

MASTER'S THESIS WIP

A COMPREHENSIVE STUDY OF THE PHYLOGENETICS OF SARS-COV-2 IN EUROPE



Cecilia Valenzuela

cEvo Group Meeting, Jan 5, 2020

**FIRST, PREPARE AND BE READY.
SECOND, DETECT, PROTECT AND TREAT.
THIRD, REDUCE TRANSMISSION.
FOURTH, INNOVATE AND LEARN.**

**WHO Director-General's opening
remarks at the media briefing on
COVID-19 – 11 March 2020**

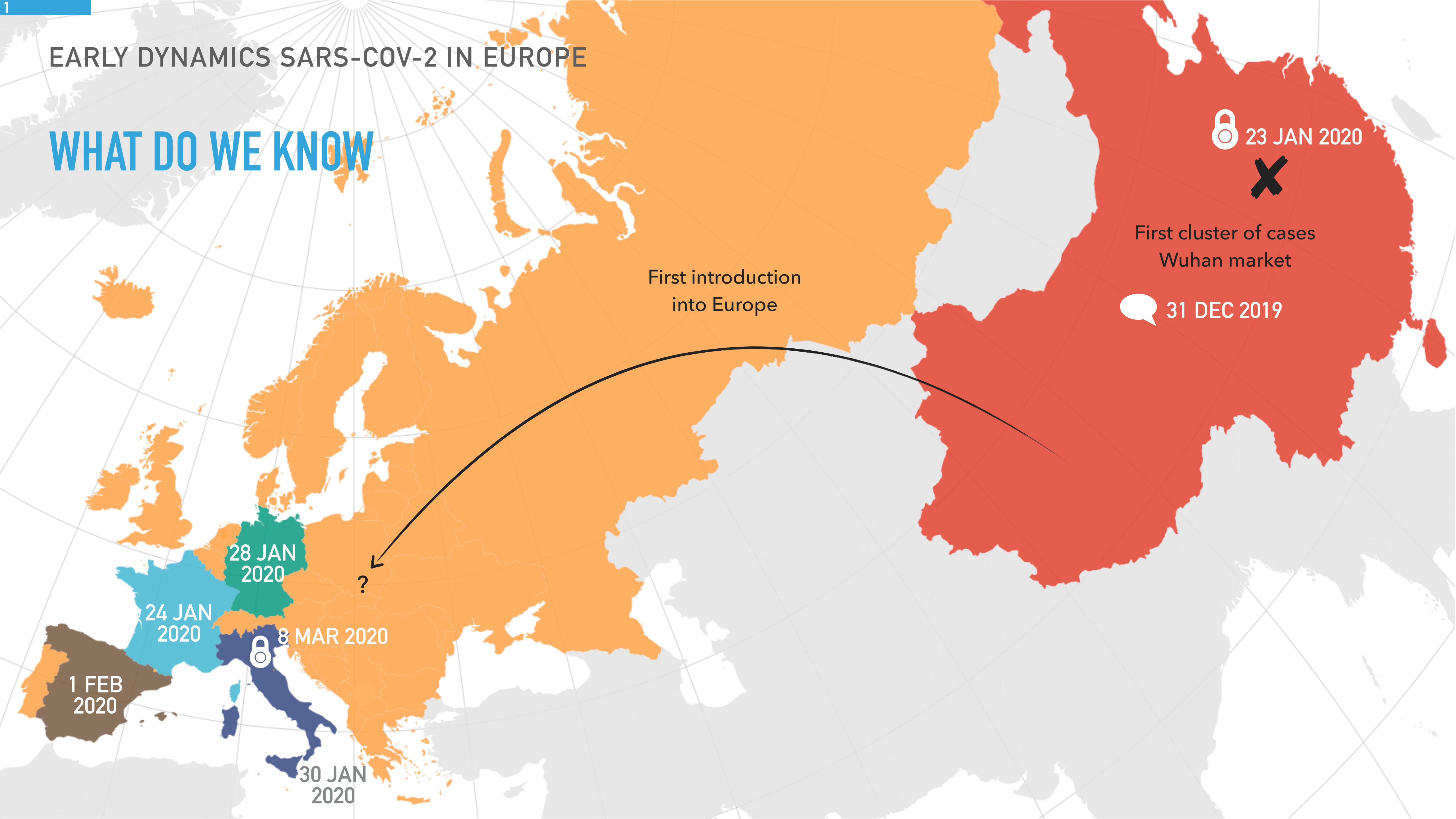
**HOW MANY CASES DID
WE MISS?**

**HOW FAST DID WE
REACT?**

**HOW WAS THE SPATIAL
SPREAD IN EUROPE?**

EARLY DYNAMICS SARS-COV-2 IN EUROPE

WHAT DO WE KNOW



EARLY DYNAMICS SARS-COV-2 IN EUROPE

INITIAL ANALYSIS



The origin and early spread of SARS-CoV-2 in Europe

Sarah A. Nadeau, Timothy G. Vaughan, Jérémie Sciré, Jana S. Huisman, Tanja Stadler

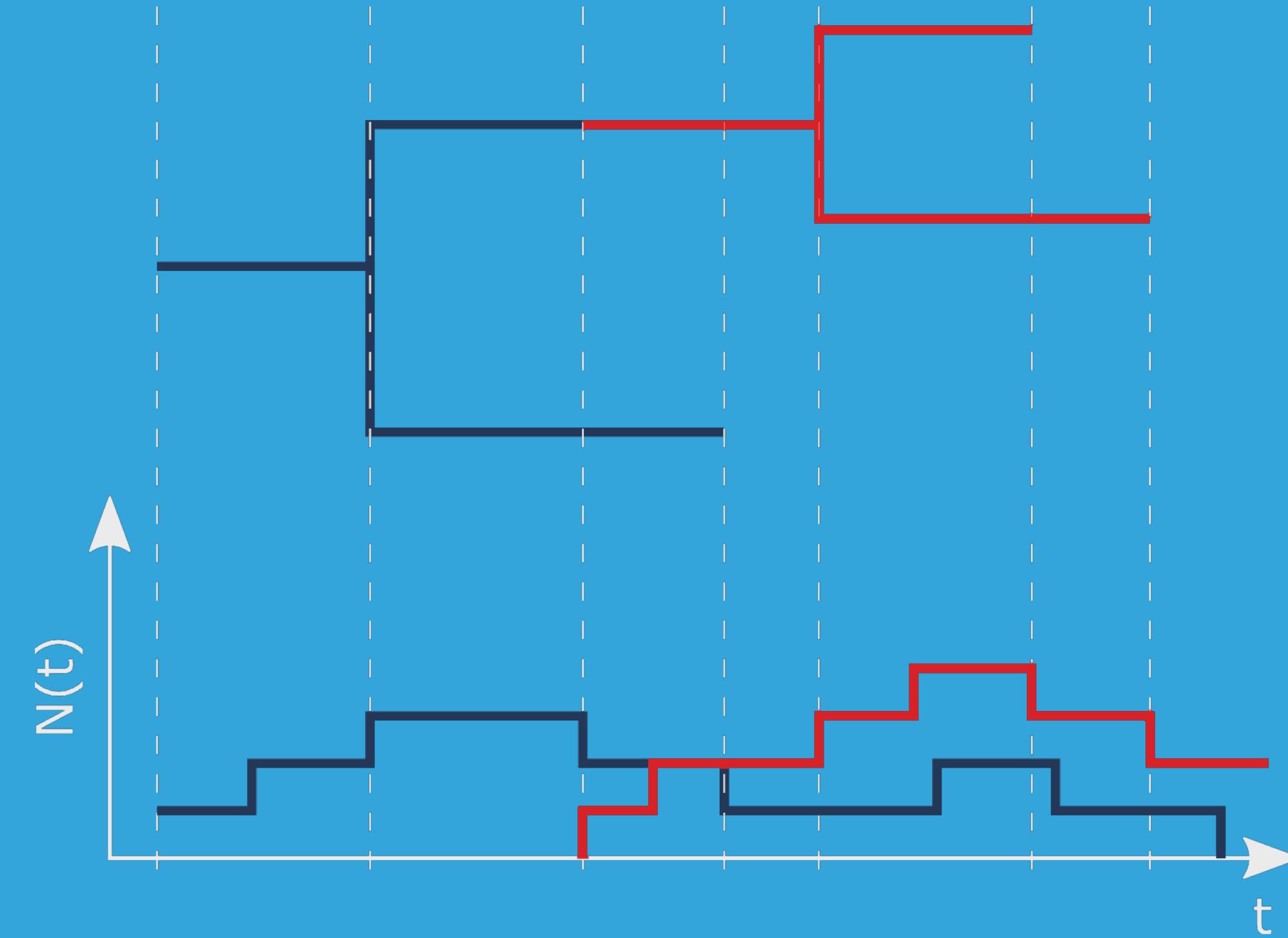
doi: <https://doi.org/10.1101/2020.06.10.20127738>

Significance Statement We estimate the origin and spread of SARS-CoV-2 in Europe prior to border closures based on viral genome sequences using a phylogenetic model with geographic structure. We confirm that the predominant European outbreak most likely started in Italy and spread from there. This outbreak was probably seeded by a transmission event in either Hubei or Germany. In particular, we find that before the first border closures in Europe, the rate of new cases occurring from within-country transmission was within the estimated bounds on the rate of new migration cases.

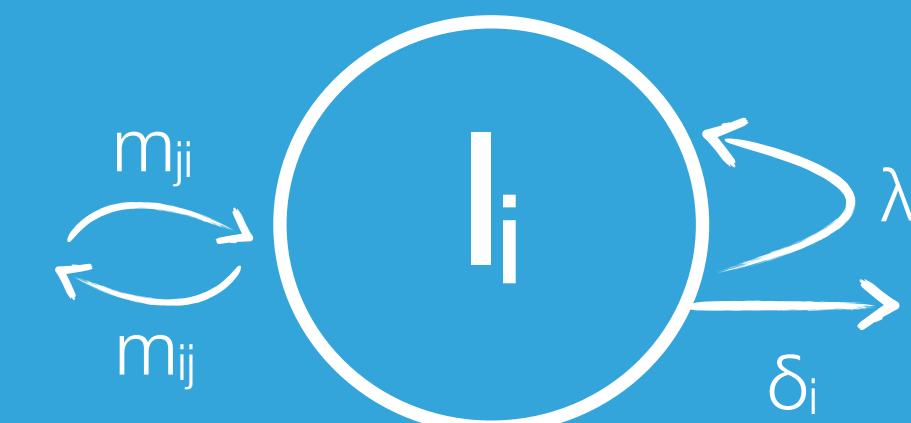
- + Add new sequences
- + Use travel and geographic data for migration rates
- + Infer multi-type epidemic trajectories with BDMM-Prime

BDMM-PRIME EPIDEMIC TRAJECTORIES

(Tim's work)



Multi-type Birth Death Model



One trajectory for each each set of parameters + coloured tree = for each step in the MCMC

PHYLOGENETIC ANALYSIS

SEQUENCE DATA

POPULATION MODEL

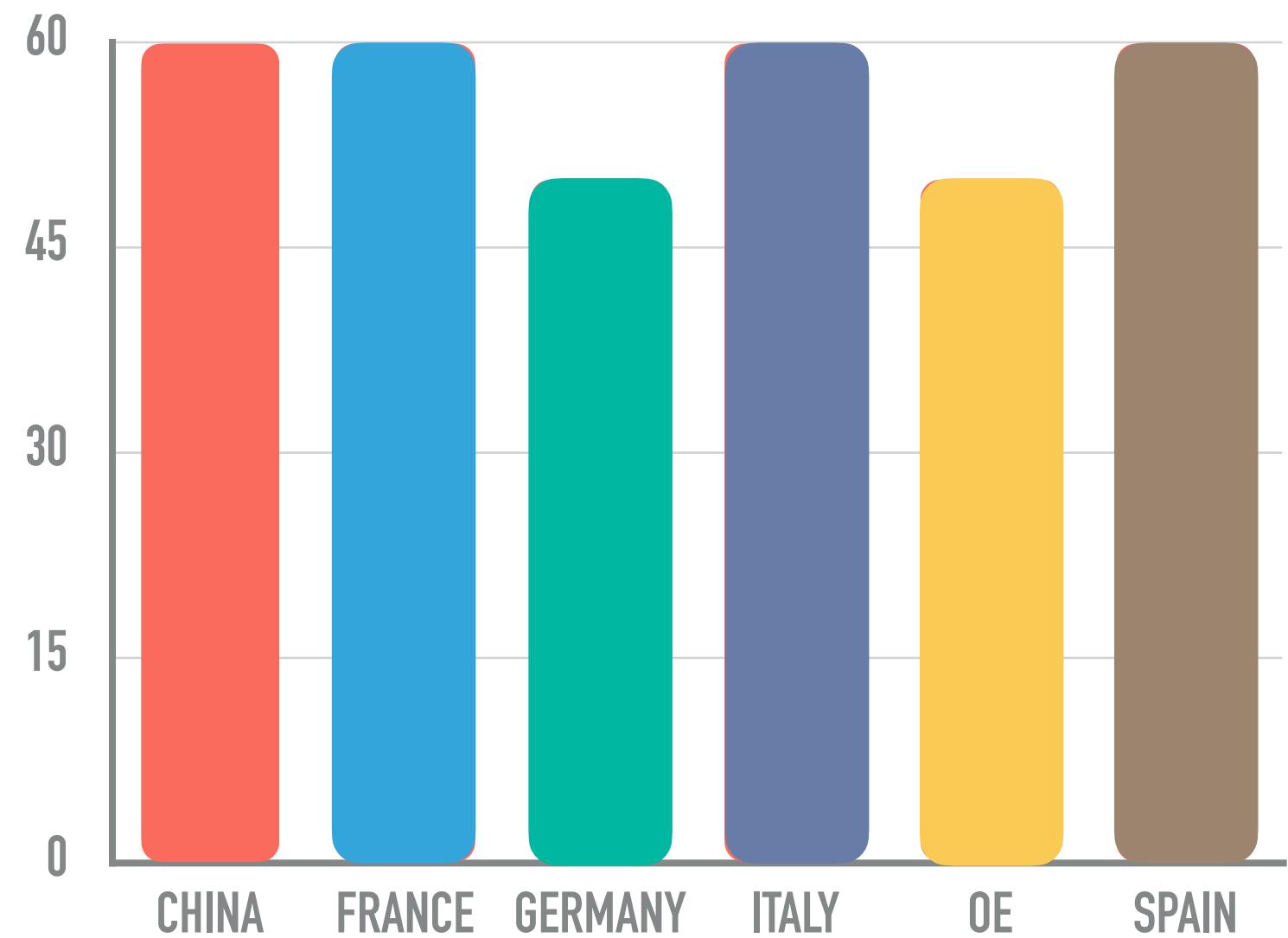
SUBSTITUTION MODEL

CLOCK MODEL

SEQUENCE PROCESSING

- ▶ Nextstrain pipeline and Sarah's excluded sequences

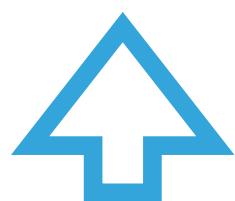
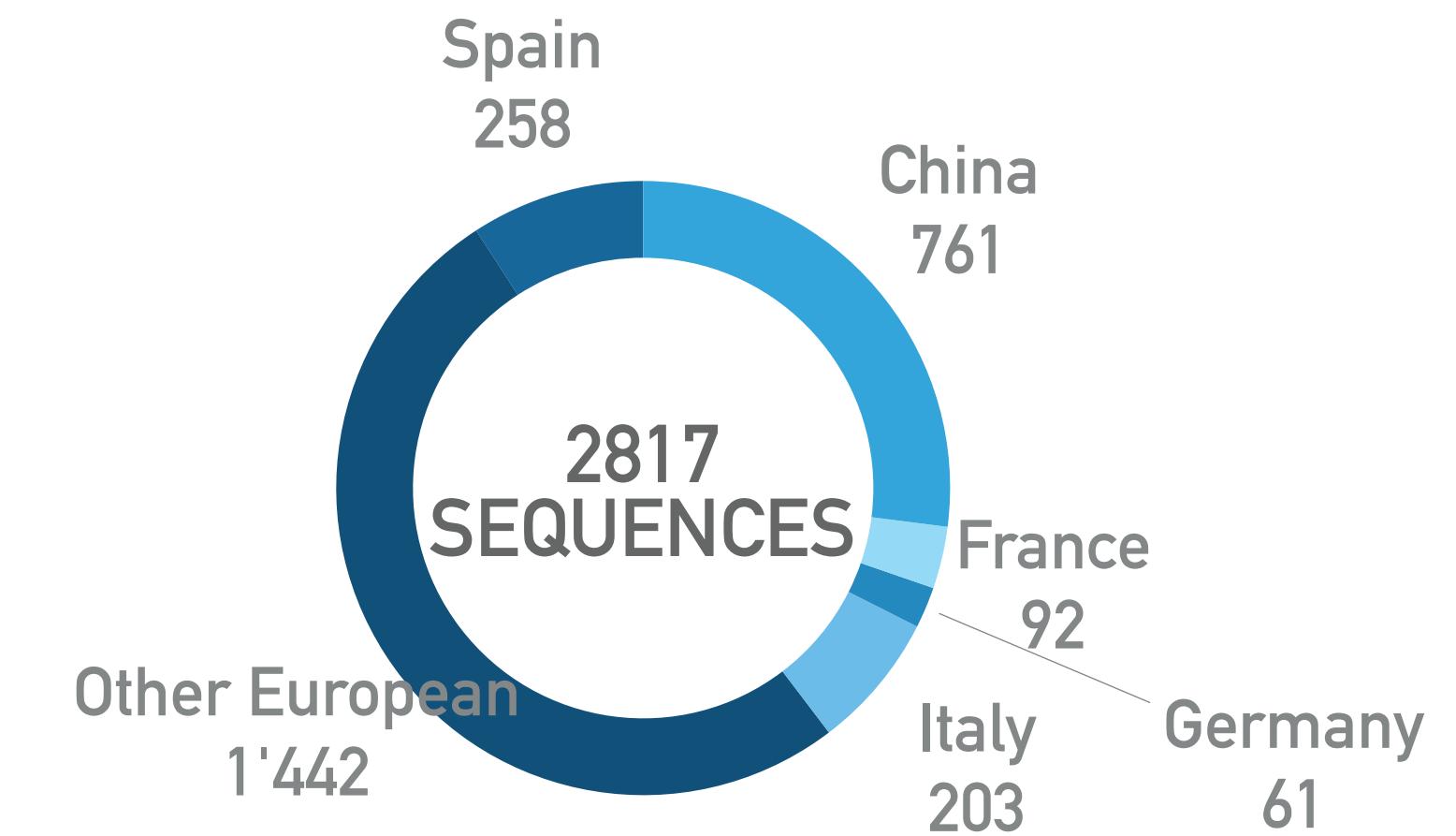
Nextstrain build for novel coronavirus, nextstrain/ncov (Sept 2020)



