GRC MEETING: 2016 - 2017

Comparative Genomics of Soil-adapted *E. coli*

Nick Waters

March 27, 2017

Department of Microbiology
School of Natural Sciences
National University of Ireland, Galway

Overview

Outline

- Genome Assembly
- Phylogenetics, MLST, and Collection Overview
- Virulence
- Plans

Genome Assembly

State of Available Bacterial Genomes

Table: Bacterial Genome Completion as of 4/1/17

Total	Complete genome	Chromosome	Contig	Scaffold
85799	6255	1143	38429	39972

State of Sequencing Data Availability

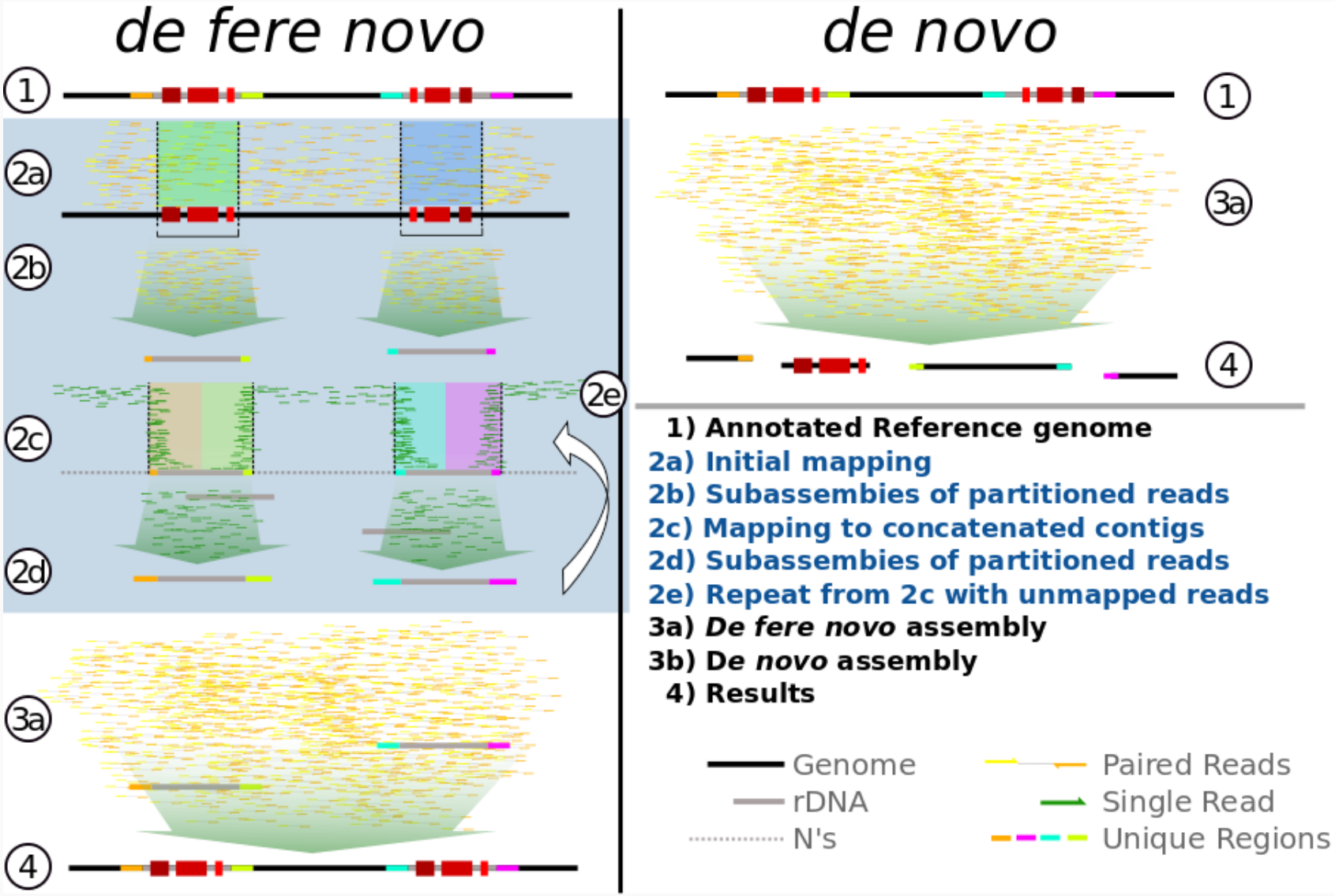
Table: Hits per Search Term in NCBI's SRA

