



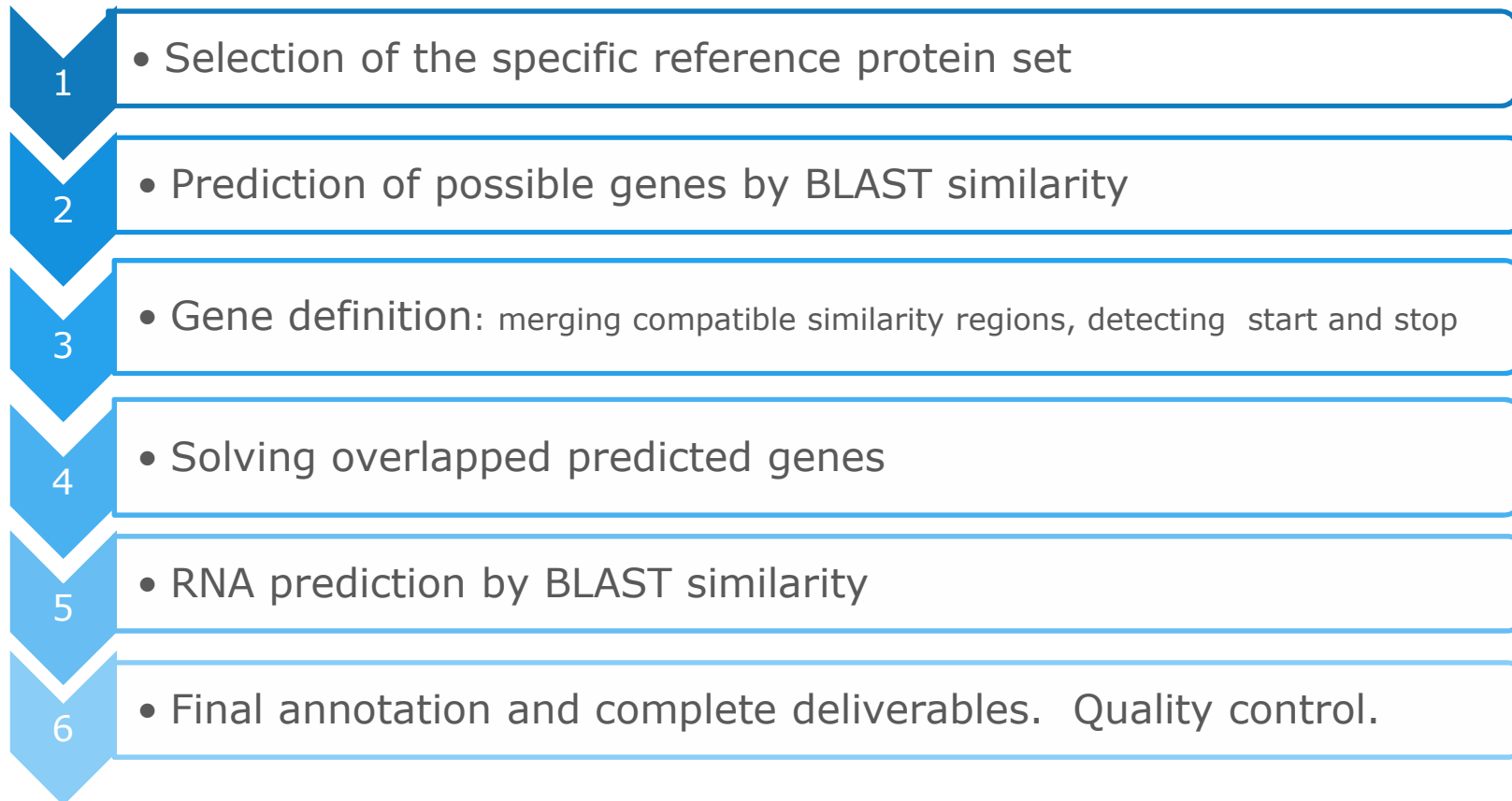
BG7

BACTERIAL GENOME ANNOTATION SERVICE

www.era7bioinformatics.com



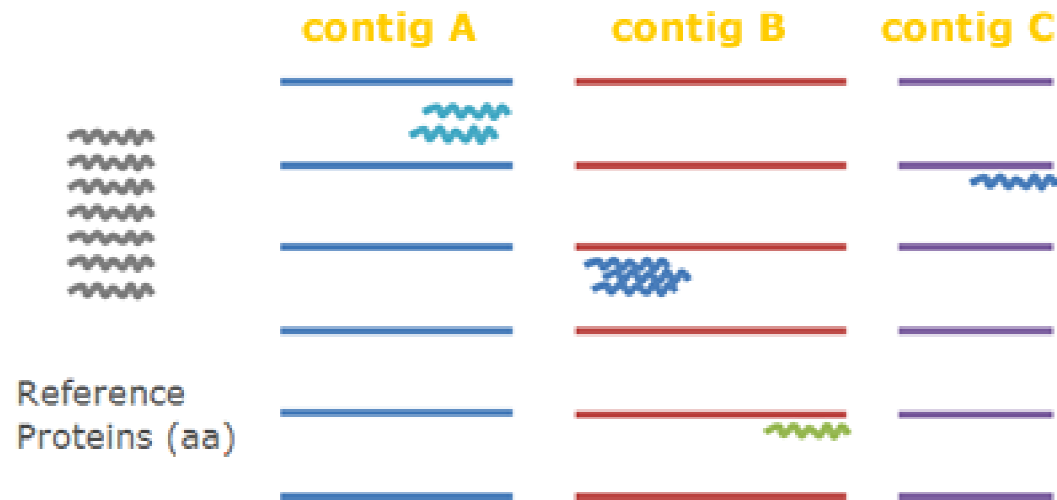
BG7 algorithm



●●● BG7 algorithm

1. Selection of the specific reference protein set

2. Prediction of possible genes by BLAST similarity

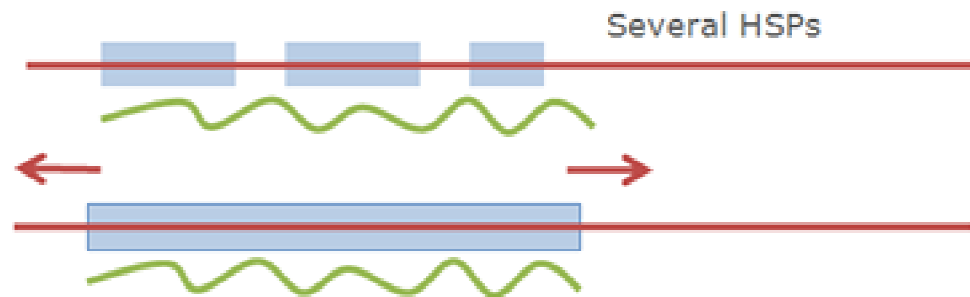


●●● BG7 algorithm

3. Gene definition: merging compatible similarity regions, detecting start and stop

ORF definition