SUBSAMPLING

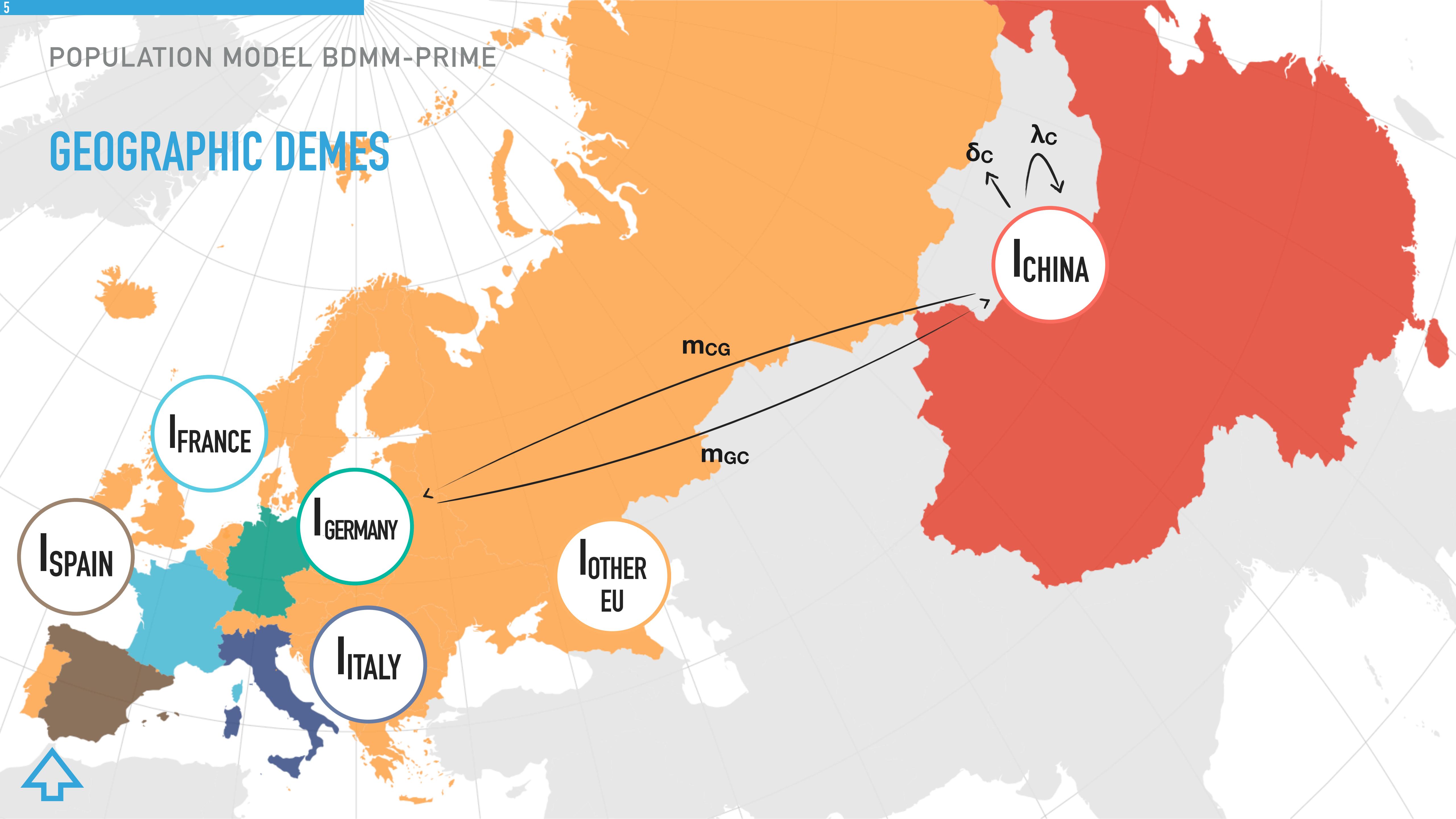
- ▶ Assumption: constant sampling proportion
- ▶ Subsampling scheme - proportional to daily cases

340 sequences



POPULATION MODEL BDMM-PRIME

GEOGRAPHIC DEMES

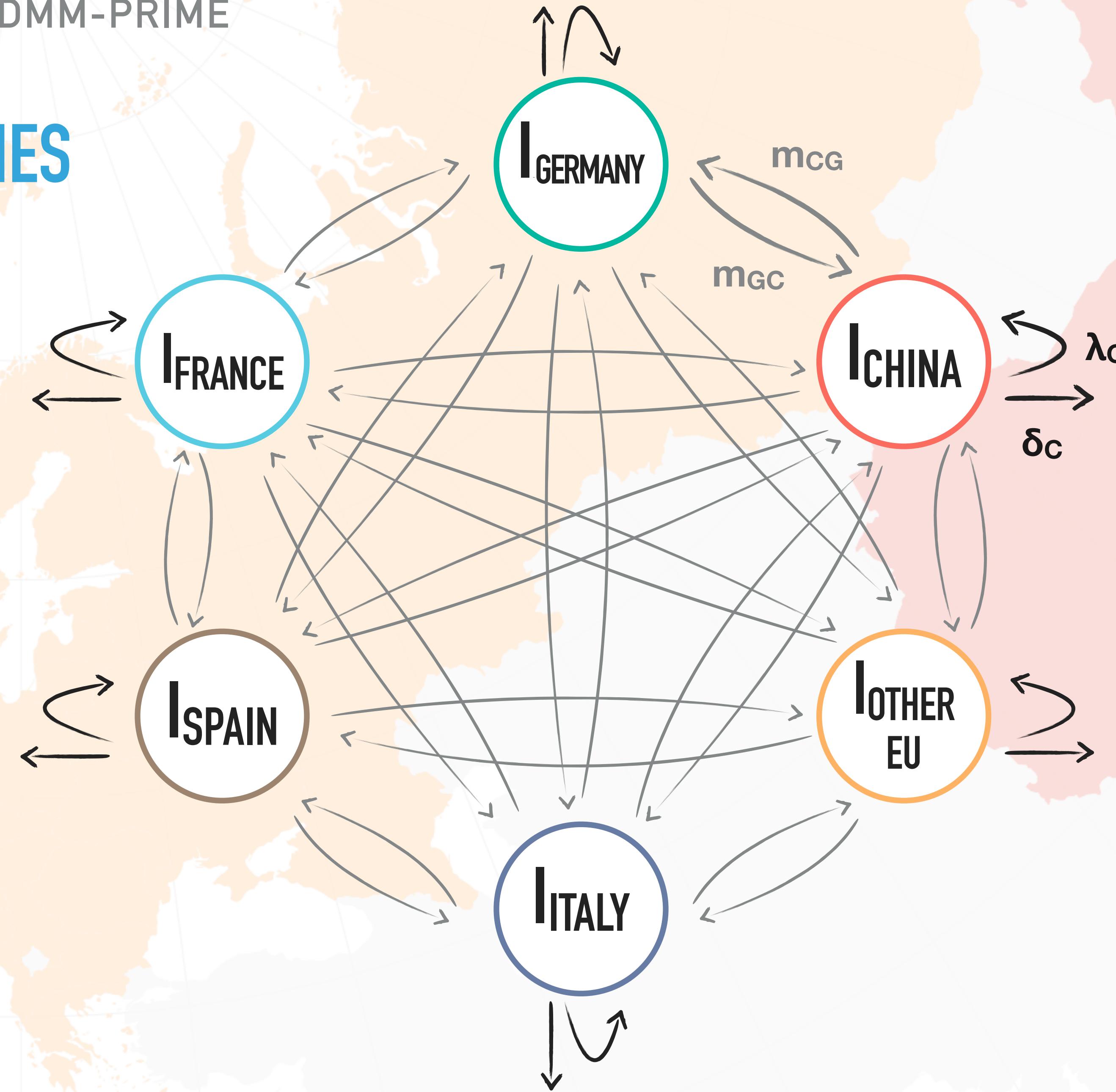


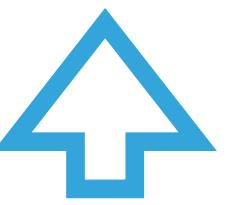
POPULATION MODEL BDMM-PRIME

GEOGRAPHIC DEMES

Epi parameterisation

- ▶ $R_{0i} = \lambda_i / \delta_i$
- ▶ Sampling proportion p_i
- ▶ Migration rates m_{ij}





TIMELINE

 R_0 p

China

Europe

 m

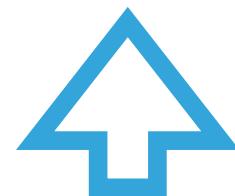
EPOCH 1

EPOCH 2

EPOCH 3

**First sample**First available sequence
from Wuhan**Hubei Lockdown**Decrease migration
out of ChinaChange in flight
passenger dataChange in flight
passenger data**Lombardy Lockdown**

Last sample



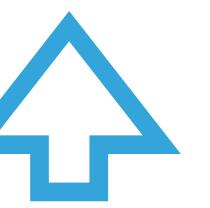
Origin December 23

January 23

February 1

March 1

March 8

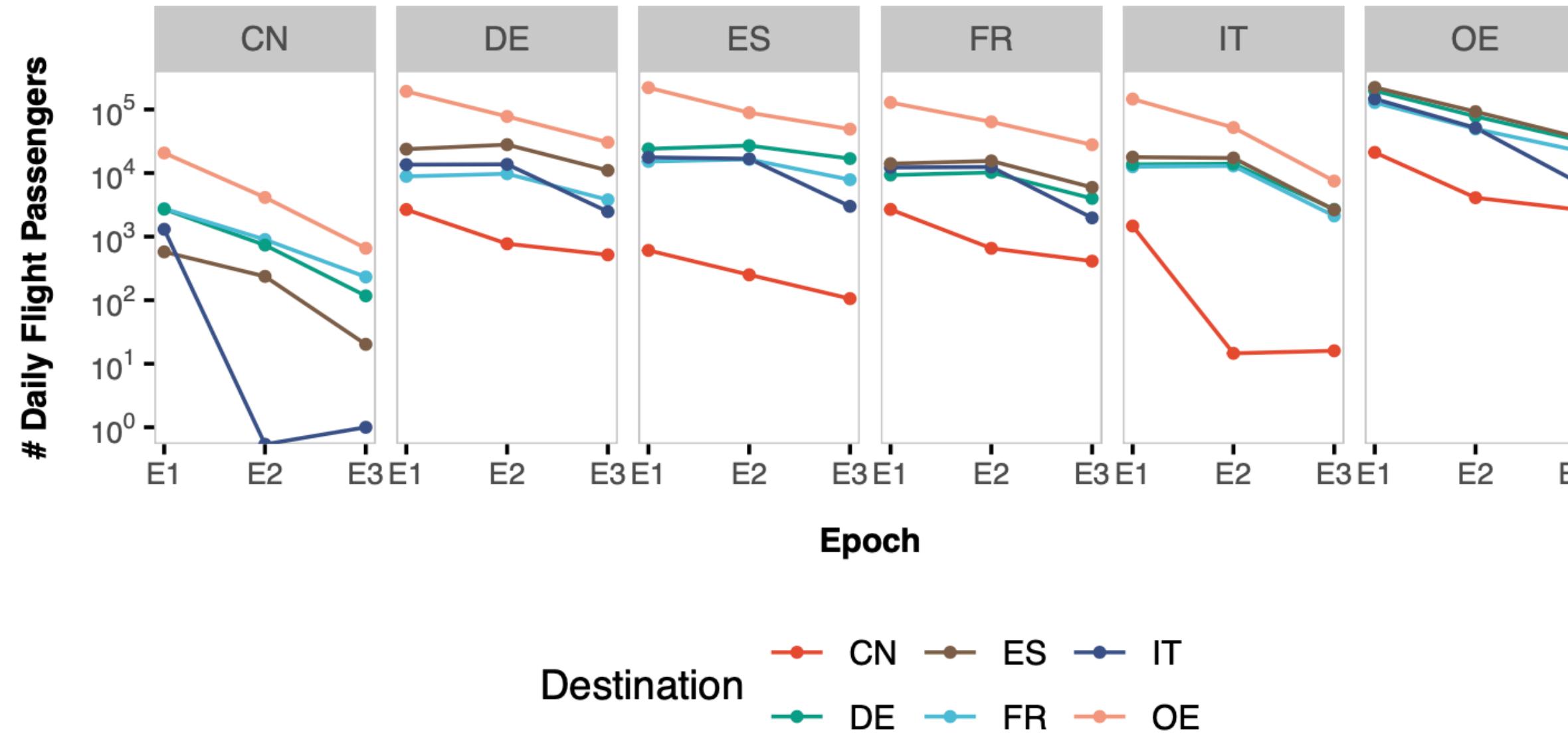


GLM MIGRATION RATES - WIP

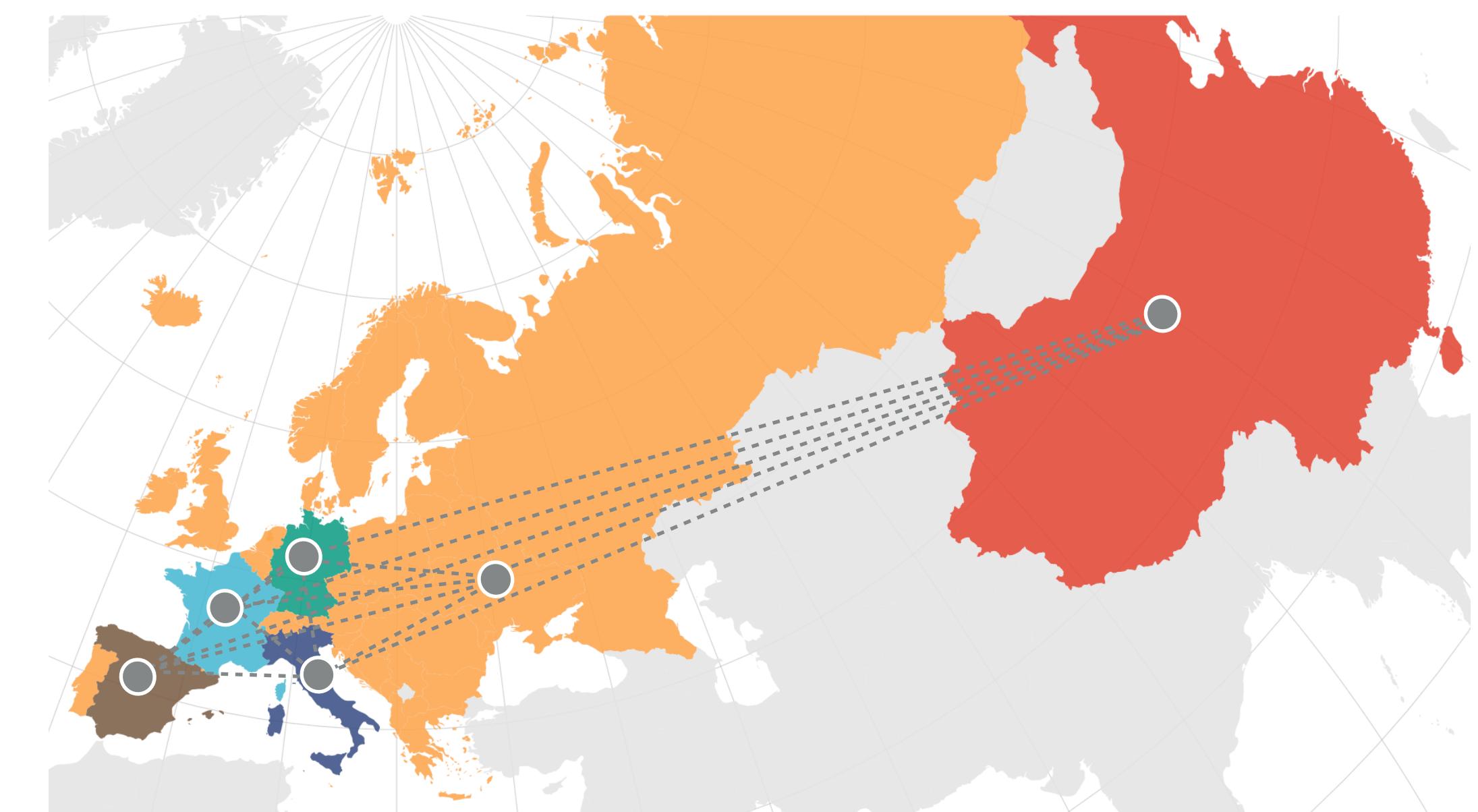
Lemey et al, 2014

$$m_{ij} = c \exp (\beta_{\text{flight passengers}} \text{airplane icon } ij + \beta_{\text{geo distance}} \text{globe icon } ij)$$