Search term	hits	percentage
'illumina'	2242225	(94.27)
'pacbio'	21131	(0.89)
'ion'	30560	(1.28)
'roche'	42445	(1.78)
'oxford'	12301	(0.52)
'solid'	29791	(1.25)
Total	2378453	(100)

Improving Illumina Short Read Assembly

1. Within a taxonomic group, GC content is largely conserved.
2. Within a taxonomic group, genome size is largely conserved.
3. Bacterial genomes are dense.
4. Nucleotide order is not random.

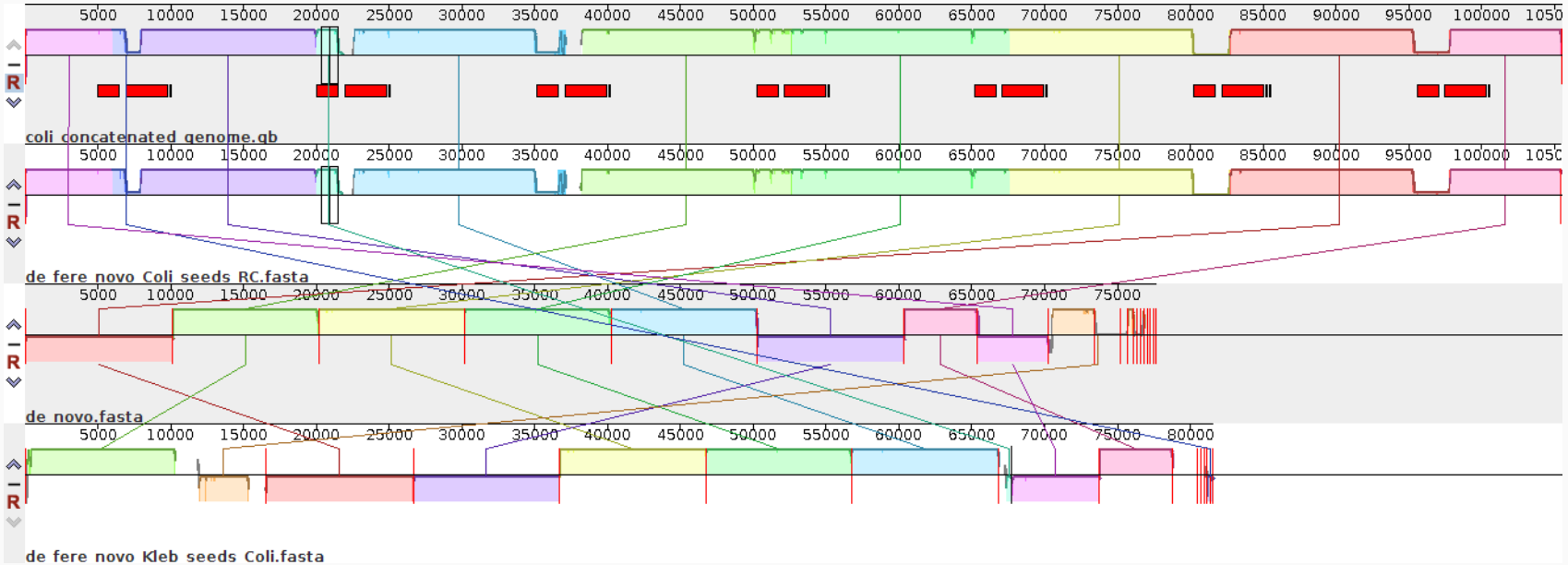
riboSeed


de fere novo
de novo
1
2a
2b
2c
2d
2e
3a
4
1) Annotated Reference genome
2a) Initial mapping
2b) Subassembies of partitioned reads
2c) Mapping to concatenated contigs
2d) Subassembies of partitioned reads
2e) Repeat from 2c with unmapped reads
3a) De fere novo assembly
3b) De novo assembly
4) Results
Genome
rDNA
N's
Paired Reads
Single Read
Unique Regions

