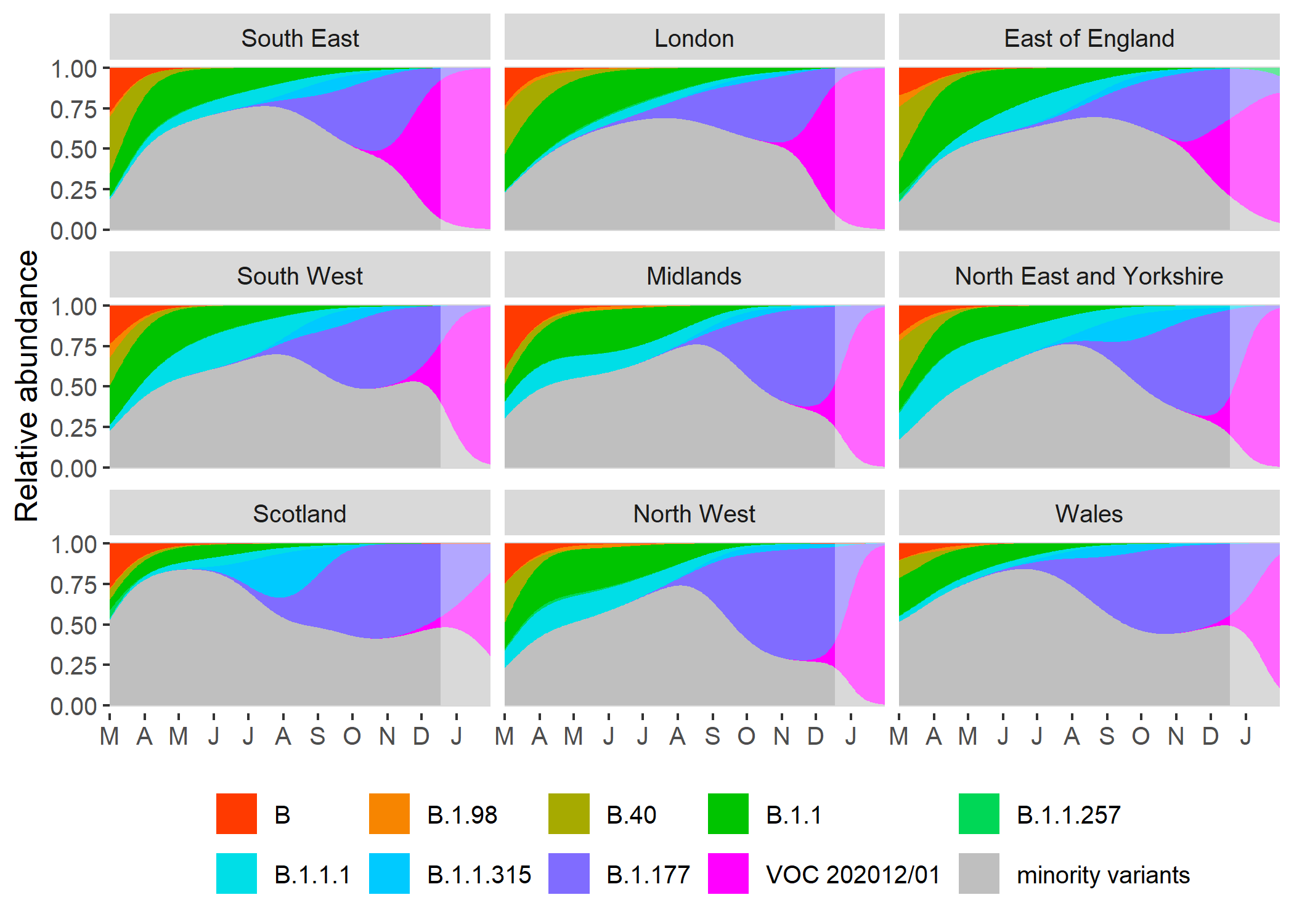
