Supported by tetfund TERTIARY EDUCATION TRUST FUND

BIOINFORMATIC INSIGHT INTO FACTORS INFLUENCING MYCOLACTONE POLYKETIDE SYNTHASE A CONTROL IN *MYCOBACTERIUM ULCERANS*

[1]Popoola, O.D., [2]Adeleke, A.J., [3]Kolawole, R.M., [1]Banjo, O.A., Adesetan, T.O., [4]Makanjuola, S.O., [1]Coker, M.O. and [1]Thomas, B.T.

[1]Department of Microbiology, Olabisi Onabanjo University, Ago Iwoye, Ogun State, Nigeria
[2]Department of Microbiology, Modibbo Adama University, Yola, Nigeria
[3]Department of Cell Biology and Genetics, University of Lagos, Akoka, Lagos, Nigeria
[4]Department of Medical Microbiology, Olabisi Onabanjo University,Sagamu, Ogun State, Nigeria
Correspondence to : benthoa2013@gmail.com

Received: March 20, 2022 **Accepted:** June 18, 2022

Abstract: The linkage between the severity of *Mycobacterium ulcerans* infection and mycolactone polyketide synthase A has been strongly correlated with the presence of sigma factors but recent updates in the field of mycobacteriology are now linking such regulation to the presence of nutrient stimulants and hence spurring parallel reports. This study was therefore aimed at providing a bioinformatic insight into the understanding of factors controlling the expression of this mycolactone polyketide synthase A. Results obtained connote that eleven (11) and fourteen (14) of the isolate's mycolactone polyketide synthase A are controlled by *Sig A* and *Sig D* factors respectively as shown by the presence of those transcriptional binding motifs while nine (9) and six (6) of the isolates lack *Sig A* and *Sig D* factors respectively. Two of the isolates mycolactone polyketide synthase A were found to be controlled by neither *Sig A* nor *Sig D* factor. When the association between sigma factor, protein structure and the mycolactone polyketide synthase A fragments expression was evaluated, no significant statistical association was found despite the fact that larger number of isolates were controlled by *Sig D* and bias toward random coil than any other protein structures. The nucleotide substitution favours guanine (60%) than cytosine (40%) and provide evidence that where the percentage of alpha helix is at equilibrium with that of random coil, the substitution rate of guanine and cytosine were also at equilibrium. The transition/transversion bias ratio, the maximum likelihood for the computation and the tajimas neutrality test were estimated to be 0.40, -8864.235 and 8 respectively. Results of this study have shown that even though both *Sig A* and *Sig D* factors play important role in mycolactone polyketide synthase A control, other factors beyond these factors are involved.

Keywords: Bioinformatics, Mycolactone, *Mycobacterium ulcerans*, Regulation

Introduction

Mycobacterium ulcerans, which has now been declared as a reemerging disease by the world health organization is known for causing buruli ulcer, a severe infectious skin disease that occur mainly in the humid tropical zones especially in West African countries (Wagner *et al.*,2008). This organism is well adapted to the tropics, with an elevated trend that makes them third mycobacteriosis to both tuberculosis and leprosy (Leao *et al.*, 2006). The distribution of the disease caused by this organism is witnessing significant increase in Western and Central Africa while countries with national buruli ulcer surveillance programmes such Ghana and Republic of Benin have reported higher incidence of buruli ulcer than both leprosy and tuberculosis (Amofah *et al.*, 2012; Debacker *et al.*, 2014).

The severity of the massive tissue destructions forming large painless ulcers with undermined edges that could touch the bone tissue are known to be induced by a bacterial toxin known as mycolactone (George *et al.*,2000). This mycolactone is a polyketide toxin that has intense cytotoxic activity in vitro and is capable of affecting numerous cell types (Deshayes *et al.*, 2013). Generally, this toxin is thought to have immunomodulatory activities by decreasing the efficiency of the immune system. However, these macrolides designated as mycolactone A and B have been reported to be controlled by several factors including sigma factors (Tobias *et al.*, 2009) and nutrient factors (Deshayes *et al.*, 2013), thereby spurring parallel reports on the above subject matter. This study was therefore aimed at providing a bioinformatic insight into the understanding of factors controlling the expression of this mycolactone polyketide synthase A.

