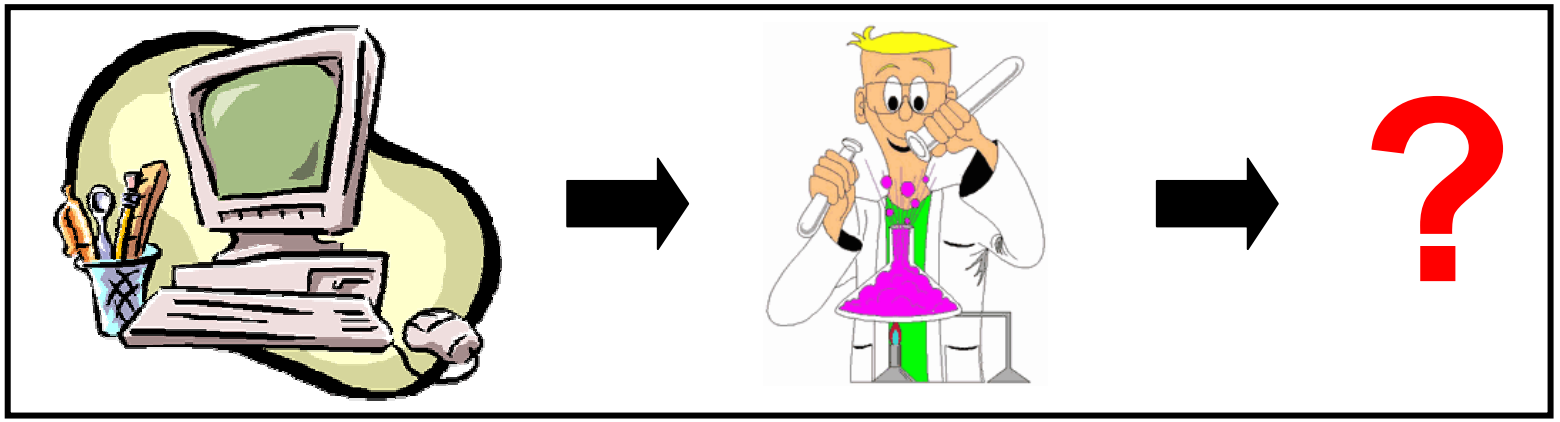
Experimental validation of the *in silico* predictions

Concept

RVM predictions | Lab-based validation | Verdict



- Predict in silico candidate GIs
- Design primers

- PCR
- Electrophoresis
- Sequencing

In silico predictions for real?

In silico Approach

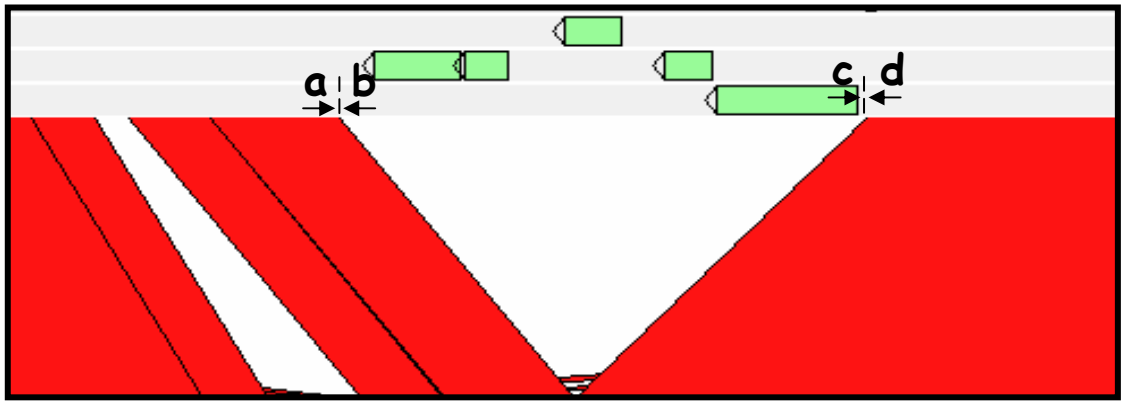
1. Genome sequence of Stenotrophomonas maltophilia strain K279a
2. Ran alien_hunter
3. The predicted candidate GIs were structurally annotated
4. Their structural annotation was used as input to the relevance vector machine (RVM) classifier
5. For the classification purposes, the 3 genus-specific structural GI models of Salmonella, Staphylococcus and Streptococcus previously described, as well as a model trained on all three datasets were exploited
6. A sample of eight predictions with both highly and less probable GI structures with a score range of 0.2371–0.9997 formed the test-dataset of this analysis

Test dataset

Location	Region	IVOM	INSP	SIZE	DENS	REPEATS	INT	PHAGE	RNA
60416..70829	R1	0.38128	1	10,413	1.3444	1	1	1	0
3089398..3127169	R16	0.74458	0	37,771	1.0060	1	1	1	1
299814..335480	R4	0.32642	0	35,666	1.2897	1	1	1	1
1323939..1367750	R12	0.55018	0	43,811	1.2325	1	1	1	0
1720046..1724493	R14	0.72176	0	4,447	1.7986	1	0	0	1
1945379..2002745	R15	0.28154	0	57,366	1.1854	1	1	1	1
3913072..3931089	R20	0.16626	0	18,017	0.6666	1	0	0	0
631285..661659	R7	0.27377	0	30,375	0.8559	0	0	0	0

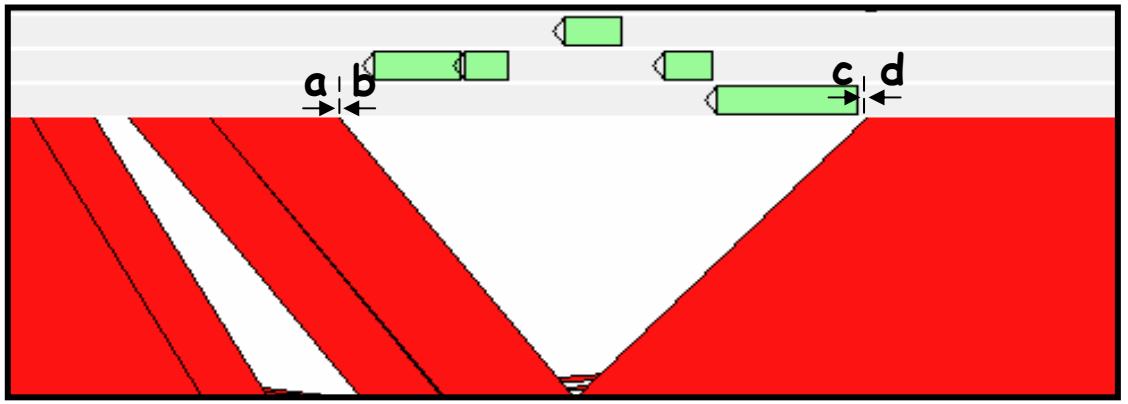
Region	Salm model	Staph model	Strep model	all3 model
R1	0.9918	0.9991	0.9994	0.9997
R16	0.9995	1.0000	0.9965	0.9992
R4	0.9944	0.9959	0.9804	0.9948
R12	0.9851	0.9997	0.9922	0.9903
R14	0.9978	0.9999	0.9005	0.9890
R15	0.9786	0.9826	0.9765	0.9835
R20	0.5023	0.2109	0.4742	0.4983
R7	0.3070	0.5223	0.1368	0.2371