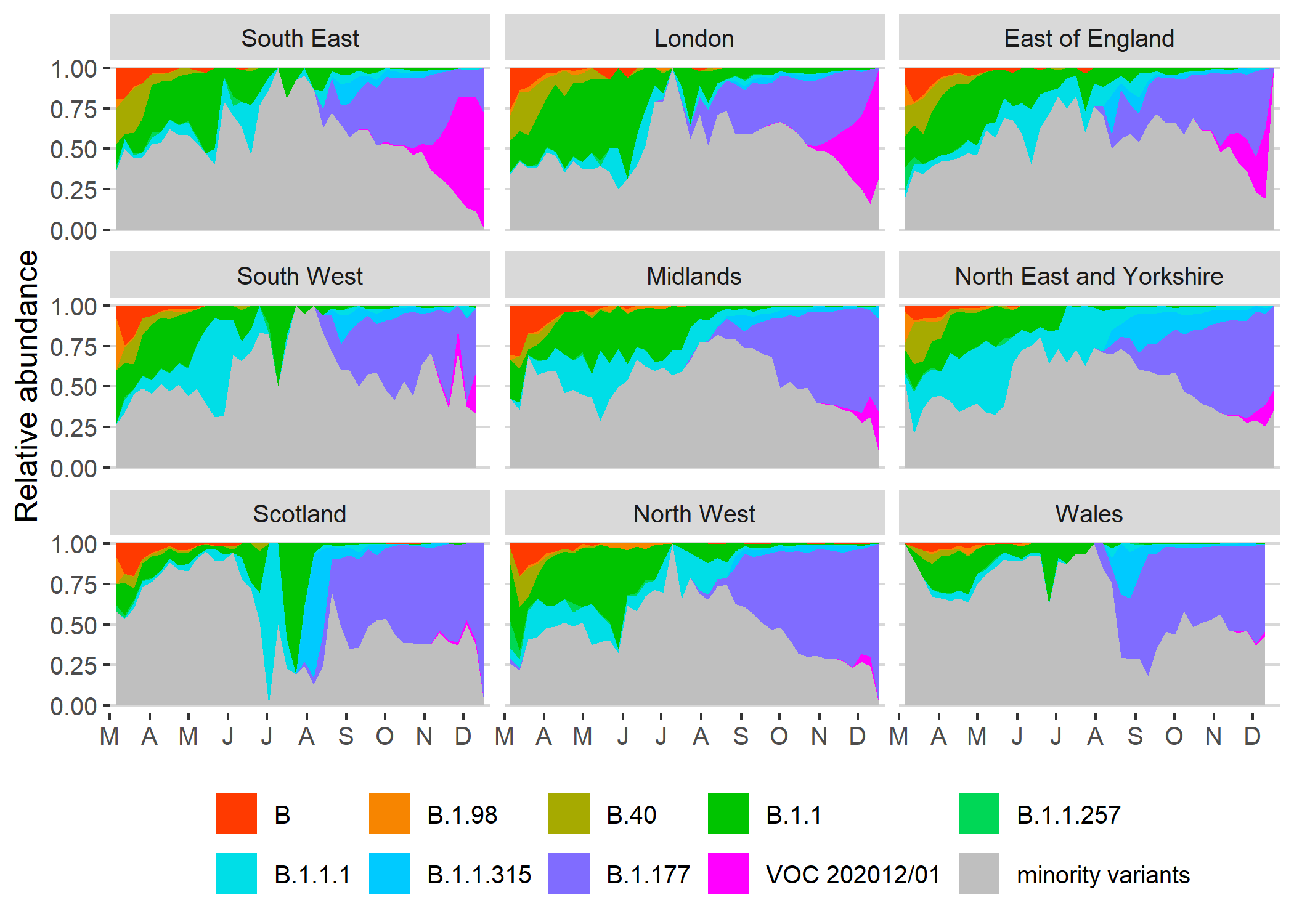
**FIGURES**