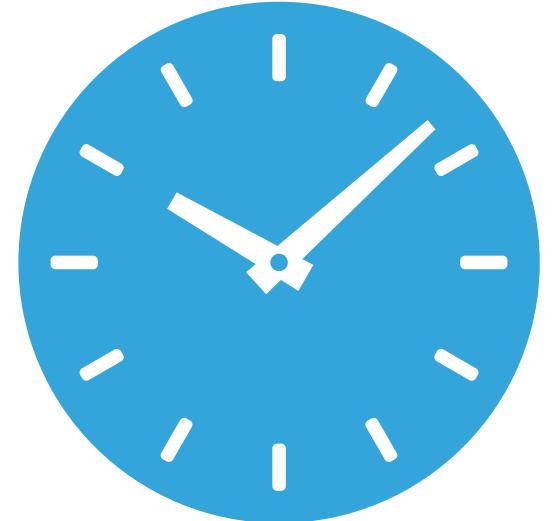
Daily flight passengers between countries, Eurostats



Great circle pairwise distance between centroids



PHYLOGENETIC ANALYSIS

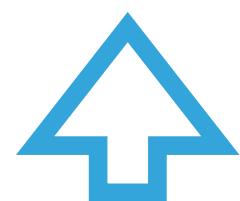
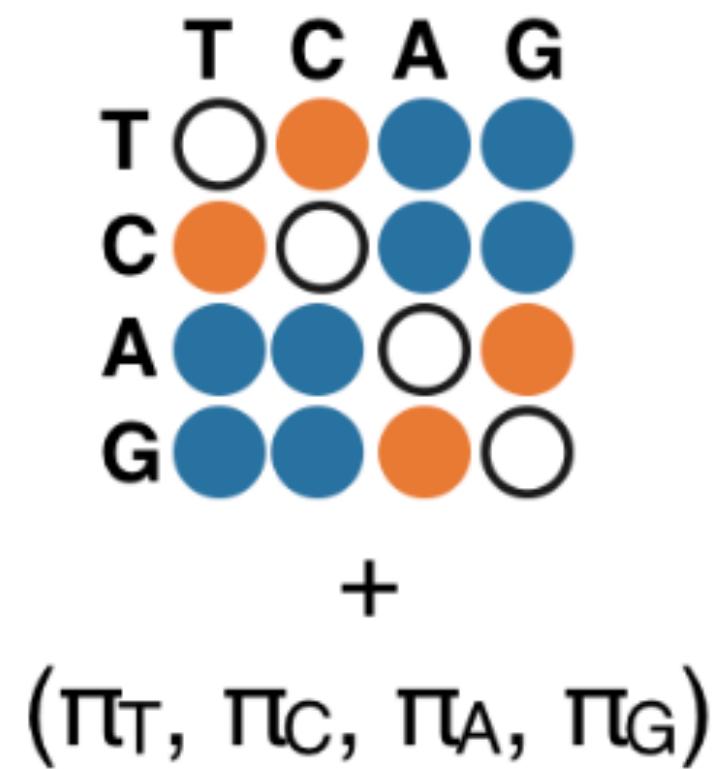


MOLECULAR CLOCK MODEL

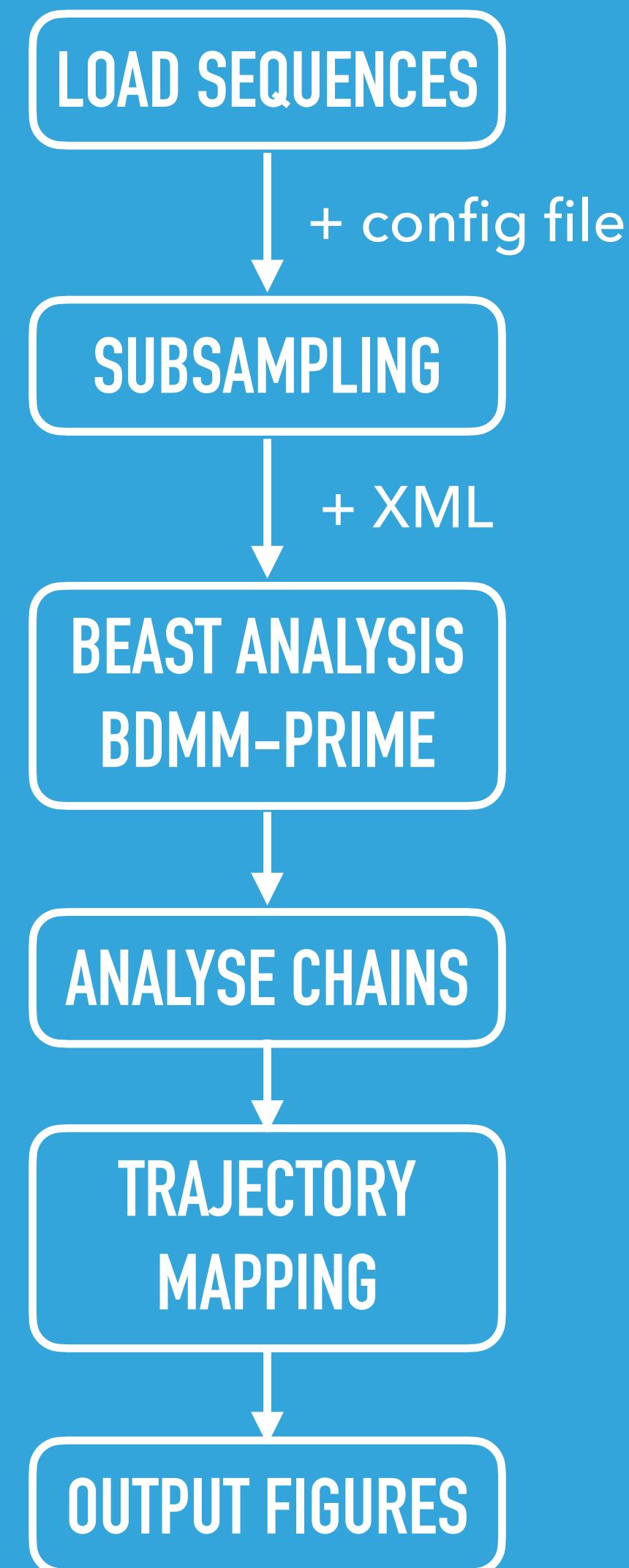
- ▶ Fixed strict clock model 0.0008, approx 24 mutations/year
Nextstrain build for novel coronavirus, nextstrain/ncov

