

A comprehensive study of the early phylodynamics of SARS-CoV-2 in Europe

December 1, 2020

Abstract

This project focus on the spatial dynamics of the early spread of SARS-CoV-2 in Europe. We apply a novel approach based on the Multi-type Birth Death phylodynamic model to infer structured population dynamics jointly with between-subpopulation transmission rates from viral genome sequences. The inferred epidemic trajectories for the combined outbreak responsible for the observed sequence data will allow us to better understand the entry into and early spread of SARS-CoV-2 in Europe.

The early dynamics of the epidemic in Europe surprised the whole world/ From isolated cases detected in some, in a few weeks we faced an unprecedented epidemic with case numbers and death counts stressing well established health infrastructures. Understanding what happened during those first month could prepare us to be less surprised in similar future events. Complex dynamics, typical epi data, case identification and travel history, very likely not all cases reported. Genetic sequences are a valuable source of information than together with population dynamic model could inform us about the first steps of the coronavirus in Europe. We have integrated sequence data, flight data and bayesian inference priors? to infer population trajectories - case counts and migration dynamics between the most relevant countries at the start of the European pandemic.

Introduction

Importance of understanding the spread of the virus to prevent future outbreaks

About phylodynamics/phylogeographics? Use of genetic sequences as a source of information combined with other sources of information as travel data

About what we know of the introduction of sars-cov-2 in Europe and early dynamics till 8 March

About what we know of the case counts in Europe? Can sequences help?

Introduce/formulate the questions: case counts, first introductions, ?migration vs within region transmission?, migration patterns, ?border closures.

Results

Total case counts

Figures 1, 2 and 3.

Inferred population trajectories that account for all cases in the country. It is not an independent estimate since we are using informative priors for the sampling rates (upperbound computed as sequences/reported cases) and the subsampling scheme also includes the number of cases in the country (but that is a weaker link I think).

Comparison with ECDC case counts: delay in notifications and underreporting, for some of the countries at the end of the period the reported cases are very closed to the trajectories, e.g. Italy, Germany. Hubei-China huge difference in tendency and absolute numbers.

Measure the (mean?) difference between inferred and reported case counts and compare among countries. By day or in general? Absolute or relative to the number of cases that day?

First introductions

Figures 4, 5 and 6.

We want to answer the question of when and where the first case in Europe occurred. No clear, Germany? Figure 4 A. We can measure the time between first case and first reported case and compare between countries.

And for each country where this first case came from, more support for Hubei introduction in all countries Figure 4 B.

Measure expected day of introduction.

In Figure 5, we are looking at the ordered set of introduction into European countries. For each trajectory, we have the first introduction into each country happening in a determined order, i.e. the first European country with a covid case, the second country with a covid case... So we are visualizing the number of trajectories that share the same order of first cases. More trajectories where the first case was to Germany, no clear order after the first country. Spain and Other European last one.

In Figure 6 we are looking at the ordered set of countries spreading covid. In the first position we have Hubei for all trajectories since it is the origin of the epidemic. In second position we will have the first European country from where a covid case was imported to other country. No clear dynamics except that more support for Spain being the last country having a first exported covid case. Germany first one.

Working on these with new analyses

ticks for every month

Explain this better

Migration vs within-region transmission

Figure 8

This question is addressed on the original paper from Sarah. Maybe could be interesting to look at this perspective since it is important for border closure measures.

Not sure if we can have any meaningful results insights on the topic. Migrations started very early in our analysis to all european countries. Withing regions transmission account for most of the cases in the countries from late january onwards (when first cases were being reported in Europe).

Think
about
results

Measure the time between the first outcoming migration from the country (specially interesting for Hubei) and first reported case. This could be interesting to say if a extreme measure closing borders with the first case could be effective to impede transmission to other countries: percentage of trajectories where transmission to Europe would have been avoided. For other countries we can look at how many migrations events could have been avoided (and how many not) if the country closed borders after first reported case.

Measure the time between first birth/migration and first reported case and compare between countries. How well did the countries detecting the first cases, there were already within region transmission?

Measure the time between the first migration into a country and the first birth in the country. Or the first outcoming migration from the country and first birth. This could be interesting to say if we should focus or not the screening and testing capacities to detect incoming migrations or a more general strategy to find cases in the population. Is it different for each country?

Migrations patterns

Figures 7 and 9.

Similar information in both plots, but in the chord plots instead of a daily evolution the time period is split on three (as in the GLM analysis). Chord plots are nicer and easier to understand I think, but barplots shows the great detail of the results of the model. Another advantage of the chord plot is that it shows mean absolute values and not only relative values.