Principle



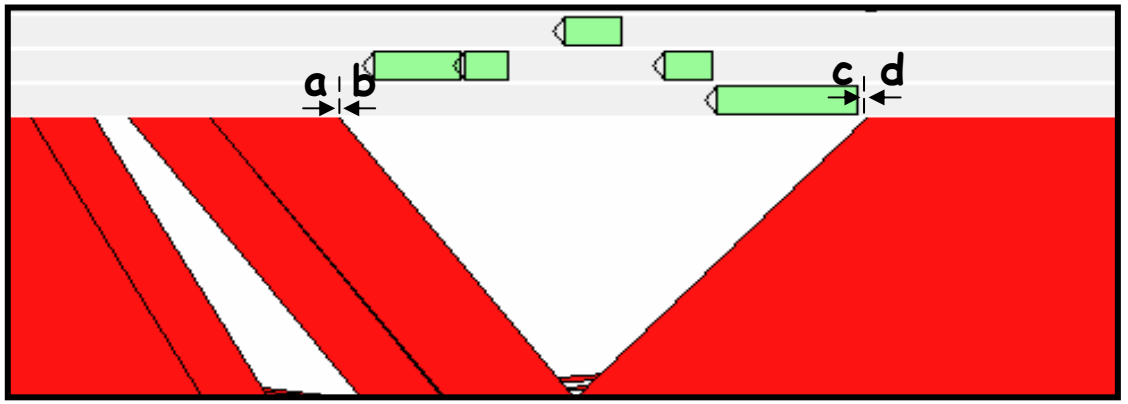
strain	PCR product			Inferred GI distribution
	ab	cd	ad	
1	+	+	-	
2	-	-	+	
3	+	-	-	
4	-	+	-	

Principle



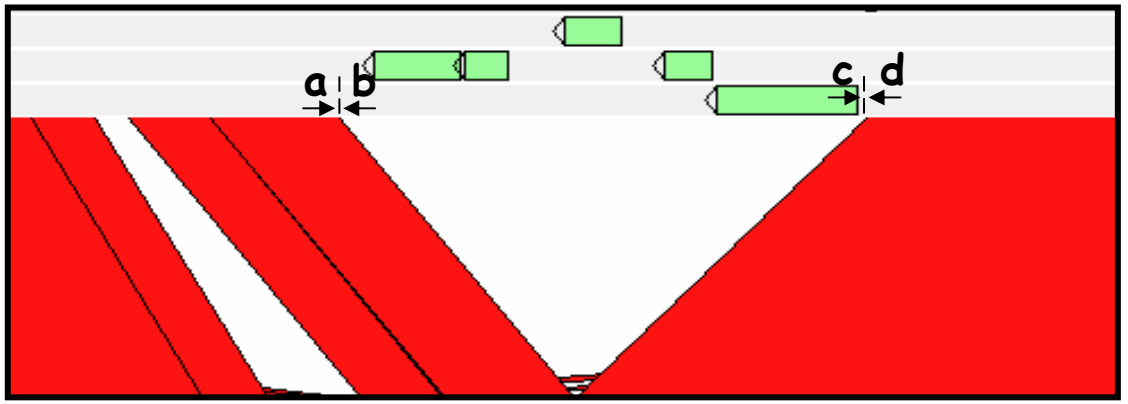
strain	PCR product			Inferred GI distribution
	ab	cd	ad	
1	+	+	-	Present
2	-	-	+	
3	+	-	-	
4	-	+	-	

Principle



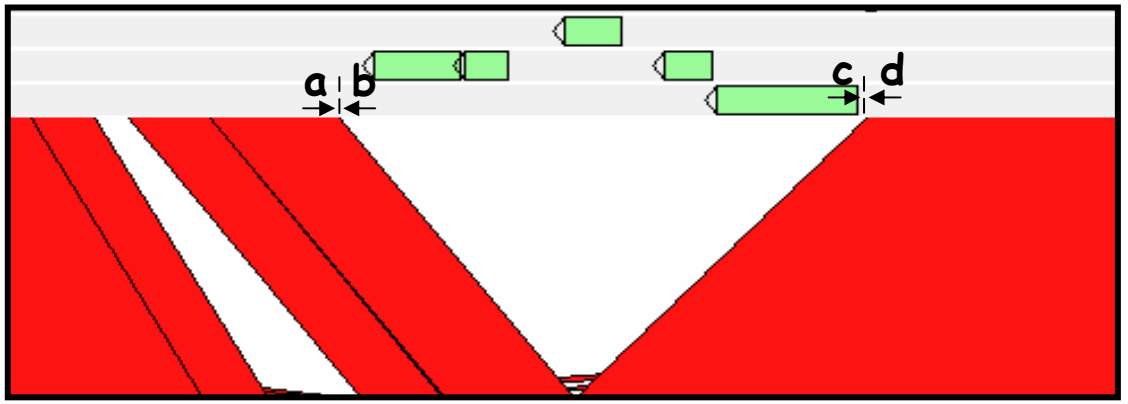
strain	PCR product			Inferred GI distribution
	ab	cd	ad	
1	+	+	-	Present
2	-	-	+	Absent
3	+	-	-	
4	-	+	-	

Principle



strain	PCR product			Inferred GI distribution
	ab	cd	ad	
1	+	+	-	Present
2	-	-	+	Absent
3	+	-	-	Ambiguous
4	-	+	-	

Principle



strain	PCR product			Inferred GI distribution
	ab	cd	ad	
1	+	+	-	Present
2	-	-	+	Absent
3	+	-	-	Ambiguous
4	-	+	-	Ambiguous

Polymerase Chain Reaction - PCR

The purpose of a Polymerase Chain Reaction (PCR) is the amplification of DNA fragments to a very large number of copies. The PCR protocol consists of three major steps, each of which is repeated 30-40 times:

1. Denaturation At this step, double-stranded DNA templates denature to single strands of DNA, a process caused by the increased temperature that disrupts the hydrogen bonds between complementary bases of the DNA molecule; the denaturation occurs at a temperature, which is partially determined by the G+C% content of the DNA templates. At this step the reaction is heated for 30-45 seconds at a temperature of 94-95 ^{o}C.