NUCLEOTIDE SUBSTITUTION MODEL

- ▶ HKY model
- ▶ 4 Gamma rate categories
- ▶ Empirical frequencies



REPRODUCIBLE WORKFLOW



snakemake
A framework for reproducible data analysis

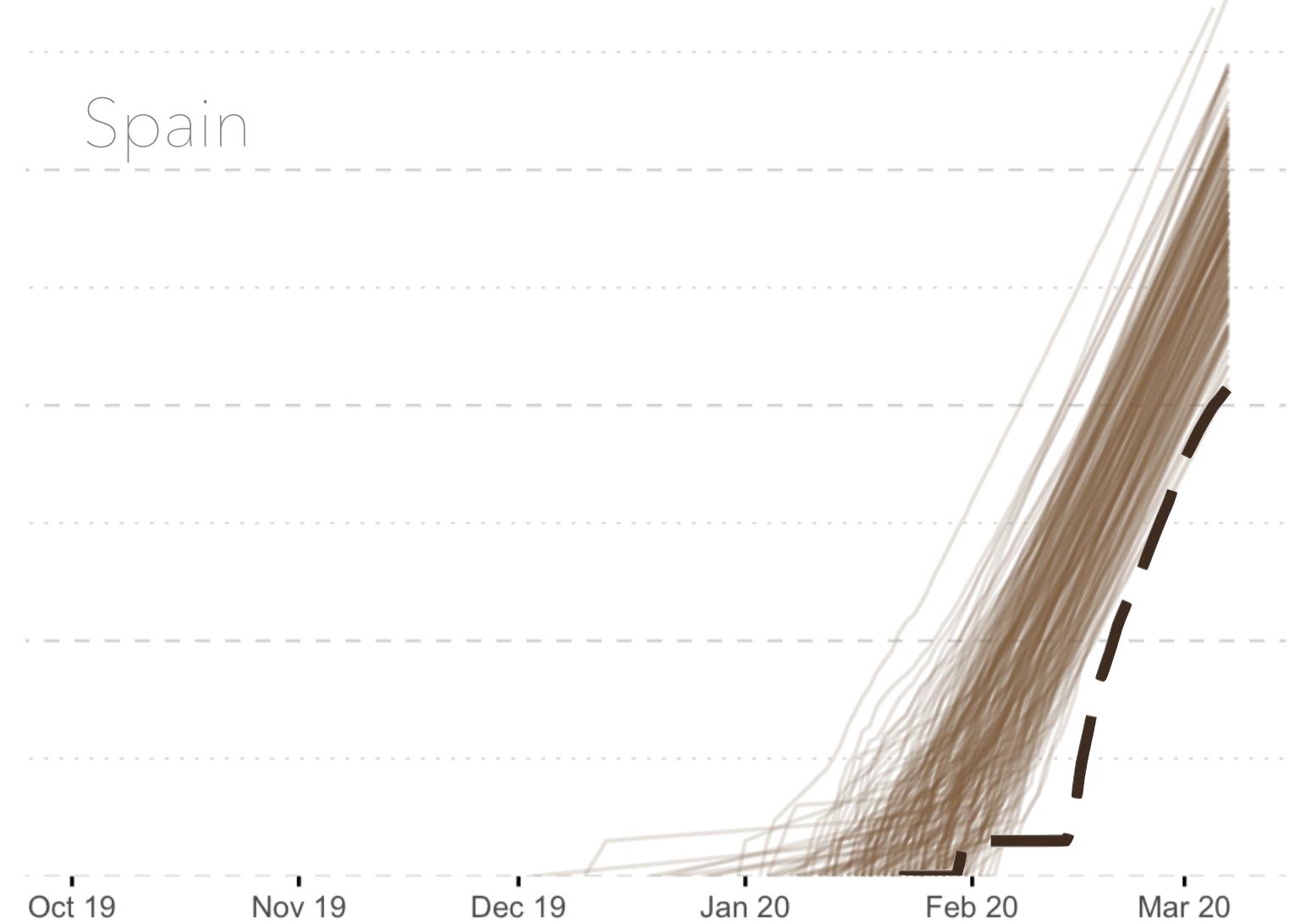
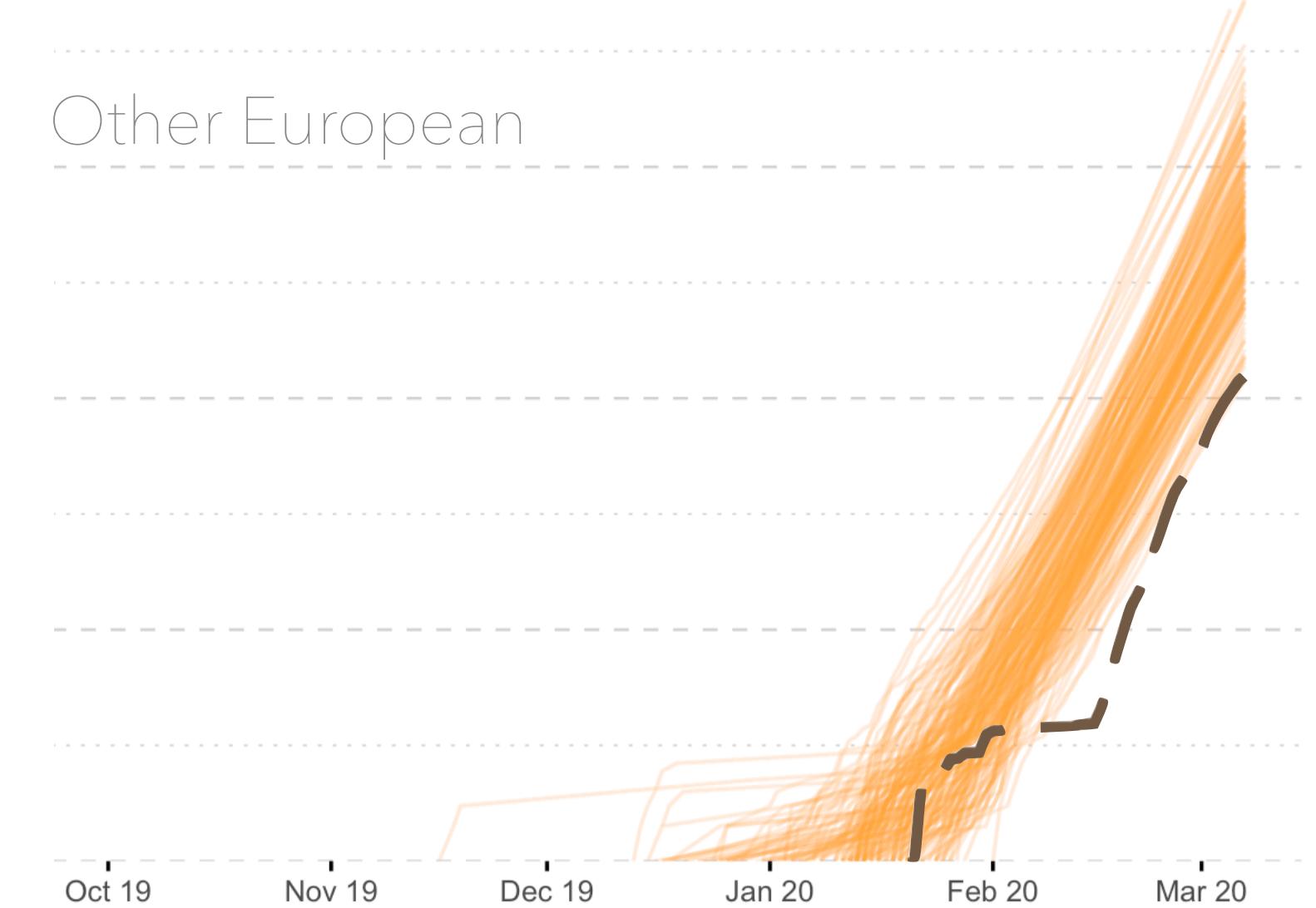
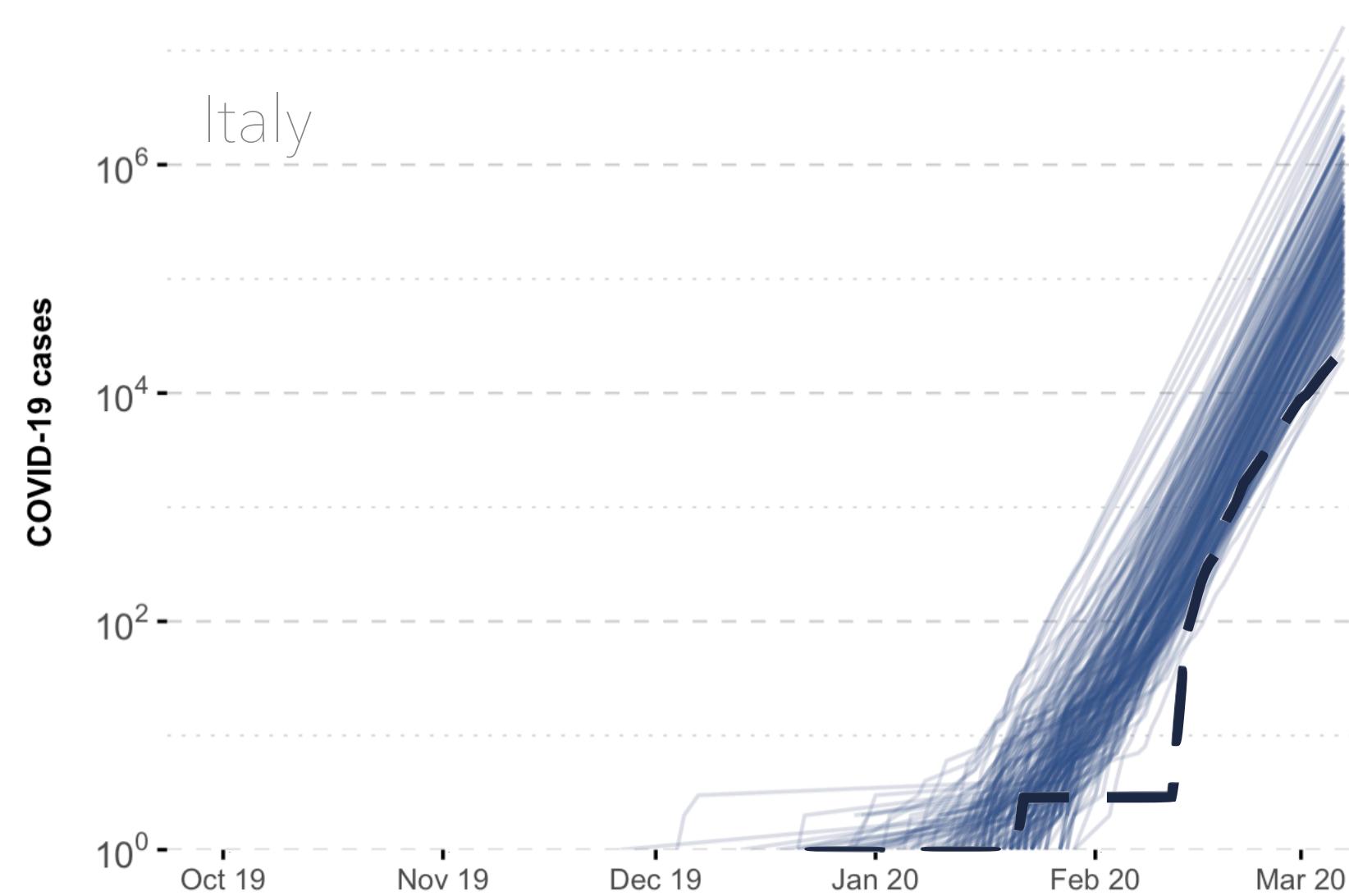
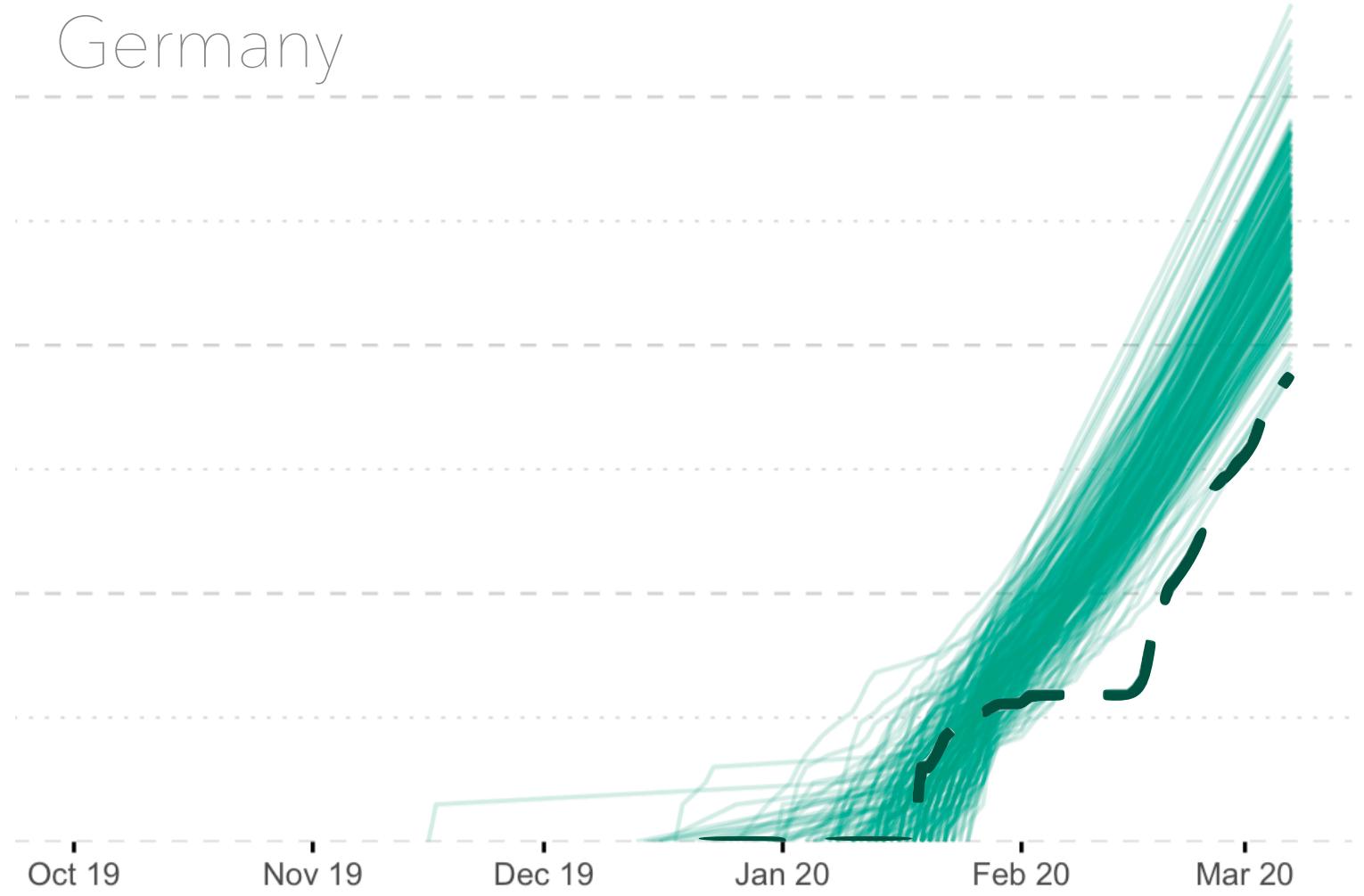
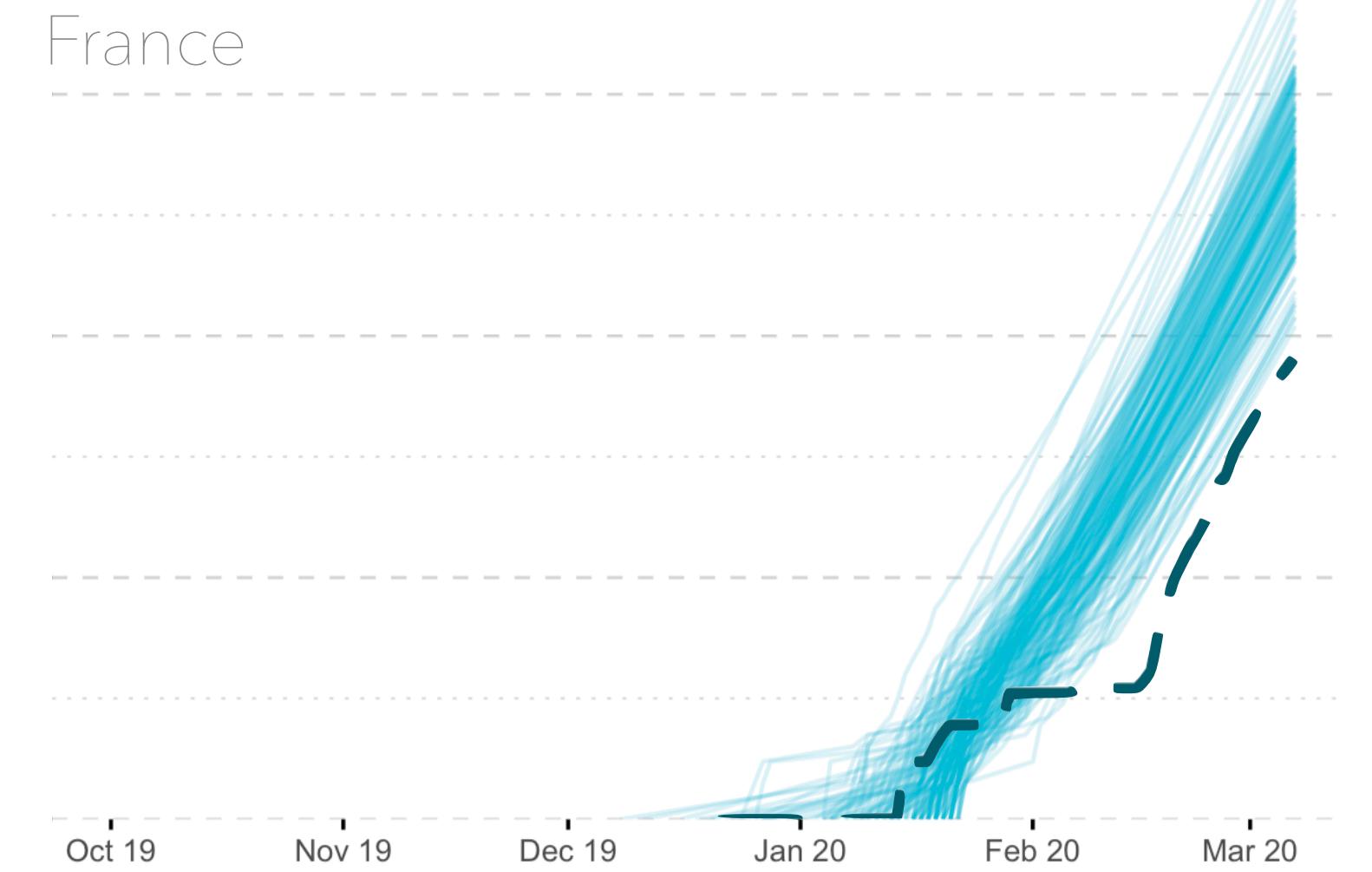
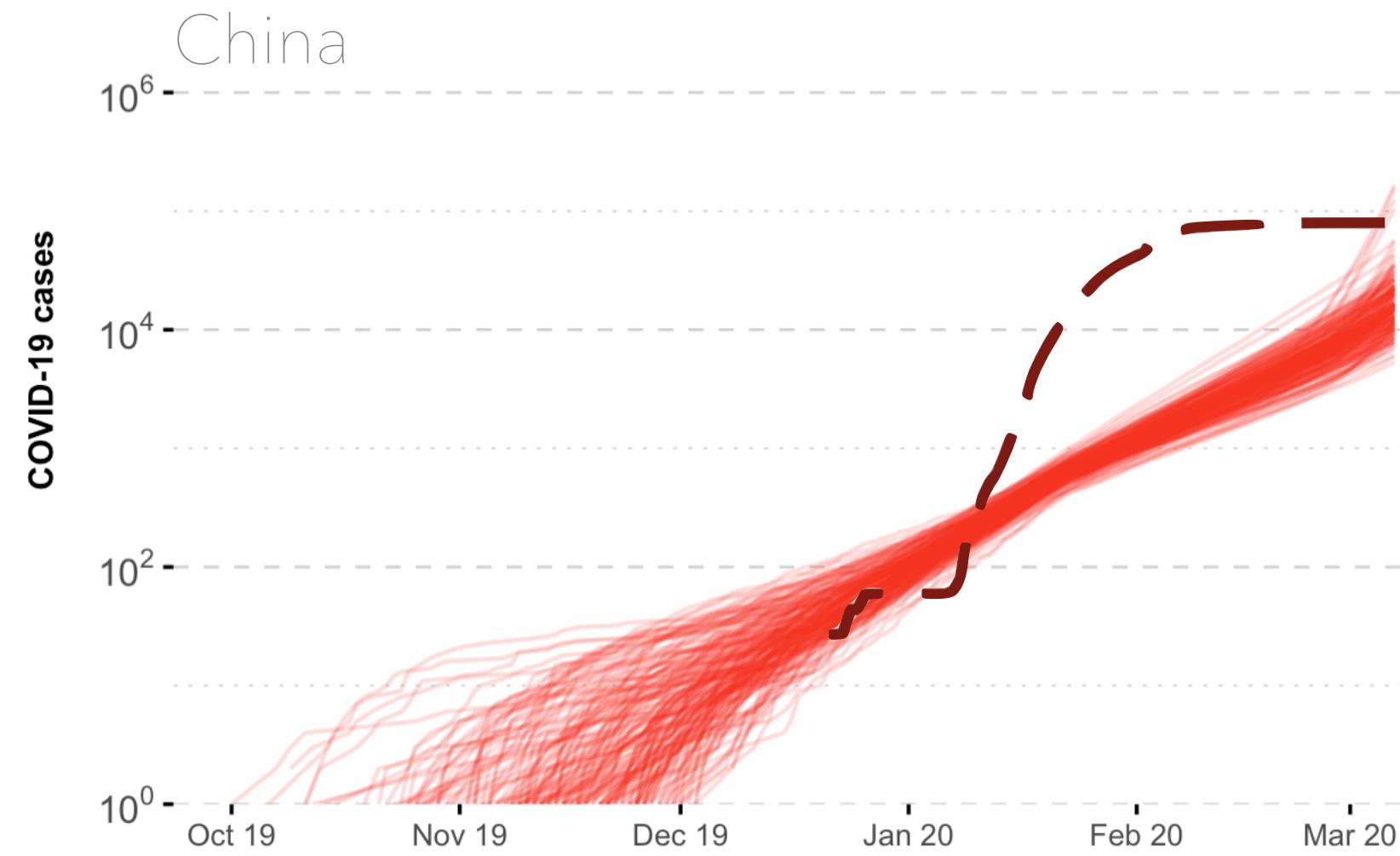
RESULTS

**HOW MANY CASES DID
WE MISS?**

TOTAL CASE COUNTS

Inferred case counts greater than ECDC reported case counts.

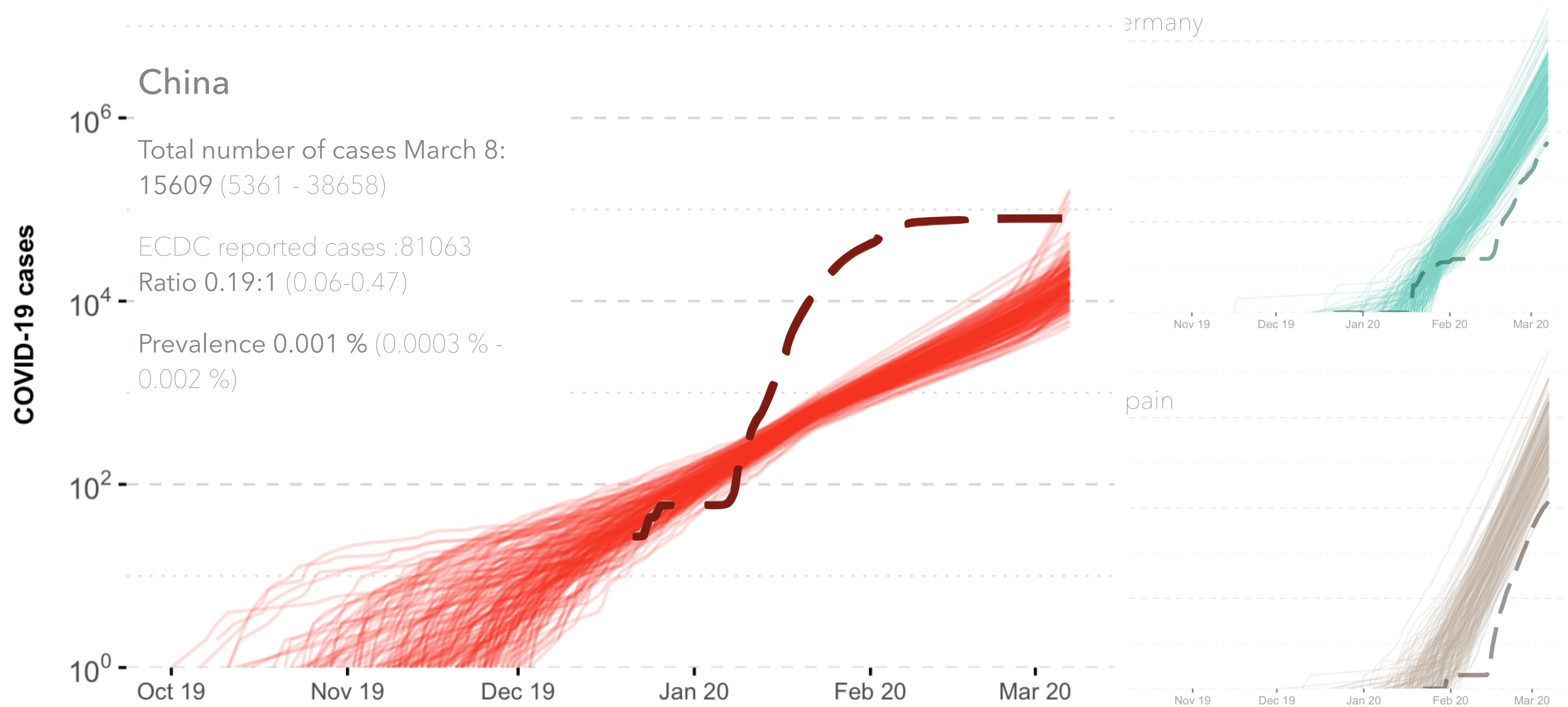
— Reported cases to ECDC
— Epidemic trajectory



TOTAL CASE COUNTS

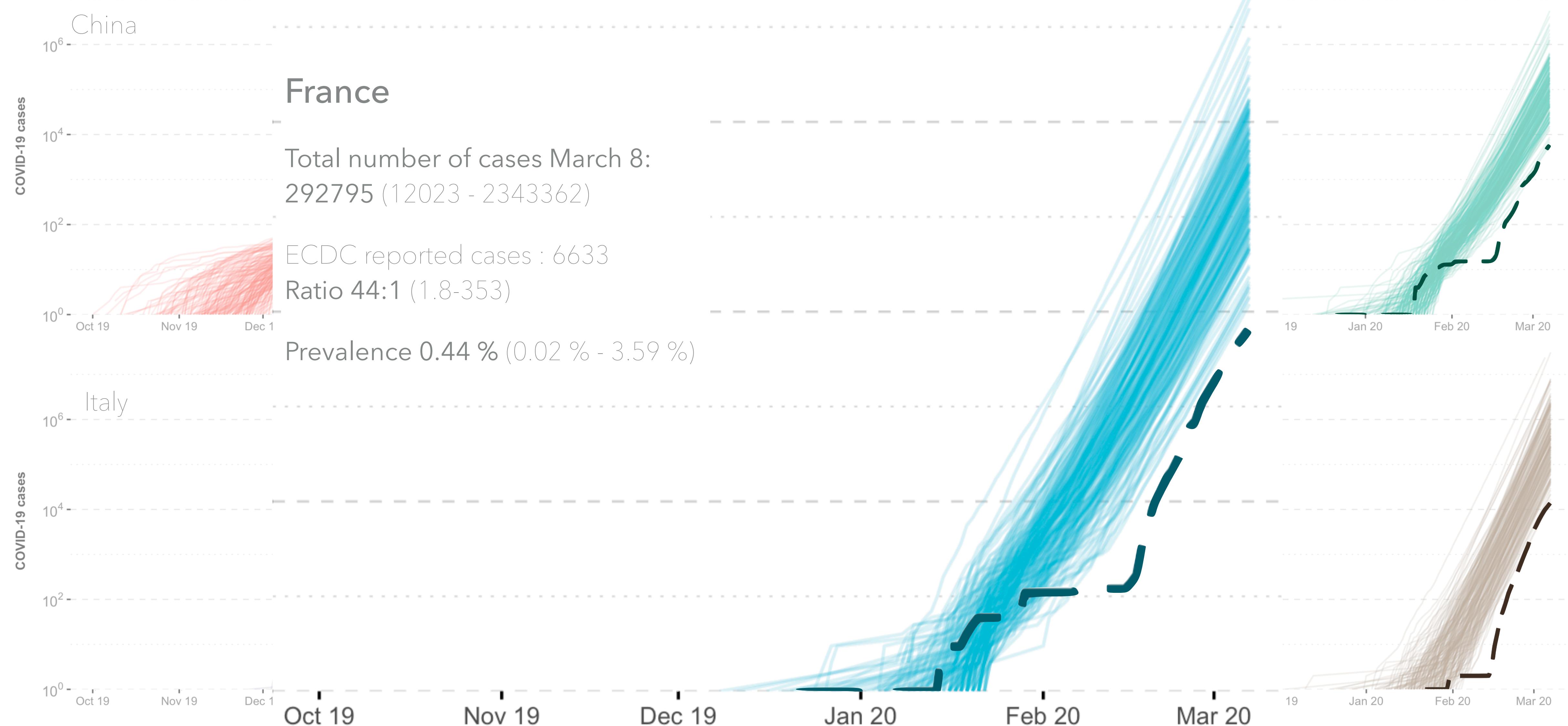
Inferred case counts greater than ECDC reported case counts.