Hubei-China is the majority source of migrations for all countries till February and then some patterns emerge (we expect more interesting results with the added info in GLM analysis).

Legend
in plot
7 and
change
white
text

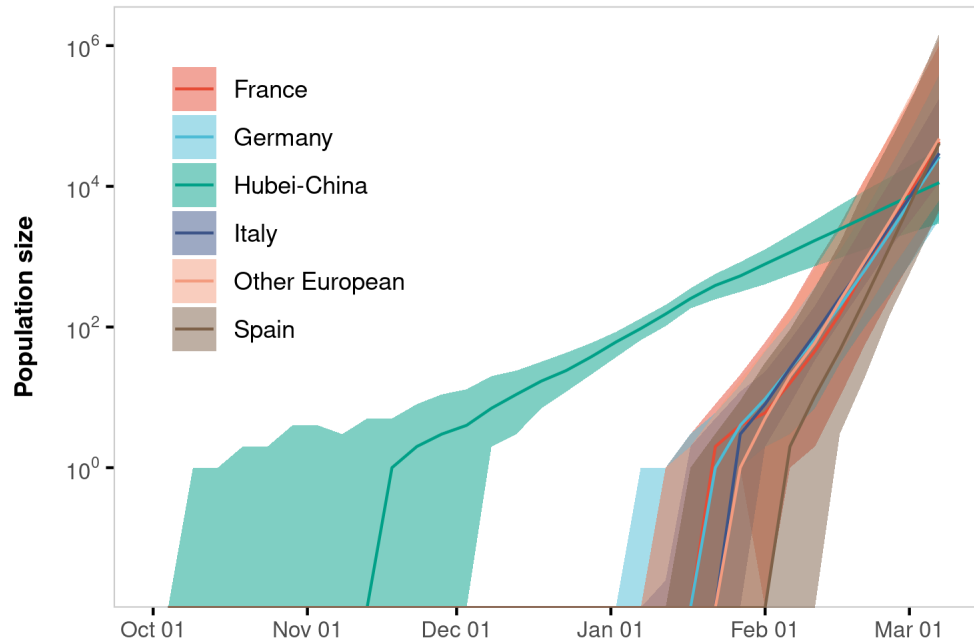


Figure 1: Inferred population size summary statistics for each deme over time. The line represents the median population trajectory and the interval is the 95% center credible interval from a subsampled set of inferred population trajectories.

GLM predictors

Which channels were the main sources of transmission across national borders.

Epidemiological parameters

Not sure about this section. But maybe it make sense for completeness to include, briefly, the values of estimate R_0 , migration and sampling rates.

Discussion

Case counts discussion and reporting rates, second wave in Europe. Other estimations in other studies.

First introduction other studies.

Migration patterns other studies.

Discuss assumptions of the model, caveats and possible improvements.