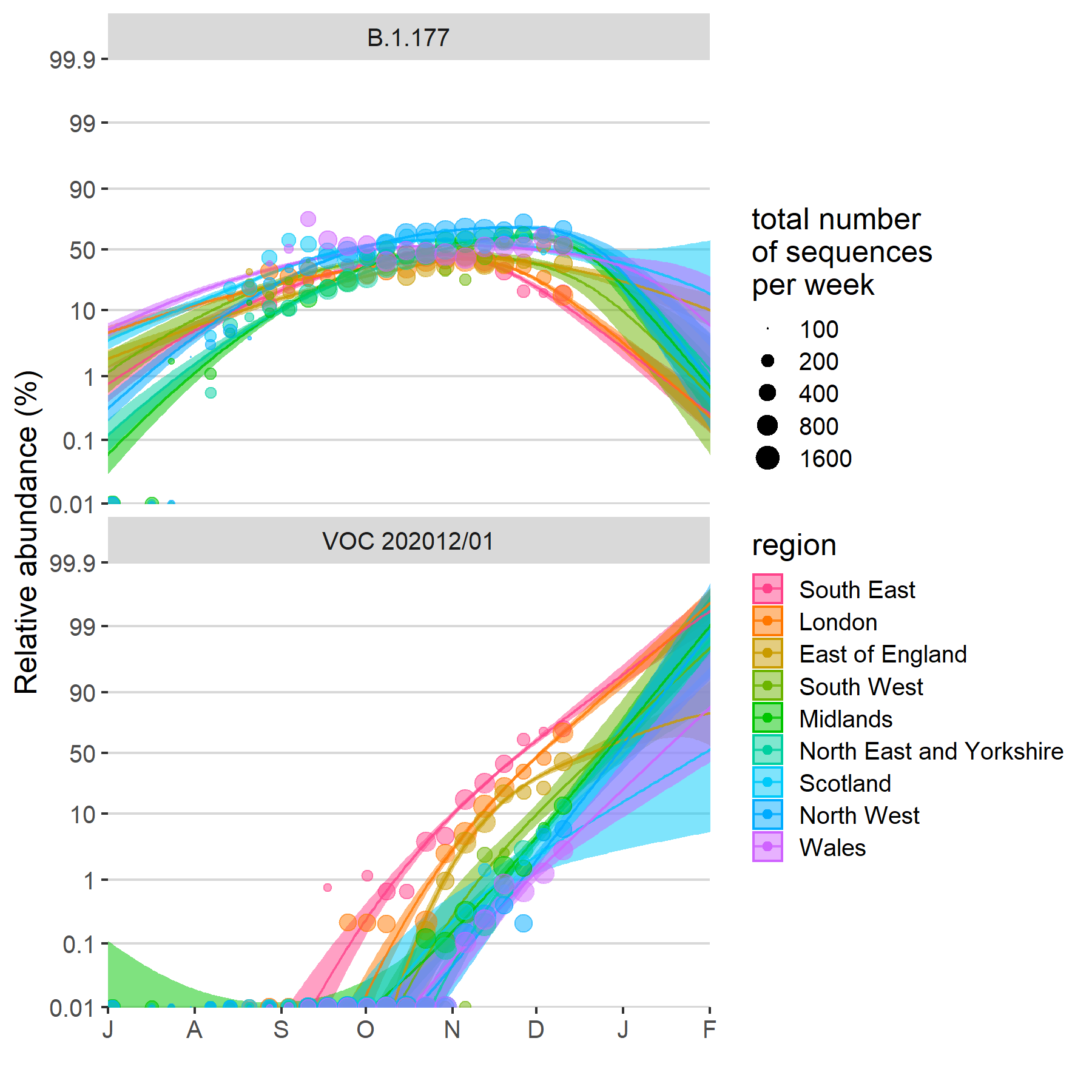
****

**Fig. 1.** Muller plots ofthe relative abundances of the major SARS-CoV2 variants over time in different regions in the UK, based on a multinomial spline fit to the COG-UK sequence data (model 1 in Table 1). This model allows for different rates of spread of the different variants per region as well as some time & region-specific shifts in the rate of displacement, e.g. due to differential influx from other regions. A model extrapolation until the end of January is shown (shaded area), and the expected spread of VOC 202012/01 is highlighted in magenta. The minority variants shown in grey are 440 circulating SARS-CoV2 strains that never reached >15% in any week overall. 95% confidence intervals on the estimated relative abundance of variants B.1.177 and the VOC 202012/01 are shown in Fig. S2.

