Multinuclear non-heme iron dependent oxidative enzymes: Landscape of their substrates, partner proteins and biosynthetic gene clusters

R. Antoine,[a#] L. Leprevost,[a] S. Jünger,[b] S. Zirah,[b] G. Lippens,[c]
Y. Li,[b] S. Dubiley,[c] F. Jacob-Dubuisson[a#]

[a] Univ. Lille, U1018-UMR9017-CIIL-Center for infection and Immunity of Lille, Lille, FRANCE

[b] Unit Molecules of Communication and Adaptation of Microorganisms (MCAM), UMR 7245 CNRS, MNHN, Alliance Sorbonne University, Paris, FRANCE

[c] Toulouse Biotechnology Institute, CNRS/INRAE/INSA/UPS, 31077 Toulouse, FRANCE

Running head: Substrates and gene clusters of MNIO enzymes

[#] for correspondence: rudy.antoine@pasteur-lille.fr; francoise.jacob@ibl.cnrs.fr

Keywords: ribosomally synthesized post translationally modified peptides (RiPPs); multinuclear non heme iron-dependent oxidative (MNIO) enzymes; biosynthetic gene clusters; genome mining; sequence similarity network (SSN) analysis

Abstract. Enzymes of the multinuclear non-heme iron-dependent oxidase (MNIO) superfamily catalyze various modification reactions on the precursors of ribosomally synthesized, post-translationally modified peptides (RiPPs). We recently identified two large families of MNIO-modified RiPPs called bufferins, which enhance bacterial growth under copper stress by chelating the excess metal ions. Here, we explored the diversity of potential MNIO substrates by performing extensive *in silico* studies. Analyses of MNIO-coding biosynthetic gene clusters (BGCs) identified all major groups of putative precursors characterized by specific Cys-containing motifs throughout the eubacterial phylogenetic tree. The precursors of most MNIO-modified RiPPs harbor N-terminal Sec-dependent signal peptides, a rare feature among bacterial RiPPs. Some precursors are very long relative to those of typical RiPPs, indicating that MNIO enzymes could modify both peptide and protein substrates. We also identified a distinct family of integral membrane proteins with large predicted extra-cytoplasmic domains mostly found in Terrabacteria, frequently but not systematically associated with MNIOs. Most MNIO BGCs harbor genes coding for DUF2063 domain-containing proteins or structurally related proteins, serving as partners of the enzymes for precursor modification. We uncovered specific associations between subgroups of MNIO enzymes, families of precursors, and types of partner proteins, indicating some specificity in these systems. This study depicts the global landscape of potential MNIO-dependent natural products by unveiling the major groups of peptides and proteins genetically associated with MNIOs. It reveals a treasure trove of potential new RiPP precursors which likely represent a widespread bacterial strategy to deal with copper stress, and most likely other stresses, in natural environments.

Impact statement. The multinuclear non-heme iron-dependent oxidative (MNIO) enzymes belong to an emerging superfamily of modification enzymes that catalyze various oxidation reactions on ribosomally synthesized, post-translationally modified peptides (RiPP) precursors.

53 The largest families of MNIO-modified RiPPs, called bufferins, are involved in copper
54 homeostasis. In this work we performed extensive *in silico* analyses to explore the diversity of
55 RiPP precursors genetically associated with MNIO enzymes and identified all major families.
56 Some precursors are much larger than typical RiPP precursors, indicating that MNIO enzymes
57 may also modify proteins. We defined subtypes of MNIO enzymes and partner proteins
58 dedicated to specific families of precursors. Our analyses of the biosynthetic gene clusters
59 unveiled functions beyond copper homeostasis, likely for the response to other metal stresses.
60 The global landscape of MNIO-modified RiPP precursors will be a basis for investigations into
61 new RiPP families and will undoubtedly lead to the discovery of new modifications and new
62 functions.
63

64 **Introduction**

65 Peptide-based natural products can be synthesized by ribosomes and then enzymatically
66 modified, like ribosomally synthesized post-translationally modified peptides (RiPPs) or be
67 fully synthesized by non-ribosomal enzymatic complexes (1, 2). RiPPs are highly prevalent
68 natural products in bacteria, with several dozen families described over the last decades (2-5).
69 Some eukaryotic organisms, particularly plants, cone snails and fungi, also produce RiPPs (6).
70 Bacterial RiPPs are encoded in biosynthetic gene clusters (BGCs) that comprise at least one
71 precursor gene and one or more gene(s) coding for modification enzymes. RiPPs are peptides
72 of typically less than 10 kDa. Their precursors harbor a core region that undergoes enzymatic
73 installation of post-translational modifications (PTMs), generally flanked at its N-terminus by
74 a leader peptide (7). This leader peptide can be recognized by a system-specific 'RiPP
75 recognition element' (RRE) domain that serves as a docking platform for the modification
76 enzyme(s) (8, 9). RiPP BGCs frequently comprise genes coding for maturation proteases and

77 export systems (10, 11), as well as for cognate immunity proteins in the case of RiPP
78 bacteriocins (12).

79 RiPPs modified by enzymes of the multinuclear non-heme iron oxidative (MNIO) enzymes
80 superfamily represent a vast, emerging RiPP class, as described in an excellent review
81 published recently (13) (Table 1). MNIO enzymes catalyze various 2-or 4-electron oxidation
82 reactions targeting mainly Cys residues and frequently work with partner proteins, notably with
83 Pfam or TIGRfam signatures 'DUF2063' (Domain of Unknown Function), 'DUF1772' or
84 'MbnC' (14-16). The first MNIO-modified RiPPs identified, methanobactins, are chalkophores
85 that scavenge copper in the milieu for the assembly of cuproenzymes called methane
86 monooxygenases in Methanobacteria (16-18). Methanobactins were identified in non-
87 methanogenic bacteria as well (19). In complex with its partner protein MbnC, the MNIO
88 enzyme MbnB oxidizes specific internal Cys residues of the MbnA methanobactin precursor to
89 form copper-chelating oxazolone thio-amide groups (18). Other MNIO enzymes modify C-
90 terminal Cys residues first added to precursor scaffold peptides, thus generating thia-amino acid
91 derivatives called 'pearlins' (20, 21). Yet another type of MNIO-modified RiPP, chryseobasin
92 harbors an imidazolidinedione heterocycle, a macrocycle, thioaminals and a thiomethyl group
93 (15). Non-Cys modifying MNIO enzymes were described to exert their catalytic activity on a
94 C-terminal Asp or Asn residue, generating a C-terminally amidated peptide or an aminopyruvic
95 acid, respectively (19, 22).

96 We have described two large families of MNIO-modified RiPPs involved in copper
97 homeostasis and called them bufferins (14). Bufferins belong either to the Buf1 (DUF2282)
98 family or to another family, Buf2, that we have named BUF_6/12Cys. Copper is a necessary
99 transition metal that is toxic in excess (23), and bufferins use Cys-derived thiooxazole
100 heterocycles to sequester this metal in the periplasm (14). The recently reported oxazolin
101 represents another family of MNIO-modified RiPPs that bind copper, but it appears to do so

102 using oxazolone thioamide groups, similar to methanobactins (24). Remarkably for bacterial
103 RiPP precursors, the precursors of Buf1 and Buf2 bufferins, oxazolin and chryseobasin feature
104 Sec-dependent N-terminal signal peptides.

105 Thousands of MNIOs have been identified *in silico* (14, 15, 19, 25, 26). In this work we
106 explored the global diversity of the MNIO BGCs by genome mining approaches and identified
107 new potential MNIO substrates and partner proteins.