Materials and Methods

Twenty Mycobacterium ulcerans mycolactone polyketide synthase A were retrieved from the gene bank at National Center for Biotechnology Information (NCBI) by using nucleotide Basic Local Alignment Search Tool (BLAST) program (http:blast.ncbi.gov.Blast.cgi). These DNA sequences were subsequently aligned using the CLUSTAL W in Molecular Evolutionary Genetics Analysis (MEGA). The aligned sequences were searched for putative transcriptional binding sites of Sig A and Sig D using the find motif in MEGA 17 (Tamura et al., 2007; Tobias et al., 2009). The highlighted binding sites were counted while their position was also noted. The significance of the presence or absence of sigma factor on the studied mycolactone gene fragments was determined using chi square analysis. The secondary structures of these mycolactone polyketide synthase A were determined in a polyphasic manner. First, the DNA sequences of the mycolactone polyketide synthase A were converted into amino acid sequences using the translate option in MEGA which automatically translate the gene sequences into amino acid sequences. Secondly, the amino acid sequences were then used for predicting the secondary structure of the mycolactone polyketide synthase A using Self Optimized Prediction method with Alignment (SOPMA) tool (Geourjon and Deleage, 1995).

The substitution pattern and rates were estimated under the Kimura (1981) parameter model. The twenty nucleotide sequences of Mycobacterium ulcerans were used for computing the maximum likelihood values. Codon positions considered were the first, second, third and the non coding region. All positions containing gaps and missing data were eliminated while evolutionary analyses were conducted in MEGA X. The maximum likelihood of gamma parameter for sites rate were estimated under the Juke and Cantor (1969) Model. The molecular phylogenetic analysis was inferred using the maximum likelihood method based on the Hasegawa-Kishino-Yano model. The initial trees for the heuristic search were computed automatically by Neighbor-Join and Bio NJ algorithms to a matrix of pair wise distances estimated using the maximum composite likelihood approach and then the topology with superior logarithmic-likelihood value were selected.

The prevalence of the putative binding sites was calculated using the formula;
P= n/N while the association between sigma factors, protein structure and mycolactone polyketide synthase A were evaluated using chi square analysis.

Results

The table 1 depicts the presence of Sig A binding sites in the mycolactone polyketide synthase A fragments of the studied Mycobacterium ulcerans. As shown in this table, 3, 5, 6, 1, 1 and 5 of the isolate's mycolactone polyketide synthase A have putative transcriptional binding sites (TFS) corresponding to TTGAA, TAAATT, CCCATT, TACAGC and GGGCCC respectively. Nine of the isolate's mycolactone polyketide synthase A lack the entire Sig A transcriptional binding factor (AY505302.1, AY505298.1, AY394445.2, AY356360.1, AY356359.1, AY358358.1, AY394448.1, AY394443.1 and AY394447.1). Five of the isolate's mycolactone polyketide synthase A have one each of the transcriptional binding sites (AY505303.1, AY505302.1, AY505300.1, AY394444.1 and AY394442.1). Four of the isolates have two transcriptional binding sites while one isolate (AY743331.1) has three TFS and (AY394440.1) has 5 Sig A binding sites. The relative distribution of the protein structure revealed significant bias towards random coil than alpha helix than other protein structures. It was also observed that all the isolates lacking Sig A factors were biased towards random coil. In one isolate (AY394447.1), where the relative abundance of random coil was at equilibrium with alpha helix, there was no putative Sig A transcriptional binding factor. The putative binding site of Sig D transcriptional factor in the mycloactone gene of Myclolactone ulcerans are represented in table 2. As shown in this table, fourteen of the twenty isolates possesss the Sig D factors corresponding to GCCCGC while three have binding sites corresponding to ACAGGC but none of the isolates have the transcriptional binding site corresponding to GTATCGAC. Of the 14 isolates having Sig D binding sites corresponding to GCCCGC, six have single binding sites, 4 have double binding sites, two have 3 binding sites while only one had four binding sites. The distribution of the transcriptional binding sites in the mycolactone polyketide synthase A of Mycobacterium ulcerans is depicted in table 3. The Sig D TFS corresponding to NNANN 1 and 2 occupies the highest prevalence with a distribution pattern of 55.3% of the total number of 47 TFS. This was followed by 27.7% of Sig A factor corresponding to TATAMTC (1-4) while 8 of the isolates (17%) have Sig A factors corresponding to TTGACN 1 and 2. All the isolates however lack the Sig D factor corresponding to GTANCGSS. The association between sigma factor expression and mycolactone polyketide synthase A control was analyzed using chi square (table 4) and despite having more isolates having the putative *Sig D* factor (16) as against 11 *Sig A* factors, the observation was found not to be statistically significant ($p>0.05$). Nine and four of the studied isolates lack *Sig A* and *Sig D factors* respectively in their mycolactone polyketide synthase A. When the association between protein structure abundance and mycolactone polyketide synthase A control was analyzed, no significant statistical association was found although many isolates demonstrated bias towards random coil than other protein structures (table 5). The relative abundance of protein secondary structure of the mycolactone polyketide synthase A of the studied *Mycobacterium ulcerans* revealed significant bias towards random coil than alpha helix, beta turns and extended strand. In no cases were beta turns and extended strands relatively higher than both alpha helix and random coil. However, one isolate namely AY394447.1 displayed an equal percentage of alpha helix and random coil and the isolate was also found to lack both *Sig A* and *Sig D* factors (table 6). In table 7, the nucleotide substitution frequencies of the studied mycolactone polyketide synthase A shows that "guanine'' was the most frequently substituted in the 12 of the 20 mycolactone polyketide synthase A while cytosine followed closely with the most substitution in 8 isolates. Isolate AY394442.1, where there was equal number of alpha helix and random coil above had equal substitution rate for both cytosine and guanine.

