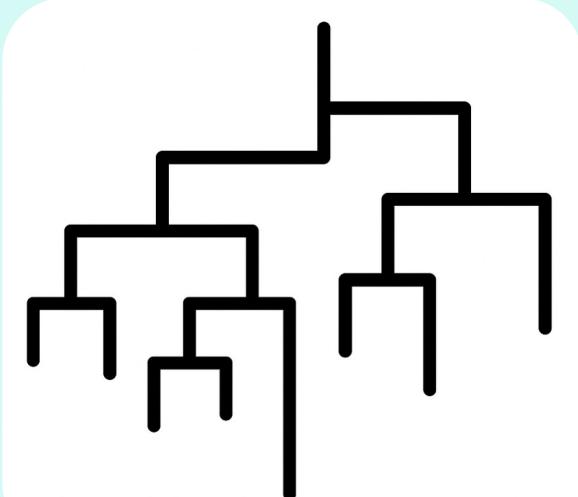


- Maximum Likelihood, Bayesian, Neighbor-joining
- Not scalable when we have $> 10^4$ sequences



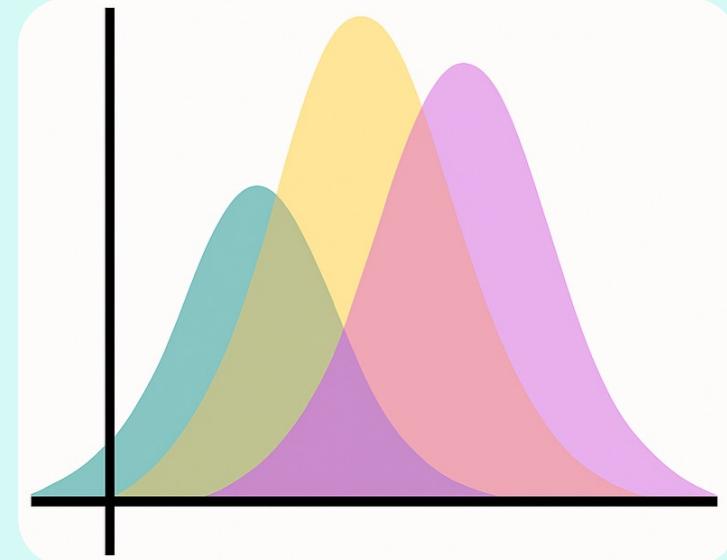
Phylogeny vs. Statistical Analysis

Classical Phylogenetic Methods



- Maximum Likelihood, Bayesian, Neighbor-joining
- Not scalable when we have $> 10^4$ sequences

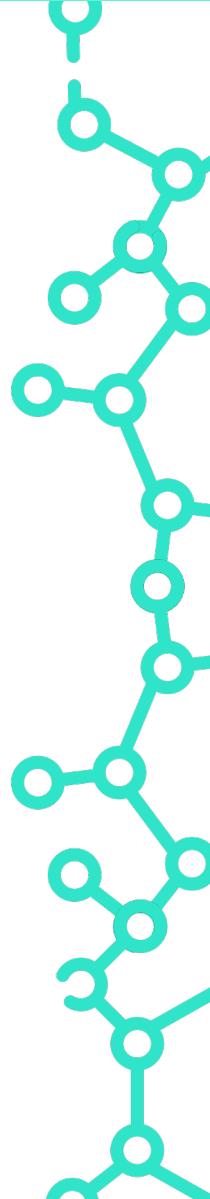
Statistical Properties in Data



- Captures probabilistic structure and patterns
- Scales up several orders of magnitude



Workflow and Data Requirements



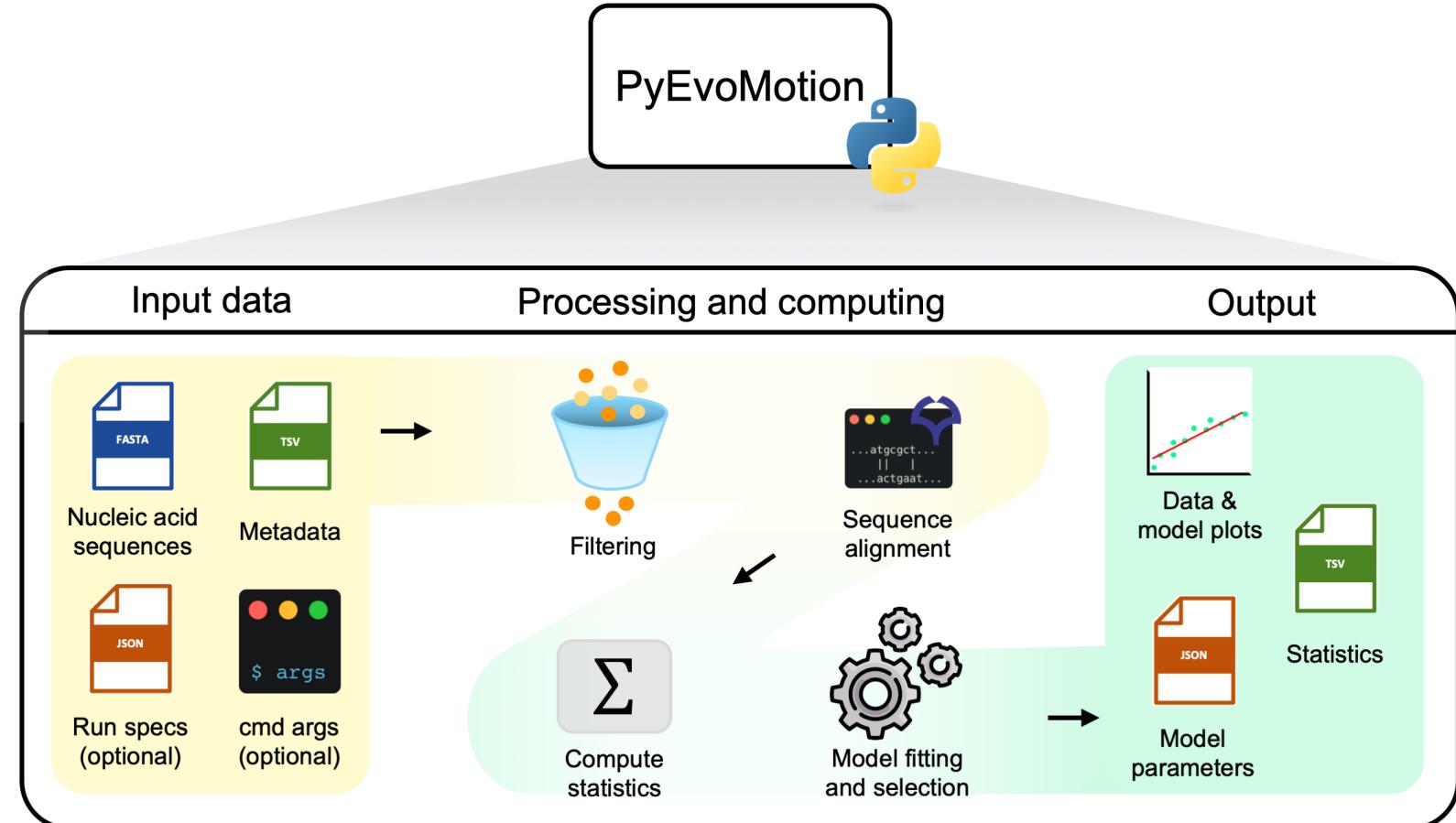
Inputs:

- .fasta file (genomic sequences)
- .tsv file (metadata, including collection dates)

Configurable options (time intervals, filter, mutations types)

Output:

- Graphical plots
- Statistics (.tsv, .json)

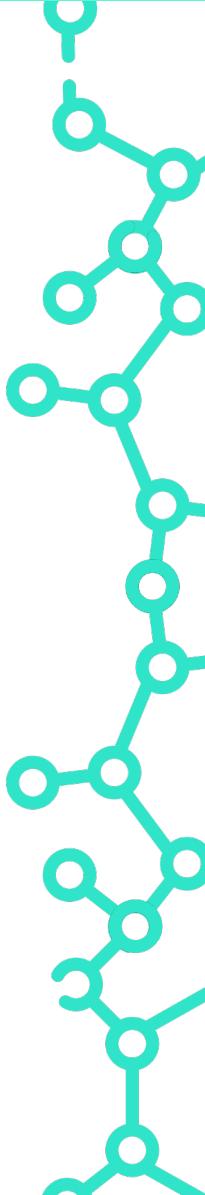


Sequence Alignment and Mutation Statistics

Alignment

Reference ⇒ first entry in .fasta file

Alignment via MAFFT using a subprocess call



Sequence Alignment and Mutation Statistics

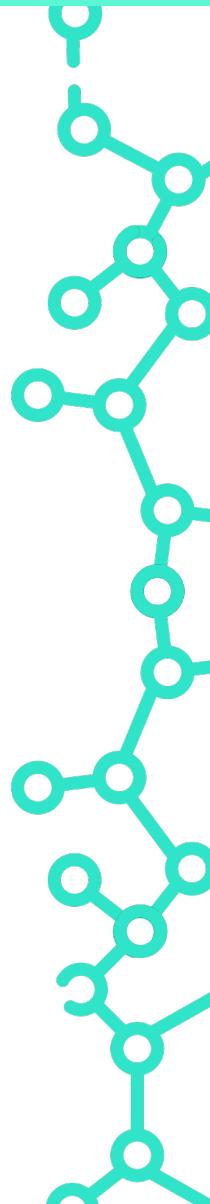
Alignment

Reference ⇒ first entry in .fasta file

Alignment via MAFFT using a subprocess call

```
● ● ●  
AlignIO.read(  
    StringIO(cls._run_mafft({...})),  
    "fasta"  
)
```

High throughput parsing via BioPython



Sequence Alignment and Mutation Statistics

Alignment

Reference \Rightarrow first entry in .fasta file

Alignment via MAFFT using a subprocess call



```
AlignIO.read(  
    StringIO(cls._run_mafft({...})),  
    "fasta"  
)
```

