Genomic Analysis of *Mesorhizobium loti* Strain TONO Reveals Dehalogenases for Bioremediation

Sefatullah Zakary 1,2, Habeebat Adekilekun Oyewusi 1,3, Fahruil Huyop 1*  

1 Department of Biosciences, Faculty of Sciences, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia  
2 Department of Botany, Faculty of Biology, Kabul University, Deh Buri, Kabul 1006, Afghanistan  
3 Department of Biochemistry, School of Science and Computer Studies, Federal Polytechnic Ado Ekiti, Ado Ekiti PMB 5351, Ekiti State, Nigeria

**ABSTRACT**

Halogenated compounds are extensively utilized in different industrial applications such as pesticides and herbicides and cause severe environmental problems because of their toxicity and persistence. Degradation of these compounds by the biological method is a significant method to reduce these recalcitrant. *Mesorhizobium loti* is important for nitrogen fixation in legume roots. Up to now, there is no report to indicate *M. loti* can produce dehalogenase enzymes. Thus, a total of twenty-five genomes of *M. loti* strains from the National Center for Biotechnology Information (NCBI) were analyzed. These strains notably carry dehalogenase genes and were further investigated. The relative ratio of haloalkane and haloacid dehalogenase type II or L-type from all twenty-five genomes was 26% and 74%, respectively, suggesting type II dehalogenase is common. Surprisingly, only *M. loti* strain TONO carries four dehalogenases and therefore it was further characterized. The chromosome of *M. loti* strain TONO contains four haloacid dehalogenase type II genes namely, dehLt1 (MLTONO_2099), dehLt2 (MLTONO_3660), dehLt3 (MLTONO_4143), and dehLt4 (MLTONO_6945), and their corresponding enzymes were designated as DehLt1, DehLt2, DehLt3, and DehLt4, respectively. The only haloalkane dehalogenase gene (MLTONO_4828) was located upstream of the dehLt3 gene and its amino acid share 88% identity with DmlA of *Mesorhizobium japonicum* MAFF 303099. The putative haloacid permease gene designated as dehrPt (MLTONO_0284) was located downstream of dehLt1 and its amino acids show 69% identity with haloacid permease of *Rhizobium sp.* RC1. The gene encoding helix-turn-helix (HTH) motif family DNA-binding protein regulator and LysR family transcriptional regulator genes were also identified, possibly for regulatory functions. The genomic studies as such, have good potential to be screened for new type of dehalogenases based on basic molecular structure and functions analysis.

**Keywords**: Genomic analysis, Haloacid dehalogenase, Haloacid permease, Halogenated organic compounds, *Mesorhizobium loti*

**Introduction**

A significant group of halogenated organic compounds such as chlorinated hydrocarbons was produced and extensively utilized in agriculture and various industrial applications. This resulted in severe environmental dilemmas due to their toxicity and persistence [1]. Fortunately, many bacterial species produce dehalogenases that could breakdown these toxic compounds and convert them into harmless products [2]. Most of these bacteria species used organohalides as their sole source of carbon and energy [3]. Dehalogenase catalyzes halogenated organic pollutants’ by cleaving of carbon–halogen bonds. Many dehalogenases were identified, for example haloalkane dehalogenase [4], fluoroacetate dehalogenase [5], haloalkanoic acid dehalogenase [6-8], and 4-chlorobenzoyl-CoA dehalogenase [9]. Haloalkanoic acid dehalogenase catalyzes haloalkanoic acids
and produces hydroxy-acid and a halide [10].

The 2-haloacid dehalogenases were classified into three groups based on the substrate specificities [11], whereas the genetic approach by Hill et al. [10] classified dehalogenases into groups I and II. Group II was stereoselective dehalogenases acting only on L-2-haloacids. While group I dehalogenases were non stereospecific dehalogenases can react with both D and L form of substrates producing L- and D-hydroxy acids, respectively or can retain the chemical’s isomeric pattern [10]. D-specific dehalogenase was categorised as group I as well. Slater and colleagues [8] showed that a dehalogenase crude extract from Pseudomonas putida PP3 grown on 2-chloropropionate, dehalogenate only a small amount 3 chloropropionate. Recently, these types of dehalogenases have received more attention since not many dehalogenases that remove chloride from the β-carbon of chloroalkanoic acids [12, 13].