Table 1: Presence of *Sig A* motifs in *Mycobacterium ulcerans* Mycolactone Polyketide Synthase A

S/N	ISOLATE ACCESSION NUMBER	TTGACN1 TTGCAA	TTGACN1 TTGACT	TATAMT1 TAAATT	TATAMT2 CCCATT	TATAMT3 TACAGC	TATAMT4 GGGCCC	SUM TOTAL	PROTEIN STRUCTURE
1	AY743331.1	-	4569_4576	-	-	460_467	338_345	3	A
2	AY505303.1	-	-	54_61	-	-	-	1	R
3	AY505302.1	-	-	406_413	-	-	-	1	R
4	AY505301.1	-	-	-	-	-	-	0	R
5	AY505300.1	889 896	-	-	-	-	-	1	A
6	AY505299.1	-	-	404_411	-	-	54_61	2	A
7	AY505298.1	-	-	-	-	-	-	0	R
8	AY505296.1	802 809	2001_2008	-	-	-	-	2	A
9	AY505295.1	-	-	379_386	-	-	27_34	2	R
10	AY505294.1	-	-	356_363	-	-	4_11	2	A
11	AY394445.2	-	-	-	-	-	-	0	R
12	AY356360.	-	-	-	-	-	-	0	R

	1							
13	AY356359.1	-	-	-	-	-	0	R
14	AY358358.1	-	-	-	-	-	0	R
15	AY394448.1	-	-	-	-	-	0	R
16	AY394444.1	-	-	211_218	-	-	1	R
17	AY394443.1	-	-	-	-	-	0	R
18	AY394442.1	-	516_523	-	-	-	1	R
19	AY394447.1	-	-	-	-	-	0	R/A[E]
20	AY394440.1	1236_1243	262_269,498_505	-	1013_1020	-	2583_2590	5	R

A = Alpha helix, R = Random coil, [E] = equal

Table 2: Presence of *Sig D* motifs in *Mycobacterium ulcerans* Mycolactone Polyketide Synthase A