2. Annealing The reaction temperature is lowered, usually at an annealing temperature 3-5 ^{o}C lower than the melting temperature (Tm) at which the primers dissociate from the DNA template. At this step, primers are moving following a Brownian motion and hydrogen bonds are constantly formed and broken between the single-stranded DNA primers and the single-stranded DNA template. Once stable primer-template hybrids have formed, the DNA polymerase catalyzes the template-dependent synthesis of DNA. The optimal annealing temperature for a given primer set and DNA template can be determined by using a range of different temperatures e.g. 50-65 ^{o}C (gradient PCR).

3. Extension At a temperature of 72-78 ^{o}C a thermostable DNA polymerase synthesizes new DNA strands that are complementary to the template DNA, by adding Deoxynucleoside triphosphates (dNTPs) in a 5' to 3' direction (relative to the primer) reading the template DNA in the opposite direction (i.e. 3' to 5'). A commonly used thermostable DNA polymerase, is the Taq polymerase, an enzyme originally isolated from, and named after, the bacterium Thermus aquaticus (Chien et al., 1976). The speed of polymerization (following a geometric growth) of the template DNA by the Taq polymerase is ~ 2000 nucleotides/minute

Demo: http://learn.genetics.utah.edu/content/labs/pcr/

Gel Electrophoresis

✓The principle of this protocol is the separation of nucleic acids or proteins based on their charge and mass

✓Using an electric field, the macromolecules can be separated on a gel, with a rate of migration that depends on many factors, including the applied voltage, the hydrophobicity, size and shape of the molecules, the agarose gel concentration and the ionic strength of the buffer solution

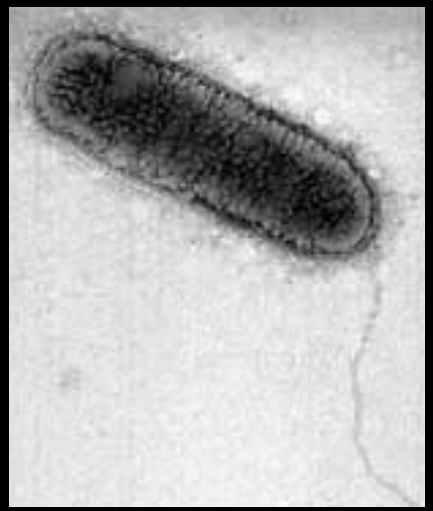
Demo: http://learn.genetics.utah.edu/content/labs/gel/

Stenotrophomonas maltophilia

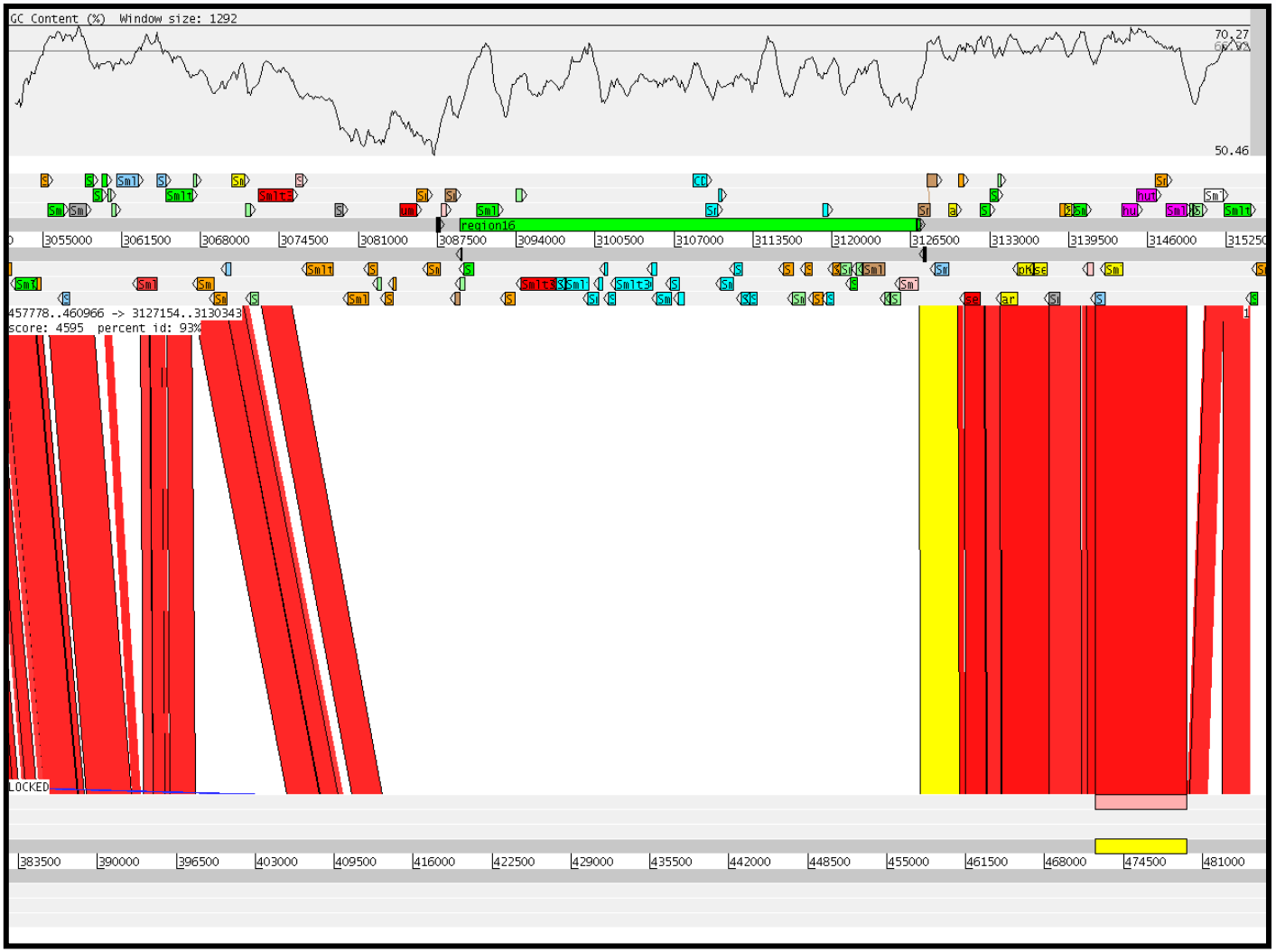
- ✓ Nonfermentative gram-negative bacillus (*Pseudomonas maltophilia* or *Xanthomonas maltophilia*)
- ✓ Important nosocomial pathogen
- ✓ Little is known about virulence factors associated with the bacterium
- ✓ Uncertainty about the route(s) of acquisition of *S. maltophilia*
- ✓ Person-to-person transmission may be an infrequent occurrence in the nosocomial setting
- ✓ Resistance to many currently available broad-spectrum antimicrobial agents