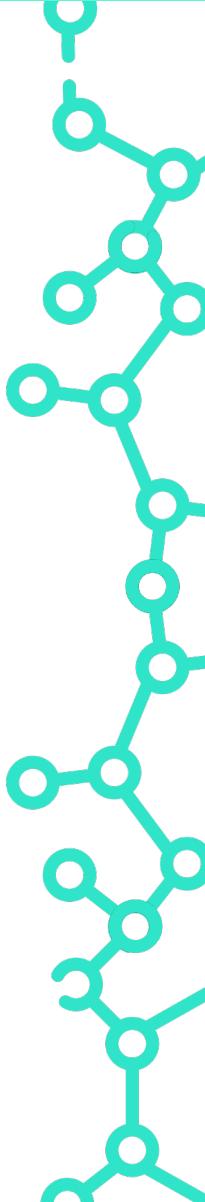
High throughput parsing via BioPython

Mutation Statistics per time interval

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} m_{k,i} \quad s_k^2 = \frac{1}{N_k - 1} \sum_{i=1}^{N_k} (m_{k,i} - \mu_k)^2$$

where

- N_k : the number of sequences in the time interval k
- $m_{k,i}$: mutations in sequence i at time interval k
- μ_k : mean ; s_k^2 : unbiased variance



Sequence Alignment and Mutation Statistics

Alignment

Reference \Rightarrow first entry in .fasta file

Alignment via MAFFT using a subprocess call



```
AlignIO.read(  
    StringIO(cls._run_mafft({...})),  
    "fasta"  
)
```

High throughput parsing via BioPython

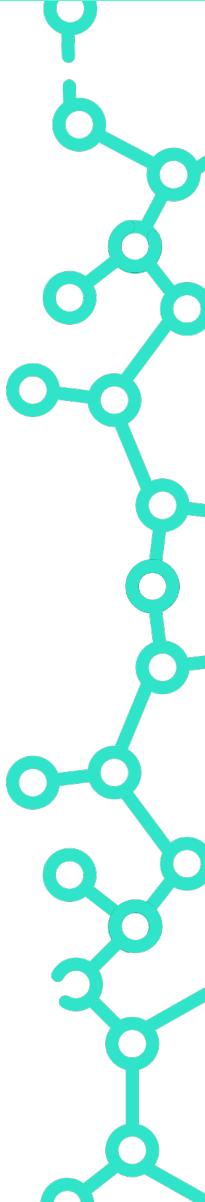
Mutation Statistics per time interval

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} m_{k,i} \quad s_k^2 = \frac{1}{N_k - 1} \sum_{i=1}^{N_k} (m_{k,i} - \mu_k)^2$$

where

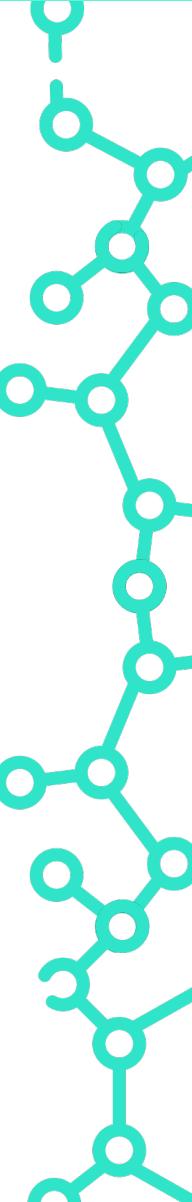
- N_k : the number of sequences in the time interval k
- $m_{k,i}$: mutations in sequence i at time interval k
- μ_k : mean ; s_k^2 : unbiased variance

Used for model fitting



Fitting Molecular Clock Models

Variance fitting to model: rescaling ($s_k^2 \rightarrow s_k^2 - s_0^2$) and time shifting ($t_0 = 0$)



Fitting Molecular Clock Models

Variance fitting to model: rescaling ($s_k^2 \rightarrow s_k^2 - s_0^2$) and time shifting ($t_0 = 0$)

Classical Brownian Motion

$$\mathbb{V}[m(t)] \propto t \Rightarrow \mathbb{V}[m(t)] = Dt$$

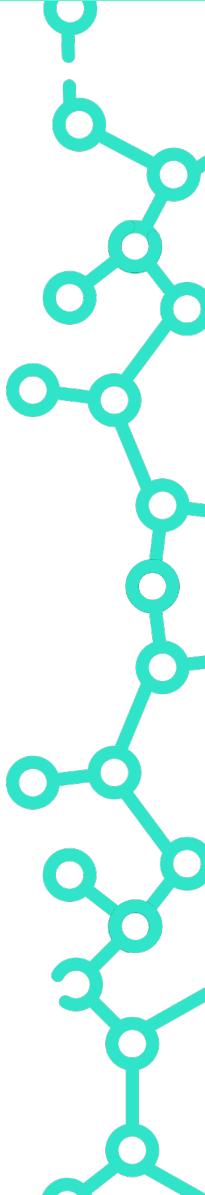
Used as null model

Fractional Brownian Motion

$$\mathbb{V}[m(t)] \propto t^\alpha \Rightarrow \mathbb{V}[m(t)] = Dt^\alpha$$

Used as challenging model

Null model is recovered by setting $\alpha = 1$



Fitting Molecular Clock Models

Variance fitting to model: rescaling ($s_k^2 \rightarrow s_k^2 - s_0^2$) and time shifting ($t_0 = 0$)

Classical Brownian Motion

$$\mathbb{V}[m(t)] \propto t \Rightarrow \mathbb{V}[m(t)] = Dt$$

Used as null model

Fractional Brownian Motion

$$\mathbb{V}[m(t)] \propto t^\alpha \Rightarrow \mathbb{V}[m(t)] = Dt^\alpha$$

Used as challenging model

Null model is recovered by setting $\alpha = 1$

Model Selection

F-tests on the residuals of the fits

$$F = \frac{\frac{RSS_1 - RSS_2}{p_2 - p_1}}{\frac{RSS_2}{n - p_2}}$$

where

- RSS_1, RSS_2 : residual sum of squares from the null and challenging model
- $p_1 = 1; p_2 = 2$: degrees of freedom
- n : number of data points

Significance level of 0.05

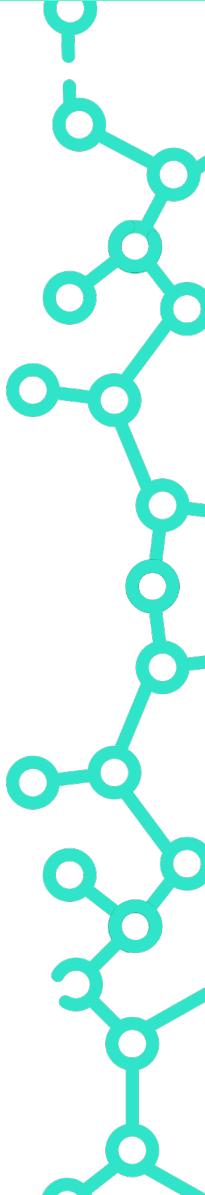


Customizability

Customizable targeted analysis available

Independent mutation processing

Analysis can be tailored to consider all mutations,
only substitutions or **only** indels



Customizability

Customizable targeted analysis available

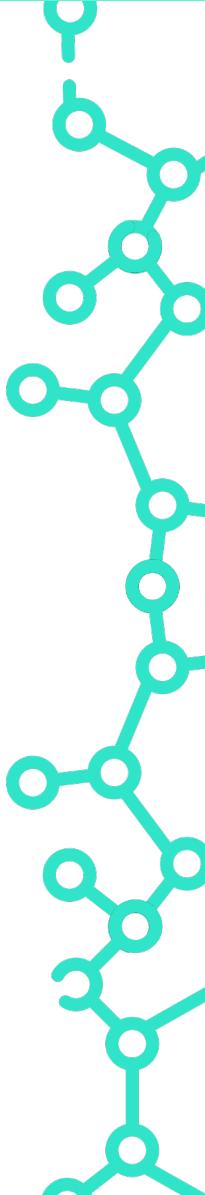
Independent mutation processing

Analysis can be tailored to consider all mutations,
only substitutions or **only** indels

Specific Genomic region targeting

Analysis can be tailored to specific genomic regions
(genes, intragenic regions, chromosomes)

Allows observation of particular intra-genome
evolutionary dynamics



Customizability

Customizable targeted analysis available

Independent mutation processing

Analysis can be tailored to consider all mutations,
only substitutions or **only** indels

Specific Genomic region targeting

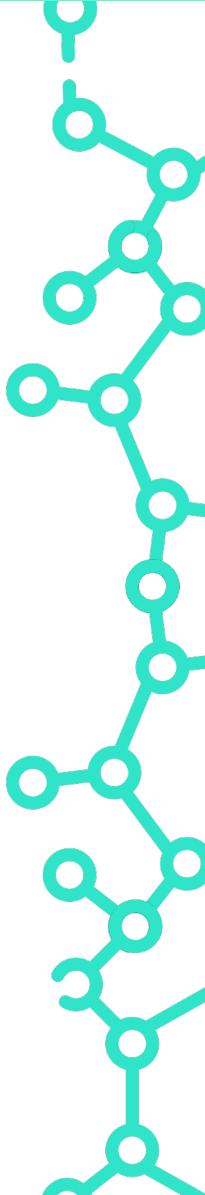
Analysis can be tailored to specific genomic regions
(genes, intragenic regions, chromosomes)

Allows observation of particular intra-genome
evolutionary dynamics

Time window customization

Allows controlling the resolution of the
evolutionary process

Useful to asses dynamics among entities with
distinct replication times



Customizability

Customizable targeted analysis available

Independent mutation processing

Analysis can be tailored to consider all mutations,
only substitutions or **only** indels

Specific Genomic region targeting

Analysis can be tailored to specific genomic regions
(genes, intragenic regions, chromosomes)