Benchmarking riboSeed

- Simulated Reads on a Simulated Genome
- Simulated Reads on Real Genomes
- Hybrid Assembly Validation
- GAGE-B Dataset

Benchmarking: Simulated Genome



Benchmarking: Simulated Reads on Sakai

Table: riboSeed on simulated Sakai reads

Coverage	Ref. rDNAs	De novo(skip, miss)	De fere novo (skip, miss)
10	7	0 (7, 0)	3 (4, 0)
20	7	0 (7, 0)	6 (1, 0)
50	7	0 (7 , 0)	6 (1, 0)
100	7	0 (7 , 0)	6 (1, 0)

Benchmarking: Hybrid Assembly

Table: riboSeed on Pseudomonas aeruginosa strain BAMCPA07-48

Coverage	Re. rDNA	De novo (skip, miss)	De fere novo (skip, miss)
200	4	1(3, 0)	4(0, 0)

riboSeed: next steps

- Finish benchmarking
- Applicability to fungal and archaeal genomes
- Improve Signal-to-noise ratio

Aggregating Metadata

Experimental Metadata

Problem: lack of single repository for the data related to the collection

- Sample Isolation data
- Phenotypic tests
- Phylotyping

In silico Metadata

- Library preparataion and sequencing QC
- Average nucleotide identity
- In silico Clermont PCR
- MLST

Results and Implications

After eliminating isolates falling beneath the 95% ANI threshold and those with poor sequencing data, the collection consists of 153 isolates.

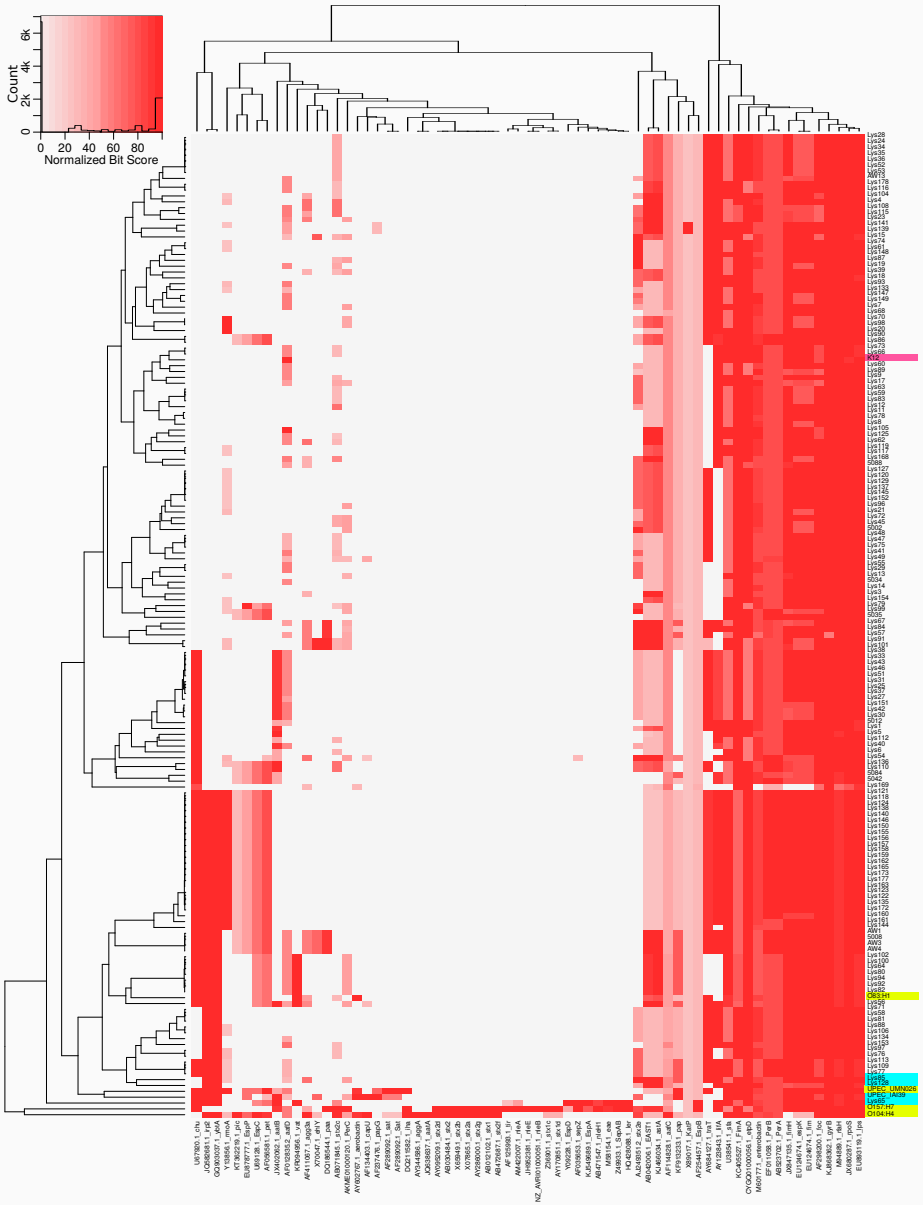
- Prevents duplication of work
- Aids automation
- Allows investigation of meta-variables