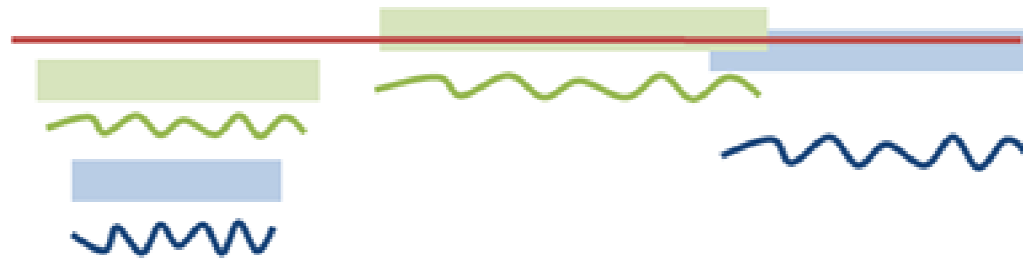
Merging HSPs and search for start/stop signals



4. Solving overlapped predicted genes

Solving conflicts

Duplicates and overlapping ORFs



5. RNA prediction by BLAST similarity

Incorporation of RNA genes

Conflicts with protein coding genes are solved

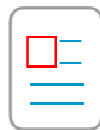


RNA gene is selected over the protein coding one

●●● Final annotation. BG7 deliverables



BG7 Genome annotation Deliverables:



Report
pdf



Genbank



Proteins
in nucleotides



Annotation
Excel



gff



Proteins
in aminoacids

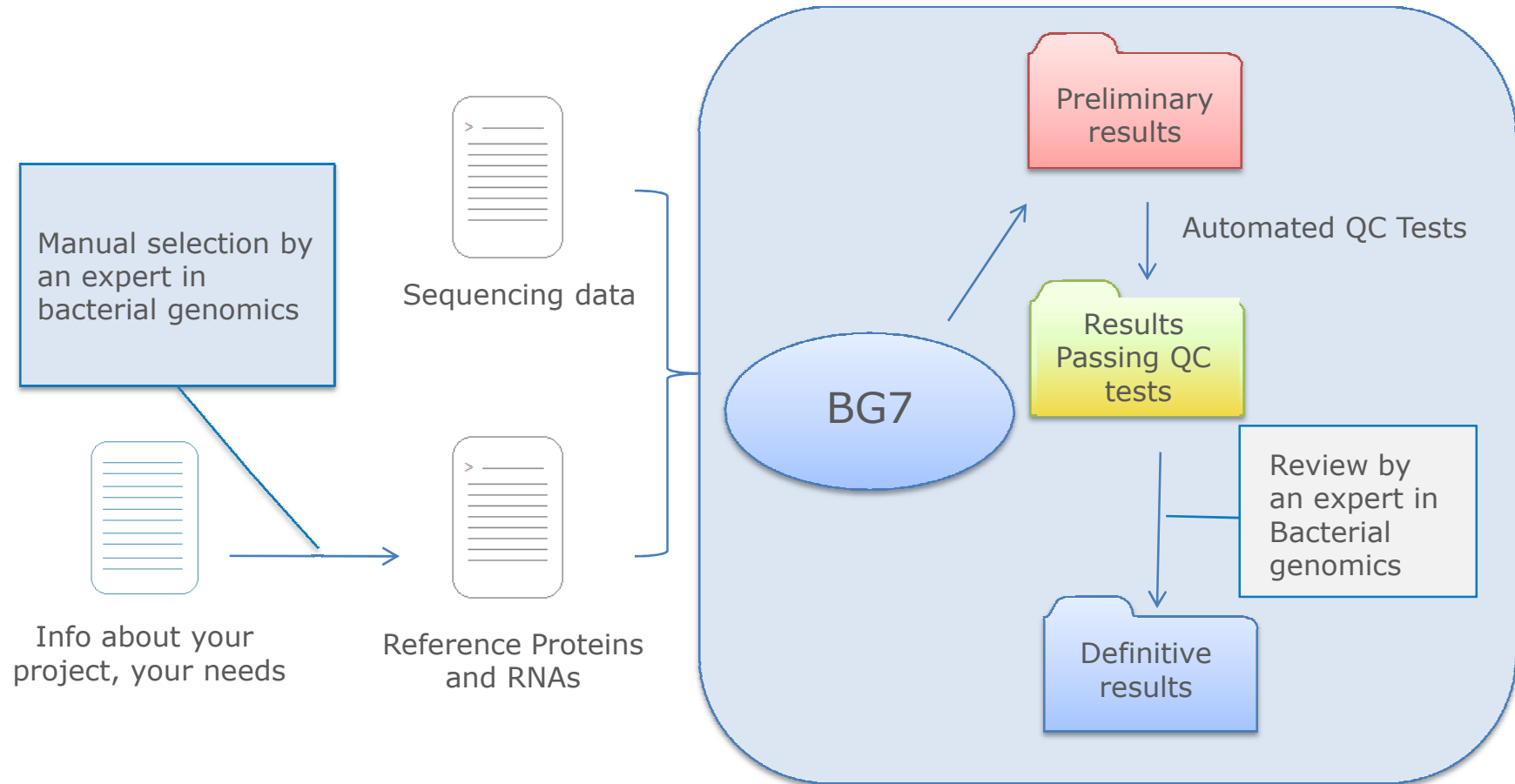


GO analysis
Excel



Non-coding
regions

●●● BG7 Service Process





●●● BG7 annotation of PacBio Data

E. coli O104:H4 Outbreak. c227-11 isolate

We annotated the PacBio assembly of the c227-11 isolate:

http://www.era7bioinformatics.com/en/other_services/BG7_annotation_by_Era7_bioinformatics/E_coli_EHEC_c227_PB1.html

