

# PROGRESS REPORT

(Or, what I've been up to)

---

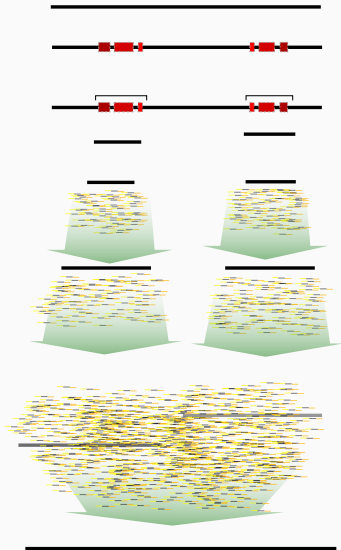
Nick Waters

December 4, 2016

National University of Ireland, Galway  
James Hutton Institute

RIBOSEED

---



**Reference genome with existing annotations**  
(or optional scanScaffolds.sh annotations  
with Barrnap)

**RiboSelect: identify and cluster ribosomal regions**

**RiboSnag: extract ribosomal regions**

**riboSeed:**

**Initial Seeding**  
**De novo assemblies**

**Next Seeding**  
**Next de novo assemblies**

**Final Assembly with seeds as trusted contigs**

**Complete Genome**



- Stochastic results, with no known indicator
- Occasional missassemblies
- Theory issues
- Time investment



- Preprocessing: no longer need properly matched .fasta's and .gb's
- Scripts to streamline sge batch submissions
- riboSnag now plots out interesting statistics
- riboSnag is now optional



- riboSeed maps to all extracted regions at once instead of individually
  - Minimizes complexity caused by multiple mappings
  - Shorter run time
- Code coverage: 61% (87% without main function)
- Object oriented refactor
- 65% rewrite



- Works on synthetic genome with synthetic reads
- Works on a few genomes with synthetic reads
- Still working on benchmarking the rest of the test strains



- Faster, more maintainable code
- Robust theory
- Thoroughly testing