— Reported cases to ECDC
— Epidemic trajectory



TOTAL CASE COUNTS

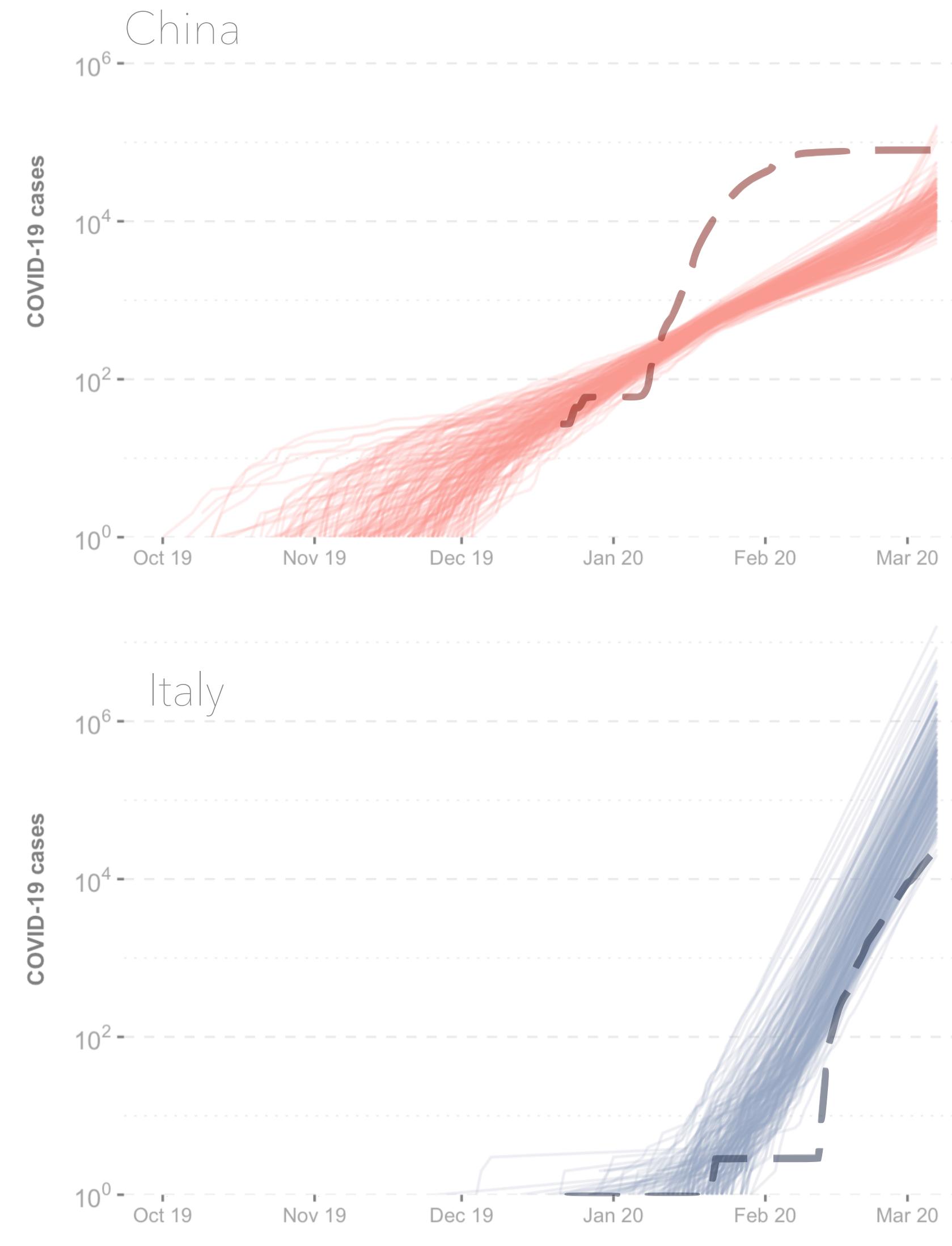
Inferred case counts greater than ECDC reported case counts.



TOTAL CASE COUNTS

Inferred case counts greater than ECDC reported case counts.

— Reported cases to ECDC
— Epidemic trajectory

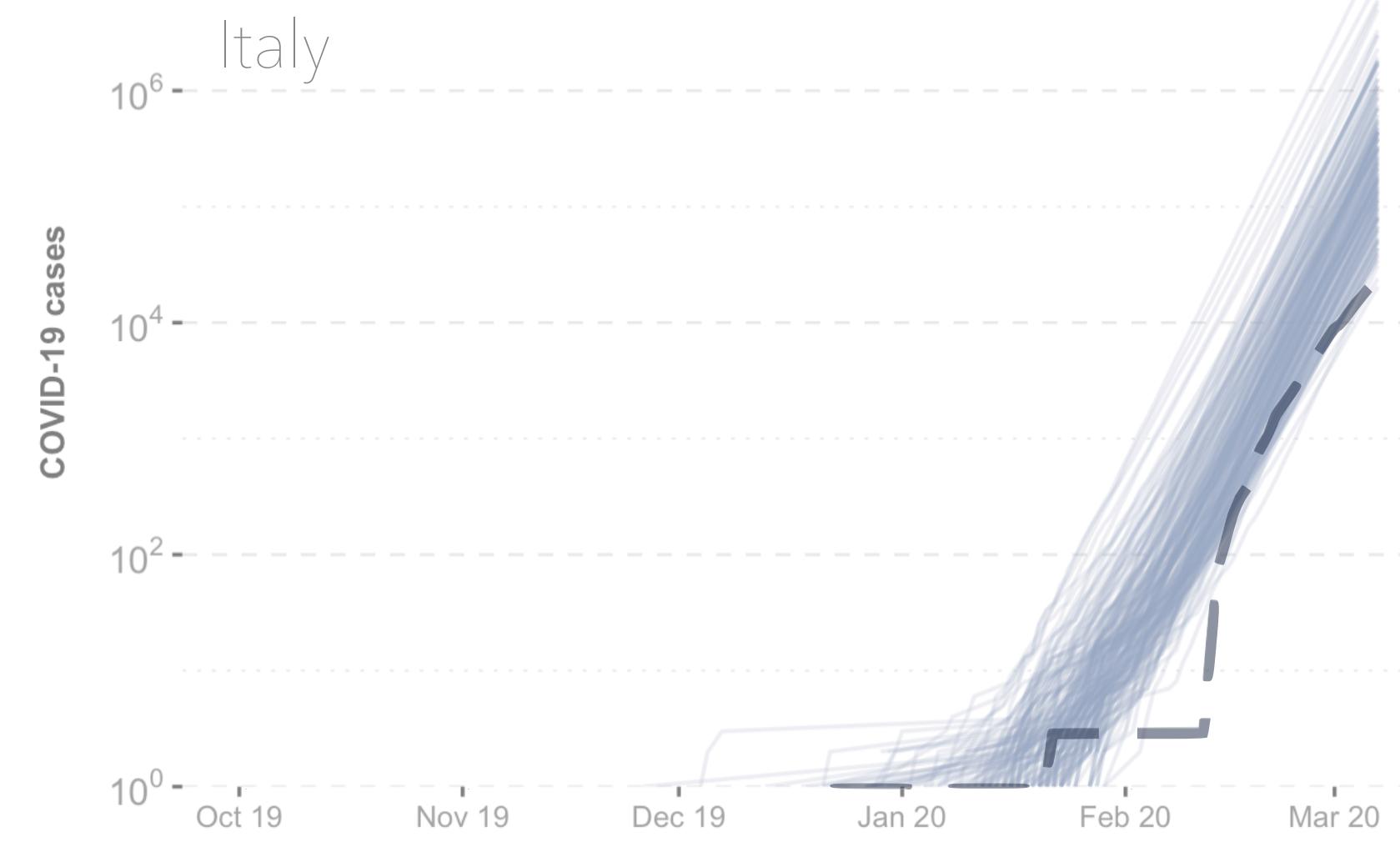


Germany

Total number of cases March 8:
101203 (5751 - 1112342)

ECDC reported cases : 6012.
Ratio 17:1 (0.95-185)

Prevalence 0.12 % (0.006 % - 1.3 %)

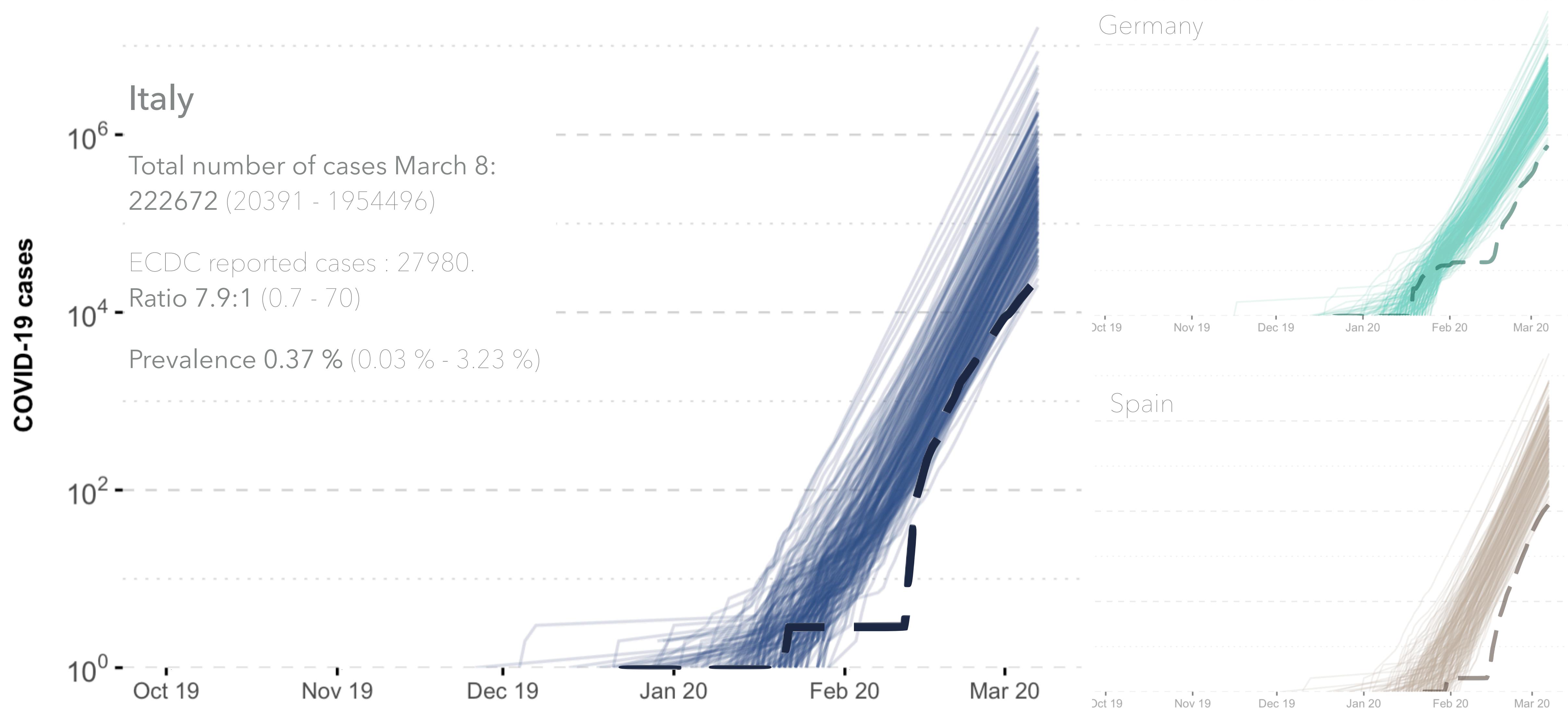


Oct 19 Nov 19 Dec 19 Jan 20 Feb 20 Mar 20

TOTAL CASE COUNTS

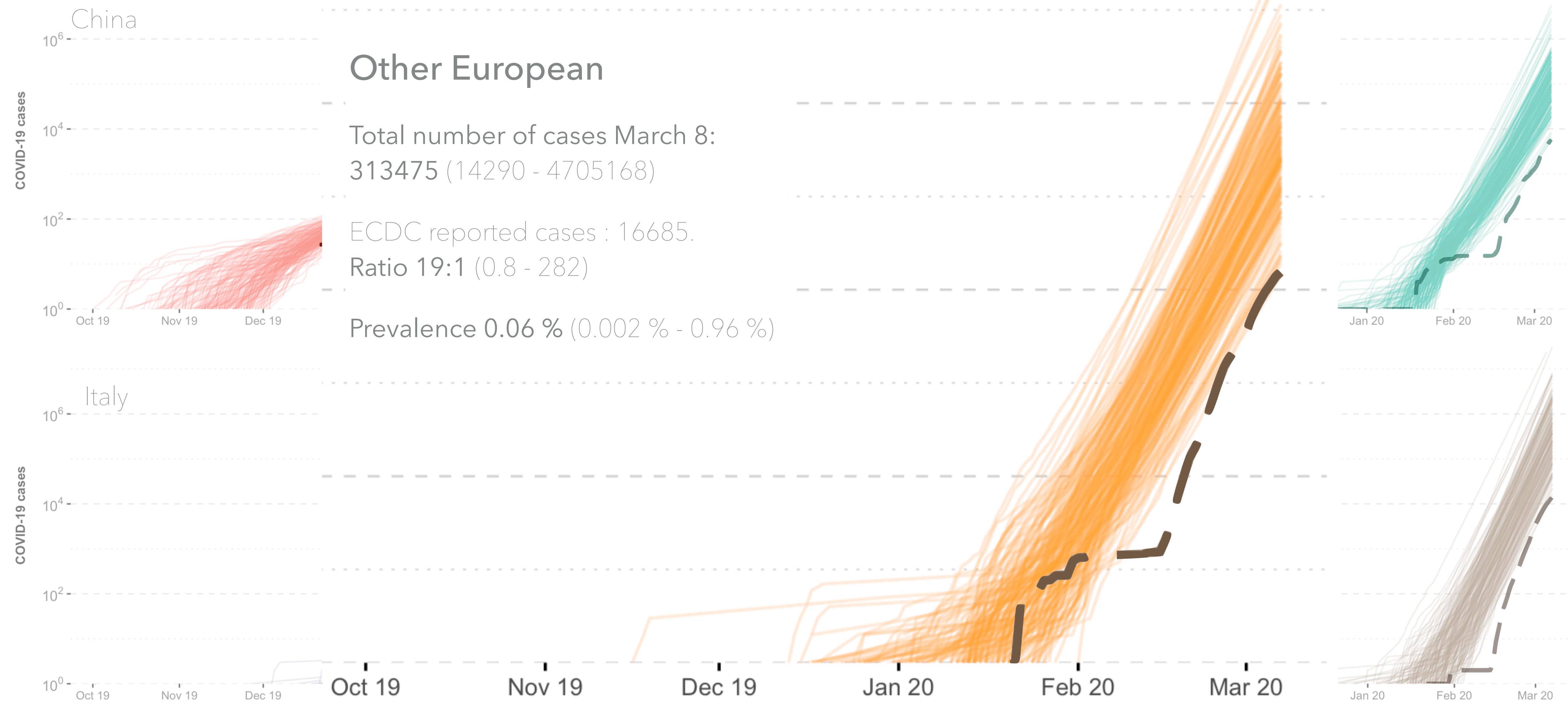
Inferred case counts greater than ECDC reported case counts.