Preliminary Virulence Profiling

Virulence

- Search literature for genes implicated in virulence
- Select representative sequences for 50 virulence factors
- Use reciprocal translated blast to find occurrences
- Filter results, visualize

Virulence




AF114828.1_aafC
KF913233.1_pap
X89017.1_KatP
AF254457.1_EspB
AY684127.1_traT
AY123843.1_lifA
U38541.1_sfa
KC405527.1_FimA
CYGQ01000056.1_etpD
M60177.1_enterobactin
EF011058.1_PerB
AB523702.1_PerA
JX847135.1_fimH
EU124674.1_espK
EU124674.1_fim
AF298200.1_foc
KJ868362.1_gyrB
M94889.1_rfaH
JX680287.1_rpoS
EU893119.1_lps
O104:H4
O157:H7
Lys65
UPEC_IAI39
UPEC_UMN026
Lys128
Lys85
Lys77
Lys109
Lys113
Lys76
Lys97
Lys153
Lys134
Lys106
Lys88
Lys81
Lys58
Lys71
Lys56
O83:H1
Lys82
Lys92
Lys94
Lys80
Lys64
Lys100
Lys102
AW4
AW3
5008
AW1
Lys144
Lys161
Lys160
Lys172
Lys135
Lys122
Lys123

Virulence: Next Steps

- Compare with recent tools (ARIBA, VirulenceFinder, etc)
- Experimentally assess phenotypes as needed

Plans for 2017-2018

Future Work

riboSeed development:

- Improve progressive QC to identify problems early
- Investigate characteristics of rDNAs that may be predictive of riboSeed success

Exploring the E. coli pangenome:

- Attempt to isolate genomic trends indicative of soil-adaptation
- Establish pangenomic context of the the collection
- Correlate metadata with pangenome

Curli loss:

- Utilize phenotypic data from Y. Soronin and others to determine additional causes of curli loss

Acknowledgments

NUIG Microbiology

- Dr. Fiona Brennan
- Dr. Florence Abram
- Matthias Waibel
- Stephen Nolan
- Camilla Thorn

James Hutton Institute, Dundee

- Dr. Leighton Pritchard
- Dr. Ashleigh Holmes

Questions?