The genetic organization of Rhizobial dehalogenases and their operon were based on gene cloning analysis. The location of dehE and dehD/dehL were far from clear. However, it was strongly believed that a single regulator gene (dehr) controls all three dehalogenases [2]. To understand how dehalogenases is regulated, soil bacteria like Mesorhizobium species is studied due to the fact that its full genome sequence readily available in the genebank [14]. Therefore, soil bacteria which is closely related to Rhizobium sp. RC1 like Mesorhizobium will be further investigated. There were several other bacteria species with the ability to degrade pollutants had been reported [15-18]. However, not many studies reported on the use of whole-genome sequencing to reveal their genetic sources of pollutant degradation potential [14]. The current study is to screen several Mesorhizobium strains for the presence of specific dehalogenases. By nature, the Mesorhizobium isolated from soil, regarded as rhizosphere bacterial of agronomic significance due to their ability to form nitrogen-fixing interactions with leguminous plants [19-21]. Analyzing dehalogenase genes from this bacterium will predict other possible functions aside from nitrogen-fixation.

Herein, we partially analyze the complete genome sequence of Mesorhizobium loti strains in view of predicting their pollutant's degradation ability just by focusing on the dehalogenase enzymes family. The metabolic novelty of M. loti strain TONO and its potential for use in bioremediation provided the motivation for complete genome sequence analysis. Here, we describe bioinformatics and phenotypic analysis that reveals diverse metabolic capabilities of strain TONO that may be useful for bioremediation of halocetate and haloacid pollutants [20]. The dehalogenase genes arrangement including their upstream and downstream regions and their non-coding sequences for the possible regulator and promoter binding sites in the complete genome sequence were investigated. The newly identified dehalogenase genes were aligned using multiple sequence comparison and phylogenetic tree. We also highlighted on the regulatory element that might present and predicted to be potential in dehalogenase regulatory mechanism for M. loti strain TONO. This approach will provide an integrative knowledge to understand the complete bioremediation process.

**Material and Methods**

**Whole genome sequence retrieval**

Altogether, twenty-five genomes of M. loti strains were retrieved, and the annotated genomes were downloaded from http://www.ncbi.nlm.nih.gov/genome (GenBank accession number listed in Table 1). The dehalogenase genes within these genomes were searched, using the dehalogenase as a keyword search. Then the number and types of dehalogenase genes were recorded.

**Phylogenetic analysis of dehalogenase amino acid sequences**

Phylogenetic analysis is a highly reliable bioinformatics approach to find out the differences and identities among genes, proteins, taxa and organisms [22]. The phylogenetic analyses of dehalogenases from M. loti strain TONO with other dehalogenases, L-DEx (Pseudomonas sp. Strain YL; AAB32245), DehH109 (P. putida 109; BAA04474), DehII (P. putida PP3; CAB38900), DehCl (Pseudomonas sp. CBS3; AAA63640), DehIVA (Pseudomonas cepacia MBA4; CAA46976), DehL (Rhizobium sp. RC1; CAA63794) were carried out. The phylogenetic tree was constructed using the CLUSTAL–W and MEGA7 software package [23, 24]. The sequence percentage identified among dehalogenases was then auto-calculated from Clustal 2.1 Identity Matrix [25].

**Multiple sequence alignment and conserved amino acid residues**

The amino acid sequence of selected dehalogenases from the genome sequence of M. loti strain TONO (GenBank accession: AP017605 to AP017608) [20] were retrieved. The amino acid residues of all type II dehalogenases were aligned
using the multiple sequence alignment method by Multalin (multalin.toulouse.inra.fr/multalin/) [26] and was enhanced by the ESPRRIPT program [27] to look for conserved amino acid residues.