Allows observation of particular intra-genome
evolutionary dynamics

Time window customization

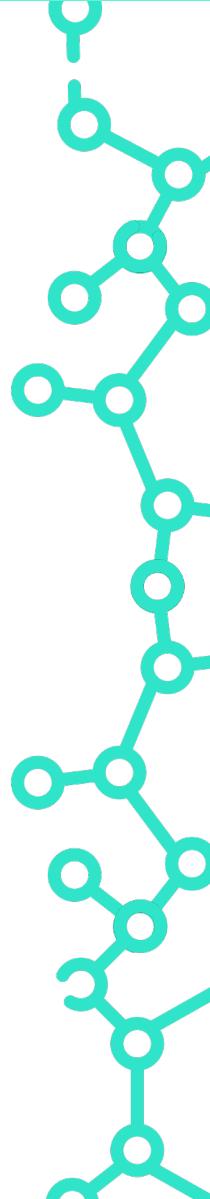
Allows controlling the resolution of the
evolutionary process

Useful to asses dynamics among entities with
distinct replication times

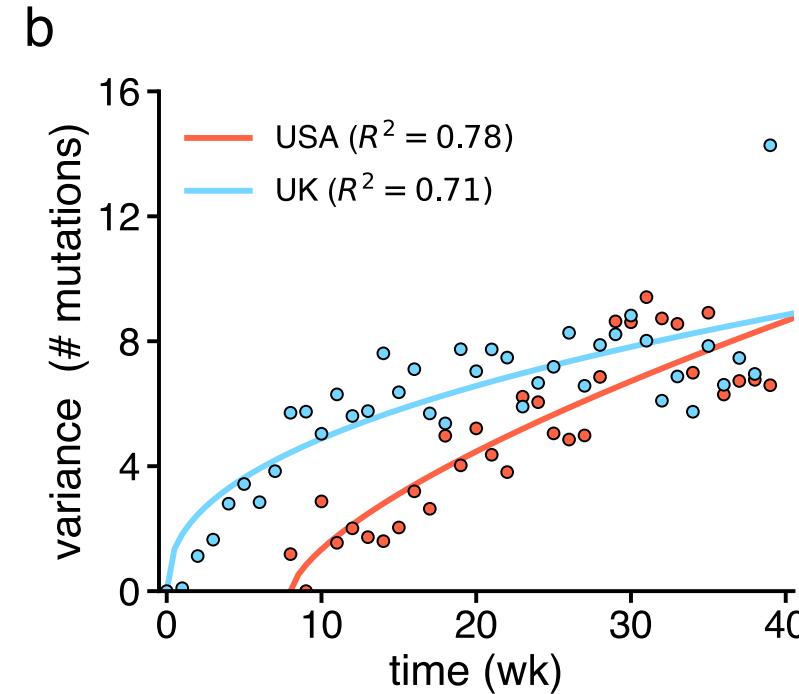
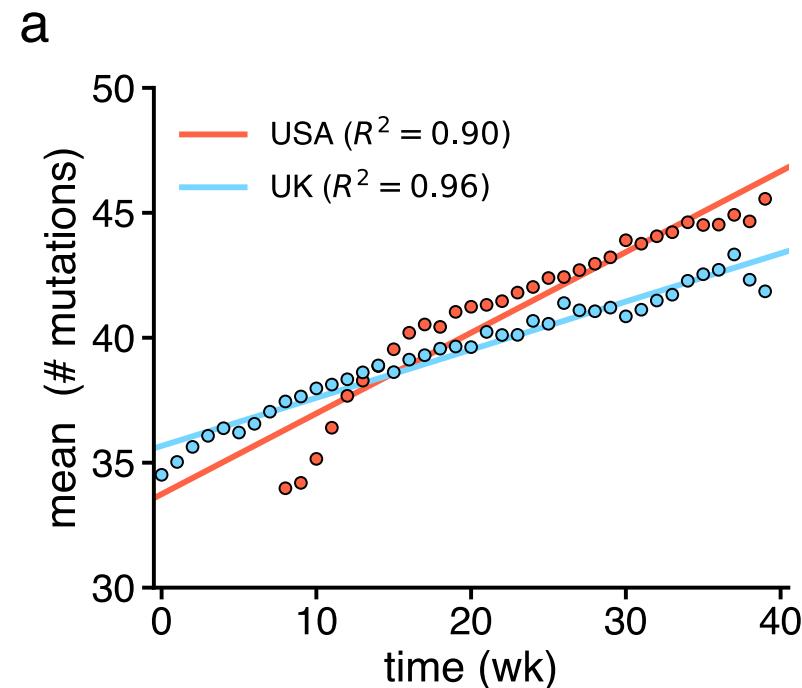
Output Flexibility

Supports standard formats for downstream
Bioinformatic processing: JSON, TSV

Graphical outputs for humans in PDF format

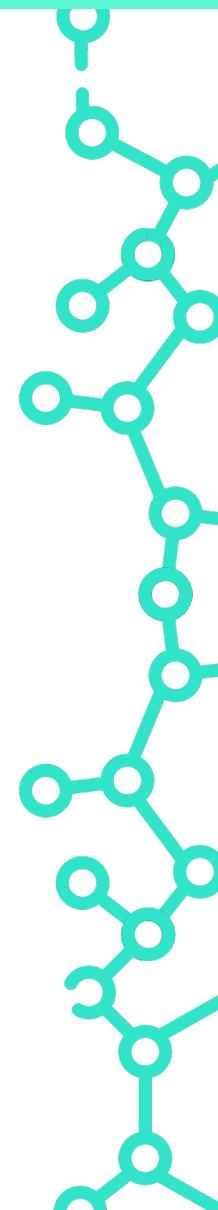


Validation: SARS-CoV-2 Alpha in the UK and the USA

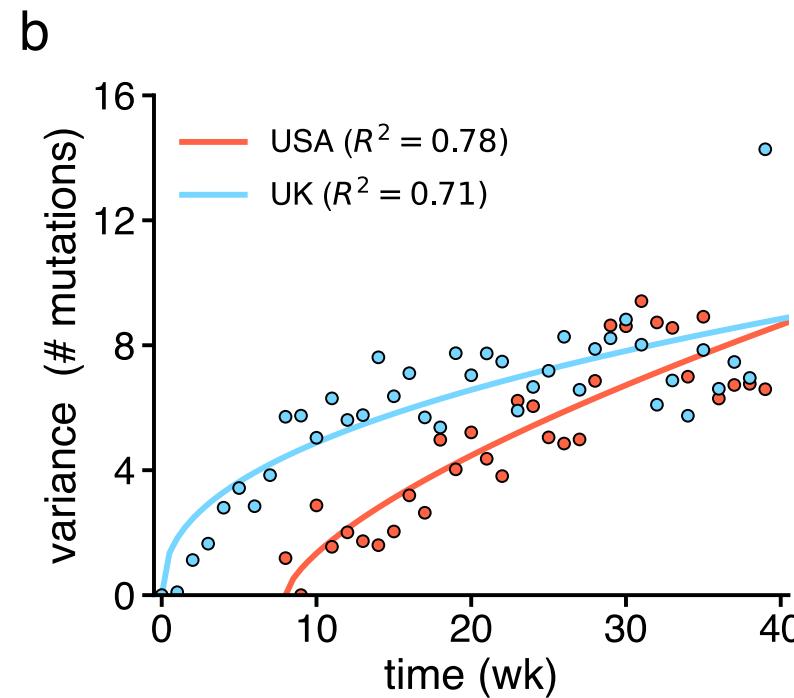
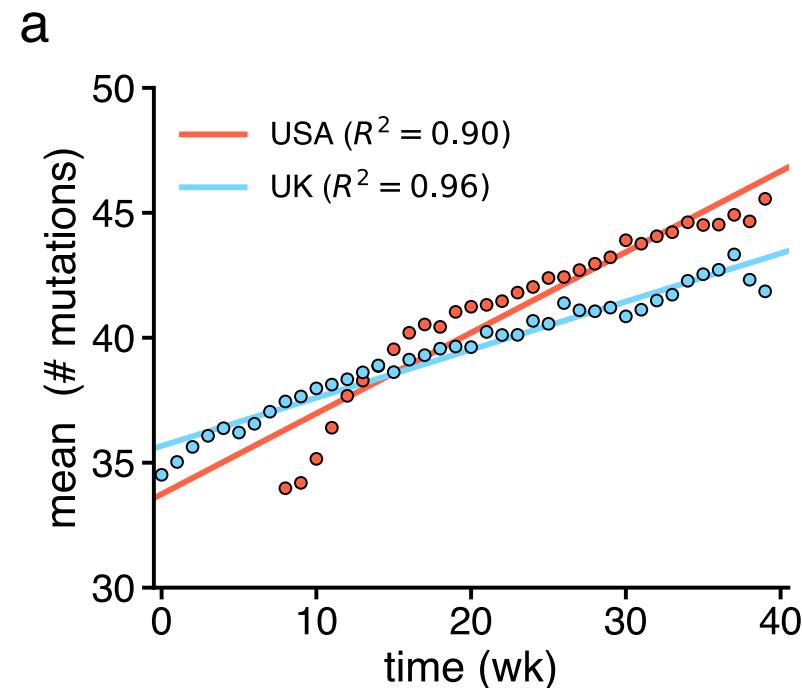