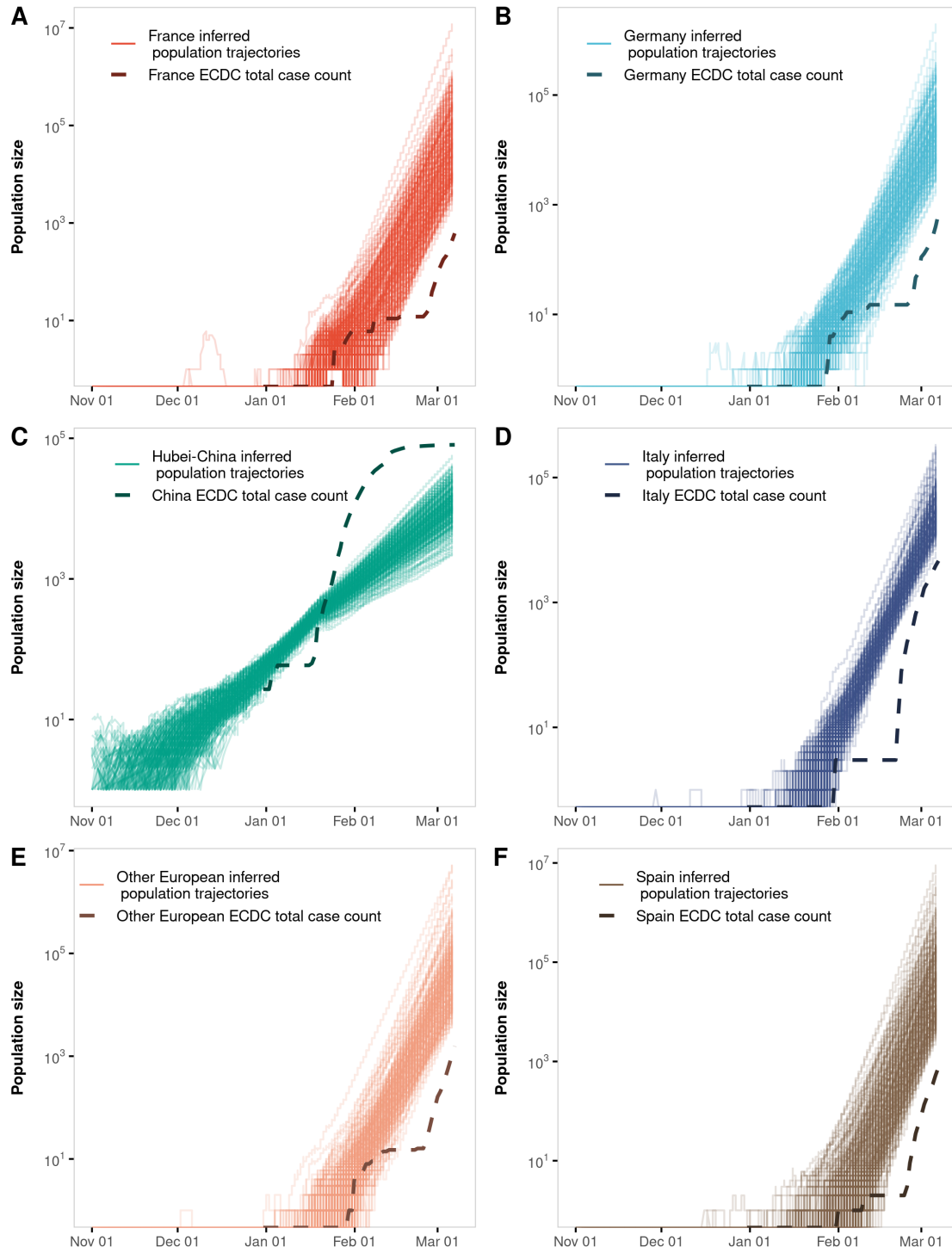


Figure 2: Inferred population size trajectories over time. A subsampled set of 200 inferred trajectories is plotted. In each subplot, the trajectories (solid lines) are compared with the ECDC total case count data (dashed line).

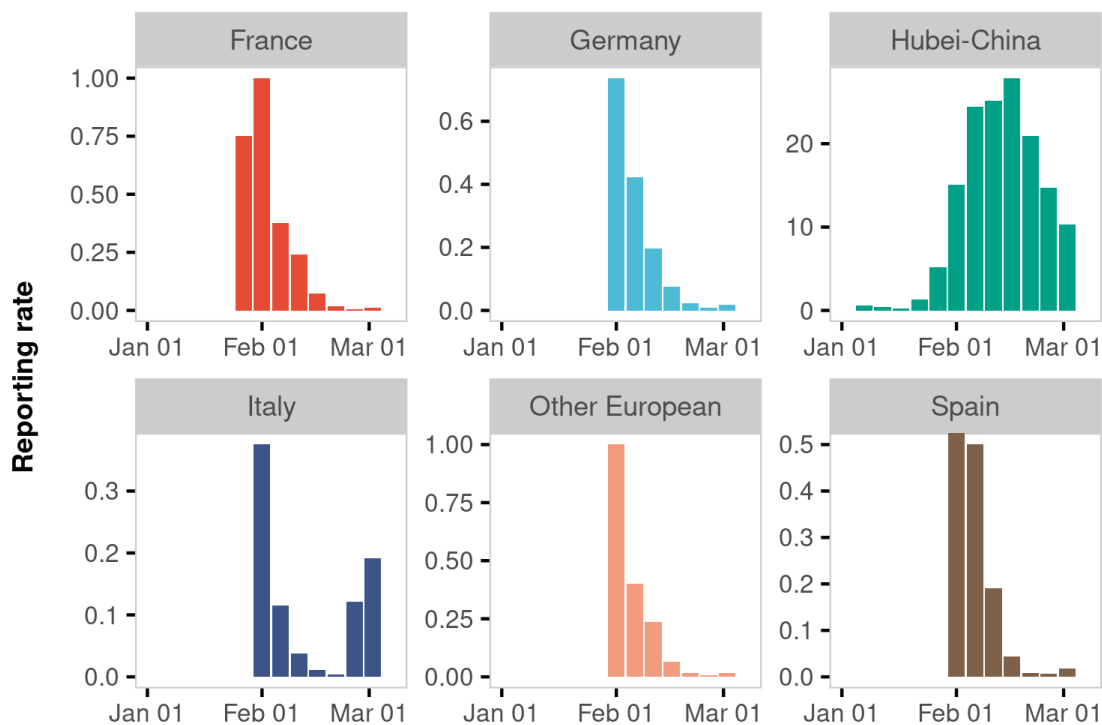


Figure 3

Conclusion

Material and Methods

Dataset and subsampling. Alignment.