**Locus prediction and genetic analysis of dehalogenase genes**

To define the potential operon of the genetic arrangement, the FGENESH program available in Softberry tools (http://www.softberry.com/) was utilised to check for the presence of the multiple genes in genomic DNA sequences. The software also predicts operon on ORFs and its neighboring genes of the bacterial genome, promoter and terminator prediction [28]. The InterPro (https://www.ebi.ac.uk/interpro/) was used for the classification of protein families.

**Results and Discussions**

**Screening of dehalogenase genes**

An advent of genome sequencing coupled with advances in bioinformatics promises invaluable insights into genomic analysis, leading to new knowledge in gene evolution, ecology, and the design of related therapeutic interventions. Previously, about 1,800 bacterial genomes have been fully sequenced, and these cover most of the major bacterial pathogens [29]. Genome sequencing can help in identifying an organism. When multiple genomes of the same or closely related species present, the ‘pan-genome’ can be designed and analysed as a single comprehensive catalog of all the sequences in the population [30]. In current study, 25 genome sequences of various M. loti strains were screened for dehalogenases. The dehalogenase genes were identified within these genomes alongside of size, G+C content (%) and protein-coding sequences (CDS) as summarised in Table 1. Haloacid dehalogenase type II (typeII) and haloalkane dehalogenase (HKD) were mostly identified (Table 1). The total percentage of the proportion of haloalkane and haloacid dehalogenase type II was 26% and 74% from the total of all dehalogenase genes, respectively. It was a common feature that many soil bacteria produce more than one dehalogenase. Many of these dehalogenases were identified in their chromosomes, although some were in the plasmids. However, no dehalogenase gene was detected in the plasmids of these strains. The presence of multiple dehaloge-
Haloacid dehalogenase type II was common compared to D-specific dehalogenases since most naturally occurring halogenated compounds exist in the L-form [15]. This is the first report on M. loti strain TONO producing four haloacid dehalogenases apart from Rhizobium sp. RC1 [31, 32]. These facts raised a fascinating question as to why M. loti strain TONO produces four haloacid dehalogenase type II. It was suggested that the ability to produce all four dehalogenases may afford the M. loti strain TONO the ability to degrade or utilize high amount of halogenated organic pollutants thus, suggesting that it might be a good candidate used for bioremediation of halogenated organic compounds or used in the synthesis of chirally active industrial chemical intermediates.

**Genetic organisation of dehalogenase genes of M. loti TONO**

The full chromosome of M. loti strain TONO sequence was further analysed to search for regulatory gene functions in relation to the control of all dehalogenases from the perspective of genome studies. M. loti, also known as Rhizobium loti, is fast-growing Rhizobia that lives in symbiosis with the genus Lotus sp. legumes. These bacteria infect leguminous plants to form root nodules, resulting in the exchange of fixed nitrogen and carbon between rhizobia and plants from the atmosphere [33]. The genetic organization of M. loti strain TONO related to dehalogenase has not been studied so far from the whole genome sequence [20]. The genome consists of a chromosome (7,856,088 bp; 62.8% G+C content) and three plasmids (80,491 bp; 60.4% G+C, 294,703 bp; 59.5% G+C and 220,869 bp; 58.9% G+C content (%)). The genome sequence was deposited in DDBJ/EMBL/GenBank under the accession number AP017605 to AP017608 and annotated by the NCBI Prokaryotic Genome Annotation Pipeline [20].

The location of dehalogenase assayres genes was investigated such as regulatory and an uptake gene. In this study, a putative uptake gene (MLTONO_0284) that encodes for Major Facilitator Superfamily protein (designated as DehrPt) was identified. The gene is possibly responsible for haloacid uptake protein. All four haloacid dehalogenase type II were present and designated as dehLt1 (MLTONO_2099), dehLt2 (MLTONO_3660), dehLt3 (MLTONO_4143), dehLt4 (MLTONO_6945). The question mark (?) indicates non-coding regions. (C) The upstream and downstream genes of dehalogenases.

