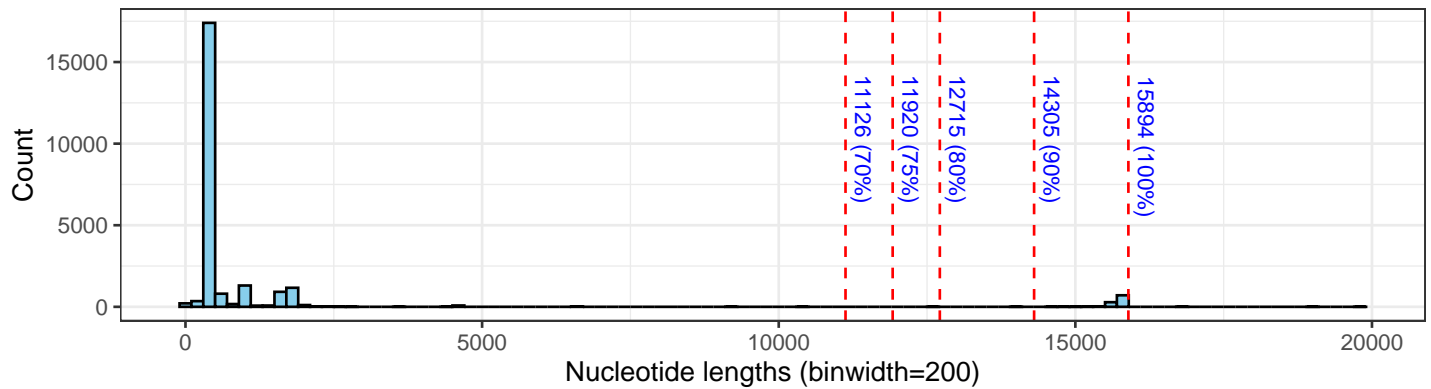


Background

Exploratory Graphics

NCBI GenBank records

Diagnostic plot to estimate min-length filter for phylogenetic analysis (all data = 23803)



Frequency and proportion (n=18065)