Model BDMM-Prime, stochastic mapping. Priors. Bayesian phylogeographic inference. Phylodynamics.

GLM model. External data.

?Same analysis but with travel data from last year, can we observe a consequence of Hubei lockdown and reduced travel?

Trajectories analysis. Subsampling and summarizing (median, 95% interval)

Snakemake workflow.

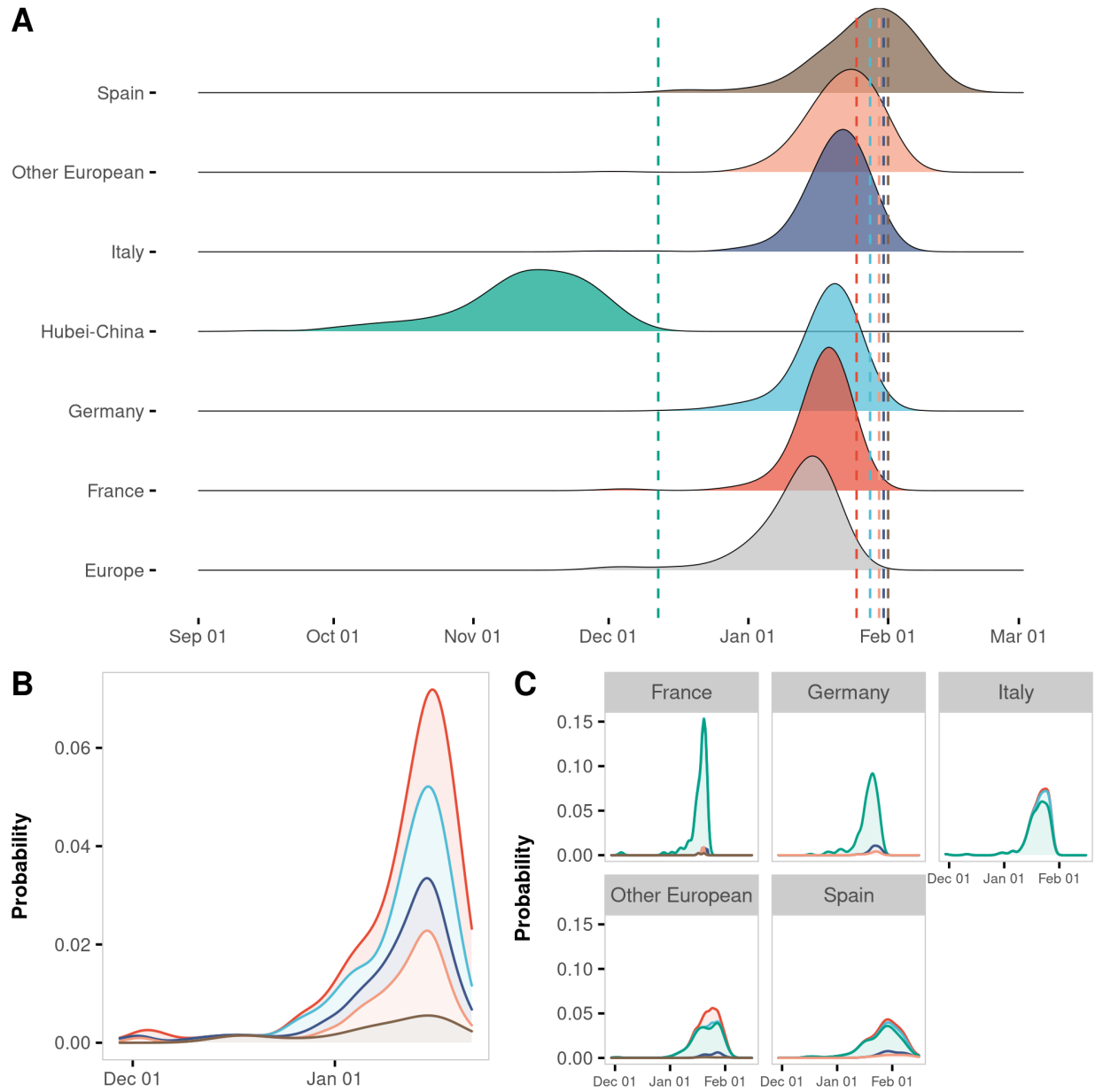


Figure 4: First introductions event time and source in each deme. **A** Probability density of the time of the first introduction for each deme from the subsampled set of trajectories. Each dotted line represents the first date when cases were reported to ECDC by deme color. In the case of Hubei, the distribution of the time of the origin is plotted, since it was defined as the origin of the epidemic in the analysis with probability 1. For the other five demes, the distribution of the time of the first migration into the corresponding deme is plotted. **B** Probability density of the source of the first introduction event in each deme for the set of subsampled trajectories.

