

coalescence

June 5, 2024

0.1 Coalescence theory

0.1.1 1) Probability distribution of TMRCA with $n=2$ (sample size)

```
[ ]: # nr of individuals in the population
N <- 500
# over 3000 generations
tmrca <- seq(1,3000,10)
```

```
[ ]: prob_tmrca <- dgeom(x=tmrca, prob=1/(2*N))
```

```
[ ]: plot(x=tmrca, y=prob_tmrca, ylab="Probability", xlab="Generations")
```

0.1.2 2) Probability distribution of TMRCA with $n=2$ and $N \rightarrow \infty$

```
[ ]: N <- 500
tmrca <- seq(1,3000,1)/(2*N) # time is in 2N generations
plot(x=tmrca*(2*N), y=dexp(x=tmrca, rate=1), ylab="Density",
     xlab="Generations", type="l")
```

0.1.3 3) Expected site frequency spectrum

```
[ ]: # with 10 haploid individuals
n <- 10
sfs <- 1 / seq(1, n-1, 1)
sfs <- sfs / sum(sfs)
```

```
[ ]: barplot(sfs, ylab="Proportion", names=seq(1,9,1), ylim=c(0,0.40), xlab="Allele_
     frequency")
```

Which class of mutations is the most common?

0.1.4 Expected folded site frequency spectrum with $n=10$

```
[ ]: folded_sfs <- c(sfs[1]+sfs[9], sfs[2]+sfs[8], sfs[3]+sfs[7], sfs[4]+sfs[6],
     sfs[5])
folded_sfs <- folded_sfs / sum(folded_sfs)
```

```
[ ]: barplot(folded_sfs, ylab="Proportion", names=seq(1,5,1), ylim=c(0,0.50),  
↳xlab="Allele frequency")
```

```
[ ]:
```