S/N	DOMAIN AND ISOLATES	GTANCGSS GTATCGAC	NNANN1 ACAGGC	NNANNN2 GCCCGC	PROTEIN STRUCTURE
1	AY743331.1	-	4395_4405	1350_1357, 1350_1357, 3696_3703, 3783_3790	A
2	AY505303.1	-	-	186_193, 483_490, 3418_3425	R
3	AY505302.1	-	-	56_63, 445_452	R
4	AY505301.1	-	43_50	591_598, 667_674, 799_806	R
5	AY505300.1	-	-	-	A
6	AY505299.1	-	-	54_61, 443_450	A
7	AY505298.1	-	-	95_102, 392_399	R
8	AY505296.1	-	1371_1378_1380_1837	-	A
9	AY505295.1	-	-	29_36,418_425	R
10	AY505294.1	-	-	6_13, 395_402	A
11	AY394445.2	-	-	329_336	R
12	AY356360.1	-	-	207_214	R
13	AY356359.1	-	-	421_428	R
14	AY358358.1	-	-	67_74	R
15	AY394448.1	-	-	-	R
16	AY394444.1	-	-	-	R
17	AY394443.1	-	-	250_257	R
18	AY394442.1	-	-	-	R
19	AY394447.1	-	-	-	R/A[E]
20	AY394440.1	-	-	2572_2579	R
	TOTAL	0	4	26	

A = Alpha helix, R = Random coil, [E] = equal

Table 3: Distribution of transcriptional binding motifs in *Mycobacterium ulcerans* Mycolactone Polyketide Synthase A

BINDING SITES	n	(%)
Sig A		
TTGACN(1-2)	8	17.0
TATAMTC(1-4)	13	27.7
Sig D		
GTANCGSS	0	0.0
NNANN(1-2)	26	55.3
Total	47	100

Nn = number, % = percentages

Table 4: Association between sigma factor expression and Mycolactone Polyketide Synthase A control

Sigma factors	Positive	Negative	Total
Sig A	11	9	20
Sig D	16	4	20
Total	27	13	40

Table 5: Influence of protein structure on Mycolactone Polyketide Synthase A control

Protein structure	**Positive**	**Negative**	**Total**
Alpha helix	5	0	5
Random coil	6	9	15
Total	11	9	20

Table 6: Secondary structure of the studied Mycolactone Polyketide Synthase A

	Secondary structure			
Organism accession number	**Alpha helix**	**Beta turn**	**Extended strand**	**Random coil**
AY743331.1	42.82	5.34	13.30	38.53
AY505303.1	36.21	6.00	12.95	44.84
AY505302.1	31.15	9.62	20.77	38.46
AY505301.1	7.84	6.72	11.57	73.88
AY505300.1	44.79	4.98	12.80	37.44
AY505299.1	51.68	6.71	16.78	24.83
AY505298.1	10.80	5.26	17.17	66.76
AY505296.1	46.32	5.88	13.24	34.56
AY505295.1	24.94	5.23	17.81	52.02
AY505294.1	57.32	7.11	14.23	21.34
AY394445.2	8.90	1.37	4.79	84.93
AY356360.1	26.50	6.01	17.15	50.33
AY356359.1	15.23	7.61	14.21	62.94
AY358358.1	29.66	12.71	16.10	41.53
AY394448.1	22.31	6.61	18.18	52.89
AY394444.1	16.50	20.39	26.21	36.89
AY394443.1	24.37	3.94	21.86	49.82
AY394442.1	7.49	3.52	9.25	79.74
AY394447.1	37.97	11.28	12.78	37.97
AY394440.1	25.46	8.33	13.52	52.69

Table 7: Nucleotide substitution patterns in the studied Mycolactone Polyketide Synthase A

	Nucleotide Substitution Frequencies				
	T(U)	C	A	G	Total
AY43331.1	16.6	37.4	21.2	24.8	4830
AY505303.1	17.3	38.4	21.4	22.9	1251
AY505302.1	24.3	25.1	15.9	34.7	824
AY505301.1	17.5	36.7	18.7	27.1	815
AY505300.1	23.4	28.0	19.0	29.5	1267
AY505299.1	24.4	24.9	15.7	34.9	896
AY505298.1	17.1	37.7	21.4	23.8	1104
AY505296.1	22.2	28.2	20.4	29.2	2040
AY505295.1	23.3	27.7	17.3	31.7	1315
AY505294.1	24.1	25.2	16.5	34.2	717
AY394445.2	16.5	37.6	21.9	24.0	442
AY356360.1	15.7	39.2	21.4	23.7	1348
AY356359.1	17.6	38.2	20.1	24.1	607