108

109 **Results**

110 *Search for MNIO-associated RiPPs*. In order to identify all types of putative substrates of
111 MNIO enzymes, we collected the members of the MNIO family (>14,000 proteins at < 90%
112 sequence identity) using the Pfam DUF692 signature. These enzymes are distributed across a
113 number of mostly bacterial phyla (Fig. 1A). Sequence similarity network (SSN) analyses
114 showed that Buf1 and Buf2-associated MNIOs represent a sizeable fraction of the largest
115 sequence cluster of the representative node network of MNIO enzymes (red and yellow dots in
116 Fig. 1B). MNIOs associated with oxazolins (hereafter called BUF_EGKC, see below) are also
117 part of this large cluster. Some of the small MNIO sequence clusters contain enzymes
118 associated with methanobactins, pearlins, methanobactin-like peptides from *Vibrio*,
119 chryseobasin and aminopyruvatides (15, 18-20, 22). As evidenced from this analysis, other
120 types of potential MNIO substrates remain to be revealed.

121 We utilized a home-made RODEO-type approach (27) to collect the BGCs of all these
122 MNIOs, and we analyzed their compositions (Table S1). Conserved proteins encoded in the
123 vicinity of MNIO-coding genes notably have the Pfam or TIGRfam signatures DUF2282,
124 DUF2063, DoxX, Sigma70_r4_2 (sigma factor), Sigma70_r2 (sigma factor), and NrsF (anti-
125 sigma factor), all found in bufferin BGCs (14). Because no signatures were available yet for

126 Buf2 or for oxazolins, we generated new hmm profiles for them, called Buf_6/12Cys and
127 Buf_EGKC, respectively (Suppl. Files S1 and S2).

128 Next, we performed two-step analyses of all collected BGCs to characterize the major types
129 of putative RiPP precursors (Fig. 2A). We first identified all potential precursors, sorted them
130 through SSN analyses and constructed hmm profiles for new families. Then we used previously
131 available and newly generated signatures to collect all members of each family from the NCBI
132 nr Protein database, retrieved the corresponding BGCs and analyzed their composition. Our
133 criteria for gene inclusion in a BGC were the following. All genes coding for domains with
134 signatures already known to be relevant to MNIO BCGs (DUF2063, DUF2282, DUF692,
135 Buf_6/12Cys, DoxX, sigma factor and anti-sigma factor, MbnC) were included irrespective of
136 intergenic distances between them. For unrelated genes, we set up a stringent intergenic
137 distance criterion (maximum 10 bp from any relevant gene) for inclusion to ensure that these
138 new genes are in operons with at least one known gene. Although somewhat arbitrary, this
139 choice proved to provide the most information while avoiding too much background noise.

140 In our initial collection of MNIO BGCs, we first collected putative precursors shorter than
141 200 residues, harboring Cys residues and a predicted Sec signal-peptide, and coded by genes
142 that precede the MNIO gene. This latter criterion was warranted by our preliminary analyses
143 showing the strict conservation of the gene order *bufABC* (encoding a bufferin precursor, an
144 MNIO protein and a DUF2063 protein, respectively) in all *buf1* BGCs. The criterion of Cys-
145 containing precursors was motivated by the fact that the overwhelming majority of MNIOs
146 modify Cys residues (13). Classification of the >11,000 identified proteins by SSN analyses
147 (Fig. 1C) revealed large sequence clusters of Buf1, Buf2 and oxazolin homologs. A sizeable
148 proportion of the latter proteins also respond to the 'EF-Hand_5' domain signature (see below).
149 The proteins of the other sequence clusters lacked specific domain signatures.

150 To enlarge our range of putative MNIO substrates, we also retrieved from our initial BGC
151 collection all proteins with Cys residues but devoid of predicted Sec signal-peptides and without
152 size restriction, whose genes precede MNIO genes. This yielded >1000 proteins, 40% of which
153 respond to the TIGRFam 'TIGR04222' signature. An SSN analysis revealed a very scattered
154 node network, with many sequence clusters harboring TIGR04222 proteins, indicating the
155 considerable sequence diversity of these proteins (purple dots in Fig. 1D). The other medium-
156 size clusters of this node network (labeled I to IV in Fig. 1D) are described in Table S2.
157
158 *General features of bona fide bufferins and their BGCs.* Next, we characterized the various
159 groups of proteins in more detail, starting with *bona fide* bufferins. In the NCBI nr Protein
160 database, we found several thousand Buf1 and Buf2 bufferins (at < 90% redundancy), mostly
161 in *Pseudomonadota,* but also in other phyla for the latter (Fig. 3A,B). Of note, Buf2 BGCs were
162 reported in uncultured Antarctic soil bacteria (28), suggesting a large environmental
163 distribution as well.
164 Genes for sigma factors and anti-sigma proteins were found in Buf1 but not in Buf2 BGCs
165 (Fig. 3C,D). In both subfamilies, sizeable proportions of precursors are coded by genes not
166 genetically associated with MNIO genes. We also identified BGCs with tandem precursor
167 genes, suggesting that some MNIO enzymes might have more than one substrate. Additional
168 conserved genes were found in extremely low proportions, confirming that the bufferin
169 precursor-, MNIO-, DUF2063- and DoxX-coding genes constitute the core of bufferin BGCs.
170 Analyses of the predicted proteins revealed that Buf1 bufferins range from 30 to 835
171 residues, including the predicted signal peptide, and that they systematically harbor the four
172 conserved Cys residues present in the founding member (14) (Fig. 3E). It is likely that the
173 smallest of these proteins are truncated. Most Buf2 bufferins are between 51 and 367 residues
174 long, with 6 conserved Cys residues in 90% of cases and 12 Cys in the remainders. An

175 AlphaFold2 model of one of the latter indicates that it consists of duplicated 6-Cys-containing
176 domains (Fig. 2F). Approximately 16% Buf2 bufferins are predicted to harbor lipoprotein
177 signal peptides and might therefore be membrane-anchored after export. We identified large
178 proteins containing bufferin domains that result from gene fusions, including chimeras with
179 domain signatures 'DUF2282|Cytochrom_C|FGE-sulfatase', 'MerC|DUF2282' and
180 'MerC|BUF_6/12Cys' (Fig. 3G,H). The Formylglycine-generating enzyme (FGE) signature
181 denotes a cuproenzyme (29), consistent with copper-related functions. MerC-containing
182 fusions are found in BGCs related to mercury detoxification (30), indicating that some bufferins
183 may participate in stress responses to other metals.
184
185 *Other bufferin-like precursors.* Proteins of the third largest cluster of the putative RiPP
186 precursors node network (Fig. 1B) contains oxazolins (24). The newly generated hmm profile
187 was called 'BUF_EGKC' (File S2), as these proteins are characterized by repeated Glu-Gly-
188 Lys-Cys motifs (Fig. 4A). Among them, hundreds of proteins contain two Ca^{2+}-binding 'EF-
189 Hand_5' domains (31, 32) fused at the C terminus of a BUF_EGKC domain (Fig. 4A). The
190 majority of the > 2,000 proteins collected with this new signature in the NCBI database (at <
191 90% sequence identity) were found in *Pseudomonadota* (Fig. 4B). Analyses of their BGCs
192 revealed that they are mostly composed of the core RiPP-, MNIO-, DUF2063 and DoxX-coding
193 genes in various combinations (Fig. 4C). Small numbers of BGCs harbor additional genes that
194 code for domains with Pfam 'oxydored_molyb' and 'Ferric_reduct' signatures, suggesting
195 metal-related or redox functions.
196 The BUF_EGKC precursor proteins are 31 to 506 residues long. An AlphaFold2-generated
197 structural model of a mid-size protein showed a β helix formed of 17-residue repeats, each
198 containing two EGKCG motifs (Fig. 4D). All Cys residues that might be putative targets of the
199 MNIO enzyme are predicted to be oriented towards the center of the helix, reminiscent of

200 copper storage proteins (33). The longest protein of the family is from *Pseudoalteromonas*
201 *piscicida*. It harbors 58 Cys-containing motifs (strain JCM_20779, gene # ATD05916.1) and is
202 predicted to form an elongated β helix. In contrast, AlphaFold2 predicted that the BUF_EGKC
203 domains of many proteins of this family are unstructured. In particular, this is the case for the
204 N-terminal domain of BUF_EGKC|EF-hand_5|EF-hand_5 chimeric proteins.