CLINICAL MANIFESTATIONS:

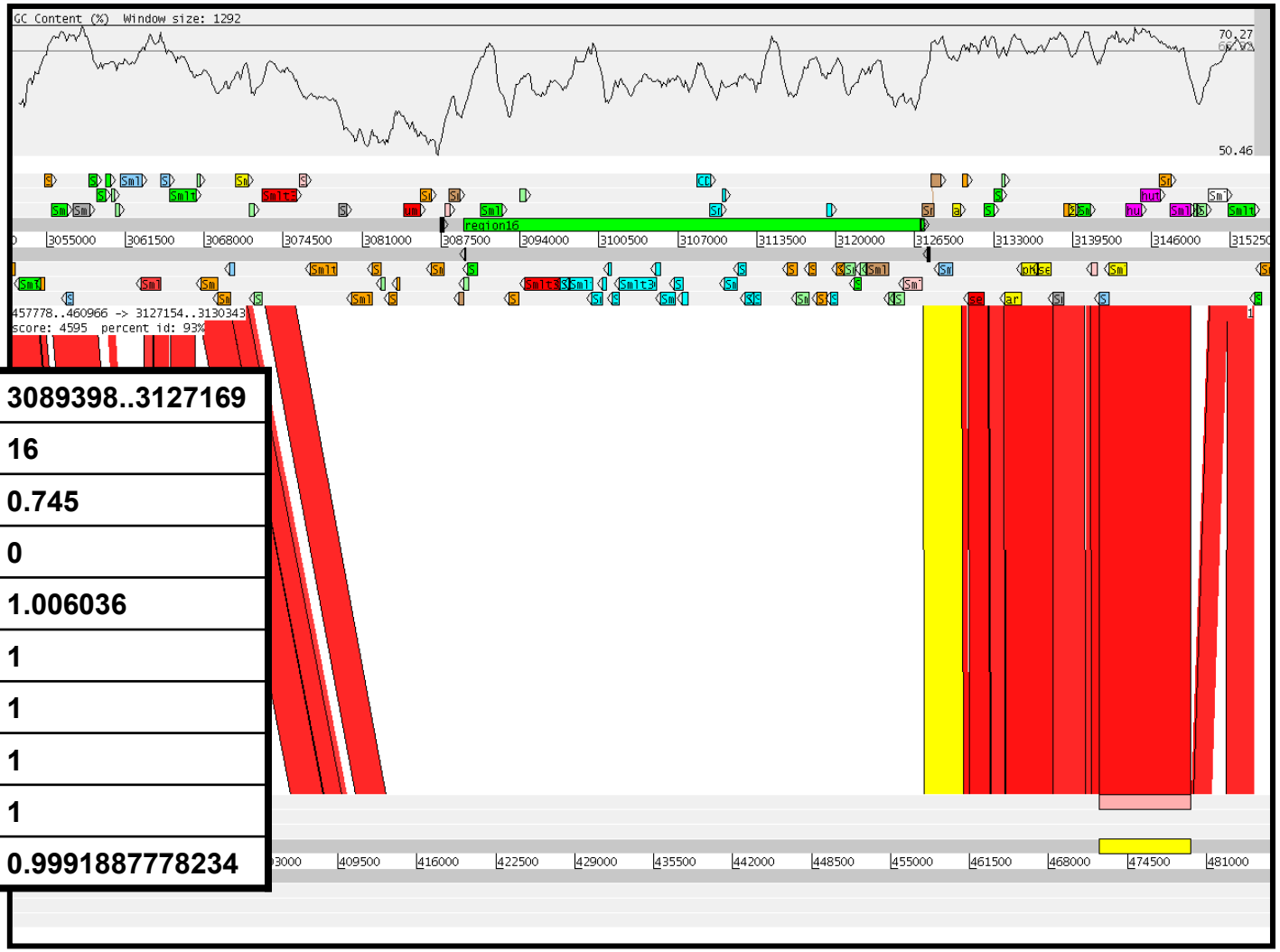
1. Bacteremia
2. Endocarditis
3. Nosocomial pneumonia (5%)
4. Meningitis is uncommon
5. Ocular infections
6. Uncommon cause of urinary tract infection
7. Frequent isolate from wounds and other skin lesions
8. Bone and joint infections are uncommon
9. Infrequent cause of gastrointestinal infection



Results (example 1)