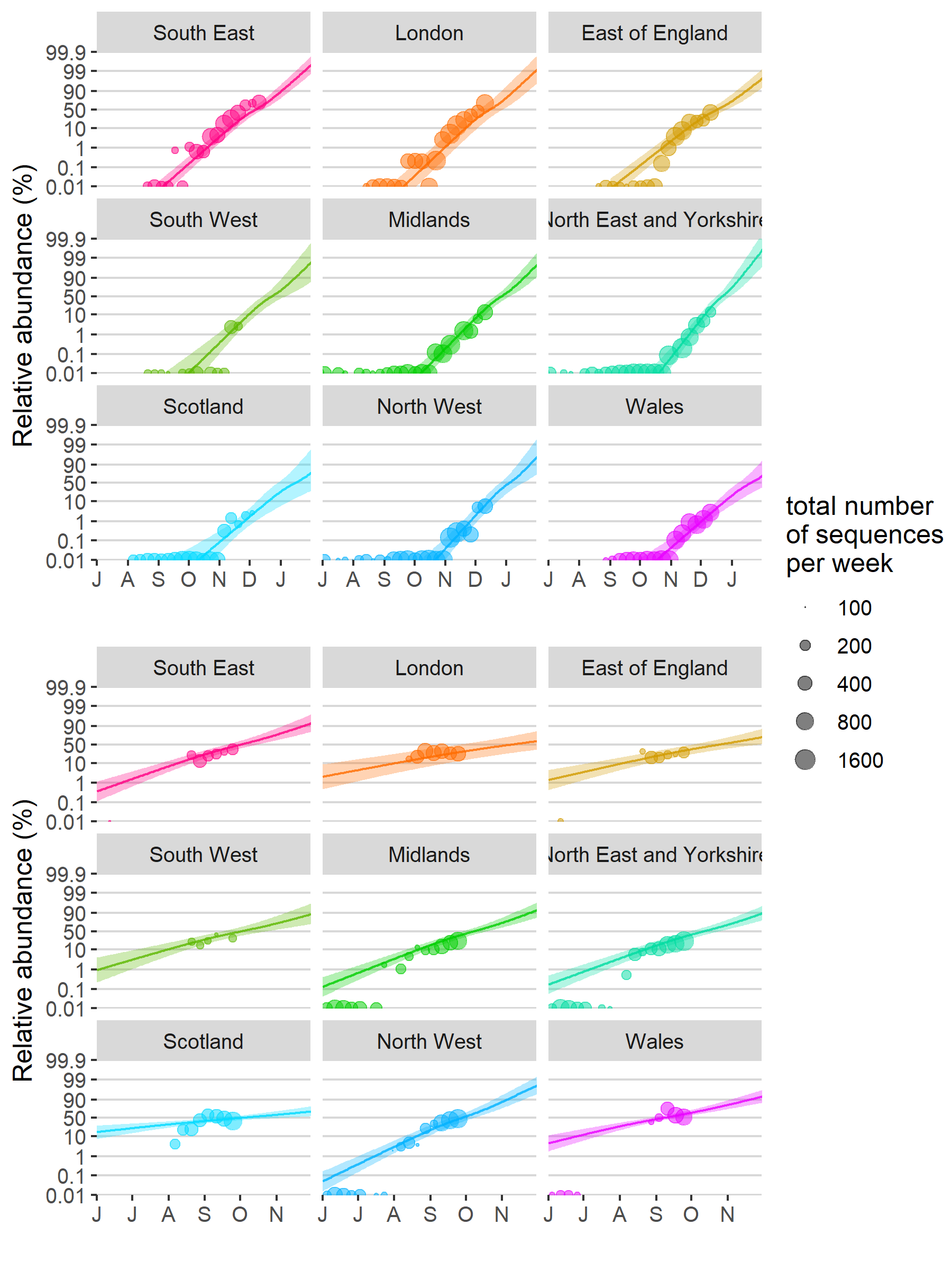
**SUPPLEMENTAL FIGURES**



**Fig. S1.** Muller plots of the relative abundance of the major SARS-CoV2 lineages (reaching at least 15% in any week overall) in different NHS regions across the UK, based on the raw COG-UK sequencing data, aggregated by week. The remaining minority variants comprise a collection of a total of 440 lineages. Note that the large fluctuations seen in July & August in some regions such as Scotland are caused by low sample size.