-- Reported cases to ECDC
— Epidemic trajectory



TOTAL CASE COUNTS

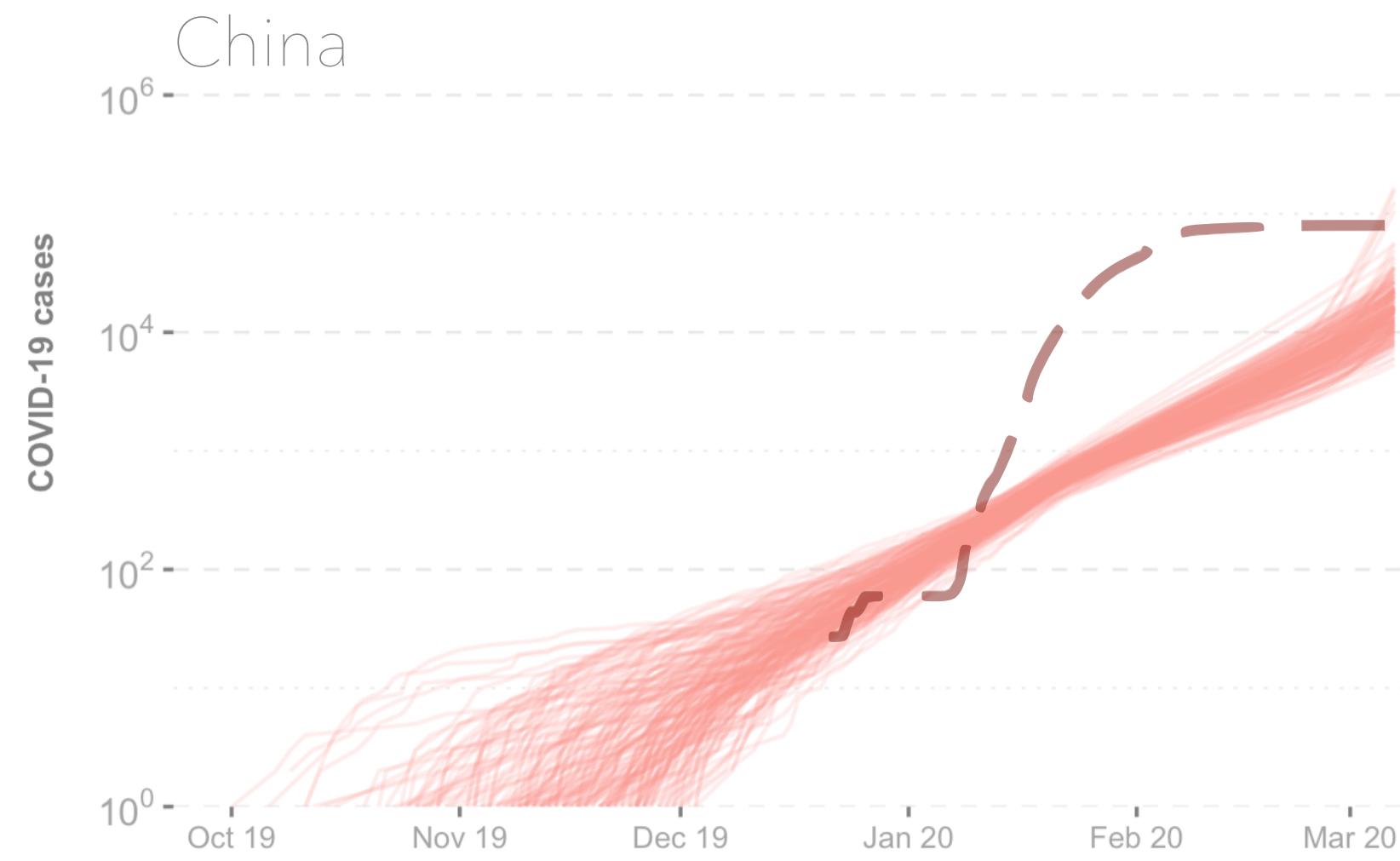
Inferred case counts greater than ECDC reported case counts.



TOTAL CASE COUNTS

Inferred case counts greater than ECDC reported case counts.

-- Reported cases to ECDC
— Epidemic trajectory

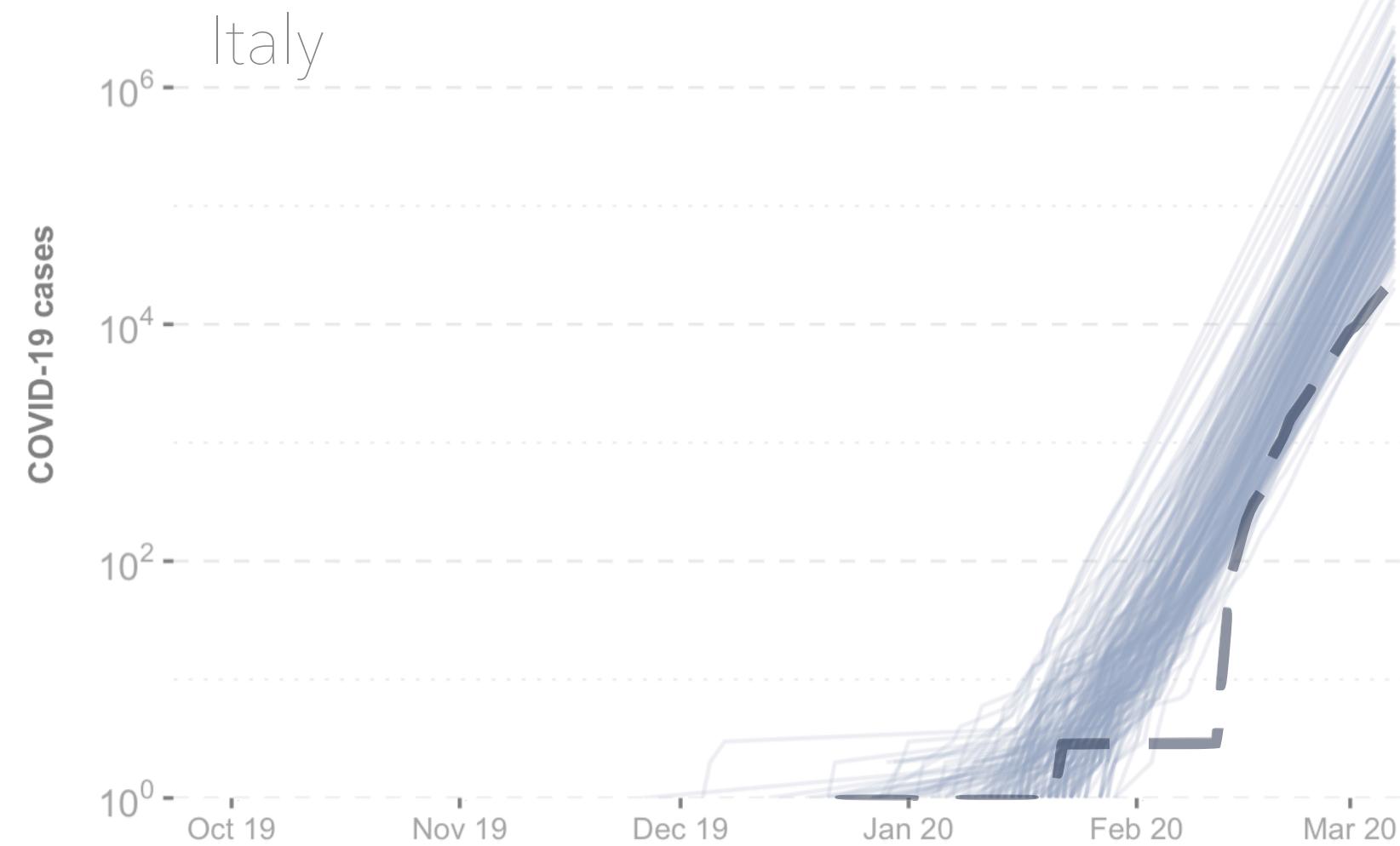


Spain

Total number of cases March 8:
361064 (14817 - 4900845)

ECDC reported cases :13994
Ratio 25:1 (1-350)

Prevalence 0.7 % (0.03 % - 10 %)