205 We also built an hmm profile called 'BUF_CxxxxC' for the fourth largest cluster of
206 precursor proteins (black dots in Fig. 1C), which contain conserved Cys residues including a
207 CxxxxC motif (Fig. 4A; File S3). The newly built signature collected 76 proteins in the nr NCBI
208 database at < 90% sequence identity, all from *Pseudomonadota*. These putative RiPP precursors
209 harbor Sec-dependent signal peptides. In addition to the core RiPP-, MNIO, DUF2063- and
210 DoxX-coding genes, 54% of these BGCs contain genes for a protein with a Pfam
211 'Lactamase_B_2' signature, which is a metallo-hydrolase domain.

212

213 *TIGR04222 proteins.* Using the TIGR04222 signature, > 3500 proteins were found in the NCBI
214 database at < 90% sequence identity, mainly from the Terrabacteria group (Fig. 5A). Analyses
215 of the genetic environments of the TIGR04222-coding genes revealed a sizeable proportion of
216 BGCs devoid of MNIO genes (Fig. 5B). Furthermore, in more than 10% cases TIGR04222
217 genes follow MNIO genes, in contrast with bufferin BGCs. No DUF2063-coding genes are
218 present in the TIGR04222 BGCs, some of which harbor conserved genes of unknown function.

219 The sizes of TIGR04222 proteins range from 70 to 655 residues. Alphafold2 models of
220 selected proteins showed that they are predicted to be composed of several transmembrane
221 segments and large periplasmic domain(s), with unstructured, C-terminal Ser- and Gly-rich
222 periplasmic regions interspersed with Cys residues (Fig. 5C). Nevertheless, more than a quarter
223 of TIGR04222 proteins are devoid of C-proximal Cys residues, and there appears to be no
224 correlation between the presence of Cys and the genetic association with MNIO enzymes.

225

226 *Other types of MNIO BGCs.* We also analyzed the BGCs that correspond to small sequence
227 clusters of the MNIO node network shown in Fig. 1B (Fig. S1). Five of these sequence clusters
228 harbor known MNIO enzymes, namely ChrH, MbnB, MovX, ApyH and TglH, previously
229 described to catalyze the installation of various PTMs on previously described RiPP precursors
230 (15, 16, 18-20, 22). In contrast, we could identify potential RiPP precursors in the BGCs of
231 only a handful of the other small MNIO sequence clusters, which lacked predicted Sec signal
232 peptides. None of the BGCs corresponding to these 18 MNIO sequence clusters encode
233 DUF2063 proteins. Some of them comprise genes that code for proteins with radical_SAM,
234 transferases or dehydratases signature domains, known to catalyze RiPP PTMs. Thus, many
235 minor types of putative MNIO-modified RiPPs remain to be characterized.

236

237 *Identification of MNIO partner proteins.* Partners of MNIO enzymes include proteins with
238 Pfam or TIGRfam DUF2063, MbnC or DUF1772 signatures, the former representing the vast
239 majority of them (62%). To identify new types of partner proteins of the MNIO enzymes, we
240 analyzed the proteins coded by genes that follow the MNIO genes in our BGC collection, based
241 on observations that the gene order *ABC* is by far the most frequent in MNIO BGCs (Fig. 2B).
242 Among the proteins thus collected, >8000 respond to the DUF2063 signature, a few dozen have
243 DUF1772, MbnC, DoxX or MNIO signatures, and >1400 others are uncharacterized. SSN
244 analyses of these latter sequences revealed that they could be sorted into several sequence
245 clusters (Fig. 6A). However, there was not enough sequence similarity within the largest protein
246 clusters to generate hmm profiles.

247 Therefore, to detect low-level similarity between these uncharacterized proteins and known
248 domains, we performed HHpred analyses. For most of them the best hits are the *Neisseria*
249 *gonorrheae* protein NGO1945 (pdb# 3DEE) and TglI (pdb#8HI7), the partner of the MNIO

250 enzyme TglH (20). Although previously predicted to be a transcription factor (34), structural
251 alignments indicated that NGO1945 contains both a DUF2063 domain and an RRE-type
252 domain and is therefore a typical MNIO partner protein (Fig. 6B). Alphafold2 models of a few
253 of these uncharacterized proteins revealed that most of them harbor an N-terminal domain
254 structurally related to a DUF2063 domain, followed in some cases by a domain with an RRE-
255 fold type, or a truncated version (Fig. 6C). Therefore, many BGCs without *bona fide* DUF2063
256 genes encode "DUF2063-like" proteins, missed by the DUF2063 signature because of their
257 lack of sequence similarity. In the light of these findings, we reanalyzed the products of
258 unknown genes in the TIGR04222 BGCs (Fig. 5B). HHpred analyses predicted that at least
259 10% of TIGR04222 BGCs encode DUF2063-like proteins. Similarly, BGCs encoding MNIO
260 enzymes from several small sequence clusters also encode DUF2063-like proteins, including
261 those of pearlins and aminopyruvatides (Fig. S1).

262 To probe the strength of the association between DUF2063 or DUF2063-like domains and
263 RRE domains, we determined the length distributions of the two types of putative MNIO
264 partner proteins. *Bona fide* DUF2063 domain-containing partner proteins have a narrow size
265 distribution corresponding to full-length DUF2063-RRE chimeras (Fig. 6D). In contrast,
266 DUF2063-like domain-containing proteins are very heterogeneous in length. The group of
267 proteins with similar lengths as the DUF2063-RRE chimeras, around 260 residues, most likely
268 harbor full-length DUF2063-like and RRE-type domains. However, a large proportion of
269 putative partner proteins are too short to comprise both domains and probably consist of a
270 DUF2063-like domain alone for the shortest of them, or a DUF2063-like domain followed by
271 a short putative RRE-type truncate or variant for those of intermediate lengths (Fig. 6 C).