Validated on datasets > 10000
sequences

Processing time ~ 1 hour on a
standard personal computer



Validation: SARS-CoV-2 Alpha in the UK and the USA

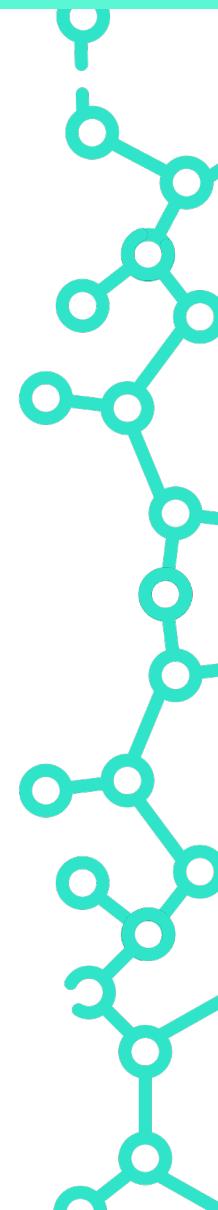


Validated on datasets > 10000 sequences

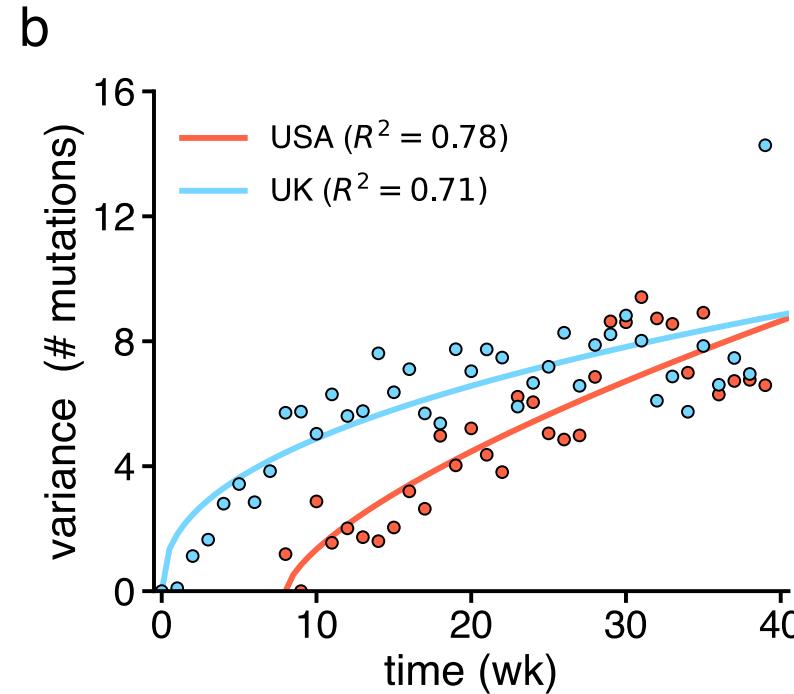
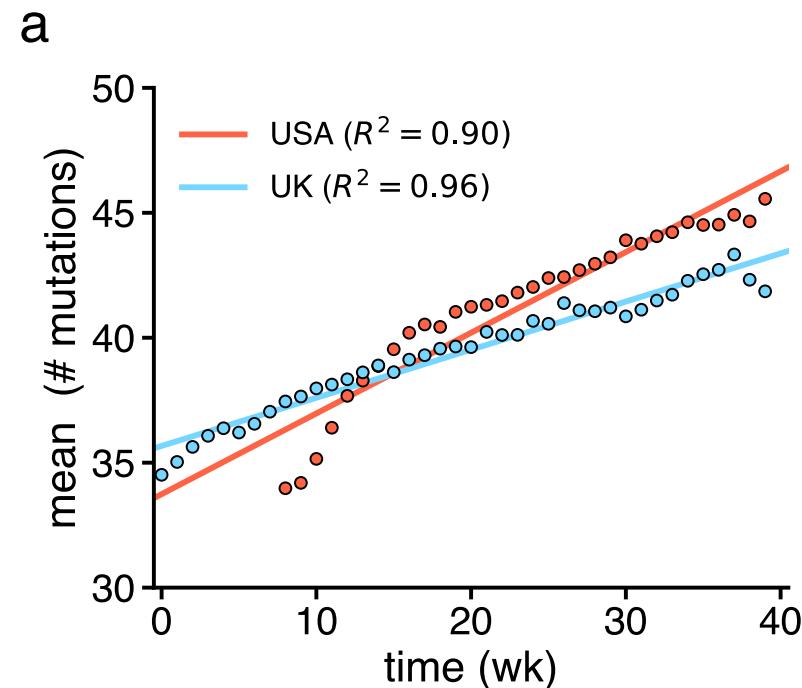
Processing time ~ 1 hour on a standard personal computer

Challenging model is often preferred

Subdiffusive nature of Alpha variant both in UK and USA



Validation: SARS-CoV-2 Alpha in the UK and the USA

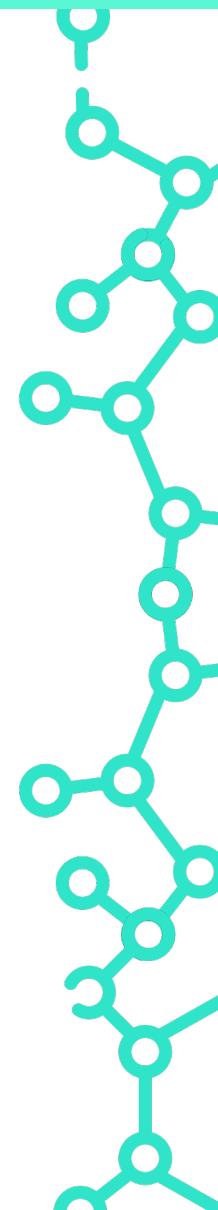


Validated on datasets > 10000 sequences

Processing time ~ 1 hour on a standard personal computer

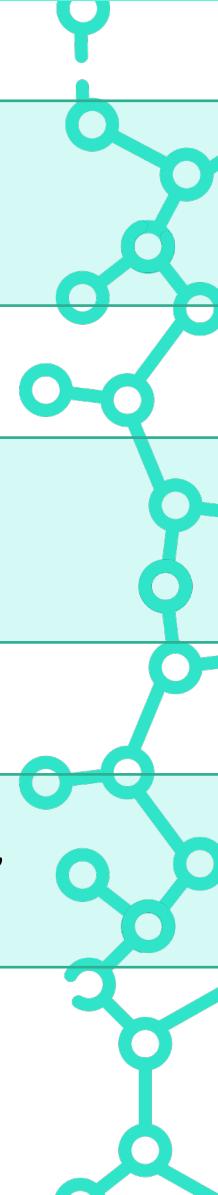
Challenging model is often preferred
Subdiffusive nature of Alpha variant both in UK and USA

Complements traditional phylogenetic approaches



Closing Remarks

- 1. Bypasses tree-based limitations:** PyEvoMotion models mutation accumulation as a stochastic process, avoiding computational bottlenecks of large-scale phylogenetics.
- 2. Statistical, variant-aware approach:** Captures non-uniform and overdispersed dynamics using Brownian and fractional Brownian models over custom time intervals.
- 3. Modular and scalable:** Despite current FFI-related performance limits, PyEvoMotion offers a flexible, open-source framework suited for pandemic-scale genomic analysis.

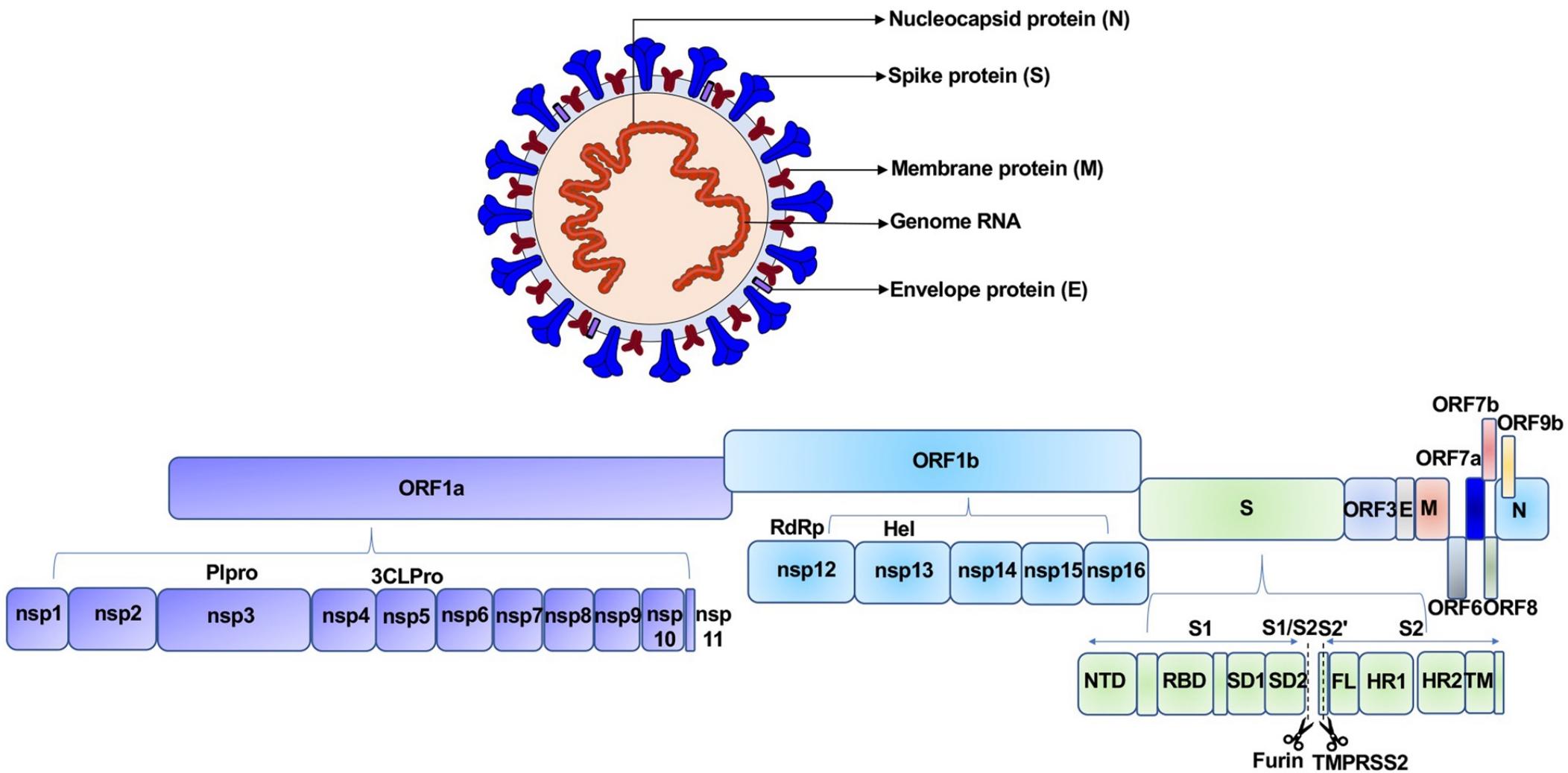


Chapter 5: Deciphering microscopic drivers of viral genome-scale molecular clock dynamics

