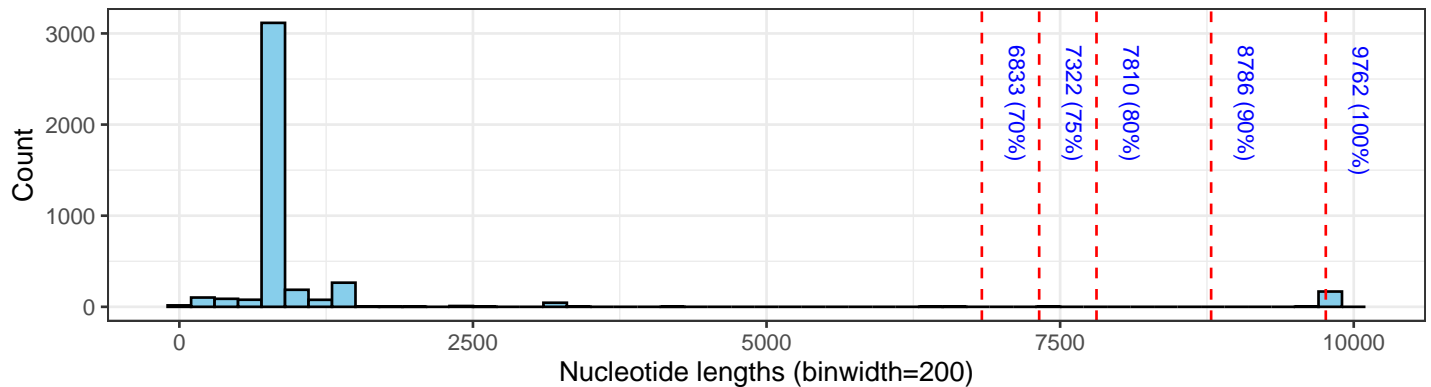


Background

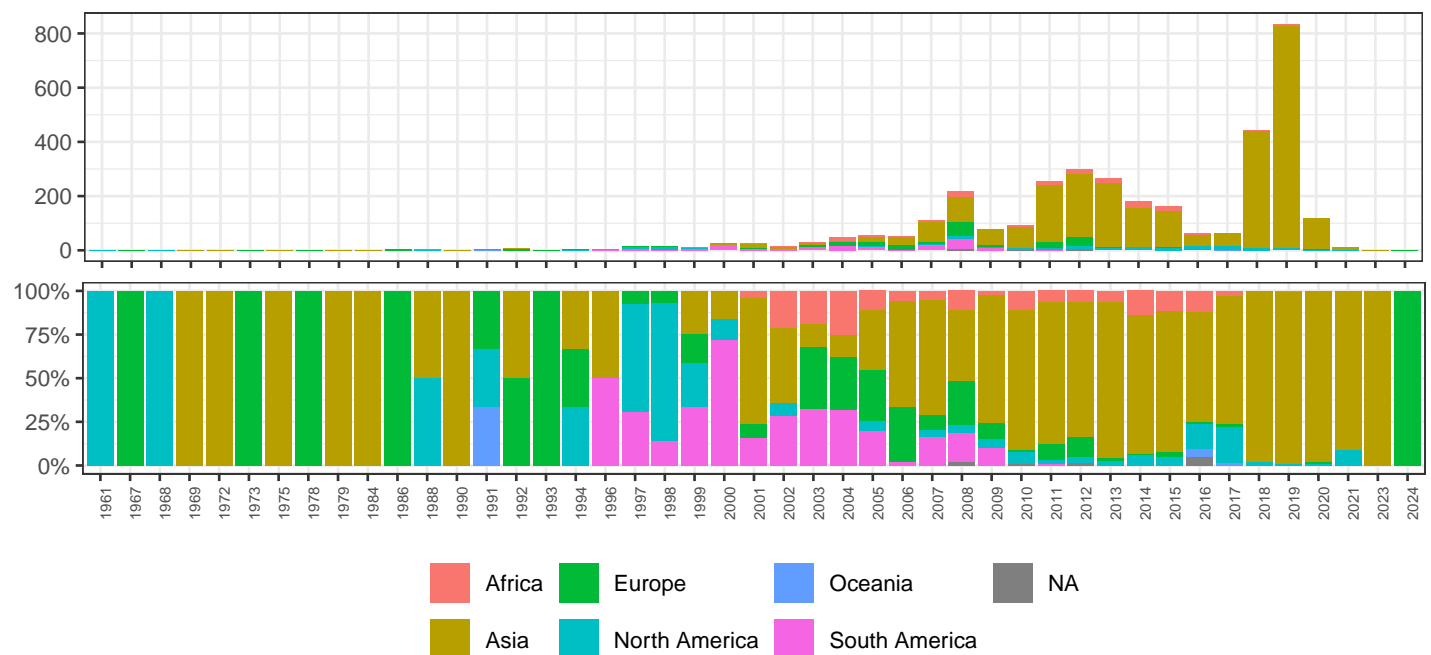
Exploratory Graphics

NCBI GenBank records

Diagnostic plot to estimate min-length filter for phylogenetic analysis (all data = 4165)



Frequency and proportion (n=3522)