---

**Figure 1.** Genetic organization of dehalogenase genes of M. loti TONO. (A) The gene location within the genome. (B) Arrows indicate the direction of transcription, dehrPt (MLTONO_0284) represents the haloacid uptake gene. The dehalogenases are designated as dehLt1 (MLTONO_2099), dehLt2 (MLTONO_3660), dehLt3 (MLTONO_4143), dehLt4 (MLTONO_6945). The question mark (?) indicates non-coding regions. (C) The upstream and downstream genes of dehalogenases.
dehLt1, dehLt2, dehLt3, haloalkane dehalogenase gene and dehLt4 were 1,619,992 bp, 489,283 bp, 719,363 bp and 2,301,883 bp, respectively (Figure 1). The genes were not located near to each other in the genome. The sequences at the upstream position of each dehLt1, dehLt2, dehLt3, and dehLt4 were 320 bp, 80 bp, 83 bp, and 155 bp, respectively, and were not encoded for any functional proteins and possibly consist of specialised promoters for each dehalogenase genes (Figure 1.).

The genome analysis by the FGENESH server showed that these dehalogenase genes were not located in a single operon to have a single regulator gene. Unlike dehalogenases from Rhizobium sp. RC1, all dehalogenases were located as one operon and regulated by a single regulator gene [2, 15]. Herein, we hypothesized that dehalogenases of M. loti strain TONO may not possibly be controlled by one regulator gene since they are not located near to each other in the genome. Therefore, further analyses of all these dehalogenases, such as site-directed mutagenesis (SDM) is needed to clearly state their functions.

The dehLt1 forms an operon with the gene (MLTONO_2098) encoding FAD-binding oxidoreductase. The adjacent upstream of dehLt1 is a gene encoding GNAT family N-acetyltransferase. There is no regulator gene adjacent to these genes except the Lrp/AsnC family transcriptional regulator (MLTONO_2107), located six genes (7,566 bp) downstream of the GNAT family N-acetyltransferase gene (Figure 1C). The dehLt2, a gene (MLTONO_3661) encoding helix-turn-helix domain-containing protein and a gene (MLTONO_3662) encoding DUF2312 domain containing protein were located in an operon (Figure 1C). Analysis by InterPro showed the DUF2312 domain-containing protein as a hypothetical protein. Besides, helix-turn-helix domain-containing protein was classified as helix-turn-helix (HTH) motif family DNA-binding protein regulator and possibly regulates dehLt2. The dehLt3
and a gene (MLTONO_4144) encoding the DMT family transporter were organized as an operon. While upstream of the latter gene, a gene (MLTONO_4145) encoding LysR family transcriptional regulator exists, and it is not included in any operon (Figure 1C). The regulation of the dehalogenase gene, {\textit{hdhA}} in \textit{Sinorhizobium meliloti} strain 1021 by LysR family transcriptional regulator, HdhR, is previously reported, which is located upstream of the \textit{hdhA}. HdhR represses the transcription of \textit{hdhA} via direct binding of HdhR to the \textit{hdhA} promoter [34]. It should be mentioned here that, the finding of this study corroborated with previous studies in which most dehalogenase regulators are located adjacent upstream of dehalogenase structural genes [15, 34-37].

The \textit{dehLt4} gene is not organized in any operon. In the annotated genome, the two genes upstream of \textit{dehLt4} are genes encoding branched-chain amino acid aminotransferase and superoxide dismutase (Figure 1C). Haloalkane dehalogenase (HKD) is organized in an operon with the TetR family transcriptional regulator C-terminal domain-containing protein gene (MLTONO_4827), peptide chain release factor 3 gene (MLTONO_4829), and IS5 family transposase gene (MLTONO_4830) (Figure 1C).

