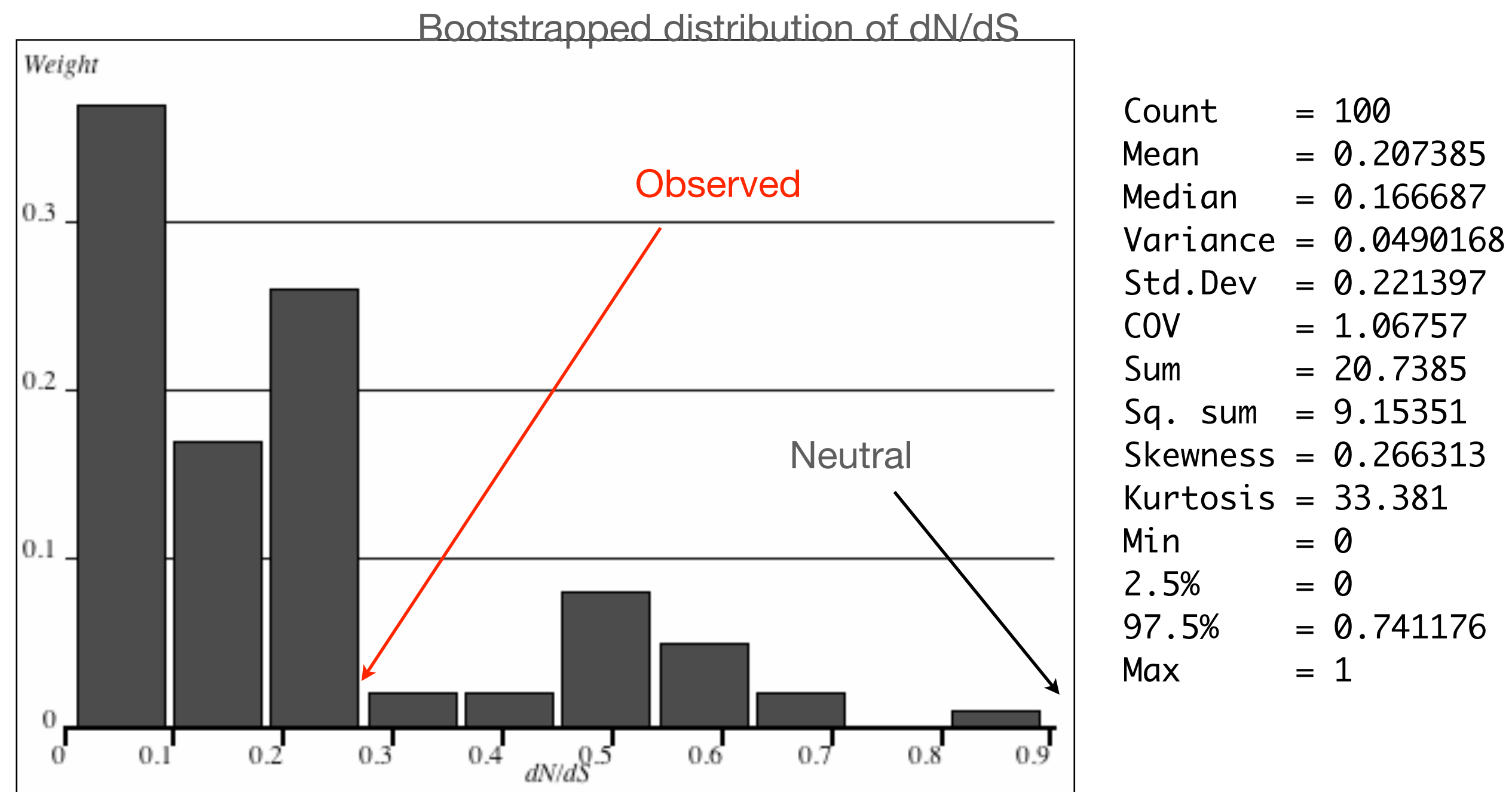


NG86 example

- How reliable is the inference based on only **6** codons?
- Obtain sampling variance via bootstrap (or by limiting approximations)
- In this case, dN/dS is **significantly** less than 1.0 ($p \sim 0.01$)



NG86 limitations: multiple substitutions

- How many synonymous and how many non-synonymous substitutions does it take to replace **CCA** with **CAG**?
- **Assume** the shortest path (minimum of 2 substitutions)
 - CCA (Proline) \Rightarrow CAA (Histidine) \Rightarrow CAG (Glutamine)
 - CCA (Proline) \Rightarrow CCG (Proline) \Rightarrow CAG (Glutamine)
- Average over the two possible paths: **0.5** synonymous and **1.5** non-synonymous substitutions.
- Intuitively, paths should **not** be equiprobable, e.g., because it should be more expensive to route evolution through (presumably) suboptimal intermediate amino-acids.