272

273 *Specificity of MNIO enzymes for RiPP precursors and partner proteins.* The discovery of
274 DUF2063-like partner proteins prompted us to investigate specific associations between the

275 subgroups of MNIO enzymes, the families of RiPP precursors, and the types of partner proteins.
276 To facilitate visualization of these relationships, we colored the nodes of the mega cluster of
277 MNIO sequences (Fig. 1B) according to the types of precursors and partner proteins encoded
278 in the same BGCs. As already apparent in Fig. 1B, specific areas of the mega cluster correspond
279 to specific precursor families (Fig. 7A). In addition, coloring of MNIO enzymes as a function
280 of their partner proteins showed that those associated with signal peptide-containing precursors
281 (i.e., Buf1, Buf2, BUF_EGKC and BUF_CxxxxC families) have authentic DUF2063 domain-
282 containing partner proteins (Fig. 7B). In contrast, those associated with signal peptide-less
283 TIGR04222 proteins appear rather to have DUF2063-like domain-containing partners. This
284 correlation was confirmed for three small MNIO sequence clusters labeled I-II-III in Fig. 7.
285 Thus, the RiPP precursors associated with the MNIO enzymes of clusters I and II have no
286 predicted Sec signal peptides, and the partners of their MNIO enzymes are DUF2063-like
287 proteins (Fig. S2). In contrast, the RiPP precursors associated with MNIO enzymes of cluster
288 III harbor predicted Sec signal peptides, and their MNIO enzymes have *bona fide* DUF2063
289 partner proteins. Altogether, the correlations between the subtypes of MNIOs and their types
290 of substrates and partners indicates a degree of specificity in these systems.
291
292 *Co-occurrence analyses.* Finally, we mined > 40,000 fully assembled bacterial genomes in the
293 NCBI database to determine the maximum occurrences of bufferin or bufferin-like BGCs in a
294 single genome. The most extreme case is that of *Legionella pneumophila subsp. fraseri*, with
295 five Buf1 (DUF2282), two Buf2 (BUF_6/12Cys)-, one BUF_EGKC-, seven MNIO-, six
296 DUF2063- and three DoxX-coding genes (Fig. 8). The presence of bufferin precursor genes not
297 associated with MNIO genes suggests that MNIO enzymes encoded at a different locus might
298 modify these precursors or that they are not modified. However, we do not know whether all
299 these BGCs are functional.

300 Our genomic analysis revealed that very few DUF2282, BUF_6/12Cys or BUF_EGKC
301 precursor genes are present without MNIO-coding genes in the same genome (Table 2). In
302 contrast, a large proportion of genomes harboring TIGR04222 genes, approximately 36%, are
303 devoid of MNIO genes, revealing a loose association between these two families.
304

305 **Discussion**

306 Here, we explored the substrate diversity of MNIO enzymes and described all major groups
307 of potential precursors in various eubacterial phyla. We revealed that the BGCs of most of the
308 putative RiPPs are composed of a small number of core genes. This limited gene set and the
309 overwhelming conservation of Cys-containing motifs in the putative precursors indicate that
310 these are the most likely targets of MNIO modifications, and that secondary PTMs are rare in
311 these families. Nevertheless, small groups of MNIO enzymes are coded in larger BGCs that
312 also comprise genes for tailoring enzymes such as radical_SAM proteins, transferases or
313 dehydratases. In fact, some MNIO-modified RiPPs were reported to carry additional PTMs
314 catalyzed by other enzymes (17, 22). Therefore, it is very likely that new PTMs remain to be
315 discovered in MNIO-modified RiPPs.

316 We have revealed that Buf1 of *C. vibrioides* is involved in copper homeostasis, and that it
317 harbors Cys-derived thiooxazole groups mediating the chelation of Cu(I) and Cu(II), thereby
318 enhancing bacterial growth in conditions of copper excess (14). Buf1 BGCs likely promote
319 survival of environmental bacteria subjected to copper stress in soils or water polluted by
320 metals, notably from anthropic sources (35), or when they are engulfed by amoeba (36).
321 Similarly, the use of copper as an antimicrobial agent in innate immunity (37) might account
322 for the presence of bufferins in bacterial pathogens of mammals. Although we have not
323 identified the chemical modifications carried by Buf2 bufferins, preliminary characterization
324 of that of *C. vibrioides* has shown that it has a similar function as Buf1, its PTMs result in the

325 same mass loss per Cys residue, and its UV/vis spectrum shows a similar 305-nm absorbance
326 maximum (14). Furthermore, the structural models of Buf1-type and Buf2-type bufferins are
327 related, and their associated MNIO enzymes belong to the same large sequence cluster,
328 suggesting that they are iso-functional. These elements argue in favor of similar PTMs and
329 functions for both bufferin families.

330 The PTMs and the function of another MNIO-modified RiPP, oxazolin, were recently
331 reported (24). Similar to Buf1, it binds Cu(I), but it appears to do so using Cys-derived
332 oxazolone thioamide groups, similar to methanobactins. We generated an hmm profile for this
333 family, which we called BUF_EGKC, and we revealed that these proteins constitute the third
334 largest cluster of sequences in the SSN node network of RiPP precursors. Interestingly, the
335 MNIO enzymes associated with these precursors are closely related to those associated with
336 bufferin precursors, and their partners are *bona fide* DUF2063 proteins, like for bufferins. This
337 is also the case for BUF_CxxxxC proteins. Whether the MNIO enzymes that belong to the large
338 sequence cluster are all iso-functional remains to be seen. Nevertheless, the discrepancy
339 between the PTMs of Buf1 and oxazolin is puzzling and warrants further investigations.
340 Because the name bufferin does not refer to a specific PTM but to a broad function, we propose
341 to include oxazolins and BUF_CxxxxC in the bufferin superfamily.

342 In contrast, TIGR04222 membrane proteins are unrelated precursors. Those that harbor C-
343 proximal Cys-rich segments predicted to be exposed to the periplasm may be modified by
344 MNIO enzymes. However, one cannot speculate on the nature of their PTMs, as the MNIO
345 enzymes and partner proteins genetically associated with TIGR04222 proteins are distant from
346 those of bufferins. Similarly, the functions of these proteins in the producing organisms remain
347 to be investigated. Some TIGR04222 proteins are not even associated with MNIOs and/or have
348 no Cys residues. Given their diversity, TIGR04222 proteins are likely to have a variety of
349 functions, and only a subset of them might be post-translationally modified.

350 The diversity of the precursors and of their BGCs suggests that some putative RiPPs
351 identified in this study likely play roles beyond adaptation to excess copper. The identification
352 of chimeras of bufferins with metal-binding domains supports their role in the homeostasis of
353 transition metals. Protection from mercury ions is suggested by MerC-bufferin fusions. As for
354 BUF_EGKC|EF-hand_5 fusion proteins, in addition to a role in copper homeostasis they might
355 contribute to Ca^{2+} homeostasis or signaling, as exemplified by EfhP in *Pseudomonas*
356 *aeruginosa* (31, 32). The presence of *doxX* genes in many bufferin BGCs suggest protection
357 from other stresses commonly encountered by bacteria. A DoxX homolog in a different genetic
358 context was shown to protect bacteria against oxidative or sulfur stress (38).

359 The partner proteins of MNIOs are an important feature of these systems. In the TglHI
360 complex, the catalytic pocket is formed by the interaction between the MNIO enzyme TglH
361 and its partner protein TglI (39), and in the MbnABC complex extended contacts between the
362 leader peptide of MbnA and MbnC position the MbnA core peptide in the catalytic site of MbnB
363 (16). Here we revealed that bufferins with signal peptides have DUF2063 domain-containing
364 partner proteins, whereas precursors devoid of signal peptides are frequently associated with
365 DUF2063-like proteins. Interestingly, authentic DUF2063-containing proteins appear to harbor
366 complete RRE domains, which is not the case of all DUF2063-like proteins. Given that
367 DUF2063-RRE proteins are almost systematically associated with signal peptide-containing
368 precursors, it is tempting to speculate that the RRE domain might interact with the signal
369 peptide, which could notably prevent premature export before installation of the PTMs. In
370 contrast, DUF2063-like proteins are associated with signal peptide-less precursors, hence the
371 possibility that an RRE-type domain has become dispensable in some cases. Furthermore, many
372 TIGR04222- and MNIO-coding BGCs appear not to harbor partner protein genes at all. How
373 or whether the corresponding precursors are modified remains to be established. Of note, the