We did it with the BG7 pipeline:

http://www.era7bioinformatics.com/en/bacterial_genome_bg7.html

As we did with the rest of E. coli O104:H4 genomes publicly available at:

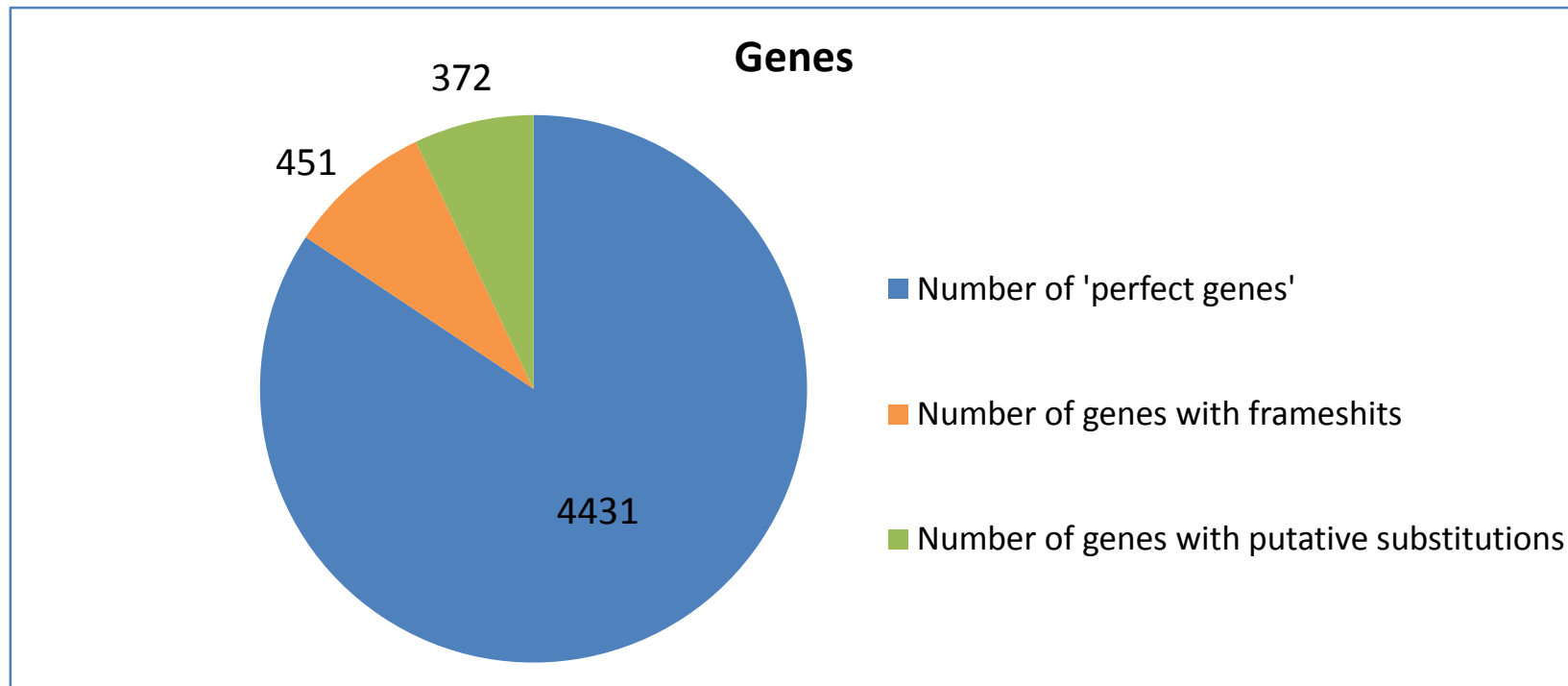
<https://github.com/ehc-outbreak-crowdsourced/BGI-data-analysis/wiki/Annotations>



