Plasmid Circular BLAST Evaluation

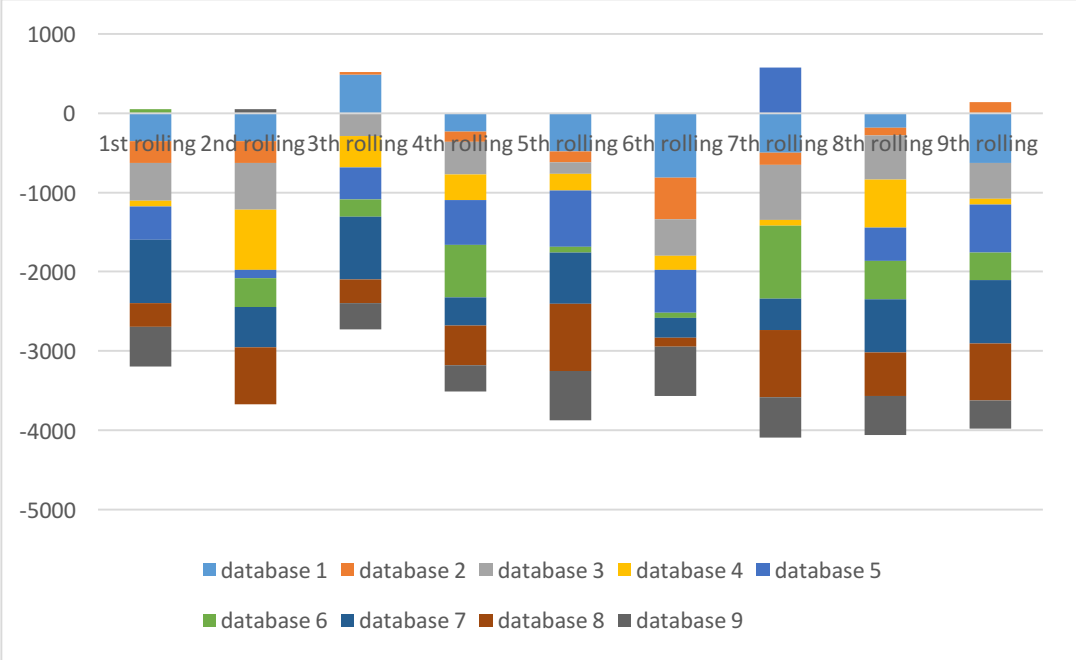
1. n^{th} rollings and n^{th} database matrix evaluation

HMM generated nine databases which are featured in three characteristics: Probability for noncoding state; Probability for coding state; transition probability. Average scores for each database refer to circular BLAST adaptability to specific database. From the matrix of 10bp ("a"), 20bp ("b"), 50bp ("c"), 100bp ("d"), 150bp ("e") queries, each database got average score hit.

Evaluation of matrix 10bp queries and database is following:

	1^{st} rolling	2^{nd} rolling	3th rolling	4^{th} rolling	5^{th} rolling	6^{th} rolling	7^{th} rolling	8^{th} rolling	9^{th} rolling
database 1	-352	-352	**486**	-230	-483	-811	-503	-186	-625
database 2	-278	-278	30	-136	-136	-531	-147	-92	***141***
database 3	-473	-584	-287	-407	**-145**	-453	-695	-564	-453
database 4	**-73**	-765	-401	-326	-215	-184	**-73**	-603	**-73**
database 5	-418	-101	-398	-564	-706	-540	**571**	-418	-604
database 6	**47**	-361	-219	-658	-73	-64	-922	-483	-350
database 7	-800	-514	-791	-361	-649	**-250**	-392	-669	-800
database 8	-298	-717	-298	-495	-848	**-112**	-848	-551	-717
database 9	-501	**49**	-335	-335	-621	-621	-510	-490	-359

Hit for nth database is bolded. Hit for all database is bolded and italic.



Database 1 and database 3 locate upper than database 7 and 9 which 1,3 are generated by HMM larger coding probability and smaller noncoding probability.

2. Circular BLAST sequence length evaluation

Database and query length are another feature of HMM sequence. We got runtime by length of query and database.

3. BLAST and circular BLAST evaluation

	1st rolling	2nd rolling	3th rolling	4th rolling	5th rolling	6th rolling	7th rolling	8th rolling	9th rolling
database 1	-352	-352	**486**	-230	-483	-811	-503	-186	-625
database 2	-278	-278	30	-136	-136	-531	-147	-92	***141***
database 3	-473	-584	-287	-407	**-145**	-453	-695	-564	-453
database 4	**-73**	-765	-401	-326	-215	-184	**-73**	-603	**-73**

database 5	-418	-101	-398	-564	-706	-540	**571**	-418	-604
database 6	**47**	-361	-219	-658	-73	-64	-922	-483	-350
database 7	-800	-514	-791	-361	-649	**-250**	-392	-669	-800
database 8	-298	-717	-298	-495	-848	**-112**	-848	-551	-717
database 9	-501	**49**	-335	-335	-621	-621	-510	-490	-359

Hit for nth database is bolded and hit for all database is bolded and italic. Only 2 database has hit at first rolling which is linear BLAST. Thus circular BLAST got more hit scores at all database. BLAST cuts plasmid to liner sequence which may not got best BLAST than other cuts. Plasmid BLAST rolls the circular sequence and records all cuts and scores.

4. **runtime evaluation**

   Runtime: O(length(plasmid)^2*length(database)) for circular plasmid BLAST.

5. **Circular query to circular/ liner database evaluation**

   Circular query rolls on fixed liner database. The run time is for the fixed database. If database is rolling, the circular query is BLAST on circular database which we got more data for plasmid BLAST. Circular query (plasmid) has circular database of NCBI. Rolling circular query on circular database will move circular query on liner database and not fix on liner database first base pair. The runtime for circular to circular database is longer and data will be more comprehensive.