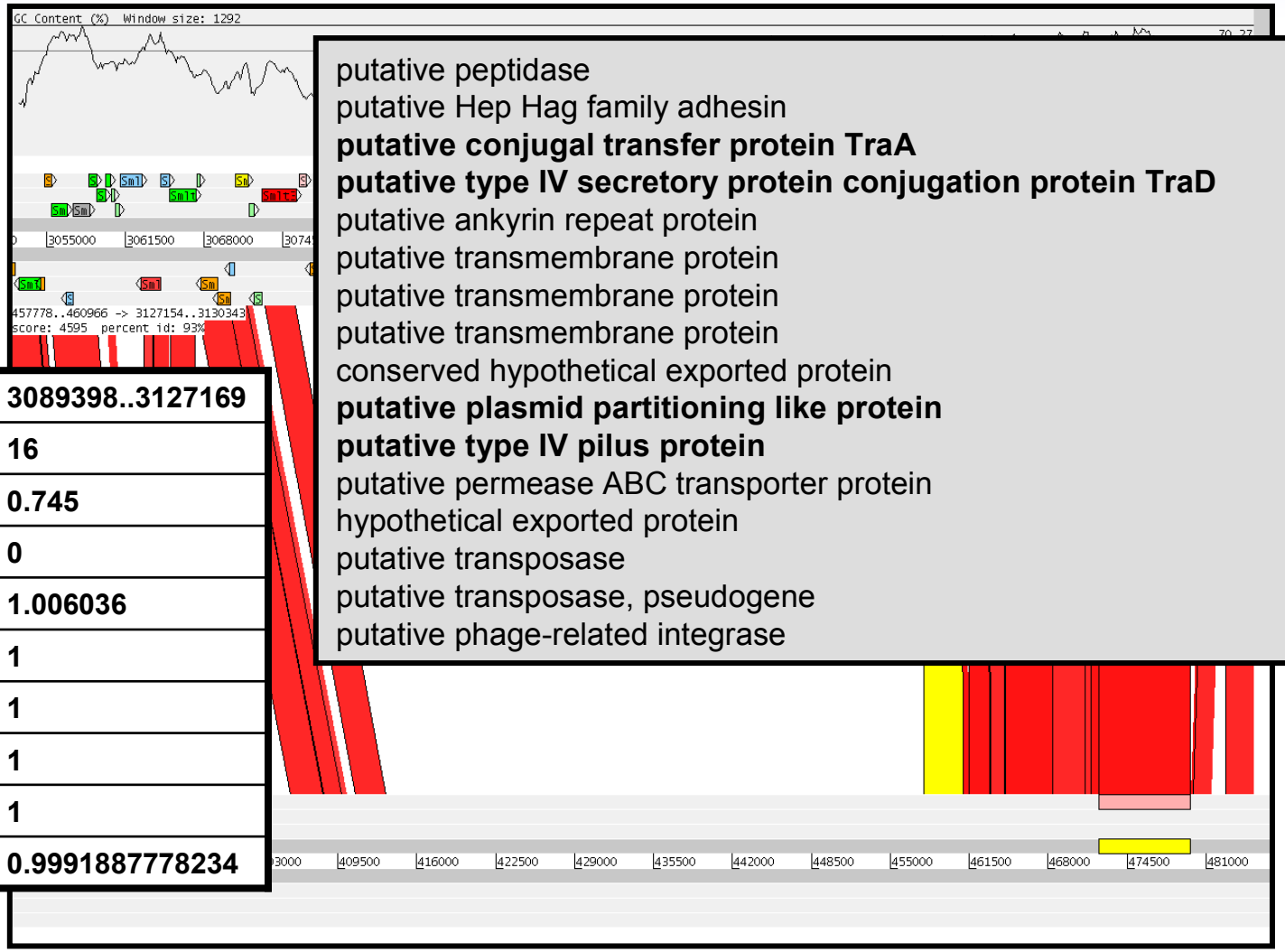


Results (example 1)



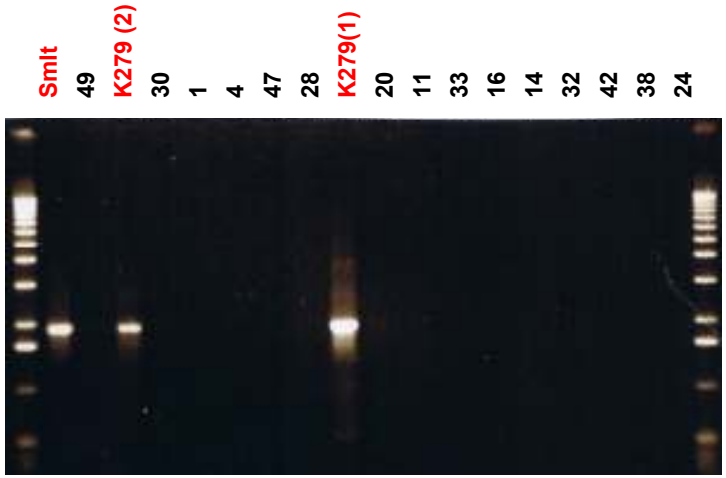
location	3089398..3127169
Region	16
IVOM	0.745
INSP	0
DENSITY	1.006036
REPEATS	1
INTEGRASE	1
PHAGE	1
RNA	1
RVM score	0.9991887778234

Results (example 1)

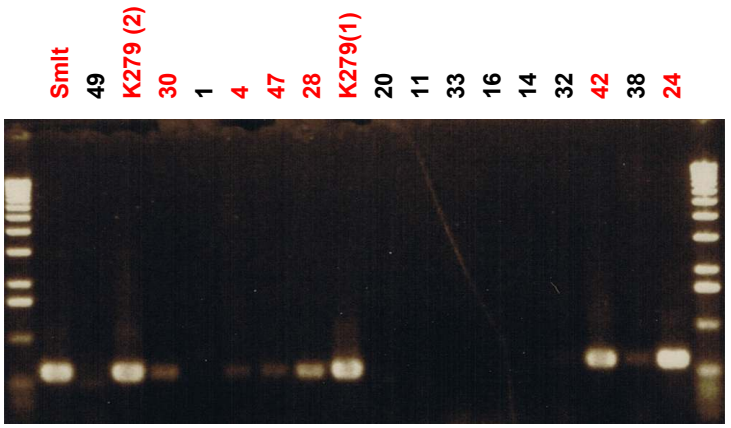


location	**3089398..3127169**
Region	**16**
IVOM	**0.745**
INSP	**0**
DENSITY	**1.006036**
REPEATS	**1**
INTEGRASE	**1**
PHAGE	**1**
RNA	**1**
RVM score	**0.9991887778234**

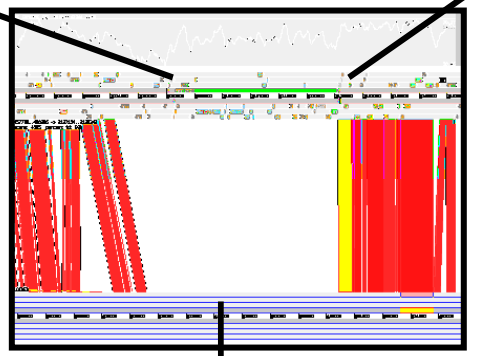
putative peptidase
putative Hep Hag family adhesin
putative conjugal transfer protein TraA
putative type IV secretory protein conjugation protein TraD
putative ankyrin repeat protein
putative transmembrane protein
putative transmembrane protein
putative transmembrane protein
conserved hypothetical exported protein
putative plasmid partitioning like protein
putative type IV pilus protein
putative permease ABC transporter protein
hypothetical exported protein
putative transposase
putative transposase, pseudogene
putative phage-related integrase