Goiriz L, Rodrigo G. (2025)
Deciphering microscopic drivers of viral genome-scale molecular clock dynamics.
(In Preparation)



SARS-CoV-2 Genome Structure



(Adapted from Zhang et al. (2021))

Gene-Level Mutation Landscape

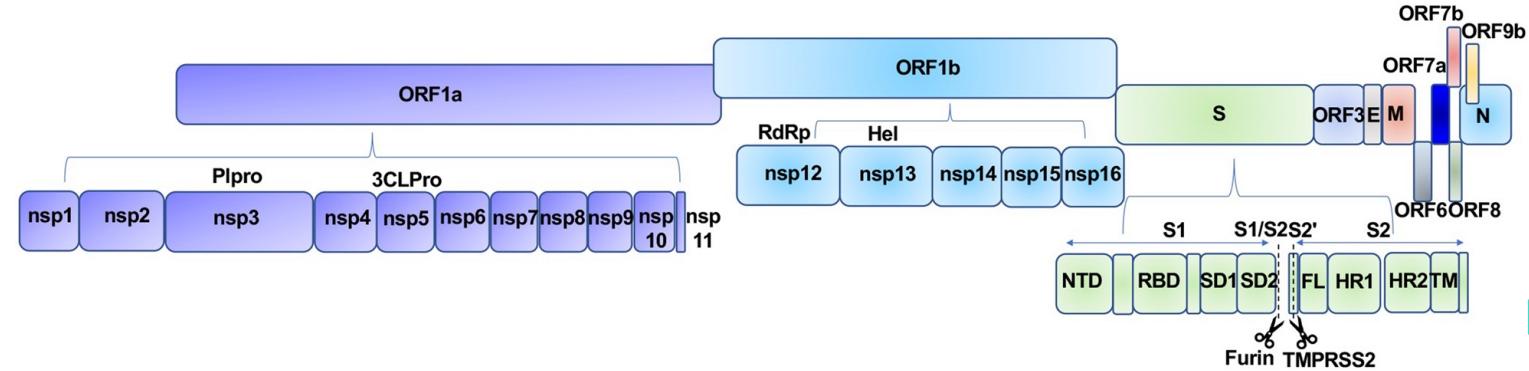
Analysis of SARS-CoV-2 gene-level mutation accumulation

27 genes analyzed

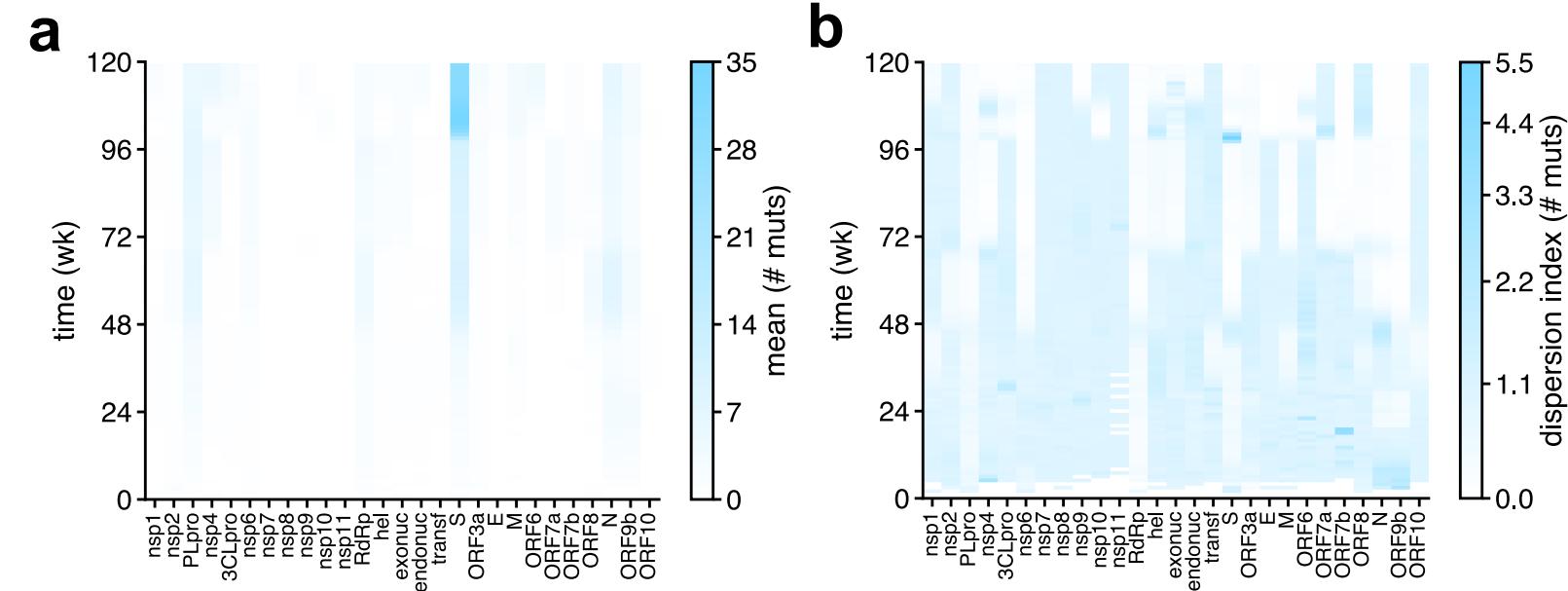
Most evolved genes:
PLpro, RdRp, S, N, ORF9b

Spike (S) diverged post-Omicron emergence

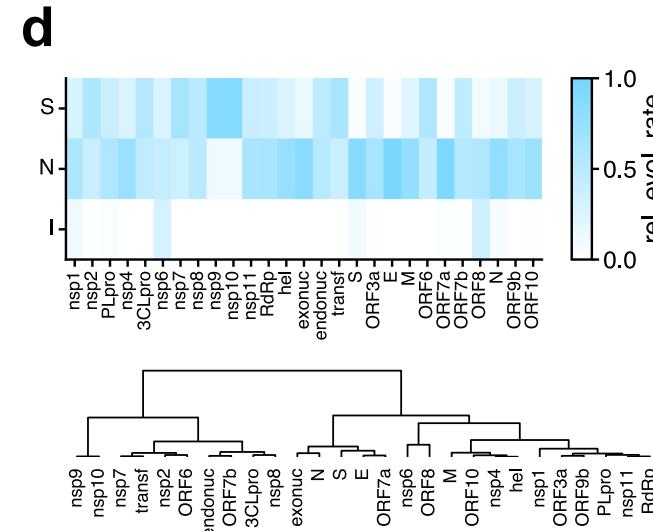
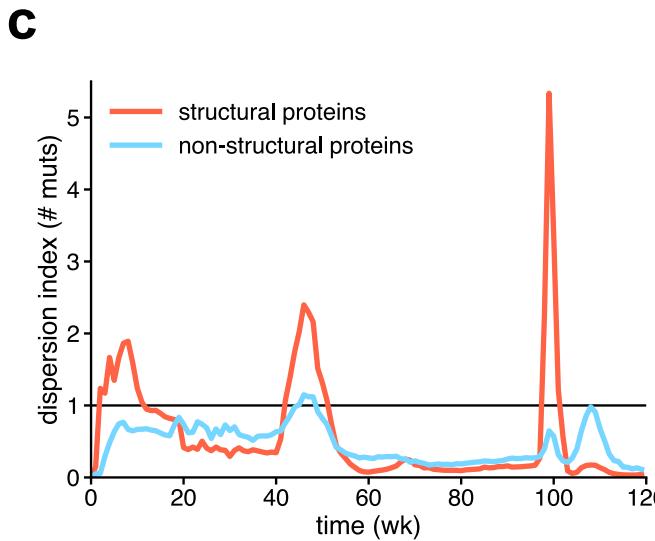
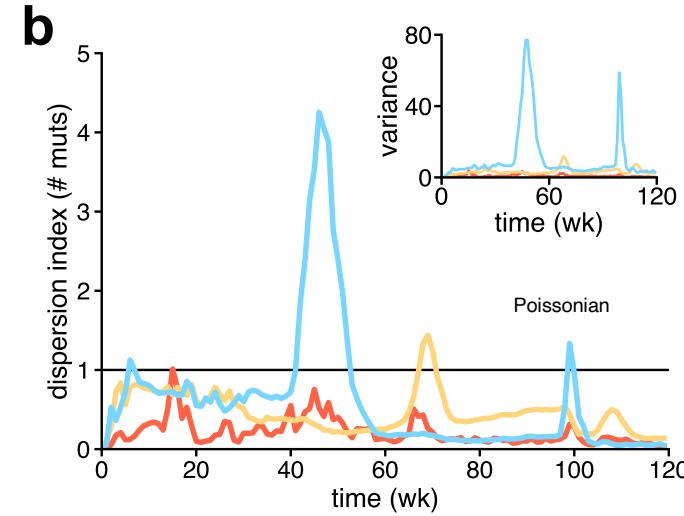
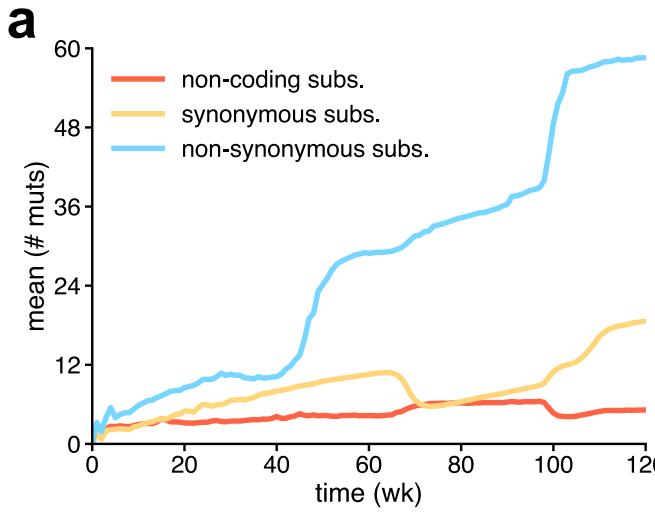
RdRp: low dispersion despite evolution



(Adapted from Zhang et al. (2021))



Gene-Level Mutation Dispersion



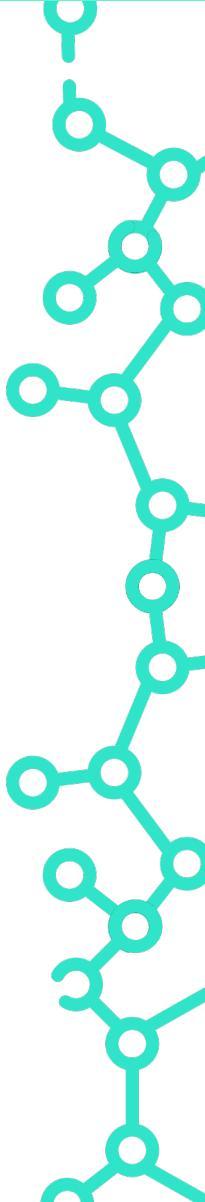
Adaptive (non-synonymous) genes:
S, E, N, ORF7a, exonuc.