374 second MNIO enzyme of methanobactin-like BGCs in *Vibrio*, MovX, appears to work without
375 a partner protein (19).
376 Bufferin precursor genes without MNIO genes are rarely present in fully assembled
377 genomes. Presumably, these rare peptides have activities that do not necessitate MNIO-
378 catalyzed PTMs, or they are no longer expressed or functional. We also found genomes with
379 several bufferin-like precursor genes but a single MNIO and DUF2063 gene pair. It is possible
380 that the lifestyle of these bacteria has fostered RiPP expansion, and that a single MNIO enzyme
381 can modify several related precursors. Alternatively, only some of them are modified, and the
382 unmodified ones serve a different purpose. Further studies are needed to address these issues.
383 Our study suggests that MNIO enzymes modify not only peptides, but also proteins that
384 respond to the same signatures as smaller precursors but contain motif repetitions or additional
385 domains. In particular, many Cys residues are found in very large precursors of the BUF_EGKC
386 family. If these residues are all modified before export, there must be a mechanism to thread
387 the precursor through the MNIO active site for its sequential processing, similar to the iterative
388 mechanisms reported for a microcin and a lanthipeptide, by which the leader peptides maintain
389 the precursors close to the active sites throughout their serial enzymatic modifications (40, 41).
390 The identification of BGCs encoding long precursors followed by two MNIO-coding genes, as
391 for the 58-Cys containing BUF_EGKC protein from *P. piscicida* also indicates that more than
392 one enzyme might be needed to extensively modify such long proteins.
393

394 **Materials and Methods**

395 *Database searches.* The non-redundant NCBI protein database was searched with various
396 hidden Markov models (Pfam (42), TIGRFAMs (43) and home-built HMM models), using the
397 hmmsearch program of the HMMER3 suite (44). Each set of results were clustered by CD-HIT
398 (45) with default parameters to remove redundant sequences sharing at least 90% sequence

399 identity, to avoid overrepresentation of well-studied bacterial genera. The size, the domain
400 composition, and the taxonomy were determined for each protein.
401
402 *Sequence similarity network analysis (SSN).* Sequence similarity networks were generated
403 using the Enzyme Function Initiative Enzyme Similarity Tool (EFI-EST) (46). The selected E-
404 value and the sequence alignment score threshold are given in each figure. Networks were
405 visualized by Cytoscape (47) by using the organic layout. The ID numbers of all entries found
406 in the SSN node networks are provided in Table S3.
407
408 *Local database construction and characterization of genetic environments.* A local DNA
409 database was constituted by downloading all available bacterial genomic sequences for both
410 Eubacteria and Archeae from the NCBI RefSeq, GenBank and WGS databases. A Python script
411 was written to extract the genetic environments of all MNIO-coding genes by retrieving the
412 sequences of the five genes on either side of the gene of interest in a Rodeo-type approach. The
413 resulting table was used to collect all genes coding for proteins of interest. To define the RiPP
414 BGCs, we imposed the following criteria. All genes encoding proteins with signatures of
415 interest (DUF2282, DUF692/MNIO, DUF2063, DoxX, BUF_EGKC, BUF_6/12Cys, NrsF,
416 Sigma70, TIGR04222, Buf_CxxxxC, MbnA) in the vicinity of the MNIO genes were
417 considered part of the BGC, irrespective of the intergenic distances. For other genes, based on
418 the analyses of model bufferin operons, we imposed a criterion of 10-bp intergenic distances or
419 less to ensure that only relevant genes organized in operons were considered. The numbers of
420 occurrences of each type of operonic structures were computed. Prediction of signal-peptides
421 and their cleavage sites was performed using SignalP V6 (48).
422

423 *Taxonomic analysis.* Krona pie charts were generated using Krona Tools (49) to visualize the
424 proportions of different taxa carrying genes coding for the different protein families.
425
426 *Generation of hmm profiles.* We generated hmm profiles for the BUF_6/12Cys, BUF_EGKC
427 and BUF_CxxxxC families according to the procedure described in (50). Briefly, a few dozen
428 proteins were picked from each cluster of interest of the Representative Node Network 70%
429 displayed with Cytoscape. We chose nodes well distributed across each cluster. Only nodes
430 with a 'number of IDs in Rep Node' equal to 1 were kept. Outliers with respect to length were
431 also removed. For each family we first determined the most frequent protein size in the cluster,
432 and we selected sequences from this major subset to generate the seed alignment. The sequences
433 were aligned with MAFFT (L-INS-i), an iterative refinement method incorporating local
434 pairwise alignment information. The non-conserved regions (mostly the N- and C-termini) were
435 edited manually in Jalview, and a hmm profile was generated with this seed alignment using
436 hmmbuild (44). We then performed new SSN analyses with all proteins collected with a given
437 hmm profile to ensure that only *bona fide* members were retained. Outliers found in small,
438 isolated SSN clusters were discarded. The hmm profile files are provided in Supplementary
439 materials (Files S1 to S3).
440
441 *Characterization of unknown proteins.* To gain insight into proteins of interest devoid of Pfam
442 or TIGRfam domain signatures, we picked several proteins in each cluster of the representative
443 node network and first looked for Interpro signatures. If no Interpro signature was identified,
444 we searched for related proteins using HHpred (51) and selected the best hits. If necessary, we
445 generated AlphaFold2 models of selected members and used DALI (52) to find related
446 structures in the RCSB Protein Data Bank (53).
447

448 *Analysis of assembled bacterial genomes.* The proteomes of the fully assembled bacterial
449 genomes were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/assembly). The
450 bacterial proteomes were searched with the hmmscan program of the HMMER3 suite (44) using
451 the hmm profile signatures DUF2282, DUF692, DUF2063, DoxX, EF-hand_5,
452 BUF_6/12_Cys, BUF_EGKC, and TIGR04222 using home-made scripts. The domains of all
453 the corresponding proteins were determined, and the occurrences of each type of domain of
454 interest were computed (Table S4).
455
456 **Data summary**
457 A local DNA database was constituted by downloading all available bacterial genomic
458 sequences for both Eubacteria and Archeae from the NCBI RefSeq, GenBank and WGS
459 databases. Protein sequences were collected from the non-redundant NCBI protein database.
460 We also mined the NCBI assembled genomes database. Accession numbers of all the protein
461 sequences used in this work and all genome identification numbers are provided in
462 Supplementary materials in the form of Excel tables.
463
464 **Funding statement**
465 This project was funded by the ANR grant CuRiPP (ANR-22-CE44-0001-02) to FJD. L.
466 Leprevost and S. Jünger were supported by PhD fellowships of Lille University and of the
467 French Ministry of Education (ED 227 MNHN-SU), respectively. S. Dubiley was supported
468 as a Visiting Scientist by a fellowship of the Collège de France.
469
470 **Conflict of interest statement**
471 The authors declare that they have no conflict of interest.
472