**Phylogenetic analysis of dehalogenases from \textit{M. loti} strain TONO**

Sequence comparison of the dehalogenase amino acid from \textit{M. loti} strain TONO with established dehalogenase amino acids suggested that all four dehalogenases were highly identical to group II. Therefore, multiple sequence alignment by CLUSTAL–W and Maximum Likelihood techniques [38] were used to identify the evolutionary relationships of all dehalogenases from \textit{M. loti} strain TONO compared to the well-established group II dehalogenase as reported by Hill et al. [10]. In Figure 2a, \textit{DehLt4} is closely related to \textit{DehCl} and \textit{DehIVa} that possesses 49% and 48% identity, respectively and has a common ancestor. The \textit{L-DEX}, \textit{DehH109}, and \textit{DehII} are closely related and share a common ancestor. The \textit{DehLt1}, \textit{DehLt2} and \textit{DehLt3} have a common ancestor in which \textit{DehLt2} and \textit{DehLt3} share 50% identity while \textit{DehLt1} have 28% identity with both \textit{DehLt2} and \textit{DehLt3} (Figure 2a). \textit{DehL} from \textit{Rhizobium} sp. \textit{RC1} is not closely related to all members of group II dehalogenase and therefore, it was outgroup in the phylogenetic analysis. Haloalkane dehalogenase is closely related to DmlA of \textit{Mesorhizobium japonicum} MAFF 303099, sharing 88% identity (Figure 2b).

**Multiple sequence alignment**

Multiple sequence alignments were used to describe the level of conserved regions within the dehalogenases of \textit{M. loti} strain TONO (Figure 3). The sequence-based identification of the important amino acid residues is sometimes difficult due to the low percentage of conserved amino acids in such enzymes. Notably, catalytically important amino acids are highly conserved in dehalogenase type II [15]. The nine catalytic important residues in \textit{L-DEX}, \textit{DhlB} and \textit{DehIVa} have been previously reported. \textit{Asp11} in \textit{DehIVa} acts as a nucleophile by attacking carbon no. 2 of the substrate during catalysis [39]. The corresponding residues in \textit{L-DEX} (Asp10) [40] and \textit{DhlB} (Asp8) [41] were reported to have a similar function. Ser176 in \textit{DehIVa} and its corresponding residue Ser175 in \textit{L-DEX}, formed a hydrogen bond with nucleophile Asp to probably maintain the orientation of its carboxyl group in a way suitable to attack the carbon no. 2 of the substrate [39, 42]. The role of Thr14 in \textit{L-DEX} is not yet identified, however its corresponding residue in \textit{DhlB} (Thr12) together with Ser171 and Asn173 were reported to firmly hold the Asp8 in a position that favors the nucleophilic attack [41]. In \textit{L-DEX}, Lys151 and Asp180 interact strongly with Asp10, L-2-chloropropionate and catalytic water. Lys151 stabilizes the substrate orientation and balance the charges around the active site. Asp180 stabilizes the Asp10 rotation, prevents Lys151 from moving toward Asp10 and fixes the catalytic water in proximity to Asp10. Asp180 may independently activate the catalytic water or in collaboration with Lys151, Ser175, and Asn177 [43]. Mutation of all these functional amino acids by site-directed mutagenesis significantly reduced the enzyme activity [44]. Arg41 in \textit{L-DEX} accepts the released chloride ion [45] and its corresponding residue in \textit{DehIVa} (Arg42) is important in the substrate “lock down” mechanism; and also acts a member of the halide- binding cradle together with Asn120 and Trp180 [39]. The alteration of Asp11 and Asn178 residues in \textit{DehIVa} turn the protein to inactive [46].

In this study, it was possible to predict a conserved catalytic residue of dehalogenases in \textit{M. loti} strain TONO by referring to \textit{DehIVa} and \textit{L-DEX} (Figure 3). The conserved amino acid residues of \textit{DehLt1}, \textit{DehLt2}, \textit{DehLt3} and \textit{DehLt4} were summarised in Table 2. The percentage of sequence identities of \textit{DehIVa} with dehalogenases from \textit{M. loti} strain TONO were: 21.00% (\textit{DehLt1}), 23.64% (\textit{DehLt2}), 22.37% (\textit{DehLt3}) and 48.18% (\textit{DehLt4}), whereas with \textit{L-DEX} were 23.62%.
Figure 3. The multiple sequence alignment of DehLt1,2,3 and 4-TONO (M. loti strain TONO) with DehIVa (accession: X66249.1) and L-DEX (accession: S74078.1).

Table 2. The important amino acid residues of dehalogenases in M. loti strain TONO compared with established L-DEX, DehIVa and DhlB.

