Pool-seq variation model

zoe elkins

Pool-seq variance

The following code adapts the analysis performed by Rode et al. 2018. Whereas their analysis was made for haploid data, the following code works for diploid data.

> Nicolas Rode, Yan Holtz, Karine Loridon, Sylvain Santoni, Joelle Ronfort, et al.. How to optimize the precision of allele and haplotype frequency estimates using pooled-sequencing data.. Molecular Ecology Resources, Wiley/Blackwell, 2018, 18 (2), pp.194-203. ff10.1111/1755-0998.12723ff. ffhal-02621308

Load dependencies

```
-- Attaching packages --------------------------------------- tidyverse 1.3.2 --
v ggplot2 3.3.6      v purrr   0.3.4
v tibble  3.1.8      v dplyr   1.0.10
v tidyr   1.2.1      v stringr 1.4.1
v readr   2.1.3      v forcats 0.5.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
```

```
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

Equation 9 function

The `echo: false` option disables the printing of code (only output is displayed).

Vstep 1-3

Run simulations and establish functions for estimating variance in allele frequency estimates due to sampling variance, unequal individual contributions to the DNA pool, and sequencing variance.

Function to sample total number of reads

Function to compute the proportion of reads with the reference allele within the pool of reads

Function to compute the mean and standard deviation of the frequency of the reference allele across many samples

```
[[1]]
[1] 0.4967572

[[2]]
[1] 0.3058073
```

Simulations