Purifying selection: nsP9, nsP10
(synonymous)

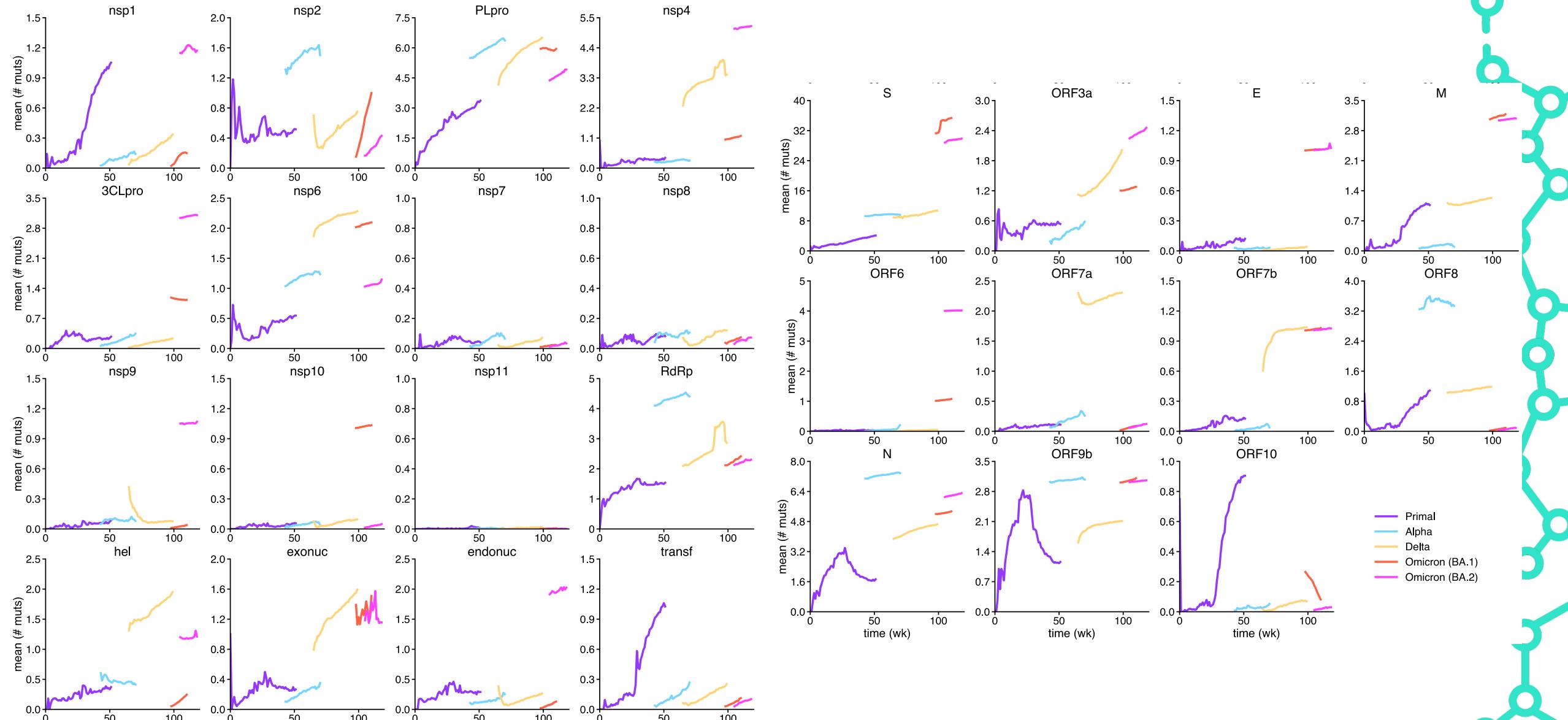
Alpha and Omicron: bursts of non-synonymous mutations

Delta: stable non-synonymous, drop
in synonymous mutations

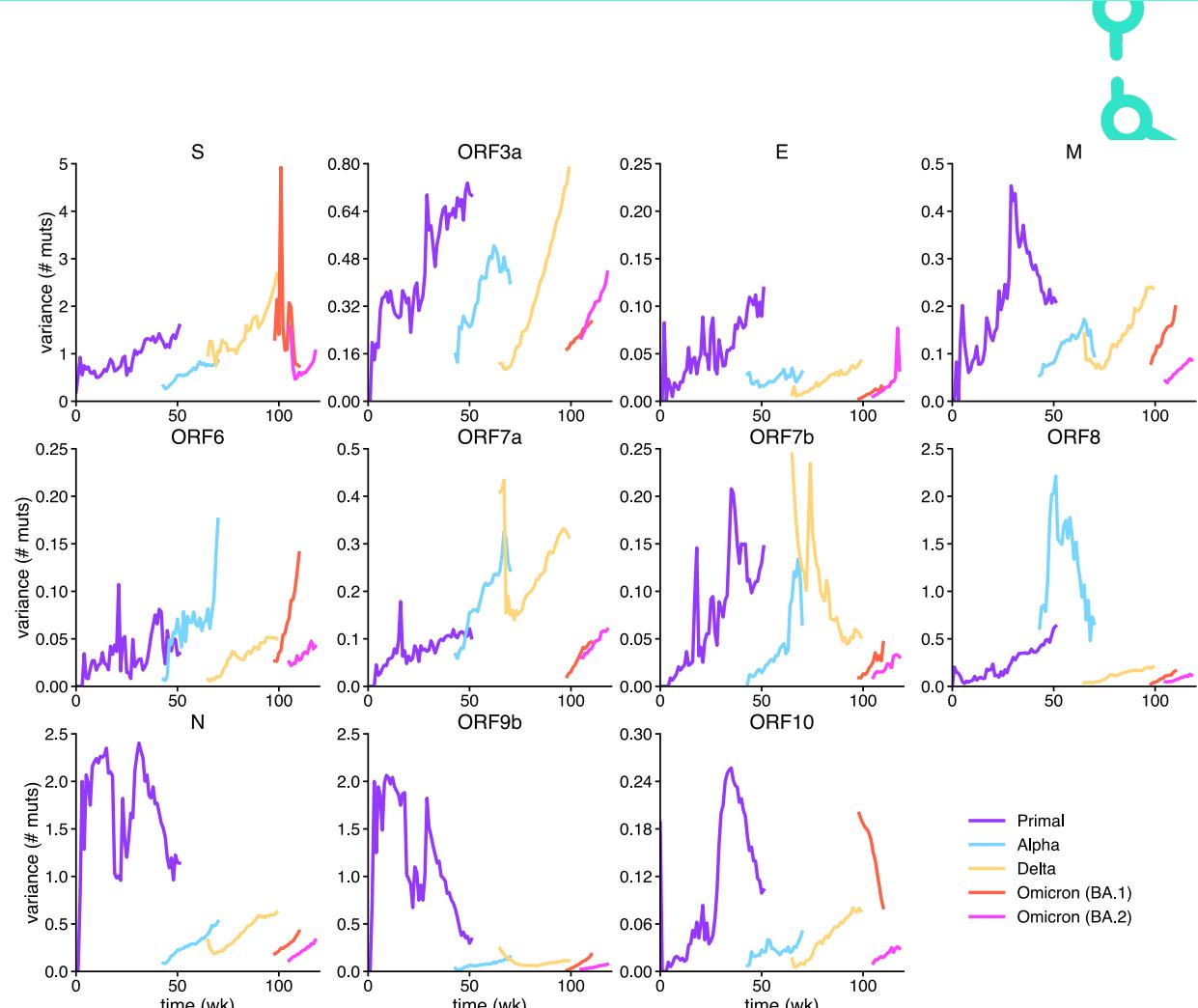
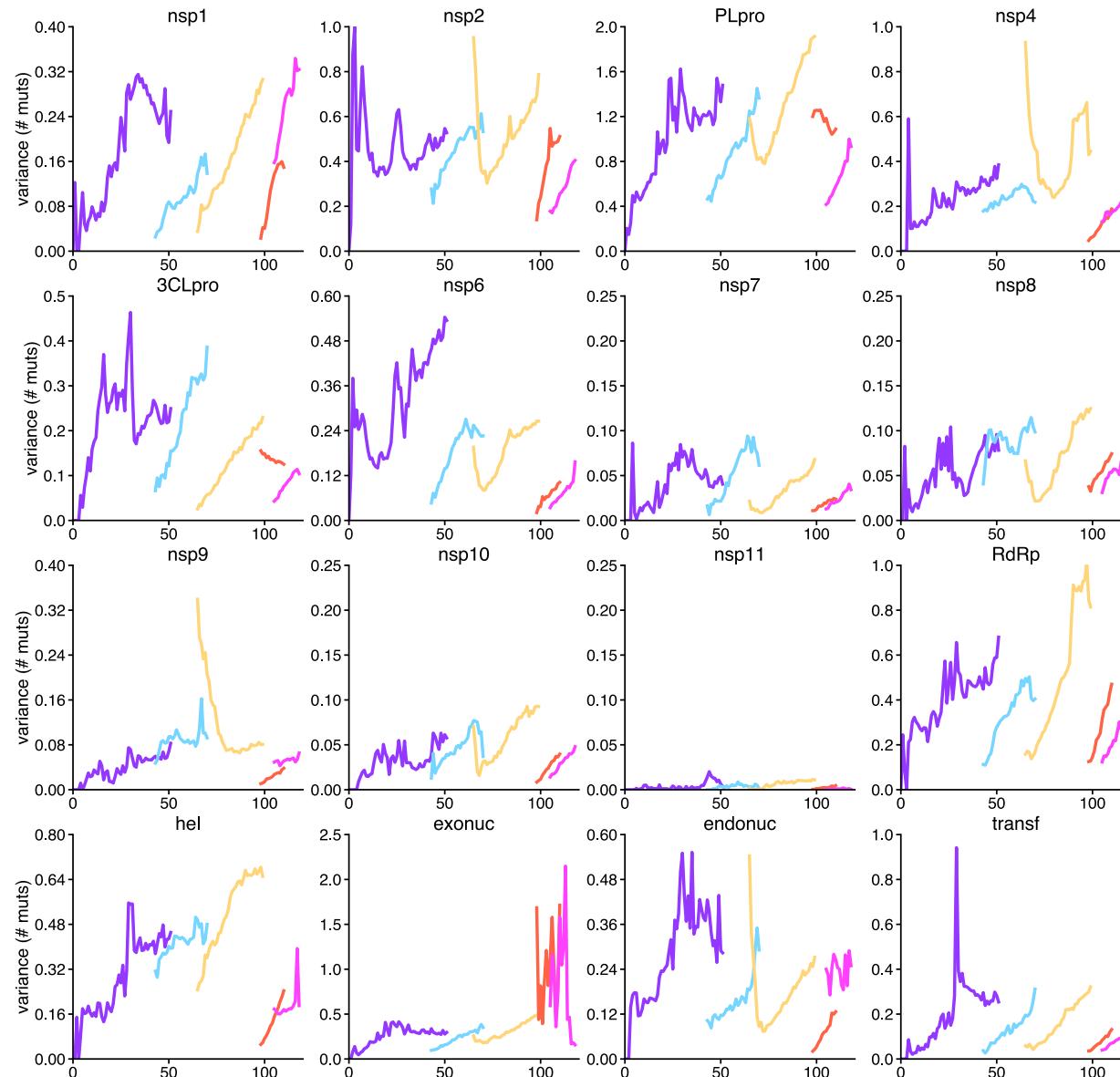
Peaks in dispersion index corroborate
episodic adaptation