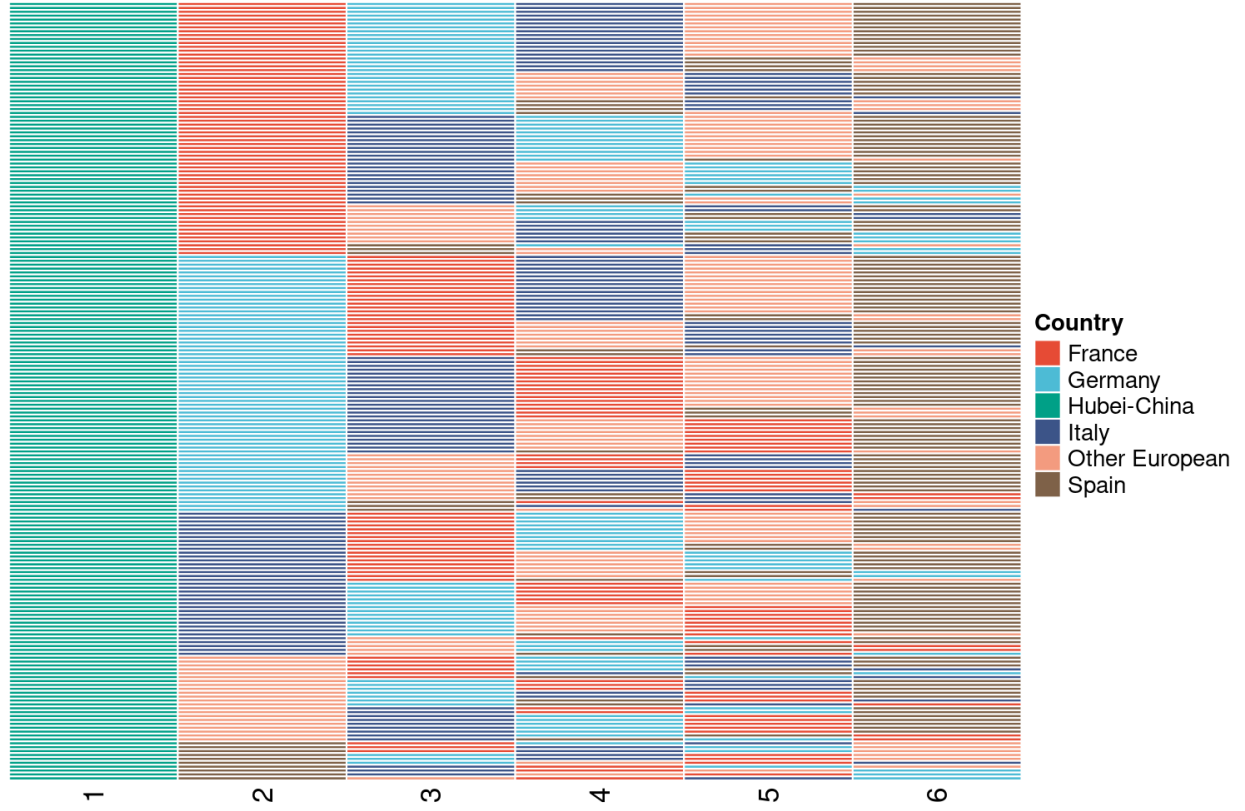


Figure 5: Ordered set of countries by its first migration, i.e. spreading event out of the country. The color corresponds to the last country with a migration to other countries. The size of the box with the country name / ribbon / flux is the number / proportion of trajectories that have that particular order of first migrations.