AY356358.1	21.7	20.1	17.7	40.5	373
AY394448.1	24.5	24.0	15.8	35.6	379
AY394447.1	21.5	19.1	12.3	47.1	325
AY394444.1	23.3	27.7	17.0	32.0	881
AY394443.1	19.7	24.1	18.7	37.4	700
AY394442.1	22.9	28.2	20.8	28.2	849
AY394440.1	19.7	29.2	21.2	29.9	3361

A = adenine, T/U = thymine/uracil, C = cytosine G = guanine

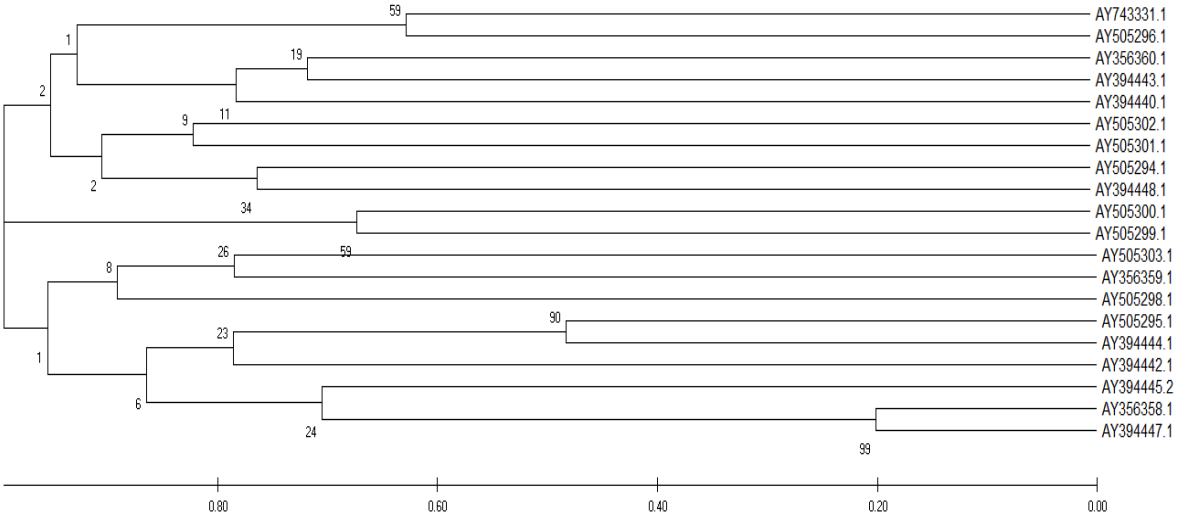

Figure 1: Phylogenetic relationship of the studied Mycolactone Polyketide Synthase A

Discussion

The use of bioinformatic tools for sequence comparison in order to deduce important gene structure and functions have been well documented (Ehrlich *et al.,* 1999; Thomas *et al.,* 2015). In this study, the comparison of the mycolactone polyketide synthase A of *Mycobacterium ulcerans* revealed variation in their homology and length. These findings may be a pointer to the fact that these parameters are not crucial among factors that influence the regulation of mycolactone polyketide synthase A. The fact that eleven (55%) of the studied mycolactone polyketide synthase A have the putative transcriptional factor corresponding to the binding motifs of *Sig A* affirmed the earlier identification of the Sig A-like promoter as a driver of the expression of the mycolactone polyketide megasynthases mls A and mls B in *M.ulcerans* (Tobias *et al.,* 2009). Some (45%) of the studied mycolactone polyketide synthase A lack the *Sig A* binding motifs to ascertain that the regulation of studied mycolactone polyketide synthase A fragments may not be totally exclusive to *Sig A* factors. However, there may be need to further confirm the hypothesis in future. In another vein, a very high proportion (70%) of these mycolactone polyketide synthase A possess the *Sig D* transcriptional binding motifs as an indication for their importance in mycolactone polyketide synthase A regulation. Consequently, mycolactone polyketide synthase A having neither the *Sig A* nor *Sig D* factors were also found in this study. This observation strongly corroborated the earlier comment quoted in other literature that other factor(s) beyond the sigma factors are also involved in mycolactone polyketide synthase A expression (Deshayes *et al.,* 2013). According to them, regulation of mycolactone polyketide synthase A is contingent upon nutrient source and plant polysaccharide was even taunted as a significant stimulants of *Mycobacterium ulcerans* growth including their toxins synthesis regulation.