Non-Linear Gene-Specific Mutation Accumulation

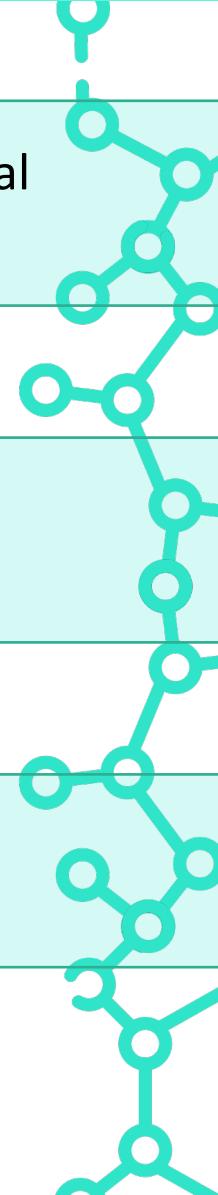


Variance Analysis Confirms Anomalous Diffusion

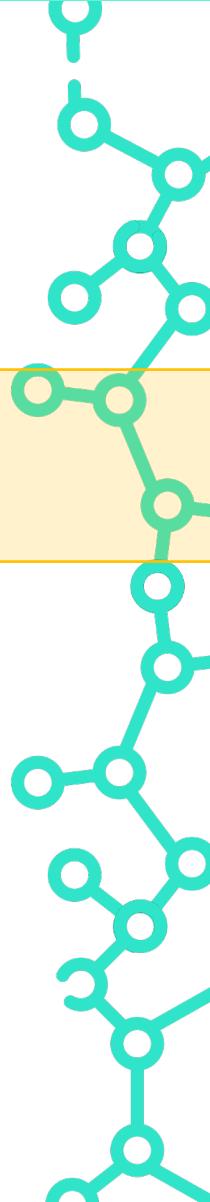


Closing Remarks

- 1. Gene-level dynamics matter:** Mutation patterns arise from lineage-specific forces acting on individual genes, beyond genome-wide averages.
- 2. Mosaic evolutionary behavior:** Some genes follow stochastic expectations, while others diverge due to selection or demography, reflecting diverse functional pressures.
- 3. Ongoing work:** Statistical descriptors like dispersion index and power-law fits uncover evolutionary subtleties, with further analyses underway.

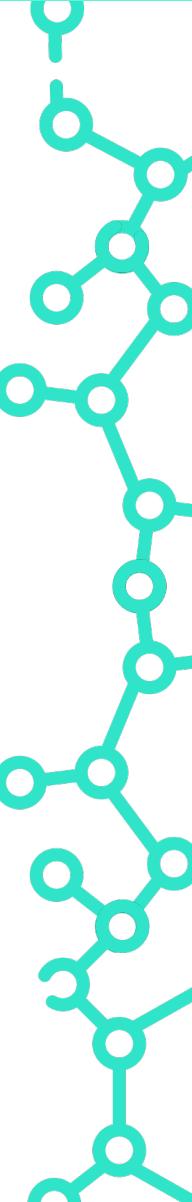
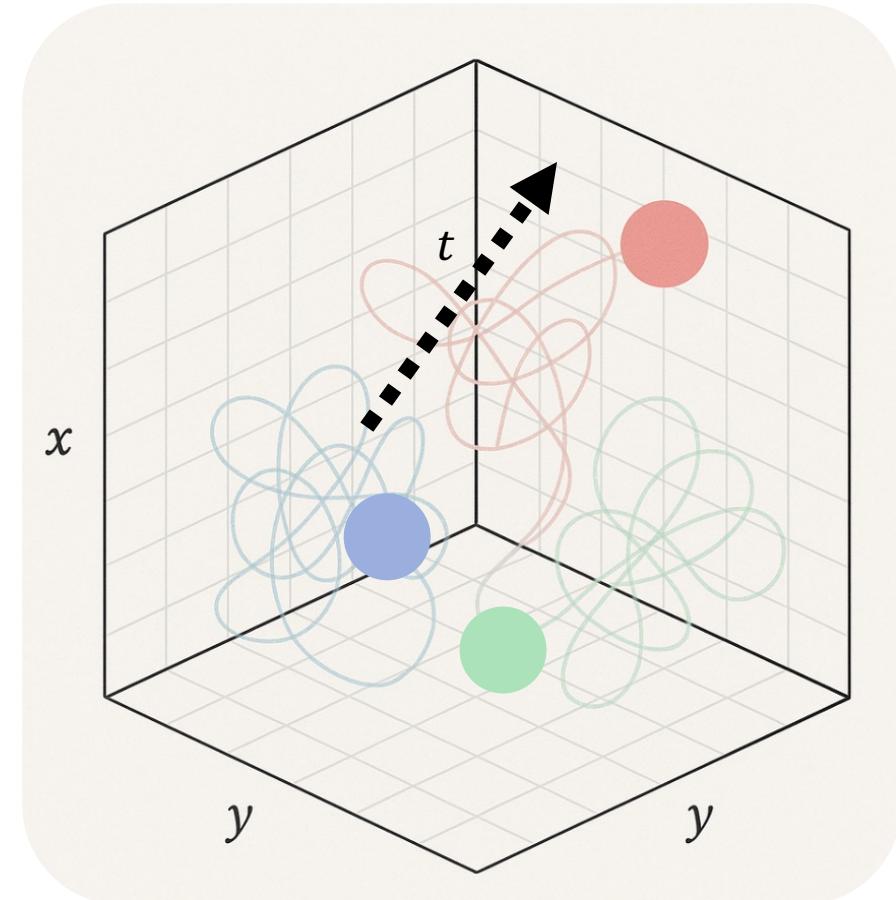


Discussion & Conclusions



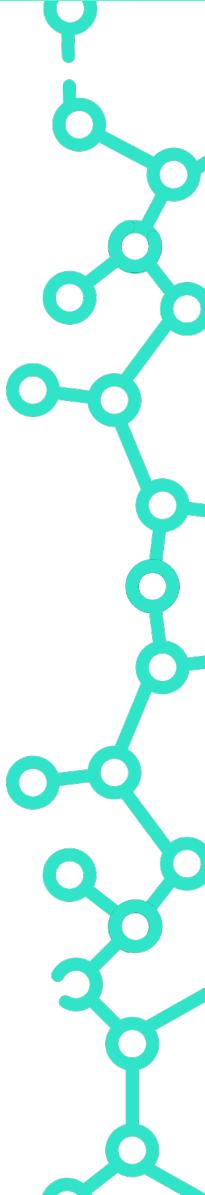
Discussion: Anomalous Diffusion in Viral Genome Evolution

- Subdiffusion (Primal, Alpha, Omicron) → slow mutation accumulation
- Superdiffusion (Delta) → rapid mutation bursts
- Fractional Brownian motion explains nonlinear variance
- Evolution behaves like a memory-influenced walk, not a simple Poisson process