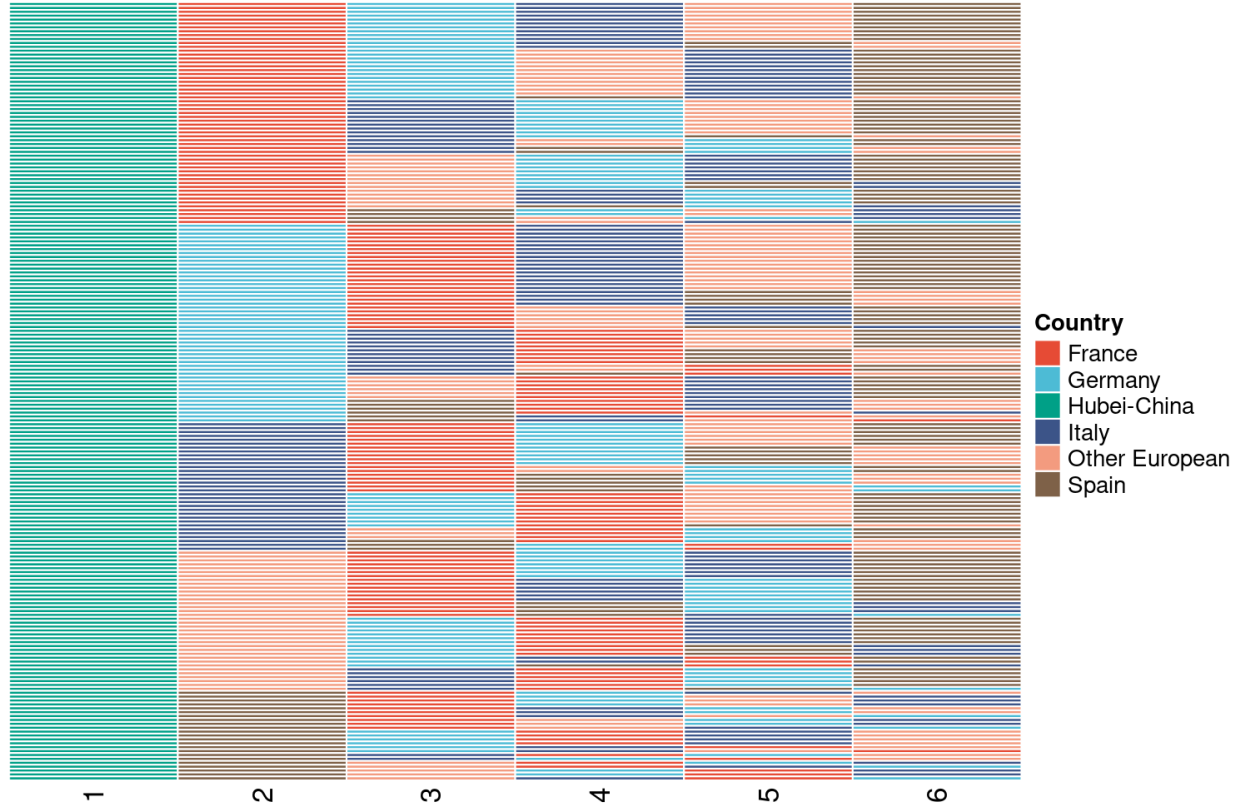


Figure 6: Ordered set of countries by its first introduction, i.e. first case in the country. The color corresponds to the first country with a case, and the size of the box with the country name / ribbon / flux is the number / proportion of trajectories that have that particular order of first introductions.