Oct 19 Nov 19

Dec 19

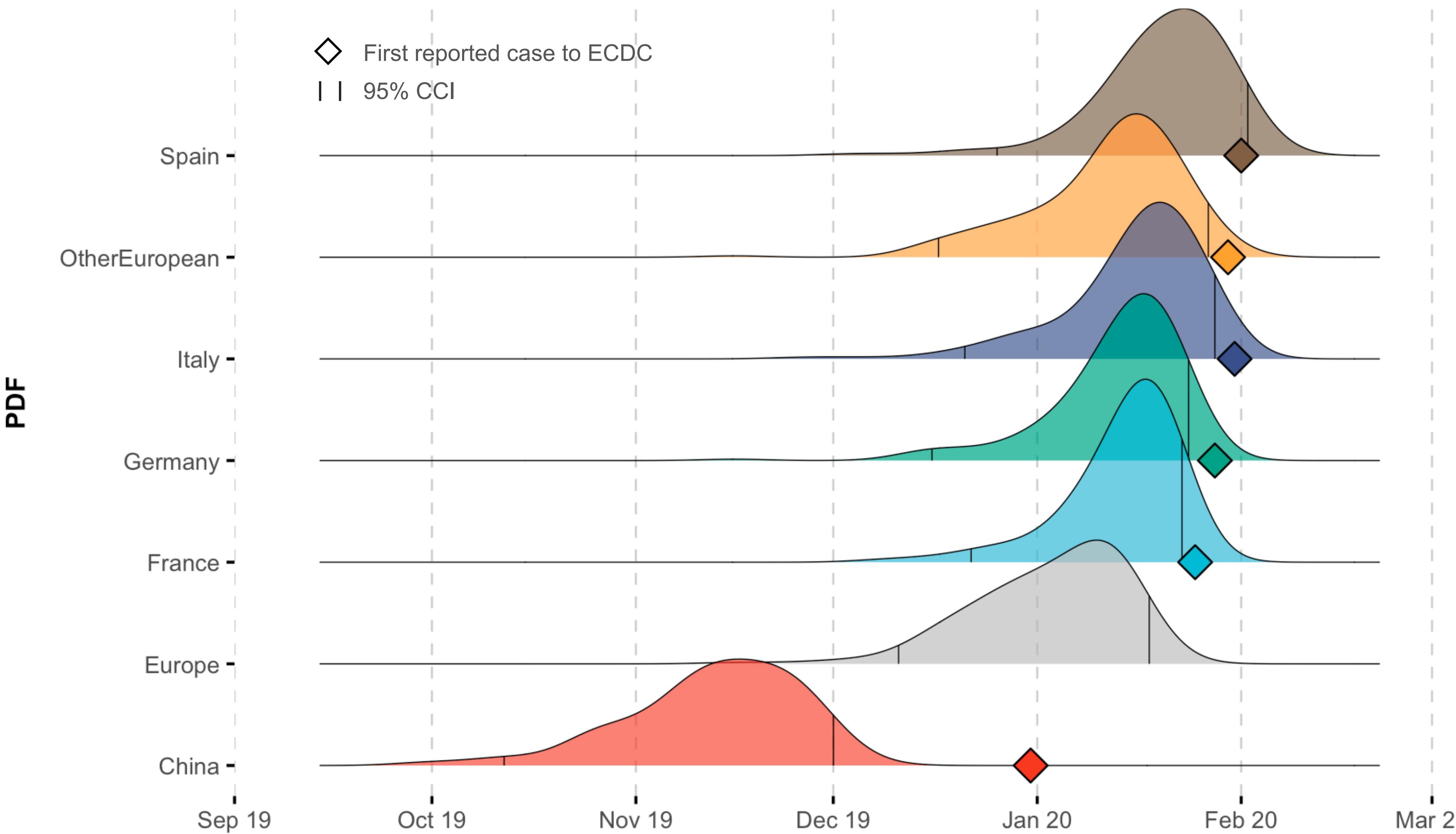
Jan 20

Feb 20

Mar 20

**HOW FAST DID WE
REACT?**

HOW FAST DID WE REACT? - FIRST CASE

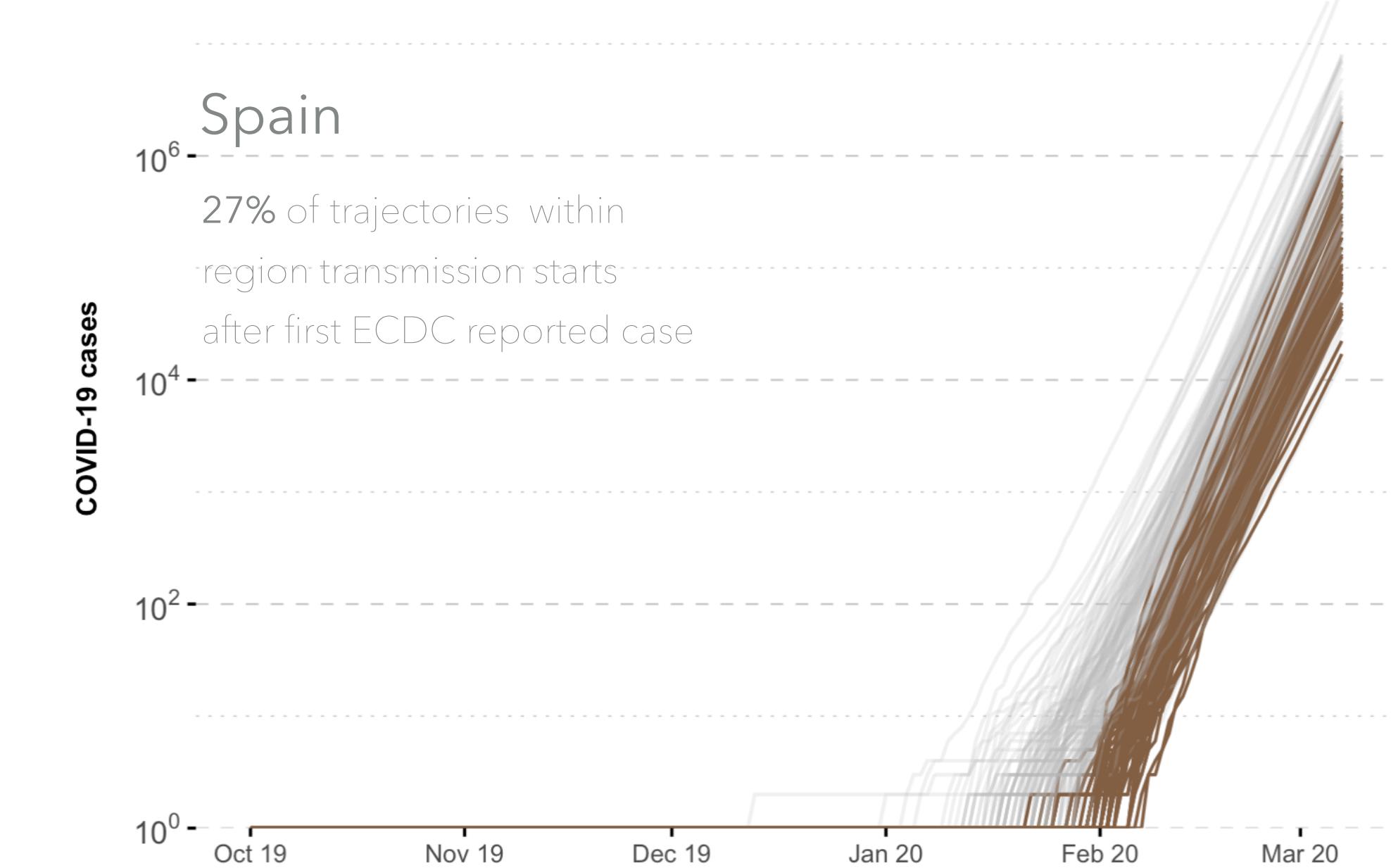
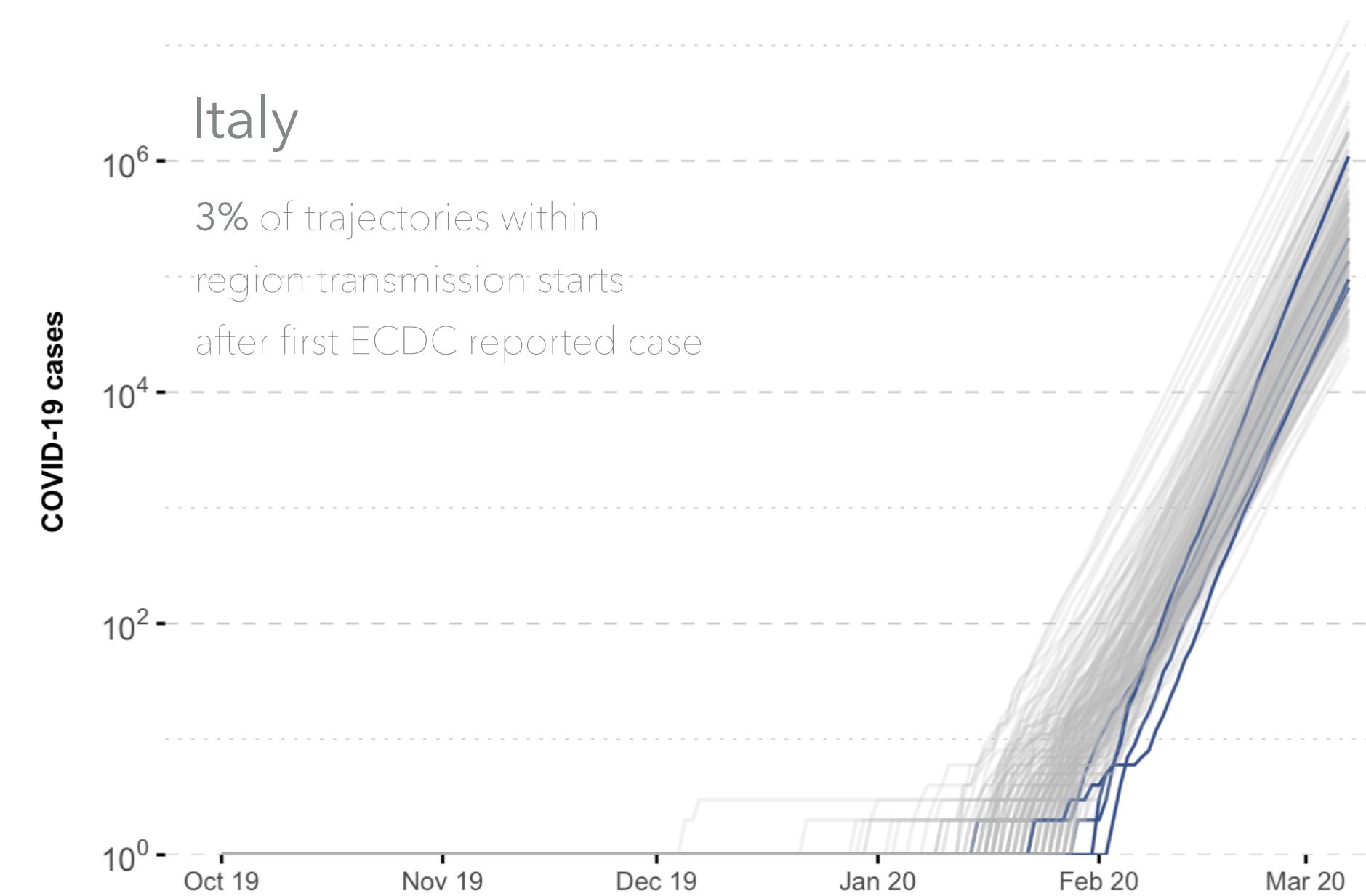
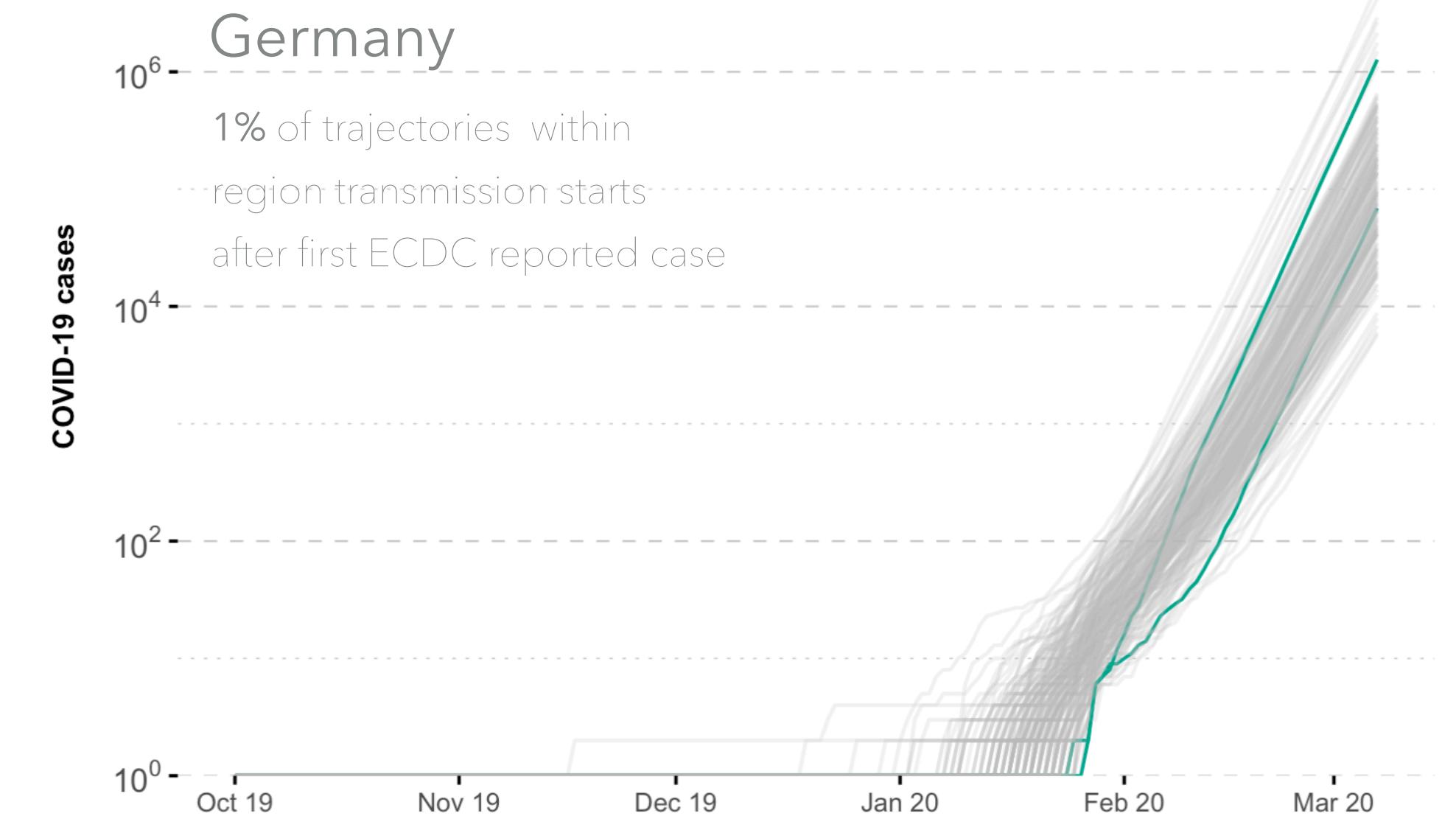
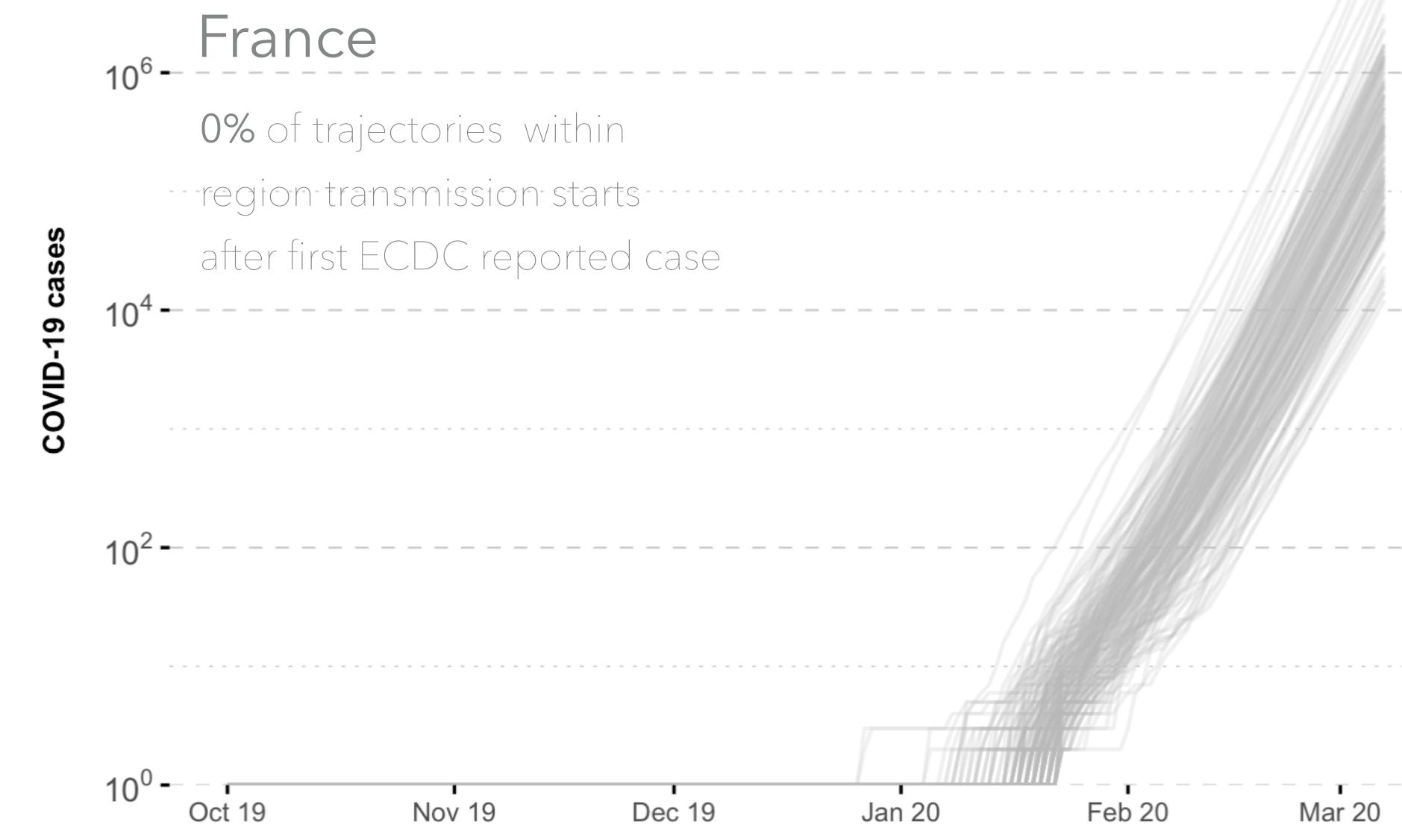


First introduction in Europe
January 5, 2020

(December 8, 2019 - January 18, 2020,
95% CCI)

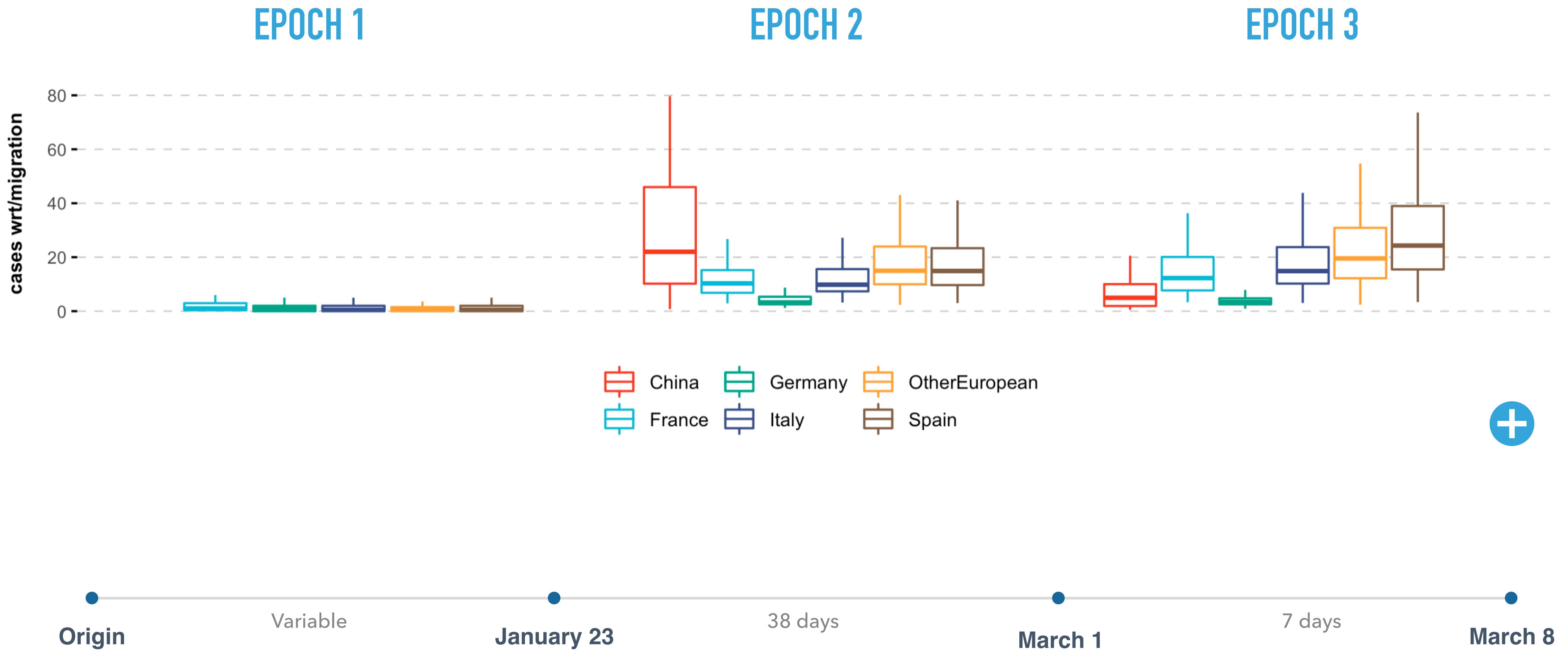
62% trajectories hint first introduction
before notification of Wuhan cluster

HOW FAST DID WE REACT? - WITHIN REGION TRANSMISSION



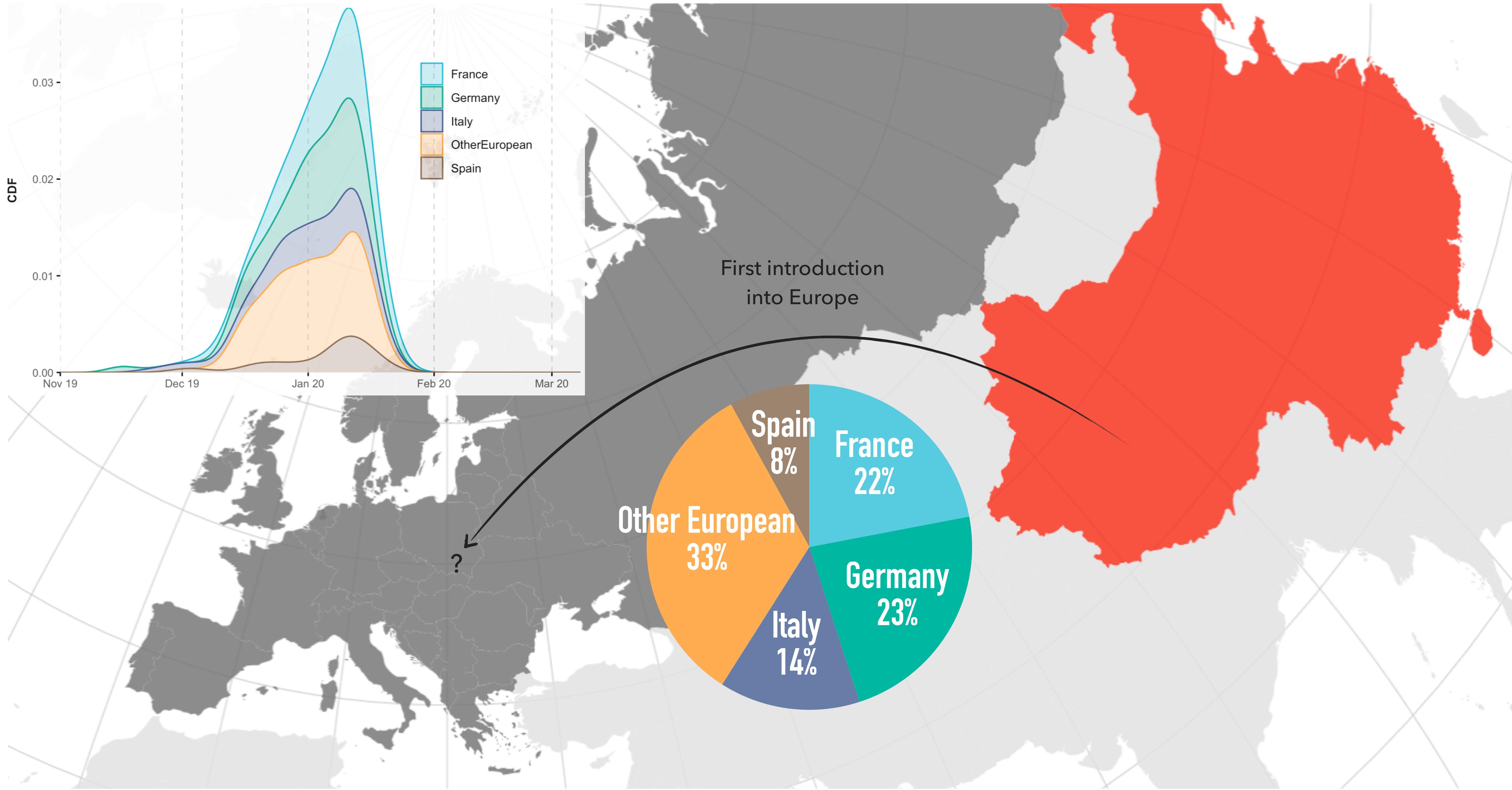
WITHIN REGION TRANSMISSION VS MIGRATION

Ratio within-region transmission cases:migrations increases as the epidemic grows with time.