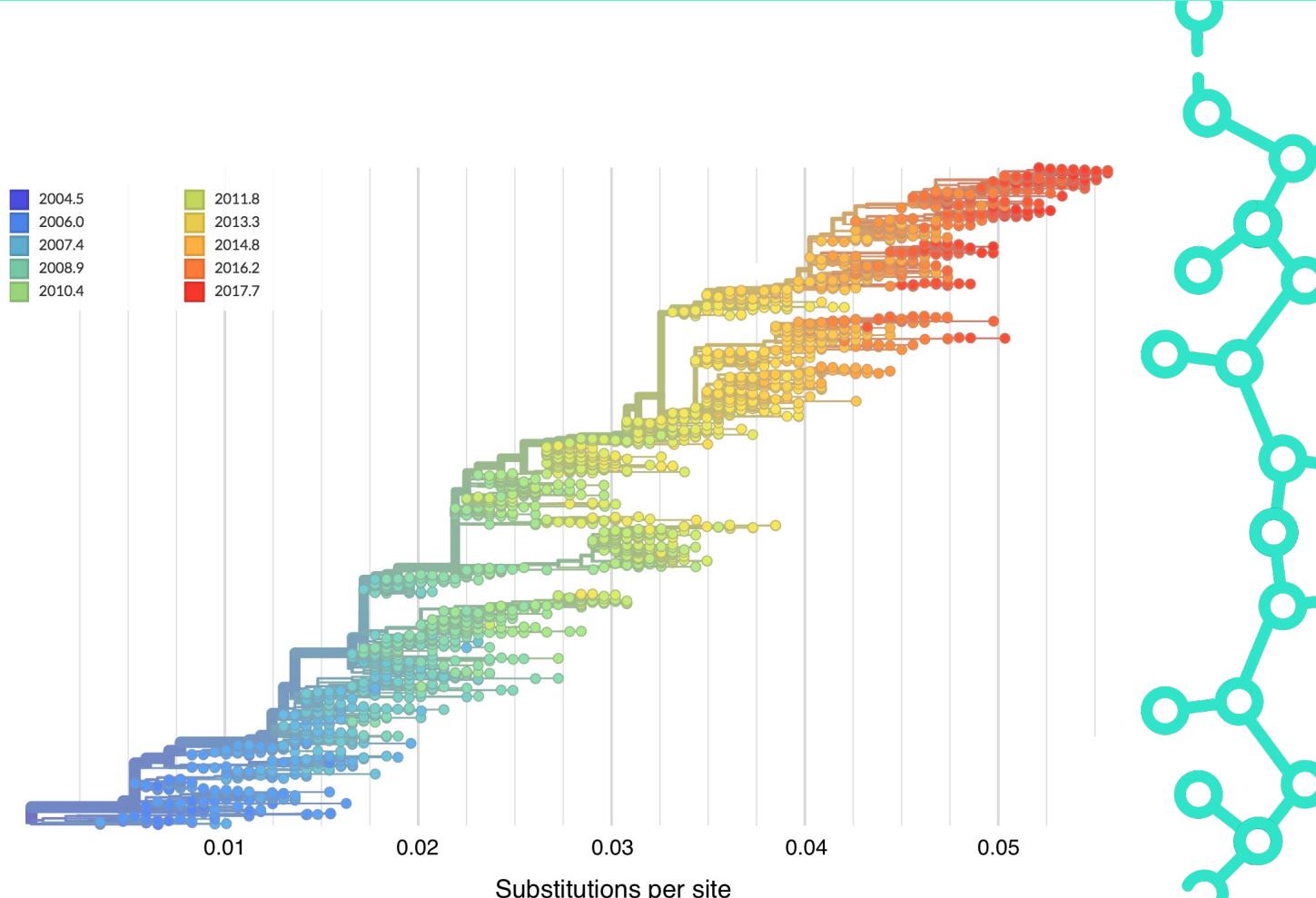
Discussion: Limitations of Modeling and Data

- Global sequence data is biased by geography, timing, and surveillance intensity
- A single diffusion exponent per variant oversimplifies biology
- Recombination (though rare) is not modeled and may confound signals
- Patterns ≠ mechanisms — results are descriptive, not mechanistic



Discussion: Integration

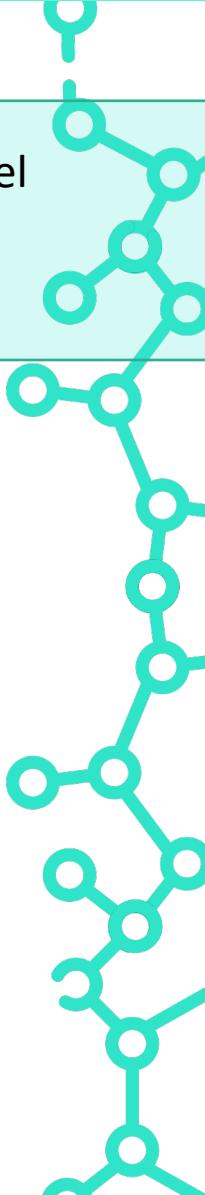
- Integrate fractional clocks into phylodynamic models
- Compare population-level vs. intrahost evolutionary dynamics
- Combine statistical insights with in vitro mutation assays
- Ultimately, unify data, experiments, and theory



(Adapted from Nextstrain (2025))

Conclusions

1. RNA Virus evolution can follow **variant-dependent, non-Brownian dynamics** best described by a stochastic model incorporating **fractional Brownian noise**, where anomalous diffusion exponents quantify **deviations** from the classical **molecular clock** through subdiffusive or superdiffusive evolution rates.



Conclusions

1. RNA Virus evolution can follow **variant-dependent, non-Brownian dynamics** best described by a stochastic model incorporating **fractional Brownian noise**, where anomalous diffusion exponents quantify **deviations** from the classical **molecular clock** through subdiffusive or superdiffusive evolution rates.
2. Variant-specific analyses reveal **distinct mutation accumulation trajectories** characterized by **bursts** in genetic **variance** during emergence and replacement events, yet converging toward **asymptotic Poissonian behavior** as captured by a reset-based dispersion index.



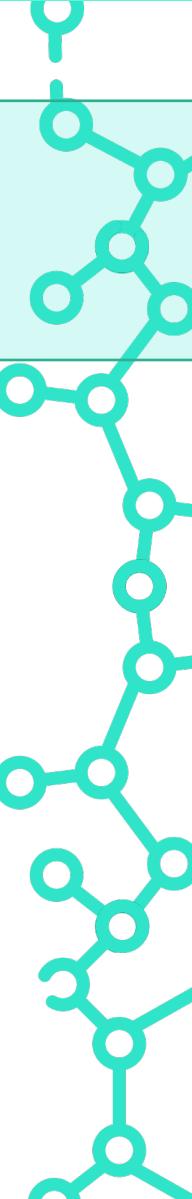
Conclusions

1. RNA Virus evolution can follow **variant-dependent, non-Brownian dynamics** best described by a stochastic model incorporating **fractional Brownian noise**, where anomalous diffusion exponents quantify **deviations** from the classical **molecular clock** through subdiffusive or superdiffusive evolution rates.
2. Variant-specific analyses reveal **distinct mutation accumulation trajectories** characterized by **bursts** in genetic **variance** during emergence and replacement events, yet converging toward **asymptotic Poissonian behavior** as captured by a reset-based dispersion index.
3. Mutation accumulation in RNA viruses exhibits strong **gene-level heterogeneity**, deviating from the uniform patterns predicted by classical molecular clock models.

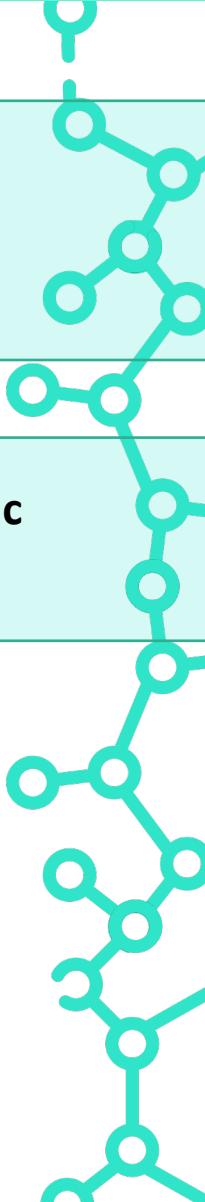


Conclusions

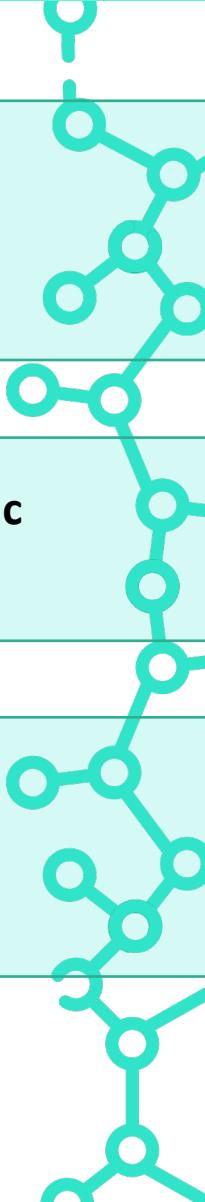
4. Genes under functional constraint exhibit **reduced variance and dispersion indicative of purifying selection**, genes involved in **host interaction** undergo **episodic adaptive bursts**, marked by high dispersion indices and non-monotonic mutation dynamics.



Conclusions

- 
4. Genes under functional constraint exhibit **reduced variance and dispersion indicative of purifying selection**, genes involved in **host interaction** undergo **episodic adaptive bursts**, marked by high dispersion indices and non-monotonic mutation dynamics.
 5. A top-down statistical analysis enables detection of **localized evolutionary dynamics**, revealing how **gene-specific contributions shape genome-scale anomalous diffusion**.

Conclusions

- 
4. Genes under functional constraint exhibit **reduced variance and dispersion indicative of purifying selection**, genes involved in **host interaction** undergo **episodic adaptive bursts**, marked by high dispersion indices and non-monotonic mutation dynamics.
 5. A top-down statistical analysis enables detection of **localized evolutionary dynamics**, revealing how **gene-specific contributions shape genome-scale anomalous diffusion**.
 6. Non-phylogenetic, population-level approaches offer scalable alternatives for modeling large viral datasets, underscoring the importance of implementing and distributing these novel models as open-source software to ensure broad accessibility and reproducibility.

BioSystems Design at Rodrigolab



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Departamento de
Matemática Aplicada



Stochastic dynamic modeling of evolving gene expression programs

All the effects of Nature are only the mathematical consequences of a small number of immutable laws

— Pierre-Simon Laplace