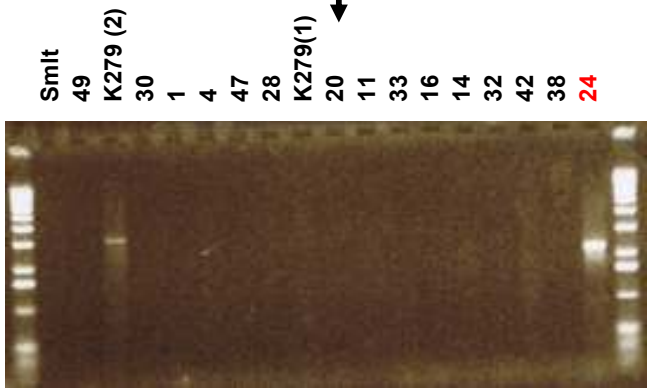


16 L: 1946bp



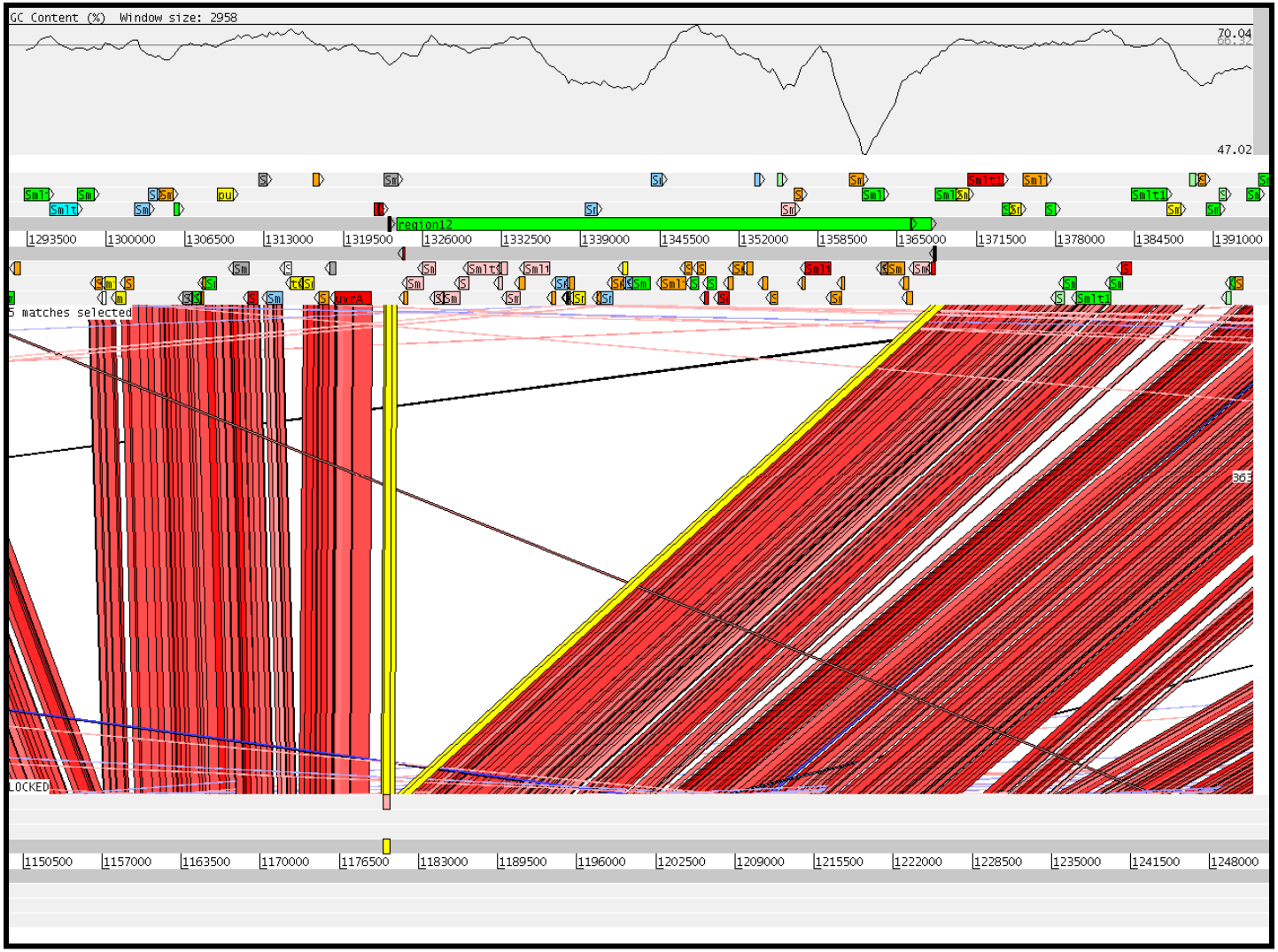
16 R: 616bp



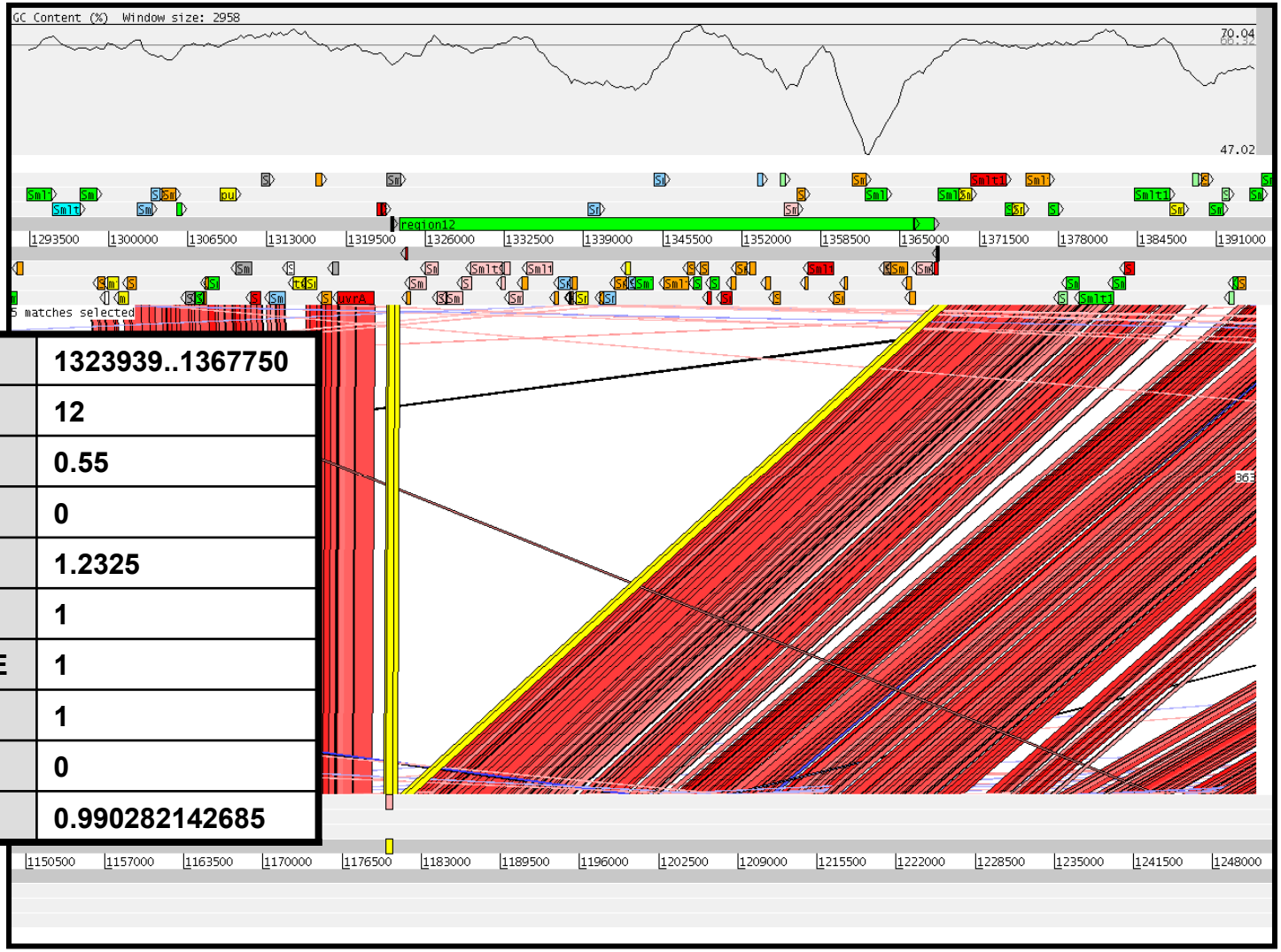


16 A-D: ~ 2276bp

Results (example 2)

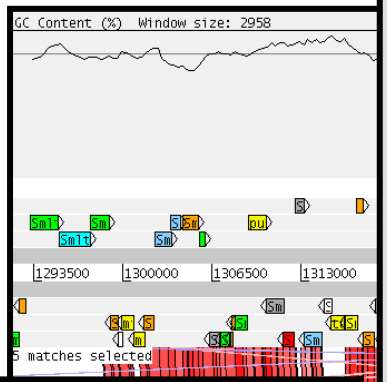


Results (example 2)



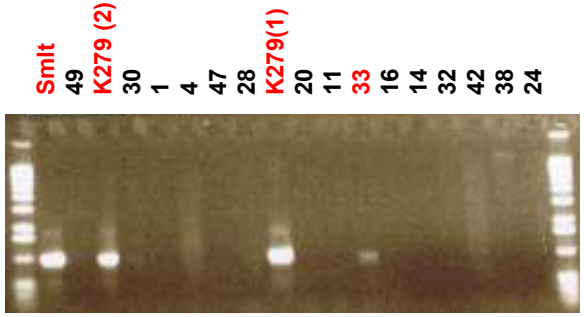
location	1323939..1367750
Region	12
IVOM	0.55