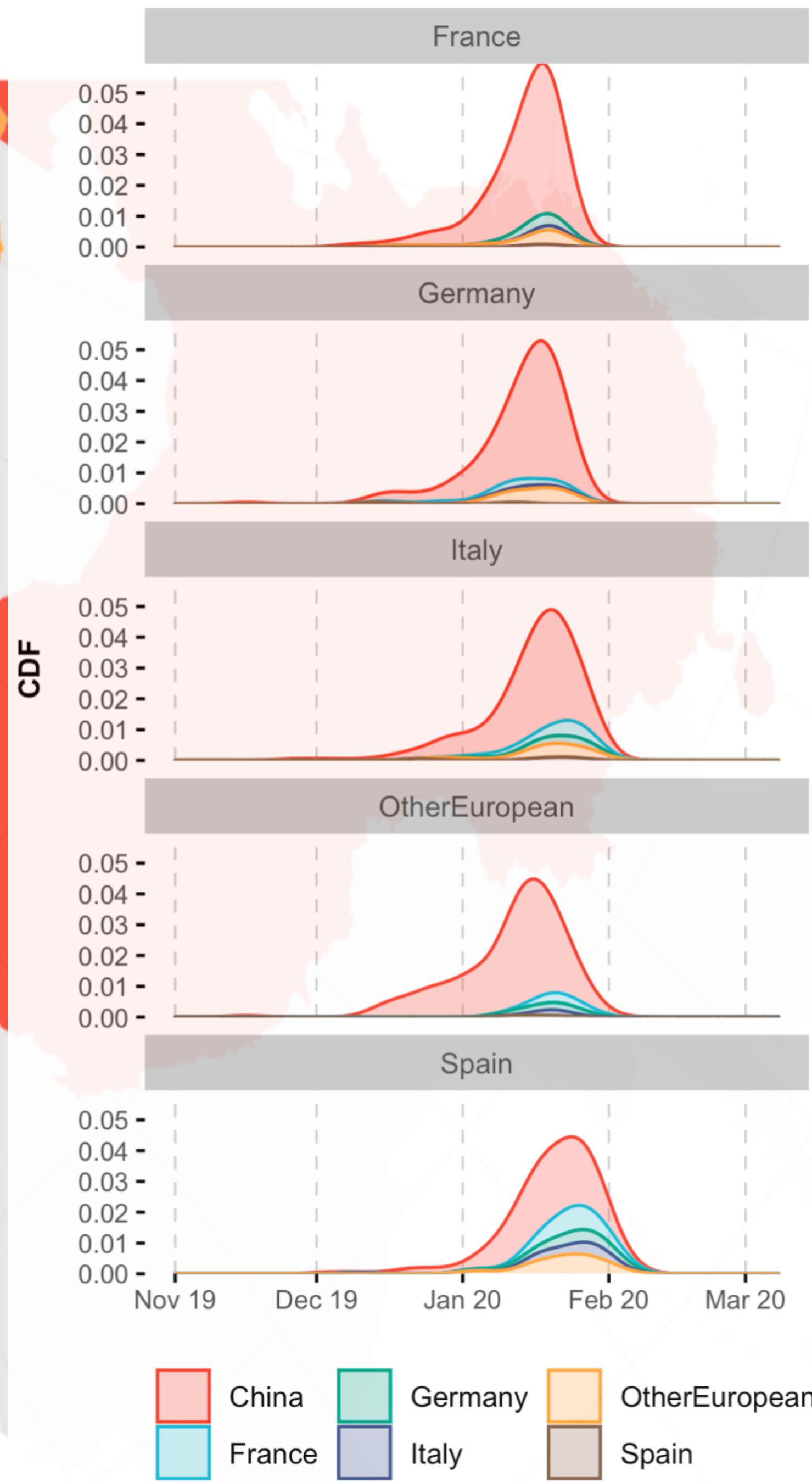
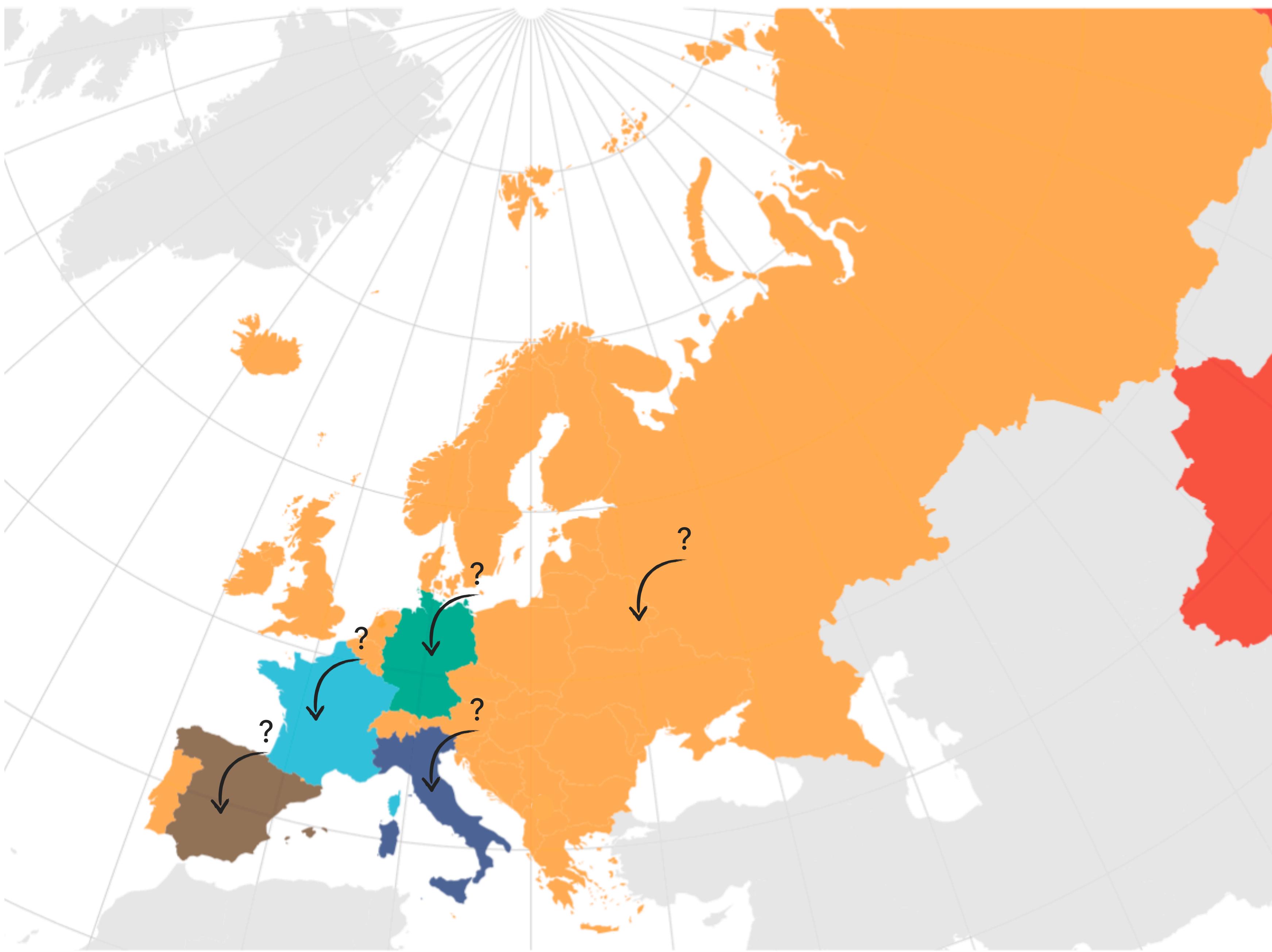


**HOW WAS THE SPATIAL
SPREAD IN EUROPE?**

HOW WAS THE INITIAL SPATIAL SPREAD IN EUROPE? - TO WHERE

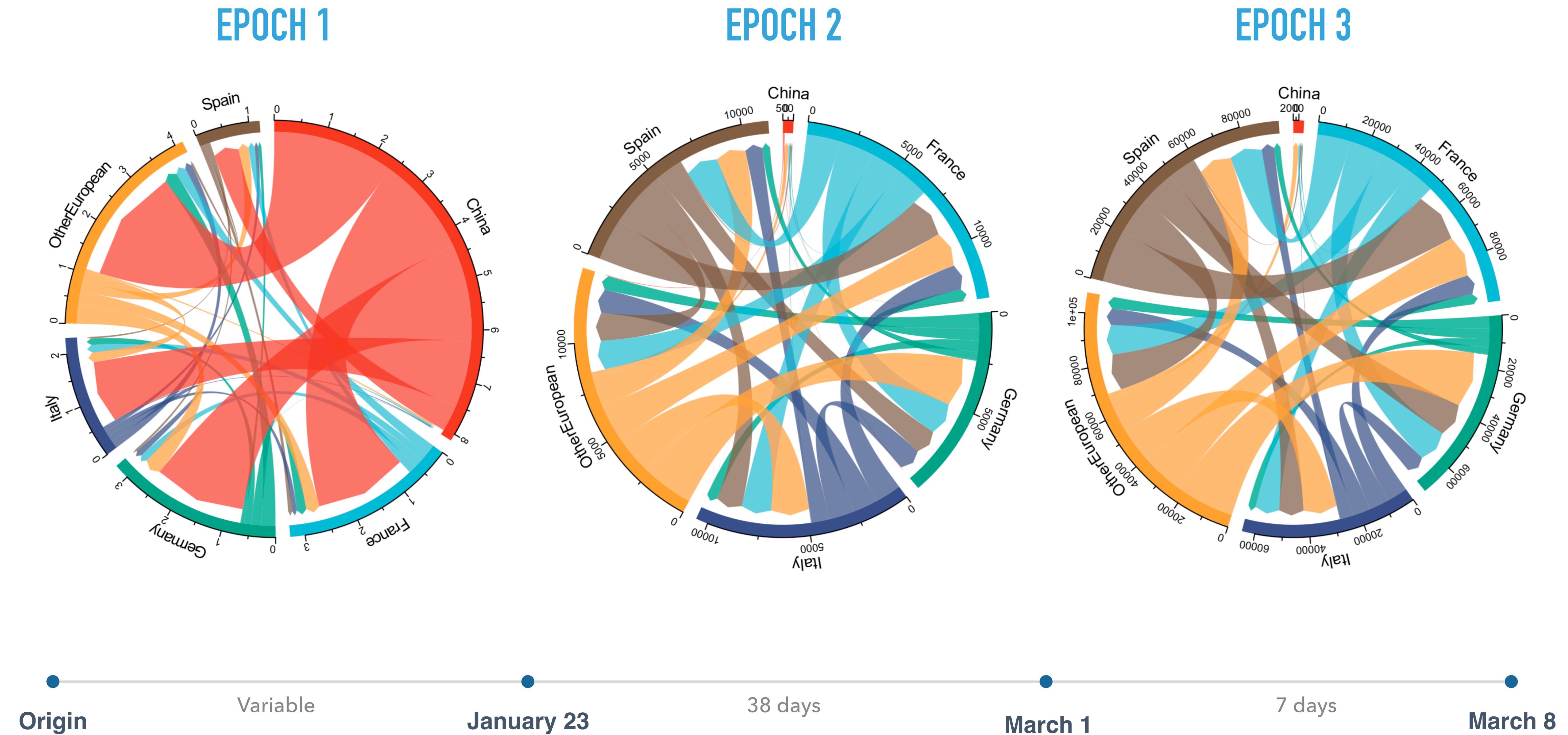


HOW WAS THE INITIAL SPATIAL SPREAD IN EUROPE? - FROM WHERE



HOW WAS THE INITIAL SPATIAL SPREAD IN EUROPE? - MIGRATION PATTERNS

Early epidemic dominated by migrations out of China, overtaken by inter-European migrations later on.



WORK IN PROGRESS GLM INDICATOR VARIABLES

NOT CONVERGING... SO FAR

$$m_{ij} = c \exp (\beta_{pas\ ij} \delta_{pas\ ij} \text{airplane icon} + \beta_{geo\ ij} \delta_{geo\ ij} \text{globe icon} + \dots)$$

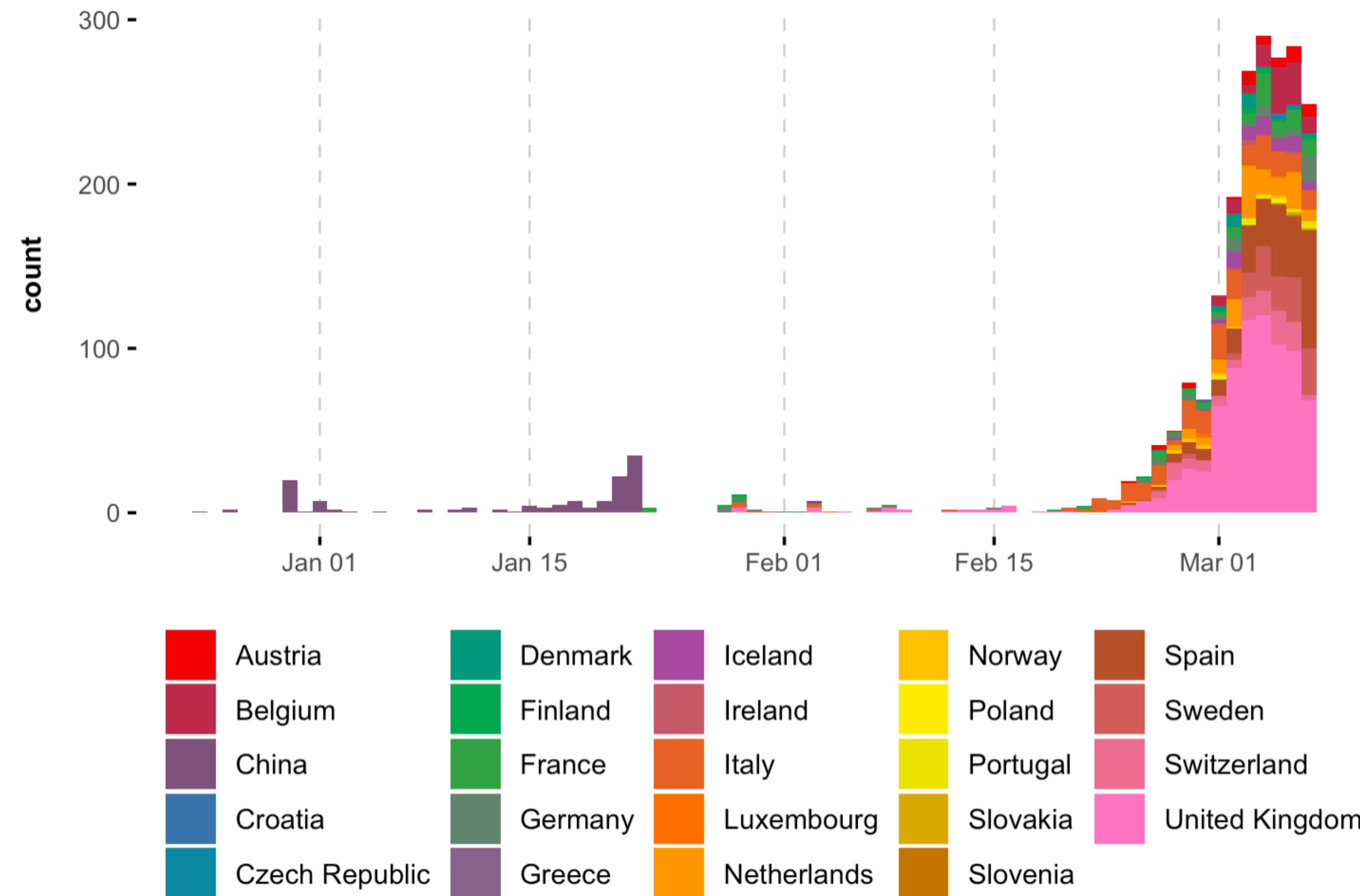
δ_{pij} binary indicator variable, switch 0-1 in/excluding the predictor

SUMMARY

- ▶ BDMM-Prime with trajectory mapping: “fast” analysis exhaustive results to describe an epidemic: case counts, introductions...
- ▶ Early SARS-Cov-2 epidemic in Europe: Underreported case numbers, first introductions from China, non detected within region transmission...

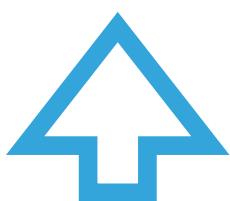
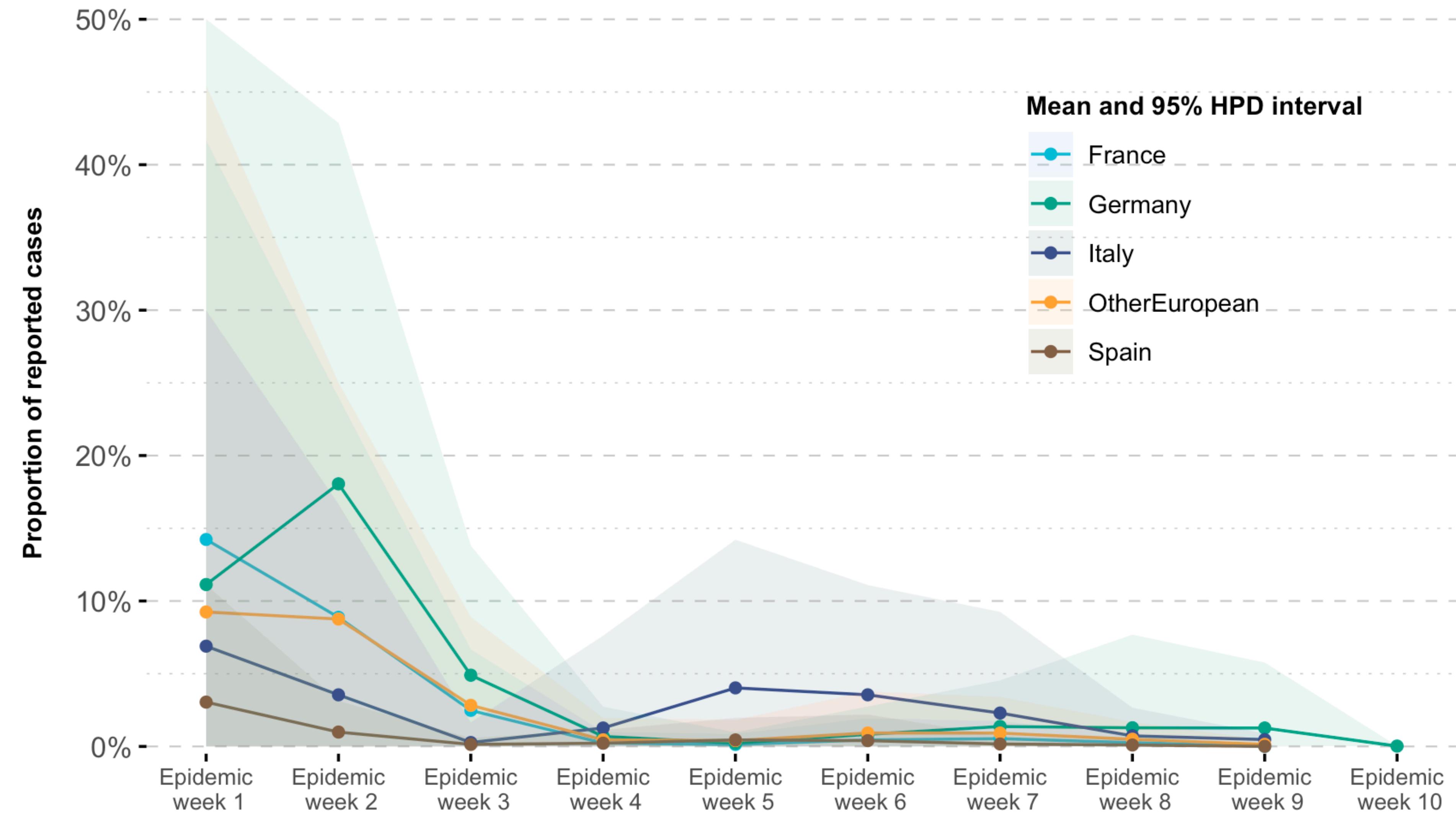
SEQUENCE DATA GISAID

All available sequences from Europe and China, from December 2019 to March 8, 2020.



PROPORTION OF REPORTED CASES BY EPIDEMIC WEEK

Mean reporting proportion under 20% for European countries.



WITHIN REGION TRANSMISSION VS MIGRATION

Ratio within-region transmission cases:migrations increases as the epidemic grows with time.