References

1. Fischbach MA, Walsh CT. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. Chem Rev. 2006;106(8):3468-96.
2. Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat Prod Rep. 2013;30(1):108-60.
3. Kloosterman AM, Cimermancic P, Elsayed SS, Du C, Hadjithomas M, Donia MS, et al. Integration of machine learning and pan-genomics expands the biosynthetic landscape of RiPP natural products. PLoS Biol. 2021;18:e3001026.
4. Li Y, Rebuffat S. The manifold roles of microbial ribosomal peptide-based natural products in physiology and ecology. J Biol Chem. 2020;295(1):34-54.
5. Montalban-Lopez M, Scott TA, Ramesh S, Rahman IR, van Heel AJ, Viel JH, et al. New developments in RiPP discovery, enzymology and engineering. Nat Prod Rep. 2021;38(1):130-239.
6. Luo S, Dong SH. Recent Advances in the Discovery and Biosynthetic Study of Eukaryotic RiPP Natural Products. Molecules. 2019;24(8).
7. Oman TJ, van der Donk W. Follow the leader: the use of leader peptides to guide natural product biosynthesis. Nat Chem Biol. 2010;6(1):9-18.
8. Burkhart BJ, Hudson GA, Dunbar KL, Mitchell DA. A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. Nat Chem Biol. 2015;11(8):564-70.
9. Chekan JR, Ongpipattanakul C, Nair SK. Steric complementarity directs sequence promiscuous leader binding in RiPP biosynthesis. Proc Natl Acad Sci U S A. 2019;116(48):24049-55.
10. Severi E, Thomas GH. Antibiotic export: transporters involved in the final step of natural product production. Microbiology (Reading). 2019;165(8):805-18.
11. Kenney GE, Rosenzweig AC. Methanobactins: Maintaining copper homeostasis in methanotrophs and beyond. J Biol Chem. 2018;293(13):4606-15.
12. Duquesne S, Destoumieux-Garzon D, Peduzzi J, Rebuffat S. Microcins, gene-encoded antibacterial peptides from enterobacteria. Nat Prod Rep. 2007;24(4):708-34.
13. Chen JY, van der Donk WA. Multinuclear non-heme iron dependent oxidative enzymes (MNIOs) involved in unusual peptide modifications. Curr Opin Chem Biol. 2024;80:102467.
14. Leprevost L, Jünger S, Lippens G, Guillaume C, Sicoli G, Oliveira L, et al. A widespread family of ribosomal peptides metallophores involved in bacterial adaptation to metal stress. BIORXIV/2024/585515.
15. Ayikpoe RS, Zhu L, Chen JY, Ting CP, van der Donk WA. Macrocyclization and Backbone Rearrangement During RiPP Biosynthesis by a SAM-Dependent Domain-of-Unknown-Function 692. ACS Cent Sci. 2023;9(5):1008-18.
16. Dou C, Long Z, Li S, Zhou D, Jin Y, Zhang L, et al. Crystal structure and catalytic mechanism of the MbnBC holoenzyme required for methanobactin biosynthesis. Cell Res. 2022;32(3):302-14.
17. Kenney GE, Rosenzweig AC. Genome mining for methanobactins. BMC Biol. 2013;11:17.
18. Kenney GE, Dassama LMK, Pandelia ME, Gizzi AS, Martinie RJ, Gao P, et al. The biosynthesis of methanobactin. Science. 2018;359(6382):1411-6.
19. Chioti VT, Clark KA, Ganley JG, Han EJ, Seyedsayamdost MR. N-Calpha Bond Cleavage Catalyzed by a Multinuclear Iron Oxygenase from a Divergent Methanobactin-like RiPP Gene Cluster. J Am Chem Soc. 2024;146(11):73132-7323.
20. Ting CP, Funk MA, Halaby SL, Zhang Z, Gonen T, van der Donk WA. Use of a scaffold peptide in the biosynthesis of amino acid-derived natural products. Science. 2019;365(6450):280-4.
21. Yu Y, van der Donk WA. Biosynthesis of a 3-thia-alpha amino acids on a carrier peptide. Proc Natl Acad Sci USA. 2022(119):e22052851119.
22. Nguyen DT, Zhu L, Gray DL, Woods TJ, Padhi C, Flatt KM, et al. Biosynthesis of macrocyclic peptides with C-terminal beta-amino-alpha-keto acid groups by three different metalloenzymes. Acs Central Science. 2024;10:2022-32.
23. Andrei A, Ozturk Y, Khalfaoui-Hassani B, Rauch J, Marckmann D, Trasnea PI, et al. Cu Homeostasis in Bacteria: The Ins and Outs. Membranes (Basel). 2020;10(9).

24. Manley OM, Shriver TJ, Xu T, Melendrez IA, Palacios P, Robson SA, et al. A multi-iron enzyme installs copper-binding oxazolone/thioamide pairs on a nontypeable Haemophilus influenzae virulence factor. Proc Natl Acad Sci U S A. 2024;121(28):e2408092121.
25. Clark KA, Seyedsayamdost MR. Bioinformatic Atlas of Radical SAM Enzyme-Modified RiPP Natural Products Reveals an Isoleucine-Tryptophan Crosslink. J Am Chem Soc. 2022;144(39):17876-88.
26. Zallot R, Oberg N, Gerlt JA. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. Biochemistry. 2019;58(41):4169-82.
27. Rivera-Millot A, Slupek S, Chatagnon J, Roy G, Saliou JM, Billon G, et al. Streamlined copper defenses make Bordetella pertussis reliant on custom-made operon. Commun Biol. 2021;4(1):46.
28. Waschulin V, Borsetto C, James R, Newsham KK, Donadio S, Corre C, et al. Biosynthetic potential of uncultured Antarctic soil bacteria revealed through long-read metagenomic sequencing. ISME J. 2022;16(1):101-11.
29. Knop M, Dang TQ, Jeschke G, Seebeck FP. Copper is a Cofactor of the Formylglycine-Generating Enzyme. ChemBioChem. 2017;18:161-5.
30. Sone Y, Nakamura R, Pan-Hou H, Itoh T, Kiyono M. Role of MerC, MerE, MerF, MerT, and/or MerP in resistance to mercurials and the transport of mercurials in Escherichia coli. Biol Pharm Bull. 2013;36(11):1835-41.
31. Kayastha BB, Kubo A, Burch-Konda J, Dohmen RL, McCoy JL, Rogers RR, et al. EF-hand protein, EfhP, specifically binds Ca(2+) and mediates Ca(2+) regulation of virulence in a human pathogen Pseudomonas aeruginosa. Sci Rep. 2022;12(1):8791.
32. Sarkisova SA, Lotlikar SR, Guragain M, Kubat R, Cloud J, Franklin MJ, et al. A Pseudomonas aeruginosa EF_Hand protein EfhP (PA4107), modulates stress responsesand virulence at high calcium concentrations. PLoS One. 2014;9(6):e98985.
33. Straw ML, Chaplin AK, Hough MA, Paps J, Bavro VN, Wilson MT, et al. A cytosolic copper storage protein provides a second level of copper tolerance in Streptomyces lividans. Metallomics. 2018;10(1):180-93.
34. Das D, Grishin NV, Kumar A, Carlton D, Bakolitsa C, Miller MD, et al. The structure of the first representative of Pfam family PF09836 reveals a two-domain organization and suggests involvement in transcriptional regulation. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2010;66(Pt 10):1174-81.
35. Staehlin BM, Gibbons JG, Rokas A, O'Halloran TV, Slot JC. Evolution of a Heavy Metal Homeostasis/Resistance Island Reflects Increasing Copper Stress in Enterobacteria. Genome Biol Evol. 2016;8(3):811-26.
36. Hao X, Luthje F, Ronn R, German NA, Li X, Huang F, et al. A role for copper in protozoan grazing - two billion years selecting for bacterial copper resistance. Mol Microbiol. 2016;102(4):628-41.
37. Sheldon JR, Skaar EP. Metals as phagocyte antimicrobial effectors. Curr Opin Immunol. 2019;60:1-9.
38. Nambi S, Long JE, Mishra BB, Baker R, Murphy KC, Olive AJ, et al. The Oxidative Stress Network of Mycobacterium tuberculosis Reveals Coordination between Radical Detoxification Systems. Cell Host Microbe. 2015;17(6):829-37.
39. Zheng Y, Xu X, Fu X, Zhou X, Dou C, Yu Y, et al. Structures of the holoenzyme TglHI required for 3-thiaglutamate biosynthesis. Structure. 2023;31(10):1220-32 e5.
40. Ghilarov D, Stevenson CEM, Travin DY, Piskunova J, Serebryakova M, Maxwell A, et al. Architecture of Microcin B17 Synthetase: An Octameric Protein Complex Converting a Ribosomally Synthesized Peptide into a DNA Gyrase Poison. Mol Cell. 2019;73(4):749-62 e5.
41. Habibi Y, Weerasinghe NW, Uggowitzer KA, Thibodeaux CJ. Partially Modified Peptide Intermediates in Lanthipeptide Biosynthesis Alter the Structure and Dynamics of a Lanthipeptide Synthetase. J Am Chem Soc. 2022;144(23):10230-40.
42. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021;49(D1):D412-D9.
43. Haft DH, Selengut JD, White O. The TIGRFAMs database of protein families. Nucleic Acids Res. 2003;31(1):371-3.

44. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011;7(10):e1002195.
45. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28(23):3150-2.
46. Oberg N, Zallot R, Gerlt JA. EFI-EST, EFI-GNT, and EFI-CGFP: Enzyme Function Initiative (EFI) Web Resource for Genomic Enzymology Tools. J Mol Biol. 2023;435(14):168018.
47. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504.
48. Teufel F, Almagro Armenteros JJ, Johansen AR, Gislason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022;40(7):1023-5.
49. Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. BMC Bioinformatics. 2011;12:385.
50. Shelton KE, Mitchell DA. Bioinformatic prediction and experimental validation of RiPP recognition elements. Methods Enzymol. 2023;679:191-233.
51. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 2005;33(Web Server issue):W244-8.
52. Holm L, Laiho A, Toronen P, Salgado M. DALI shines a light on remote homologs: One hundred discoveries. Protein Sci. 2023;32(1):e4519.
53. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000;28(1):235-42.

604 **Table 1. Known types of MNIO-modified RiPPs**
605

RiPP name	RiPP Precursor	MNIO & partner	Target residue	PTM	Ref
Methanobactin	MbnA	MbnBC	Cys	Oxazolone thioamide	18
Pearlin	TglA	TglHI	C-terminal Cys	3 thia-Glu	20
Chryseobasin	ChrA	ChrH(I) *	Cys	Peptide macrocycle, Imidazoline-2,4 dione, thiomethyl	15
Aminopyruvatide	ApyA	ApyHI	C-terminal Asp	Aminopyruvic acid	22
Methanobactin-like	MovA	MovX [#]	C-terminal Asn	C-terminal amide	19
Oxazolin	HvfA	HvfBC	Cys	Oxazolone thioamide	24
Bufferin	BufA	BufBC	Cys	Thiooxazole	14

606 * The activity of ChrH requires a SAM radical; ChrI appears to be dispensable. [#] No partner protein required.
607

608

609 **Table 2. Analysis of fully assembled bacterial genomes for bufferin-related gene**

610 **composition**

Type of RiPP precursor	# genomes with precursor genes	Max # precursor gene per genome	# genomes with bufferins but devoid of MNIO gene
DUF2282	4072	5	21
RiPP_6-12Cys	864	4	0
RiPP_EGKC	2583	4	67
TIGR04222	1372	5	490

611

612

613

614

615

616

617

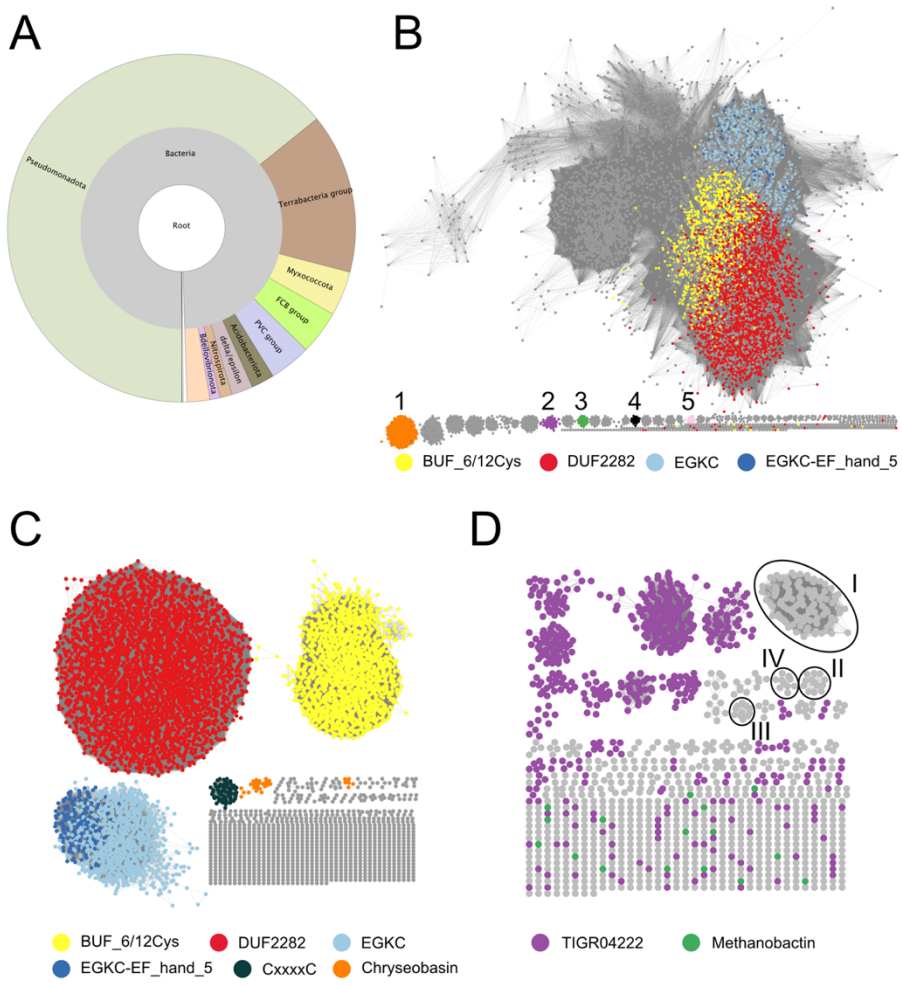

Figure 1. The MNIO superfamily and associated families of pr ecursors. A. Taxonomic distribution of the MNIO enzymes. The white sector connected to the root notably encompasses a few Eukaryota, Archea and uncultured bacteria. Whether MNIO enzymes found in Eukaryota come from bacterial symbionts is unknown. **B,C,D.** SSN analyses. **B.** Representative node network of MNIO enzymes (E value of 5, sequence alignment score threshold of 50). MNIOs genetically associated with Buf1 (DUF2282), Buf2 (BUF_6/12Cys), BUF_EGKC (see below) and BUF_EGKC|EF_hand_5 (see below) peptides are colored red, yellow, light blue and dark blue, respectively. The small clusters containing MNIOs that modify the precursors of chryseobasins, methanobactins, pearlins, aminopyruvatides and *Vibrio* methanobactin-like RiPPs are numbered 1 to 5. **C**. Representative node network of putative precursors genetically associated with MNIOs and harboring predicted signal-peptides (E-value of 10, sequence alignment score threshold of 10). The largest clusters (by decreasing sizes) correspond to precursors of the Buf1 (DUF2282), Buf2 (BUF_6/12_Cys), EGKC, EGKC|EF-hand_5, CxxxxC (see below) and chryseobasin families. The other clusters are too small to build family signatures. **D**. Representative node network of putative precursors genetically associated with MNIO enzymes and lacking a predicted signal-peptide (E value of 10, sequence alignment score threshold of 10). TIGR04222 proteins (colored purple; see below) and methanobactins were notably found in this search. Methanobactins do not form clusters in the node network, most likely because of their small size and low sequence identity. The largest four sequence clusters of unknown proteins (labeled I to IV) are described in Table S1.