1 How many rubella samples are missing strain name?

```
(number_missing_strian_name = data %>% subset(strain == accession) %>% nrow())  
## [1] 2924  
  
(total_rubella_records = nrow(data))  
## [1] 4165  
  
(percentage_missing = number_missing_strian_name/total_rubella_records)  
## [1] 0.7020408
```

Table 1: Top 25 most frequent sequence submitters with their region and countries

authors	n	regions	countries
liu et al.	542	Asia	China
zhu et al.	514	Asia	China
zhou et al.	301	NA, Europe, Oceania, North America, Asia	NA, Russia, Germany, New Zealand, US
kanbayashi et al.	189	Asia	Japan
saikusa et al.	184	Asia	Japan
chen et al.	132	Asia, North America, South America, Europe, Africa	China, USA, Israel, Panama, El Salvador
deng et al.	128	Asia	China
cheng et al.	124	Asia	Taiwan
abernathy et al.	109	Asia, South America, North America, Europe, Africa, NA	Philippines, Peru, USA, United Kingdom
woo et al.	103	NA, Asia	NA, China, Hong Kong
zheng et al.	101	NA, South America, Asia, Africa, North America	NA, Peru, China, Guyana, Côte d'Ivoire
mori et al.	81	Asia	Japan
yu et al.	63	Asia	China
wang et al.	59	Asia	China, South Korea
vauloup-fellous et al.	56	NA	NA
jin et al.	52	Africa, Europe	Ethiopia, Turkey, United Kingdom
ishikawa et al.	50	Asia	Japan
katow et al.	50	NA, Asia	NA, Japan, Myanmar, Mongolia
figueiredo et al.	49	NA, South America, Europe	NA, Brazil, Portugal, Peru, Finland, Ge
yasui et al.	49	Asia	Japan
akiyoshi et al.	41	Asia	Japan
shulga et al.	41	Europe, Asia	Russia, Kazakhstan, Ukraine
saito et al.	40	Asia	Japan
frey et al.	39	NA	NA
martinez-torres et al.	37	Europe	Spain

Table 2: Top 20 earliest records

date_adjusted	accession	strain	country	authors
1961-01-01	DQ388300	DQ388300	USA	Zhou et al.
1961-01-01	JN635281	JN635281	USA	Kirkness et al.
1967-01-01	DQ388279	DQ388279	Russia	Zhou et al.
1968-01-01	OM735619	OM735619	USA	Chen et al.
1969-01-01	FJ040181	FJ040181	Japan	Gao et al.
1972-01-01	OM735623	OM735623	Israel	Chen et al.
1973-01-01	DQ388297	DQ388297	Russia	Zhou et al.
1975-01-01	OM735621	OM735621	Israel	Chen et al.
1978-01-01	DQ388298	DQ388298	United Kingdom	Zhou et al.
1979-01-01	OM735620	OM735620	Israel	Chen et al.
1984-01-01	DQ255946	JR23	China	Wang et al.
1986-01-01	DQ388286	DQ388286	United Kingdom	Zhou et al.
1986-01-01	DQ388305	DQ388305	United Kingdom	Zhou et al.
1986-01-01	KF792832	RVi/London.GBR/86/1I	United Kingdom	Abernathy et al.
1986-01-01	OM735654	OM735654	United Kingdom	Chen et al.
1988-01-01	JN635285	JN635285	USA	Kirkness et al.
1988-01-01	OM735622	OM735622	Israel	Chen et al.
1990-01-01	OM735631	OM735631	Japan	Chen et al.
1991-01-01	DQ388281	DQ388281	New Zealand	Zhou et al.
1991-01-01	DQ388291	DQ388291	Italy	Zhou et al.

Table 3: Top 20 latest records

date_adjusted	accession	strain	country	authors
2020-02-17	LC528258	LC528258	Japan	Osawa et al.
2020-03-05	MT463288	MT463288	Belarus	Semeiko et al.
2020-03-06	MT292373	MT292373	China	Woo et al.
2020-03-12	LC718028	19-1691	Japan	Mori et al.
2020-03-12	LC718038	19-1691	Japan	Mori et al.
2020-03-22	MT292374	MT292374	China	Woo et al.
2020-08-06	OP242874	OP242874	China	Yu et al.
2021-01-01	OM899300	OM899300	China	Liu et al.
2021-01-01	OM899365	OM899365	China	Liu et al.
2021-01-01	OM899436	OM899436	China	Liu et al.
2021-01-01	OM899437	OM899437	China	Liu et al.
2021-01-01	OM899438	OM899438	China	Liu et al.
2021-01-01	OM899439	OM899439	China	Liu et al.
2021-01-01	OM899440	OM899440	China	Liu et al.
2021-01-01	OM899441	OM899441	China	Liu et al.
2021-01-01	OM899442	OM899442	China	Liu et al.
2021-01-01	OM899443	OM899443	China	Liu et al.
2021-09-21	PP079378	PP079378	USA	Chen et al.
2023-05-20	LC783969	RVs/Osaka.JPN/20.13/2[2B]	Japan	Kanbayashi et al.
2024-05-29	PP975510	km47614	Italy	Marinelli et al.