Conclusion

Results of this study have therefore shown that regulation of the studied mycolactone polyketide synthase A fragments is not only limited to the presence of sigma factors but could also be strongly linked to other factors beyond these factors. We however, would consider evaluating other arising hypothesis from this study using the entire mycolactone gene in future.

References

Amofah G, Bonsu F, Teeteh C, OKrah J, Asamoa K, Asiedu K and Addy J. Buruli ulcer in Ghana: results of a national case search. *Emerg. Infect Dis*. 2012; 8(2), 167-170.

Debacker M, Aguiar J, Steunou C, Zinsou, C, Meyers WM, Guedenon A, Scott JT, Dramix M and Portaels F. *Mycobacterium ulcerans* disease (Buruli ulcer) in rural hospital, Southern Benin,1997-2001. *Emerg. Infect. Dis*. 2014; 10(8), 1391-1398.

Deshayes C, Angala SV, Marion E, Brandli I, Babonneauu J, Preisser L, Eyangoh S, Delneste Y, Legras P, Chastellier CD, Stinear TP, Jackson M, Marsollier L. Regulation of Mycloactone, the *Mycobacterium ulcerans* toxin, depend on Nutrient Source: *PLOS Neglected Tropical Diseases.* 2013; 7(11), 1-12.

Ehrlich KC, Cary JW and Montalbano BG. Characterization of the promoter for the gene encoding the aflatoxin biosynthetic pathway regulatory protein AFLR. Biochim. Biophys Acta. 1999. 14(44): 412-417.

George KM, Pascopella L, Welly DM and Small PLC. A *Mycobacterium ulcerans* toxins, mycolactone causes apoptosis in Guinea pigs ulcers and tissue culture cells. *Infection and Immunity*. 2000; 68 (2), 877-883.

Geourjon C and Deleage G. SOPMA significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.*1995;11(6), 681-684.

Jukes TH, Cantor CR. Evolution of protein molecules. In: Munro H.N., editor. Mammalian Protein Metabolism. Academic Press; New York: 1969. pp. 21–132.

Kimura M. Estimation of evolutionary distances between homologous nucleotide sequences. Proc. Natl. Acad. Sci. Unit. States Am. 1981;78:454–458.

Leao SC, Romano MI, Garcia MJ. Tuberculosis,Leprosy and other Mycobacteriosis. Chapter BO3. In: Bioinformatics in tropical disease research: A practical and case study approach accessed on 19/11/2019 @www.ncbi.nlm.nih.gov/books/NBK6820.

Sanhueza D, Guegan JF, Jordan H and Chevillon C. Environmental Variations in *Mycobacterium ulcerans* transcriptome: Absence of Mycolactone Expression in suboptimal Enviroments. *MDPIAG*. 2019, 11(3).

Tamura K, Dudley J and Nei M. MEGA 4 Molecular Evolutionary Genetics Analysis (MEGA software version 4.0. Molecular Biology and Evolution. 2007; 24, 1596-1599.

Thomas BT, Oni MO, Agu GC, Lakunle OM, Davies A. In Silico Sequence Analysis of Pks Gene of Ochratoxin A Producing Aspergillus Section Nigri. *International Journal of Genetics.* 2015; 5(2), 48-52.

Tobias NJ, Seemann T, Pidot SJ, Porter JL, Marsollier L. Mycolactone Polyketide Synthase A expression is controlled by strong SigA-like promoters with utility in studies of *Mycobacterium ulcerans* and buruli ulcer. PLoS Negl Trop Dis . 2009; 3: e553.

Wagner T, Benbow ME, Brenden OT, and Johnson RC. Buruli ulcer disease prevalence in Benin, West Africa: associations with land use/cover and the identification of disease clusters. *International Journal of Health Geographics*. 2008; 7(25), 1-13.