INSP	0
DENSITY	1.2325
REPEATS	1
INTEGRASE	1
PHAGE	1
RNA	0
RVM score	0.990282142685

R

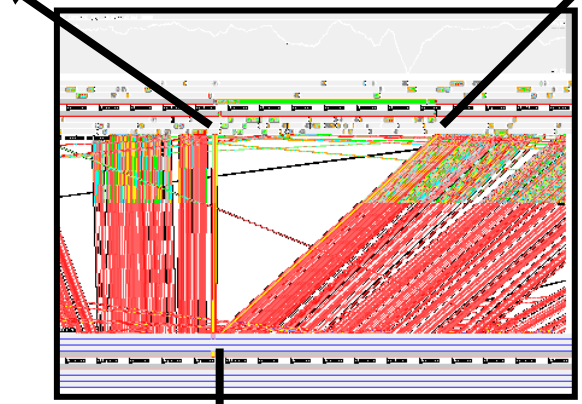


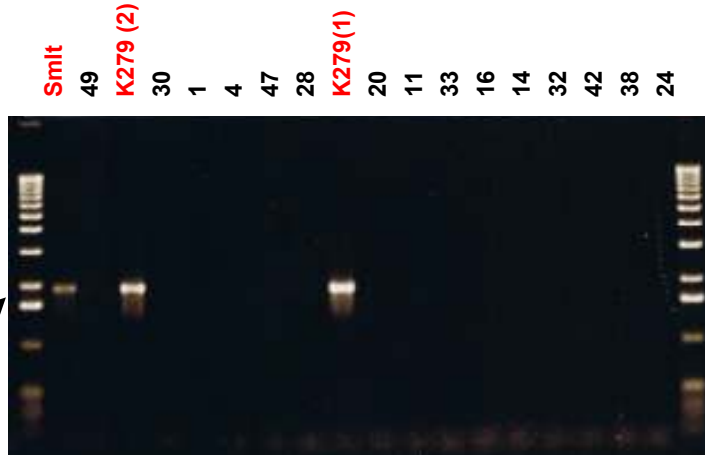
location	**1323939..1367750**
Region	**12**
IVOM	**0.55**
INSP	**0**
DENSITY	**1.2325**
REPEATS	**1**
INTEGRASE	**1**
PHAGE	**1**
RNA	**0**
RVM score	**0.990282142685**