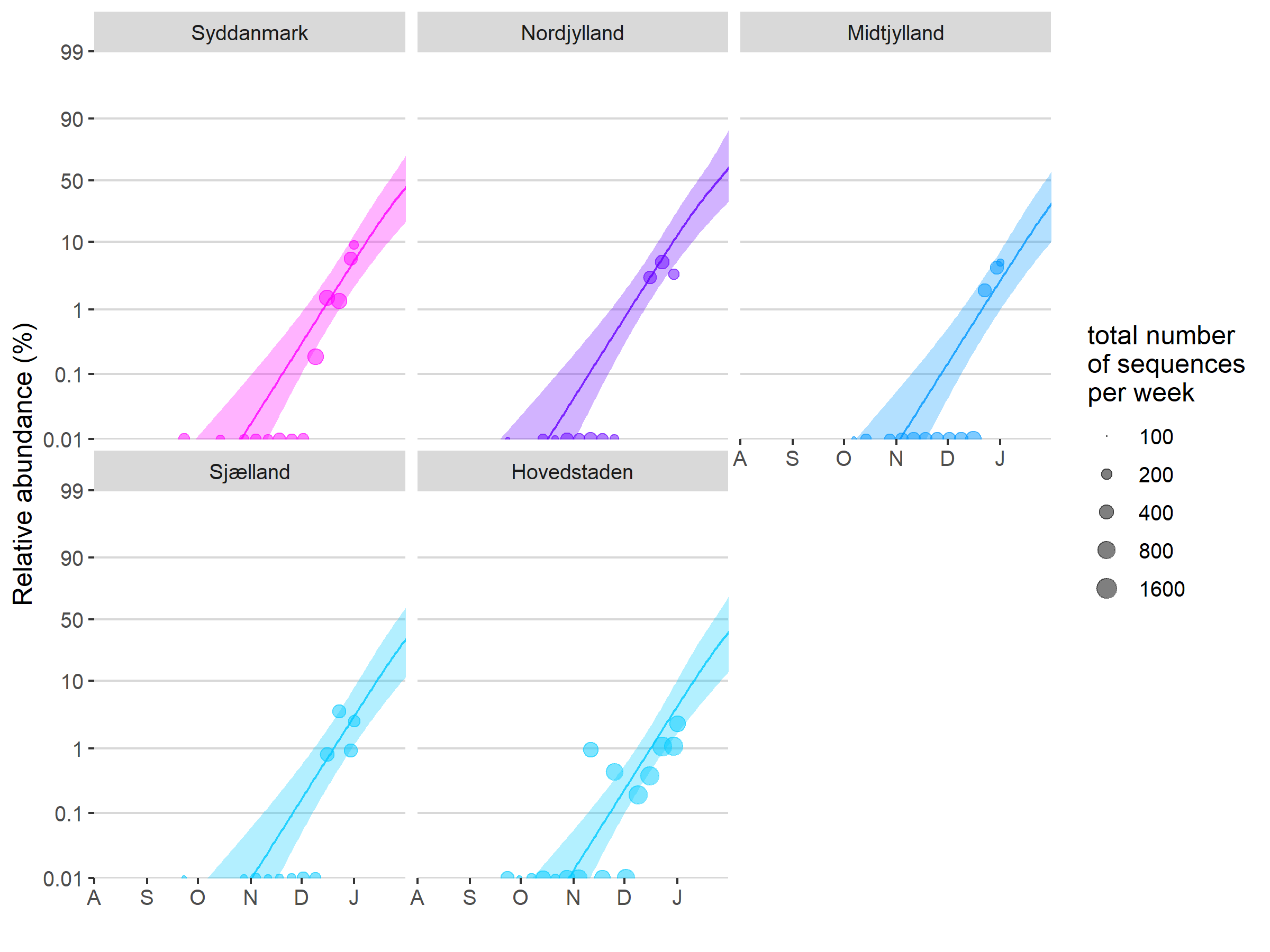


**Fig. S2.** Fitted spread of variants to B.1.177 and VOC 202012/01 estimated from a multinomial spline model by NHS region (model 1 in Table 1 and Fig. 1) with 95% confidence intervals and per-week aggregated raw proportions, shown on a logit (log(odds)) scale. The ca. 3 times faster rate of spread of VOC 202012/01 compared to B.1.177 is apparent (cf. Δr values in Table 1). The excellent linearity on a logit scale for VOC 202012/01 allows us to realistically model the spread of this variant using spatially more fine-grained binomial GLMMs (carried out the level of LTLAs), using a subset of the data from August 1 2020 onwards. Likewise, a binomial GLMM was used to model the spread of variant B.1.177 for the period between July 1st 2020 and September 30 2020, before it starting to be displaced by VOC 202012/01.

