

***Mycobacterium terramassiliense*, *Mycobacterium rhizamassiliense* and *Mycobacterium numidiamassiliense* sp. nov., three new *Mycobacterium simiae* complex species cultured from plant roots.**

A. Bouam<sup>1</sup>, N. Armstrong<sup>1</sup>, A. Levasseur<sup>1</sup> and M. Drancourt\*<sup>1</sup>

**Supplementary file 1**

Details of OriC regions predicted in the genome of strains AB308, AB215 and AB57.

***Mycobacterium terramassiliense* :**

Genome size	6029590 nt
Genome GC content	0.6839
DnaA box distribution	<a href="#">[DnaA box distribution]</a>
OriC length	647 nt
OriC AT content	0.3184
The number of DnaA box	3
The location of oriC region	3483469..3484115 nt
The location of dnaA gene	-
The extremes of GC disparity	3431648 nt (minimum), 6022463 nt (maximum)
The extremes of AT disparity	5893509 nt (minimum), 3960 nt (maximum)
The extremes of RY disparity	3431573 nt (minimum), 6025419 nt (maximum)
The extremes of MK disparity	6022453 nt (minimum), 3387045 nt (maximum)
Note	Note that the E. coli perfect DnaA box (ttatccaca) was searched for with no more than two mismatches. <a href="#">[Gene list (zcurve1.02)]</a>
Z-curves	<a href="#">[Figure1]</a> <a href="#">[Figure2]</a>
OriC Sequence	The DnaA boxes identified in the below sequence are capitalized and also marked in bold, if any.
	<pre> catggtcgtcgatatgcacggcccgccgagccgggacatcgtcgacaaggcgatcaccgc gctgctgaacctcgaacaccgcgccagggcgccgagccgcgagcggtgacggcgc cggcatcctgattcaggtgccggacgcgttctgcgaggtcgtggatttcaactgcc cgctcgggcagctacgccaccggtatcgcttcttccgcagtcgtcgaaggacgccgc gaccgcgtgcccgc<b>GGTGGAGAA</b>gatcgccgagtcgagggcctgcaggtcatcggctg gcgcaactgcccacggacgactcgtcgtgggtgcctgtcccgcgacgcgatccac ggtccggcaggtgttcatgacgggcgctccgggatggcgtggagcgacgcgctactg ggtgcgcaagcgcgccgagcacgagctgggcaccaagggcccgggccaggatggcccggg tcgcgaaaccgttacttccaagccttccggtcagacct<b>TCATCTACA</b>agggcatgct gaccaccccgagctcaaggcgttttaccttgacctgcaagacgatcggctgaccagcgc </pre>

gctgggcatcgtgcactcccgc**TTCTCCACG**aacaccttcccgtcgt

*Mycobacterium numidiamassiliense*

The information of genome and oriC region

Genome size	6248949 nt
Genome GC content	0.6585
DnaA box distribution	[ <a href="#">DnaA box distribution</a> ]
OriC length	534 nt
OriC AT content	0.3914
The number of DnaA box	6
The location of oriC region	1772264..1772797 nt
The location of dnaA gene	-
The extremes of GC disparity	1848671 nt (minimum), 5310514 nt (maximum)
The extremes of AT disparity	5327831 nt (minimum), 1897876 nt (maximum)
The extremes of RY disparity	1848674 nt (minimum), 5246343 nt (maximum)
The extremes of MK disparity	5328449 nt (minimum), 1857179 nt (maximum)
Note	Note that the E. coli perfect DnaA box (ttatccaca) was searched for with no more than two mismatches. [ <a href="#">Gene list (zcurve1.02)</a> ]
Z-curves	[ <a href="#">Figure1</a> ] [ <a href="#">Figure2</a> ]
OriC Sequence	The DnaA boxes identified in the below sequence are capitalized and also marked in bold, if any.
	tagcgggtagctgaaagggccac <b>TGTTCATAA</b> cactgtgcatcgtg <b>TGTGCATAT</b> ccttc gacacctgtgttgagtgggt <b>TGTGGAGAA</b> cc <b>TGTGGATTA</b> aaacgcggcggtgaagcatc cgtgcagatacgcgcagtacaccactgcacatttgg <b>TGTGGAGAA</b> caattcgcctcggcgt gtcgggccagggttactgcgtcgggctgttggctt <b>TCACCCACA</b> tttctgcgcgagttacg agccggttaaggcgccccatgcaccccggtcaccgaaatagttcgcacaggtccgcccgcac cgcgcgtctacctagccggctctgagcgtcacagagcgccaaacgagttccggcggcca ccctcaggggocggocgggacaaagcaaacagtgctcagagtttagctgtccgcgacgacgt ccagcgtgaactccacgcttgacctccgggtgcaggtgcaccgacaccgagtgccgaccga tggccttgatgtgcgccttgggcagccggacgatgcgcttgtccaggttcgggc