putative conjugal transfer protein
putative conjugal transfer protein
putative conjugal transfer protein
putative conjugal transfer protein
putative conjugal transfer protein
putative conjugal transfer protein
putative conjugal transfer protein
putative conjugal transfer protein
putative conjugal transfer protein
putative conjugal transfer protein
putative conjugal transfer protein
putative LysR family transcriptional regulator
putative regulator, pseudogene
putative esterase, pseudogene
putative NAD(P)H dehydrogenase [quinone]
putative LysR family transcriptional regulator
putative LysR family transcriptional regulator
putative 4-carboxymuconolactone decarboxylase
putative MerR family transcriptional regulator
putative MFS family transmembrane transporter
putative LysR family transcriptional regulator
putative transmembrane anchor conjugal transfer protein
putative parB partition protein
putative ParA/CobQ/CobB/MinD nucleotide binding domain protein
putative RepA-like replication protein
putative HTH transcriptional regulator
putative transposase
putative ParB-like nuclease domain protein
putative transmembrane protein
putative RadC DNA repair protein, pseudogene
putative prophage integrase

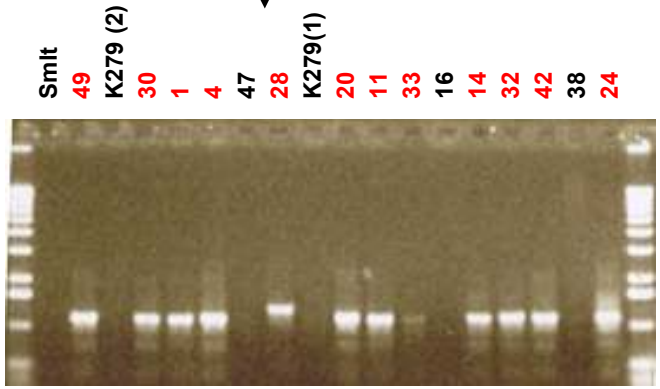


12L: 999bp

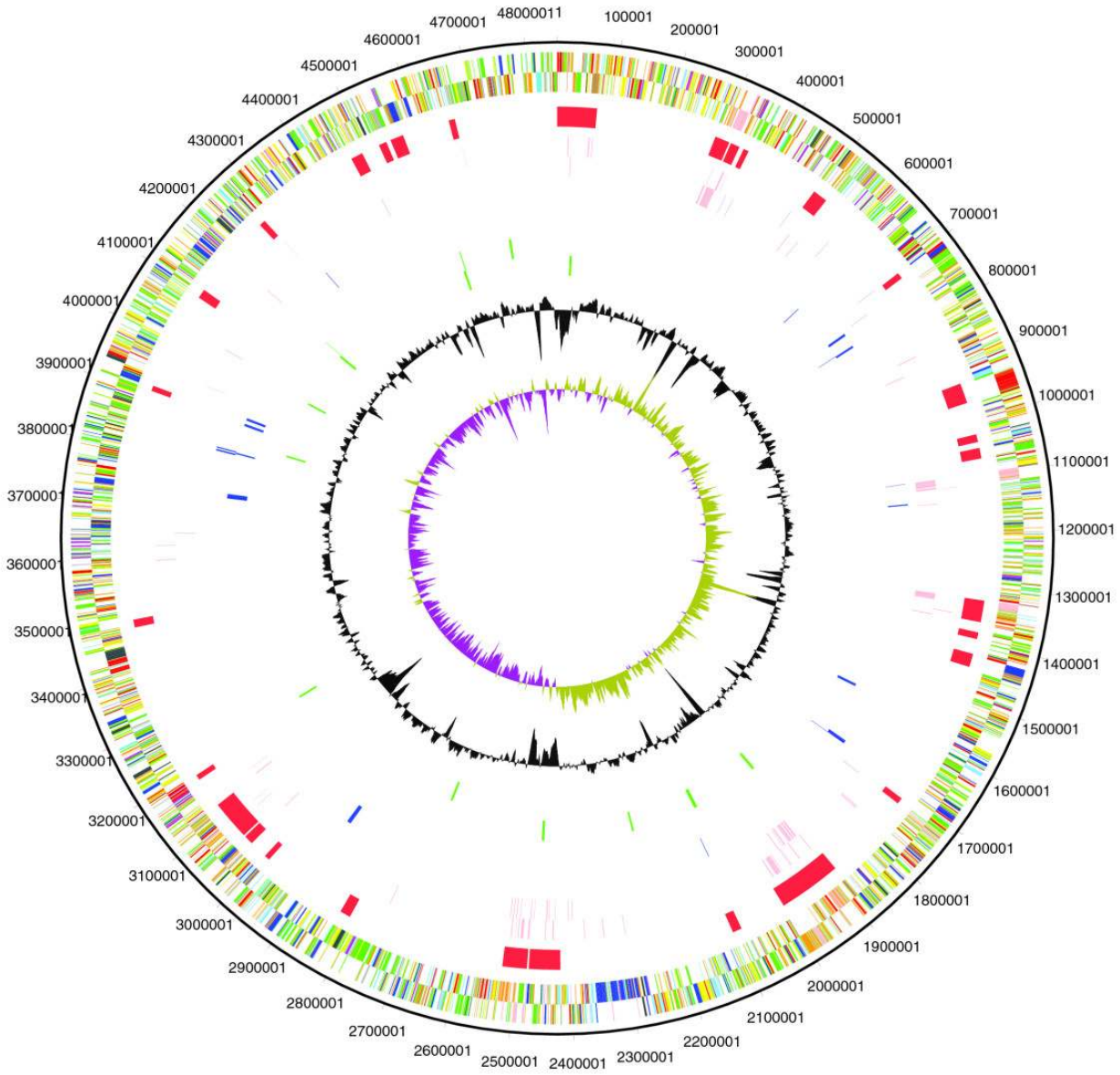




12 R: 1959bp



12 A-D: ~ 1000bp



" … β-lactams, chloramphenicol, aminoglycosides, fluoroquinolones and macrolides … Many of the resistance genes are located on small islands with no obvious mobile DNA features (determined by Alien Hunter [26]) …", ***Crossman LC et al., Genome Biol 2008***

Summary

✓Experimentally validated *in silico* predictions:

Benchmarking → SP = 0.83, SN = 1.0, AC = 0.88

✓Accurately determined predicted boundaries (verified by sequencing)

✓Interesting gene products → potential virulence factors