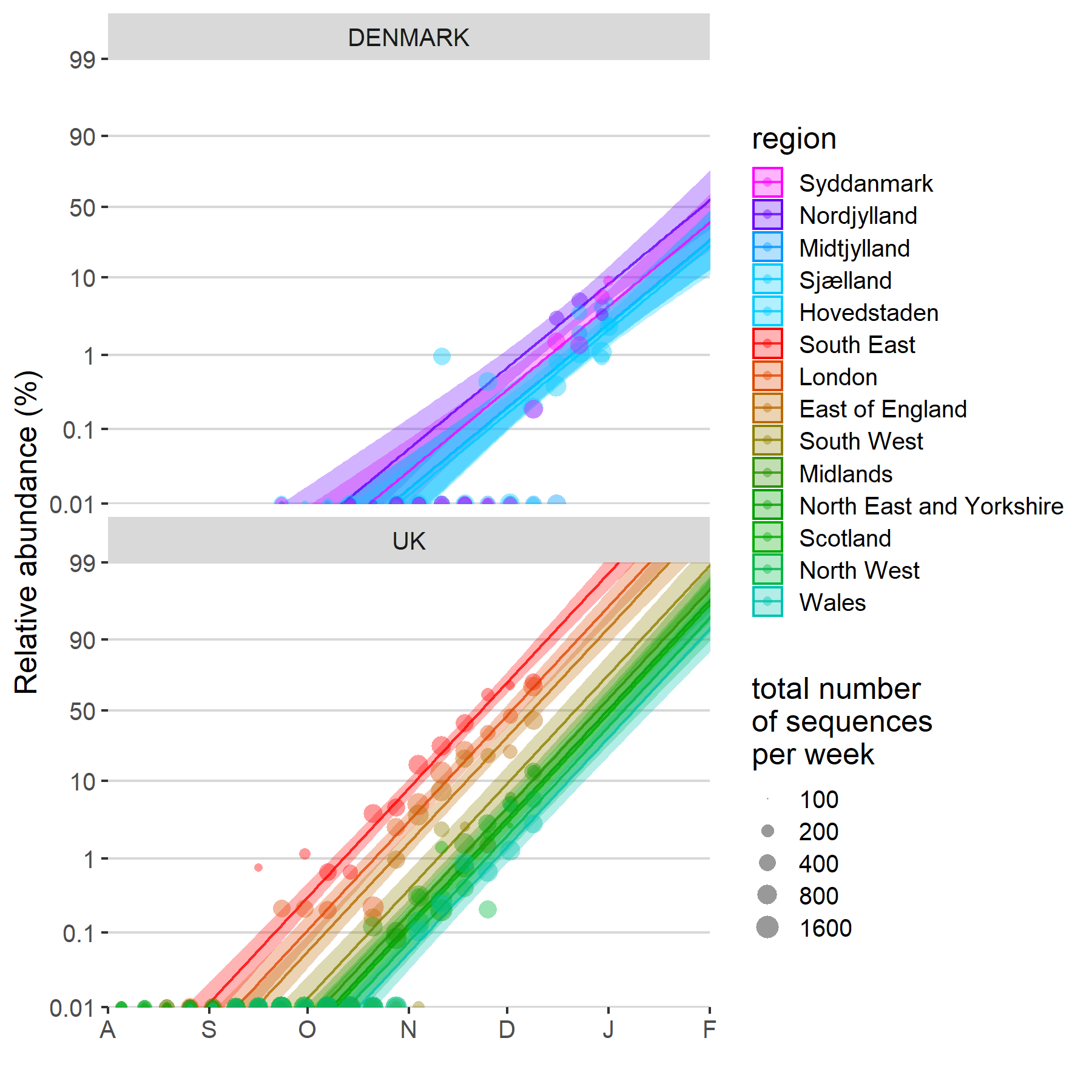


A  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
B

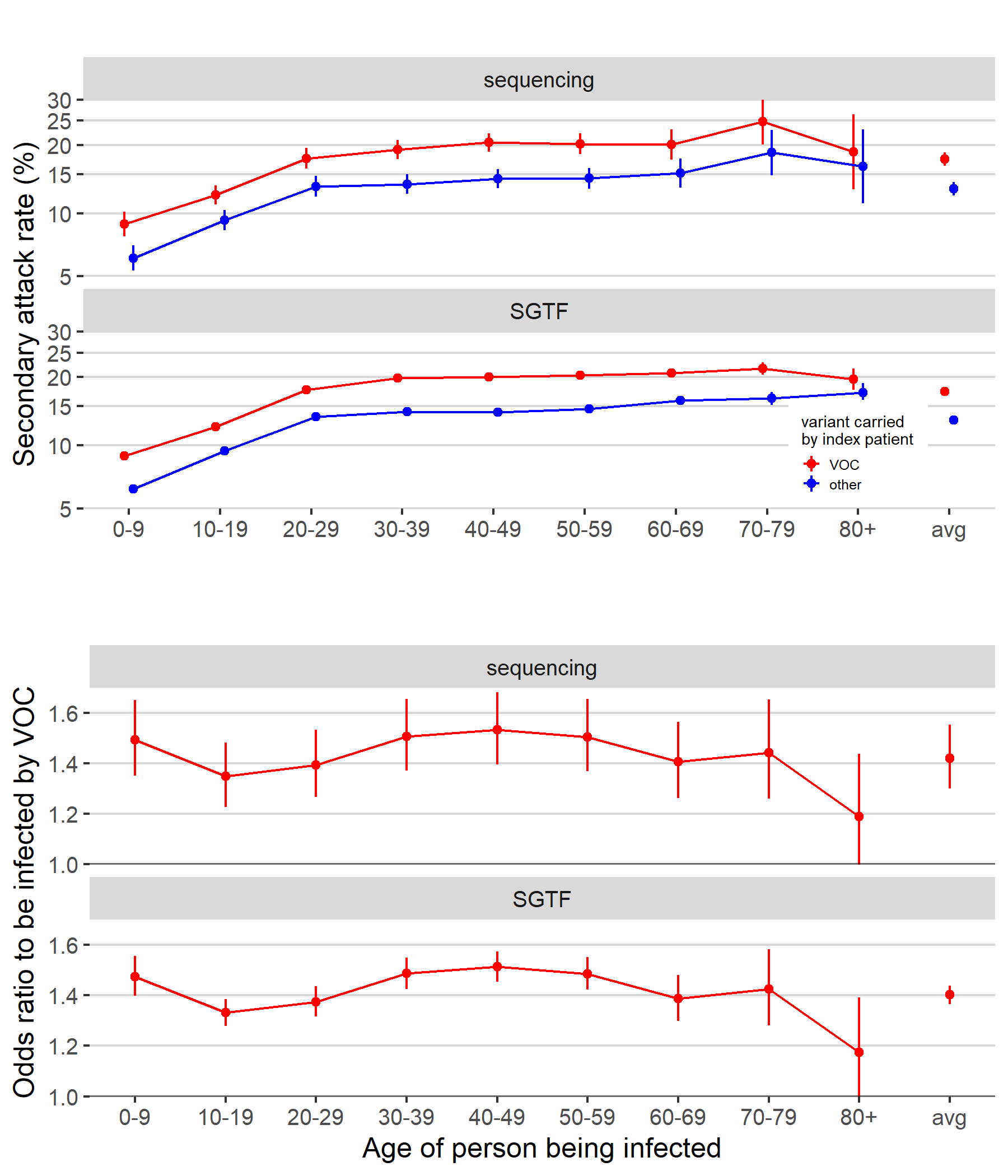
**Fig. S3.** Binomial GLMMs with separate-slopes by region (models S1 in Table S1 and 2e in Table 1) show that VOC 202012/01 has been displacing all other SARS-CoV2 at a near-constant rate across different regions in the UK (A), with pairwise Tukey posthoc tests for differences in slopes across regions only demonstrating a slightly slower rate of displacement in the East of England vs. in the North East and Yorkshire (*z* ratio = -3.68, *p* = 0.007, all other *p* > 0.05). In addition, a common slope model with a constant rate of spread in different regions had a better BIC value (model 2a in Table 1). By contrast, variant B.1.177, which in the UK became the major strain at the end of September, had a much lower competitive advantage in comparison with the minority variants that it displaced, evident from a ca. 3 times lower slope on a log(odds) scale (Table 1). In addition, pairwise Tukey posthoc tests for differences in slopes across regions demonstrate significant cross-regional variation in the rate of spread of this variant (12 out of 36 pairwise comparisons with *p* < 0.05), and a model with separate slopes per region provided the best fit based on the BIC criterion. This supports the idea that VOC 202012/01 enjoys a consistent competitive advantage, whilst the small competitive advantage enjoyed by variant B.1.177 may have been largely, though perhaps not exclusively, the result of stochastic introduction events, e.g. linked with travel to Spain, where it was first observed.