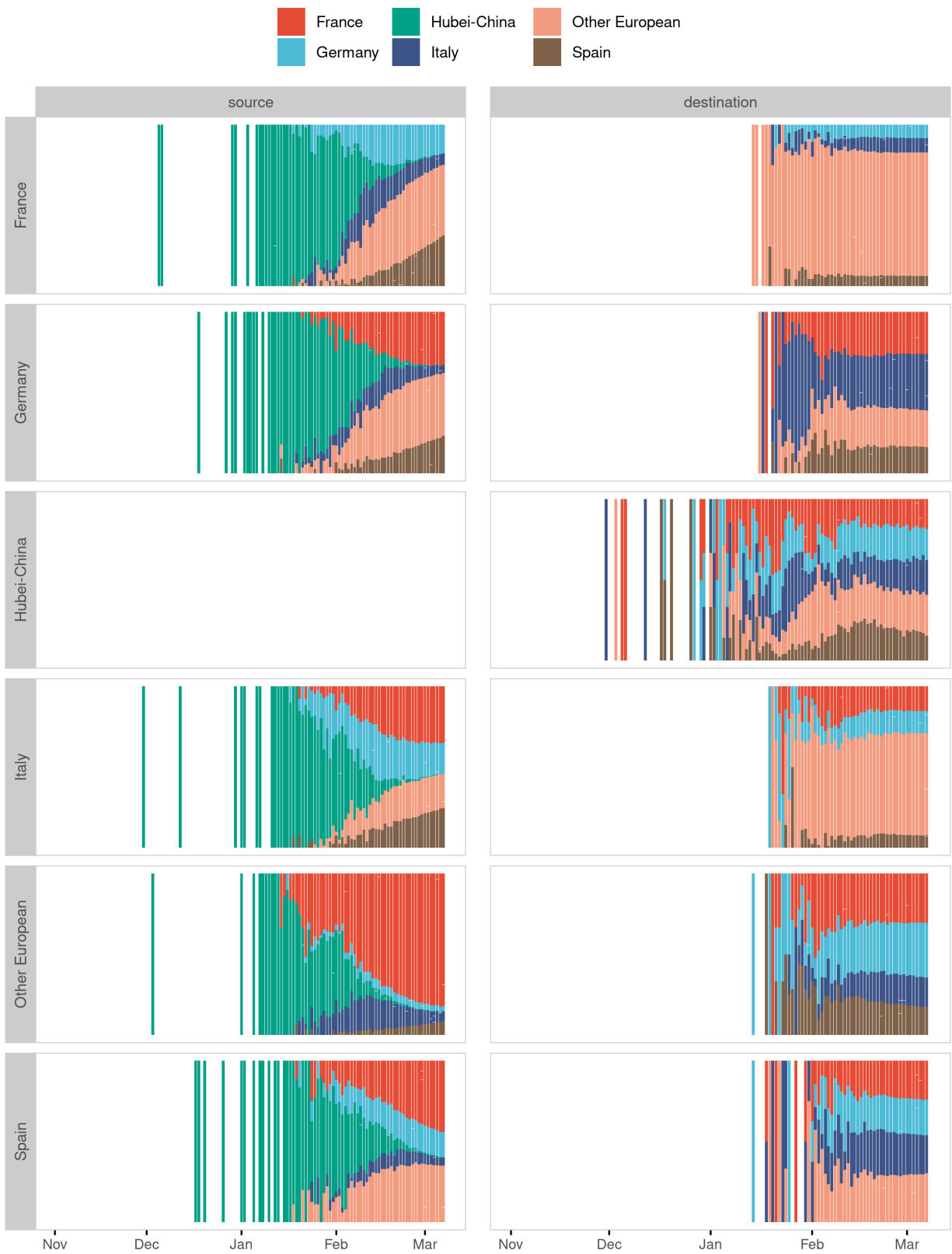


Figure 7

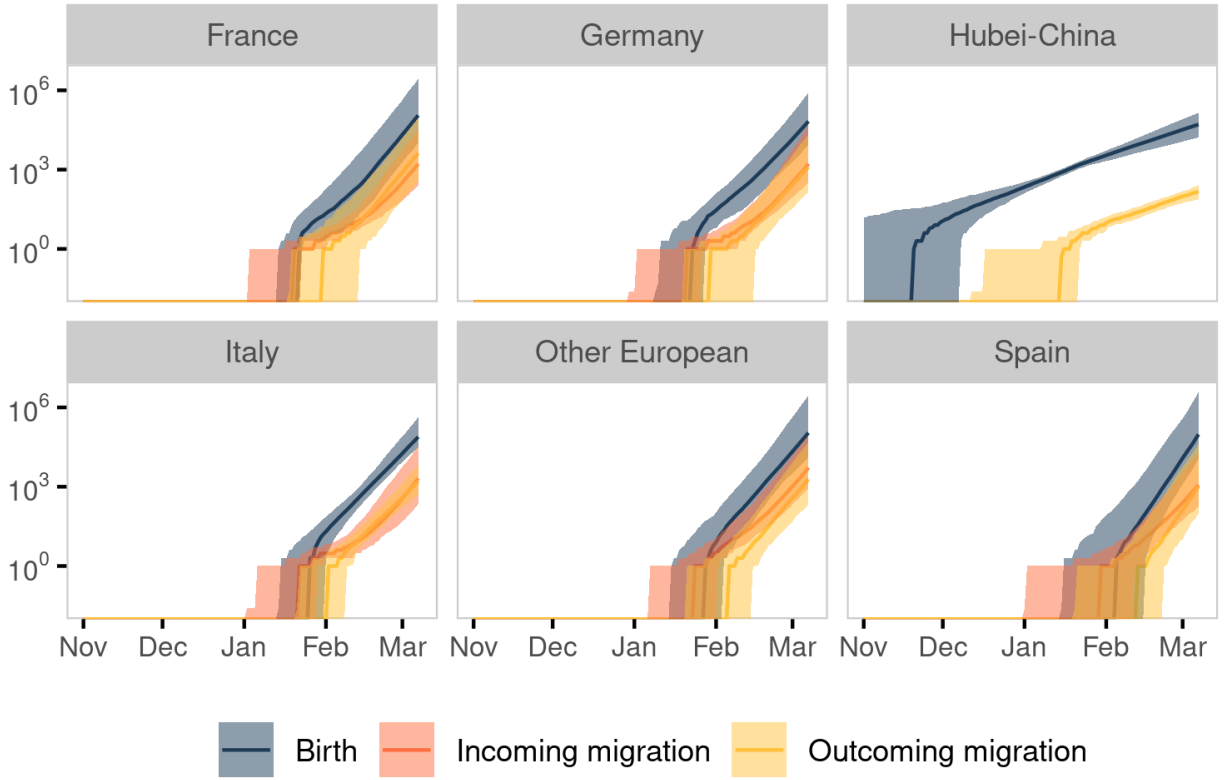


Figure 8: Mean inferred cumulative number of events (births and migrations) over time. A