<table>
<thead>
<tr>
<th>L-DEX</th>
<th>DehIVa</th>
<th>DhlB</th>
<th>DehLt1</th>
<th>DehLt2</th>
<th>DehLt3</th>
<th>DehLt4</th>
</tr>
</thead>
<tbody>
<tr>
<td>D10</td>
<td>D11</td>
<td>D8</td>
<td>D12</td>
<td>D12</td>
<td>D12</td>
<td>D9</td>
</tr>
<tr>
<td>T14</td>
<td>T15</td>
<td>T12</td>
<td>T16</td>
<td>T16</td>
<td>T16</td>
<td>T13</td>
</tr>
<tr>
<td>R41</td>
<td>R42</td>
<td>R39</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>R40</td>
</tr>
<tr>
<td>S118</td>
<td>S119</td>
<td>S114</td>
<td>S118</td>
<td>S121</td>
<td>121</td>
<td>S116</td>
</tr>
<tr>
<td>K151</td>
<td>K152</td>
<td>K147</td>
<td>K149</td>
<td>K152</td>
<td>K152</td>
<td>K149</td>
</tr>
<tr>
<td>Y157</td>
<td>Y158</td>
<td>Y135</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Y155</td>
</tr>
<tr>
<td>S175</td>
<td>S176</td>
<td>S171</td>
<td>S173</td>
<td>-</td>
<td>-</td>
<td>S173</td>
</tr>
<tr>
<td>N177</td>
<td>N178</td>
<td>N173</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>N175</td>
</tr>
<tr>
<td>D180</td>
<td>D181</td>
<td>D176</td>
<td>D79</td>
<td>D186</td>
<td>D185</td>
<td>D178</td>
</tr>
</tbody>
</table>
The sequence alignment of four haloacid dehalogenases from strain TONO were also compared with group I dehalogenases. According to Hill et al. [10] five essential catalytic residues (Thr65, Glu69, Asn117, Tyr120, and Asp194) reported by Nardi dei et al. [47] in DL-DEX were supposed to be conserved in almost all group I dehalogenases. However, the multiple sequence alignment neither show any significant identity or similarity. Therefore, all four haloacid dehalogenase sequences found in M. loti strain TONO were proposed belongs to type II.

**Analysis of an uptake protein of DehrPt from M. loti strain TONO**

According to the FGENESH, dehrPt gene (MLTONO_0284) is not organize in any operon. Pairwise analysis of DehrPt and DehrP from Rhizobium sp. RC1 [48] showed 69% amino acid sequence identity. Asp36(Asp35), Arg130(Arg130) and Gln133(Gln133) which were H+-binding sites were conserved to each other. The interacting amino acid residues of DehrP (Rhizobium sp. RC1) and DehrPt (strain TONO) depicted as Figure 4.

![Figure 4. Pairwise sequence alignment of DehrPt (strain TONO) and DehrP (Rhizobium sp. RC1; accession number CAJ98618.1). Amino acids with high consensus values were highlighted in red; In box: conserved amino acid residues.](image-url)
Glu33(Glu32), Trp34 (Trp33), Phe37(Phe36), Gln165(Gln165) (Figure 4). The sequence analysis suggested that DehrPt is equivalent to DehRP function as haloacid uptake protein in M. loti strain TONO.

Conclusion
L-2-Haloacid dehalogenases (type II) are common and this is consistent with the current study. Most dehalogenase genes were not organized as a single operon and scattered within its chromosomes and their expression is regulated differently referring to helix-turn-helix (HTH) motif family DNA-binding protein regulator and LysR family transcriptional regulator. The dehalogenase genes sequence is subjected to further analysis using bioinformatic tools for their protein structure and functions.

Acknowledgement
The authors would like to express their gratitude to Geran Universiti Penyelidikan-UTM Number Q.J130000.2414.08G59 & FRGS R.J130000.7854.5F189 for financial assistance. SZ thanks the Afghanistan Higher Education Ministry for MSc. studentship.

References


This page is intentionally left blank.