**Fig. S4.** Independent estimate of the rate at which VOC 202012/01 is displacing other variants based on the random sequencing of SARS-CoV2 strains in Denmark, reported on an aggregated per-week basis (for week 39 of 2020 until week 1 of 2021). A binomial GLMM with a common slope across regions and an observation-level random effect to take into account overdispersion fitted the data best, based on the BIC criterion, and resulted in an estimated selection rate Δr of 0.10 [0.07, 0.12] 95% CIs (Table 1, model 3a). In addition, a model with separate slopes per region showed that there were no significant differences in the slopes and implied rates of spread across regions (pairwise Tukey posthoc tests for differences in slope, all *p* > 0.05).



**Fig. S5.** Estimates of the rate at which VOC 202012/01 is displacing other variants across Denmark and the UK, based on a joint fit of the per-week aggregated data from both countries. A binomial GLMM with separate slopes per country but identical slopes per region nested within country provided the best fit based on the BIC criterion, resulting in selection rates Δr estimated for Denmark and the UK of 0.08 [0.07, 0.10] 95% Cis and 0.11 [0.10, 0.12] 95% Cis, respectively (Table 1, model 3b). The rate by which the VOC displaces resident variants was slightly but significantly lower in Denmark than in the UK (*z* ratio = 2.60, 2-sided *p* = 0.009), but this might be due to the presence of other resident strains in Denmark than in the UK.



A  
  
  
  
  
  
  
  
  
  
B

**Fig. S6.** Analysis of age-stratified secondary attack rates, based on data reported by Public Health England (data derived from the COG-UK dataset, the PHE Second Generation Surveillance System and NHS Test and Trace). A binomial GLM with data type (sequence data or S-gene target failure), age group of the person being infected, and variant (VOC 202012/01 or not) plus all first order interaction effects shows that the odds to be infected by an index patient carrying the VOC is consistently higher than by those carrying other variants (A). Sidak posthoc tests show the odds to be infected by the VOC to be significantly greater than by a non-VOC variant for nearly all age groups (for all age groups and both data types 2-sided *p* < 1E-7, except for 80+ where *p* = 0.07 and 0.06 for sequencing and SGTF data, respectively). The mean probability for secondary contacts to become infected in function of age was not significantly different across both types of data (no significant data type by age interaction effect, Type III test, *χ*28=2.90, *p* = 0.94) and there was also no difference in the estimated increased odds to be infected by a VOC vs. a non-VOC index patient (no significant data type by variant interaction effect, Type III test, *χ*21=0.09, *p* = 0.77). The mean odds ratio to be infected by an index patient carrying the VOC vs. a non-VOC variant across all age groups and both data types was 1.41 [1.34, 1.48] 95% CIs. The relative susceptibility to be infected by the VOC showed little variation in function of the age of the person being infected, with only the 40-49 category being slightly more susceptible to be infected by a VOC vs a non-VOC carrying index patient than average (measured in terms of difference in log odds ratios, Sidak age group x variant interaction contrasts, *z* ratio = 3.45, *p* = 0.01, all other *p* > 0.05).