subsampled set of 200 inferred trajectories is plotted.

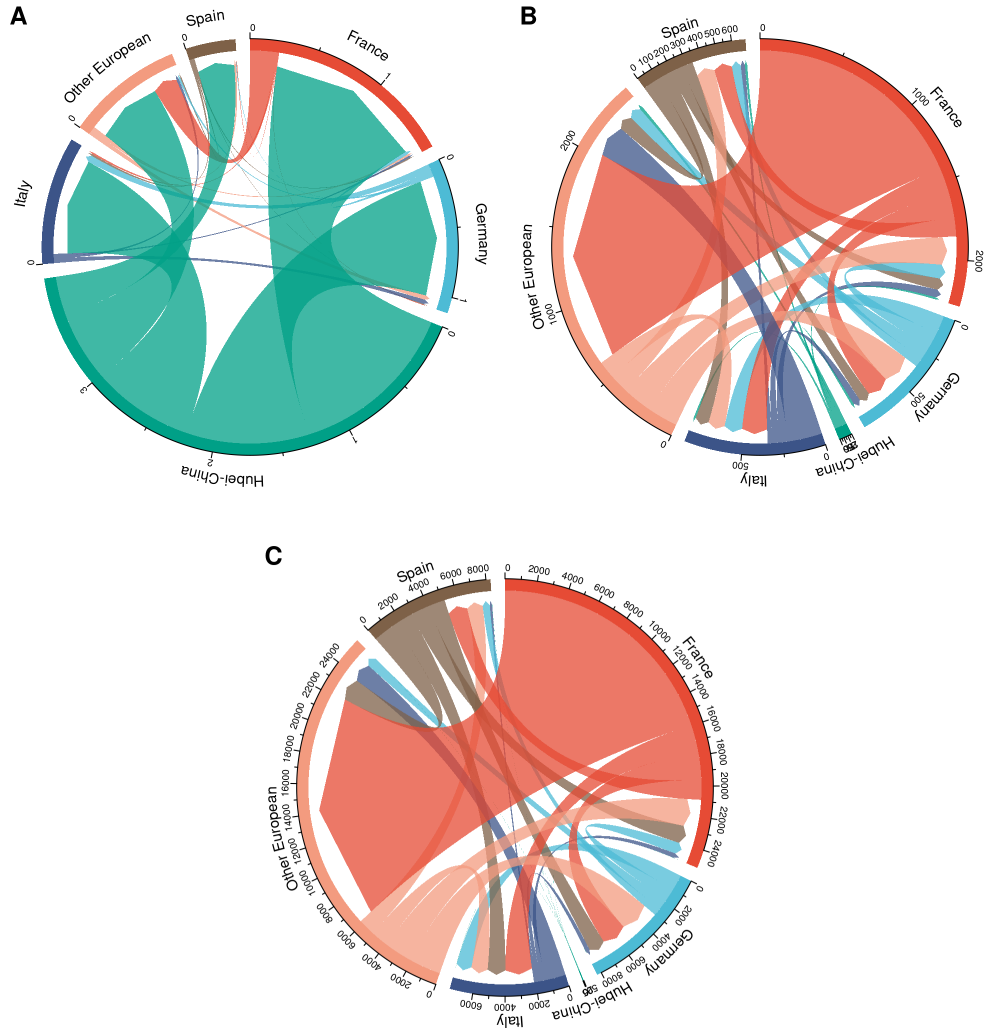


Figure 9: Migration flux among demes over the three periods defined in the analysis.