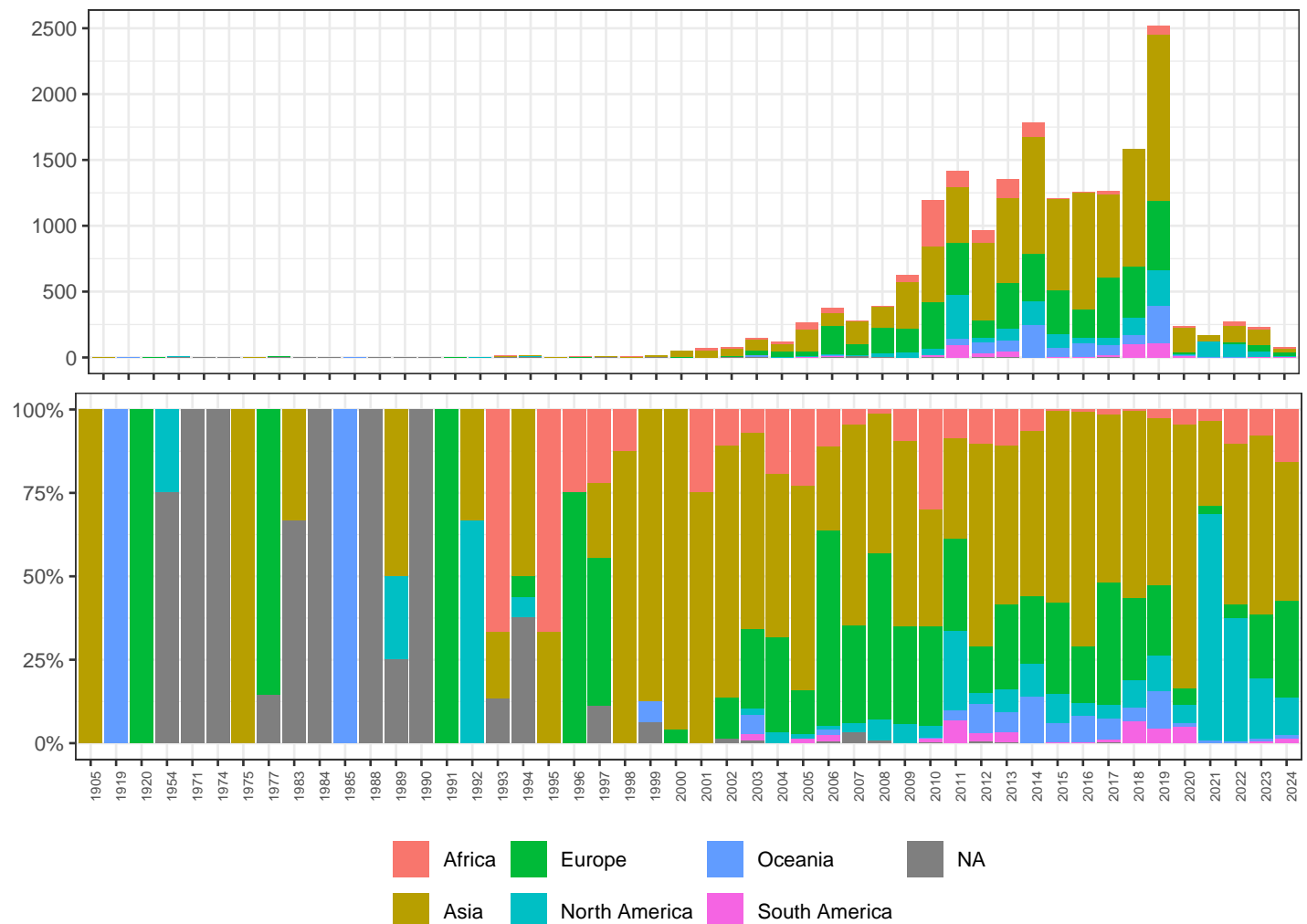


Table 1: Top 25 most frequent sequence submitters with their region and countries

authors	n	regions	countries
rota et al	1444	North America, NA, South America, Africa, Asia, Europe	USA, NA, Venezuela, Guatemala, Amer
penedos et al	954	Europe, Africa	United Kingdom, Romania, Morocco
tran et al	723	Oceania, Europe, Asia	Fiji, Australia, France, Papua New Guir
zhang et al	720	Asia, NA	China, NA
wang et al	711	Asia	China, South Korea
vaidya et al	629	Asia, NA	India, NA
cheng et al	628	Asia	Taiwan, China, Hong Kong
santibanez et al	539	Europe, NA, North America	Germany, NA, Switzerland, Slovakia, Ice
li et al	495	Asia	China, Nepal
mokhtari azad et al	475	Asia	Iran
smit et al	455	Africa, NA	Kenya, Zambia, Namibia, Malawi, South
ji et al	433	Asia	China
krishnaswamy et al	373	Asia	India
lazar et al	337	Europe	Romania
dina et al	331	Europe	France
magurano et al	302	Europe	Italy
chem et al	290	Asia	Malaysia
kessler et al	282	Europe, Africa, North America, Asia	Romania, United Kingdom, Democratic
woo et al	275	Asia, NA	China, Hong Kong, NA
hutse et al	266	Europe	Belgium
hiebert et al	262	North America, NA, Africa	Canada, NA, Uganda
zaidi et al	241	Asia	Pakistan, Afghanistan
bodewes et al	226	Europe	Netherlands
sharma et al	206	Asia, NA	India, NA
kaida et al	196	Asia	Japan

Table 2: Top 25 most frequent host

lower_field	n
homo sapiens	19515
NA	4282
chlorocebus aethiops	3
macaca fascicularis	2
saguinus oedipus	1