

Contigs determination and analysis: *Xenopus laevis*

Bharath Manicka Vasagam, The Barski Laboratory, CCHMC

March 14, 2016

Objective

The objective was to determine the contigs elimination point with the *Xenopus laevis* genome data.

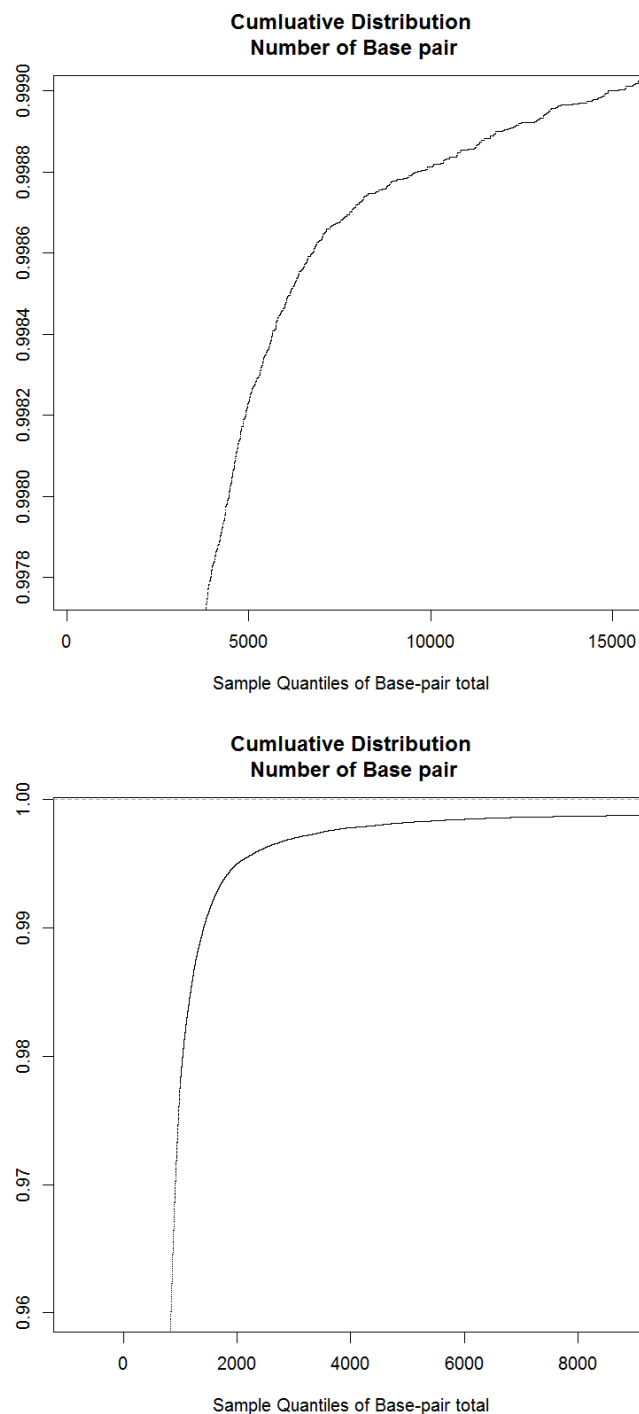
Packages used

1. data.table
2. zoom - for interactive visualisation of plot

Methodology

1. Raw Data file of base pairs of ***Xenopus laevis*** was cleaned and tabulated as Scaffold names and Number of base pairs identified from the file **fasta.R**
2. Then the output's cumulative distribution function was plotted. Sample quantiles of base-pairs, to determine the lowest base pair cut in the file **distribution.R**

zoomed in plots:



From the plots it was determined that we remove contigs below 5000, 10000 base pairs this calculation was done in the file **ques.R**

Result

If base pairs below 10,000 was removed, we would loose 862 genes or 3.56% of total genes If base pairs below 5,000 was removed, we would loose 835 genes or 3.47% of total genes If base paired below 1,000 was removed, we would loose 580 genes or 3.01% of total genes