Results of the annotation of c227 isolate sequenced with PacBio



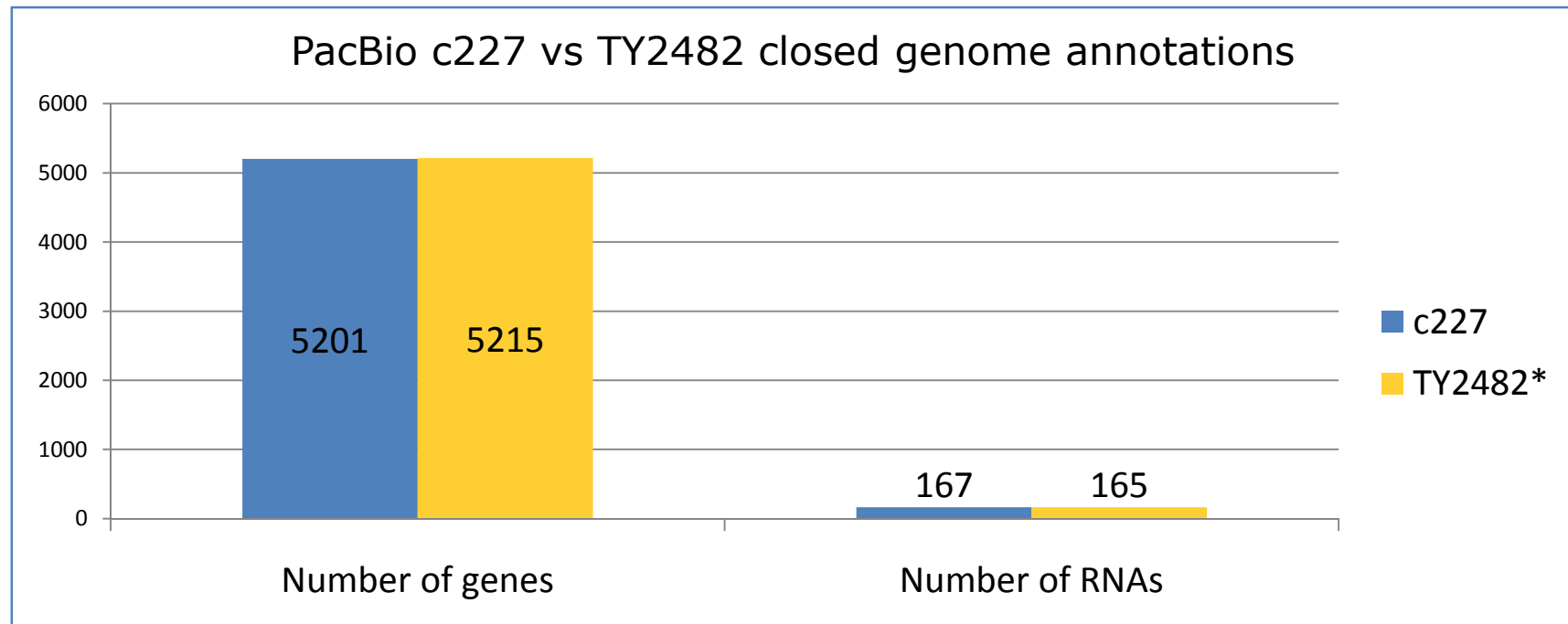
Number of contigs	Number of RNAs	Number of genes	Number of perfect genes	Number of genes with frameshifts	Number of genes with putative substitutions*
37	167	5201	4431	451	372



* non-canonical start, stop or with intragenic stops

c227 PacBio vs TY2482 closed genome annotations

Isolate	Technology	Details	Number of contigs	Number of genes	Number of RNAs
c227	PacBio	CCS and Long reads	37	5201	167
TY2482*	Illumina	Illumina PE: 500bp, 2kb and 6Kb	4	5215	165



* TY2482 closed genome (1 chromosome and 3 plasmids)