The information of genome and oriC region

Genome size	6248949 nt
Genome GC content	0.6585
DnaA box distribution	[ <a href="#">DnaA box distribution</a> ]
OriC length	923 nt
OriC AT content	0.3792
The number of DnaA box	4
The location of oriC region	1846020..1846942 nt
The location of dnaA gene	-
The extremes of GC disparity	1848671 nt (minimum), 5310514 nt (maximum)
The extremes of AT disparity	5327831 nt (minimum), 1897876 nt (maximum)
The extremes of RY disparity	1848674 nt (minimum), 5246343 nt (maximum)
The extremes of MK disparity	5328449 nt (minimum), 1857179 nt (maximum)
Note	Note that the E. coli perfect DnaA box (ttatccaca) was searched for with no more than two mismatches. [ <a href="#">Gene list (zcurve1.02)</a> ]
Z-curves	[ <a href="#">Figure1</a> ] [ <a href="#">Figure2</a> ]
OriC Sequence	The DnaA boxes identified in the below sequence are capitalized and also marked in bold, if any.
	<pre> aagcgcgcgcggcccttgccggcgcgaccggacacgatggcaccgccggcgcggggtacg catgcgcagacggaagccgtgcaacggcggcggcgccggttggtcggctgaaaggtccg cttgcccttgccacggcattctcctcgggtctgtcttgctcgaagctttcccagccatcc ggccgcccgcggttttcaccgaatcgccggctcgatcgaaaagtgggtcttgctactggccgg cgcggttccccctgacataacctgggtcgcagccgcacatcgccgactttcgggcgactgtatg agggtactgatgaggattcgccctgggtcaaaacttggcctgccccaacggcaccagcggaac tgcgacccccgaaccggcaccgcgcggcggaaaacgaaaaatgccccacggcagcttaaa gaagccgactggaatgcagcggaaacgggttggcagccgcacggaaaactgtagcttcggg caatgccggtttcagacccccagcggcgcaagacaacgaagcaaggatggcggatcaactgg ctgtctaggtgaacttctcgaggttccctccagcacggagatcgcgagccgtcgccggta ggctgcggcccgtatacgatcacc<b>CTGTCCACA</b>cc<b>TGTGTATAA</b>cta<b>TGTGGACAG</b>atgc ttcgtcgcgaagcgtgggttagctcaagcccagagtttgcgagaaccagggagatgcgt cgttgaccgatgacccccggttcgggctttaccagcgtatggaatgcagtcggttccgaac tcaatggcgcagccagcgcggacgggtctcaccaccaatcgaaacgggt<b>TCTGGACAA</b>cccgc taaccctcagcaacgagccctggctgaatctcgttcagccactcaccatcgtcgaggggt ttgccctgctgtctgtgcccagac </pre>

The information of genome and oriC region

Genome size	6248949 nt
Genome GC content	0.6585
DnaA box distribution	[ <a href="#">DnaA box distribution</a> ]
OriC length	489 nt
OriC AT content	0.4376
The number of DnaA box	5
The location of oriC region	1947253..1947741 nt
The location of dnaA gene	-
The extremes of GC disparity	1848671 nt (minimum), 5310514 nt (maximum)
The extremes of AT disparity	5327831 nt (minimum), 1897876 nt (maximum)
The extremes of RY disparity	1848674 nt (minimum), 5246343 nt (maximum)
The extremes of MK disparity	5328449 nt (minimum), 1857179 nt (maximum)
Note	Note that the E. coli perfect DnaA box (ttatccaca) was searched for with no more than two mismatches. [ <a href="#">Gene list (zcurve1.02)</a> ]
Z-curves	[ <a href="#">Figure1</a> ] [ <a href="#">Figure2</a> ]
OriC Sequence	The DnaA boxes identified in the below sequence are capitalized and also marked in bold, if any.
	agaagaacaacttagagatctaagtatcagtattaaggcctgtgcattctggggacagc ctggcttctgtgcagctaggacctcgatggggg <b>TGTGGATGA</b> cagcgcgcgtggctgtgt gcacggtggtgagcggg <b>TGGGGATGA</b> atgacagatgtacacggcggcgcggt <b>TCATCCAC</b> <b>G</b> gtggtttgctgct <b>TTTTCCACA</b> gcgctgccacaccgtcaggg <b>TGTGGCTGA</b> tgtgac agccgtgcgagaagtTTTTTgaaaaagtgtgactcgtgagtttcaaatacagcgttggg gcgctggcggatgagagtcgtgagttctttaacgtggtcgaacacctcacgtcgcctcggc catctcggacaagatctttcgttgccgacatacatgaccgtggtgtggtcgcggccgaatgc ctgtccgatcttcggcagcagaggtcagtgagttcgcggcacaggtacatcgcgatctg ccgcgactg

*Mycobacterium rhizamassiliense*:

Genome size	6015465 nt
Genome GC content	0.6722
DnaA box distribution	[ <a href="#">DnaA box distribution</a> ]
OriC length	488 nt
OriC AT content	0.4344
The number of DnaA box	4
The location of oriC region	1721223..1721710 nt
The location of dnaA gene	-
The extremes of GC disparity	1721273 nt (minimum), 4895128 nt (maximum)
The extremes of AT disparity	4894759 nt (minimum), 644303 nt (maximum)
The extremes of RY disparity	1719935 nt (minimum), 4506867 nt (maximum)
The extremes of MK disparity	4895128 nt (minimum), 1705512 nt (maximum)
Note	Note that the E. coli perfect DnaA box (ttatccaca) was searched for with no more than two mismatches. [ <a href="#">Gene list (zcurve1.02)</a> ]
Z-curves	[ <a href="#">Figure1</a> ] [ <a href="#">Figure2</a> ]
OriC Sequence	The DnaA boxes identified in the below sequence are capitalized and also marked in bold, if any. agcgtcccttcacctacaaaaatctcagtcagtggttagccgcgcccccgctagcgtg gggccccgaaacggggagtcgaagaacaaccgtagatcgtccggggccatcttgaaagc taatcgcggaaggccgacaaccgcagtgatcgagacgcttgtaagccccctggggacgaggc gtccccaaacctgttcttcaaagaagaaacgacgaagagatctcagtaccagtattaagg gctgtgcaagt <b>TGGGGAGAA</b> gtgcaactcctgtgtagctaggggagcggcgggg <b>TGTGGA</b> <b>TGA</b> cgcggtgcgcaactgtgtgtcccgtgtgggtcgac <b>TGGGGATGA</b> attacagatgtac accgatggccgattactgcacgggggttgcgtggt <b>TTTTGCACA</b> agcctgcgcatag ccagcgtgtggctggtgtgacagacgtgtgggaagttttttgaaaaagtgtgactcgtgg ggtgcaaa

The information of genome and oriC region

Genome size	6015465 nt
Genome GC content	0.6722
DnaA box distribution	[ <a href="#">DnaA box distribution</a> ]
OriC length	793 nt
OriC AT content	0.3670
The number of DnaA box	3
The location of oriC region	1723229..1724021 nt
The location of dnaA gene	-
The extremes of GC disparity	1721273 nt (minimum), 4895128 nt (maximum)
The extremes of AT disparity	4894759 nt (minimum), 644303 nt (maximum)
The extremes of RY disparity	1719935 nt (minimum), 4506867 nt (maximum)
The extremes of MK disparity	4895128 nt (minimum), 1705512 nt (maximum)
Note	Note that the E. coli perfect DnaA box (ttatccaca) was searched for with no more than two mismatches. [ <a href="#">Gene list (zcurve1.02)</a> ]
Z-curves	[ <a href="#">Figure1</a> ] [ <a href="#">Figure2</a> ]
OriC Sequence	The DnaA boxes identified in the below sequence are capitalized and also marked in bold, if any.
	cgacgcacatctccctgggttctcgcaaactctgggctgagctacaacctcgctgtgacgaca aagcaa <b>CTGTCCACA</b> tag <b>TTATACACA</b> gg <b>TGTGGACAG</b> gatgatcgatatacgggcccga cctaccggcgacggctcgcgatctccgcgctggaggcaaccccgtggaggtcgcctagac agctagtcgatccgccatcctcgttctggtcatgcccgttgacagtctgaaacggcat tgcccgaagctaacagttttccgaccggctgccaacagttcggctgcattccaatcggct tctctagcctccttgggcgctctttctctcggccgcccggacgggaggccgagcggcctg cctgagcacgggtattttccggtgacgagtggtagtcgcccgttgccaggacaagttt gaccaggcgaagtgtcatcagtaacctcatacagtcgcccgaagtcggcgatacggctg cgacccgagccatctgtgccccgaccgcccggccagcagcagttagaccactttccgac tgaccccggggttgcaaacgcgagcggccggatggcaatgcgagacacaccgaggagaa tgccgtggccaagggcaagcggacctttcagccgaataaccggcgccgagcccgtgtgca cggcttccgtctgcgatgcgtacccgcgcgggctgcccattgtgtcgggtcggcgtcg caagggtcgcgcgcgctttctgcctgatccgcgccgacaggttcgtgggcggtgctttc ggcacgcaaccgt

The information of genome and oriC region

Genome size	6015465 nt
Genome GC content	0.6722
DnaA box distribution	[ <a href="#">DnaA box distribution</a> ]
OriC length	110 nt
OriC AT content	0.4818
The number of DnaA box	3
The location of oriC region	1743438..1743547 nt
The location of dnaA gene	-
The extremes of GC disparity	1721273 nt (minimum), 4895128 nt (maximum)
The extremes of AT disparity	4894759 nt (minimum), 644303 nt (maximum)
The extremes of RY disparity	1719935 nt (minimum), 4506867 nt (maximum)
The extremes of MK disparity	4895128 nt (minimum), 1705512 nt (maximum)
Note	Note that the E. coli perfect DnaA box (ttatccaca) was searched for with no more than two mismatches. [ <a href="#">Gene list (zcurve1.02)</a> ]
Z-curves	[ <a href="#">Figure1</a> ] [ <a href="#">Figure2</a> ]
OriC Sequence	The DnaA boxes identified in the below sequence are capitalized and also marked in bold, if any.
	ccgcggtcgccgacaagctcagctaaaaggtccttgcggttgacatggtttttgcggtt <b>GTGGGTAA</b> accagcagcactccgcg <b>TGTGGATCTGTGAAAAA</b> cagctgaa