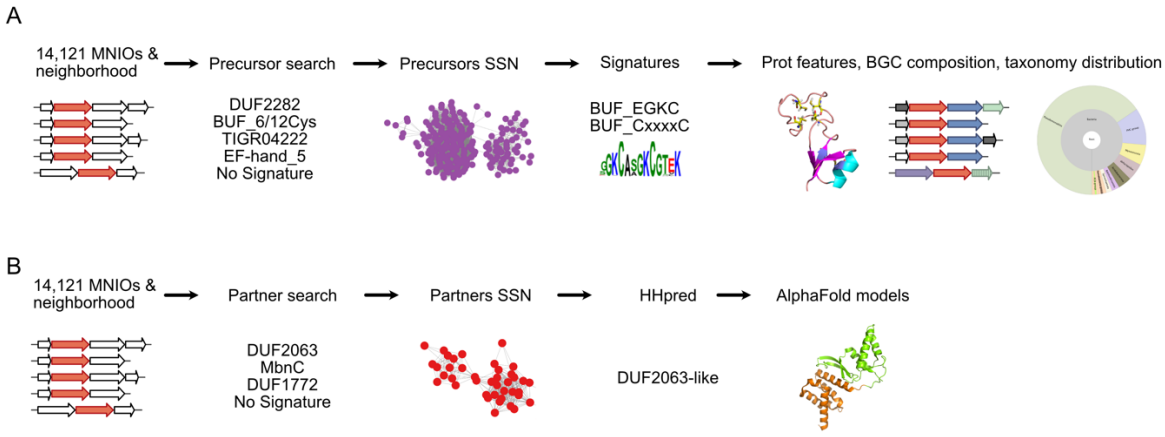

Figure 2. Flow chart for the identification of putative MNIO substrates and partners. A. Search for the major groups of MNIO substrates (putative RiPP precursors). After collecting all MNIO ORFs, the ORFs that precede them and harbor Cys residues were collected and analyzed for domain signatures. SSN analyses revealed new putative families, for which hmm signatures were generated. Available and new signatures were used to collect all family members in the NCBI nr protein database. The proteins and their BCGs were analyzed. **B**. Search for partner proteins of MNIO enzymes. The ORFs that follow MNIO ORFs were analyzed for domain signatures, and SSN analyses were performed to classify the proteins. HHpred analyses were carried out on selected signature-less proteins from the various sequence clusters of the SSN network, and AlphaFold2 models were generated. Only proteins less than 90% identical in sequence were used in all analyses.

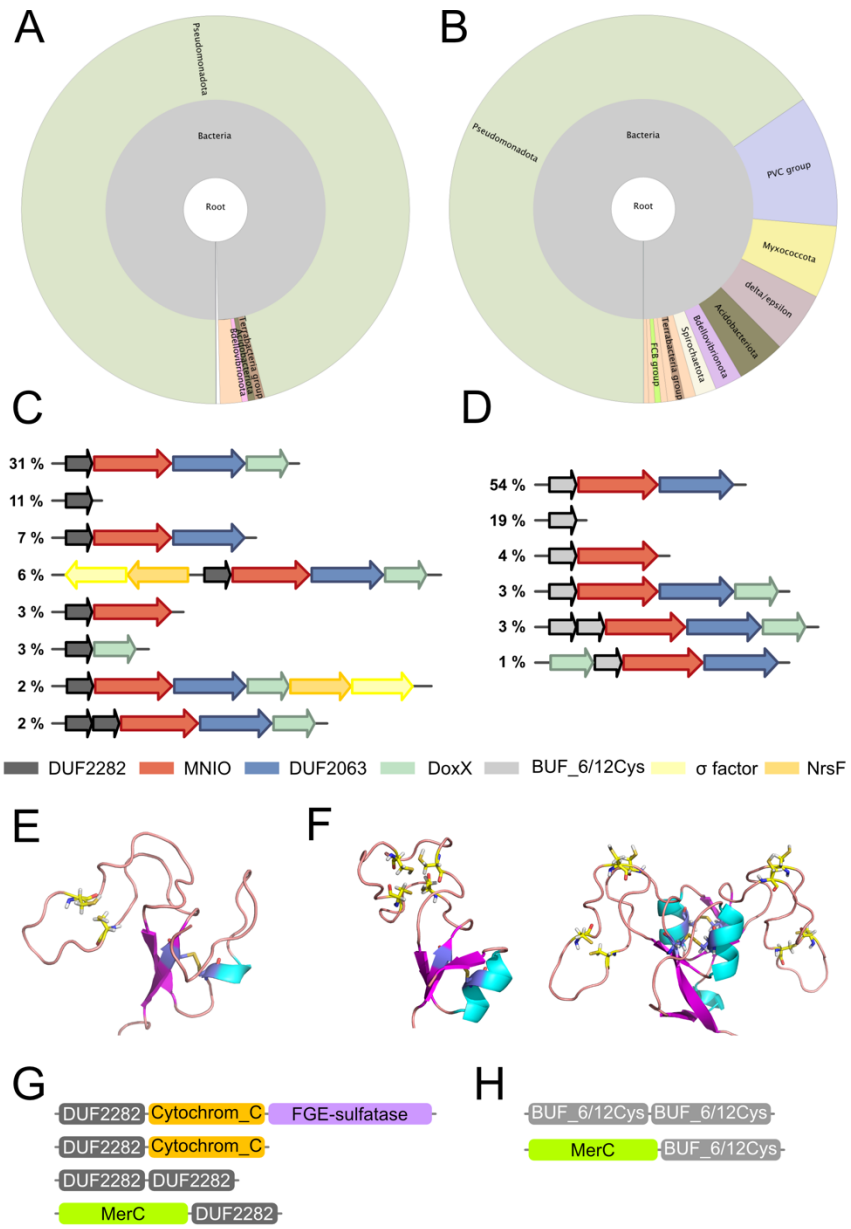

Figure 3. Features of *bona fide* bufferins. A,B. Taxonomic distributions of the Buf1 (DUF2282) (A) and Buf2 (BUF_6/12Cys) (B) bufferins. **C,D.** Major types of genetic organizations with their relative frequencies in the NCBI nr protein database. Some of the stand-alone precursor genes genuinely form single-gene transcriptional units, whereas others are flanked by genes that are not conserved to a significant extent. **E,F.** Structural models of a Buf1 bufferin (WP_011519119 of *Cupriavidus metallidurans*; 94 residues; E) and two Buf2 bufferins (WP_010947964 of *Legionella pneumophila* subsp. *fraseri*; 83 residues, and WP_077285100 of *Cognaticolwellia aestuarii*; 156 residues; F). In Buf1, the central Cys residues are the targets of PTM, whereas the first and last ones form a disulfide bond (the 4 Cys residues are in yellow stick representation). **G,H**. Fusion proteins harboring Buf1 (G) or Buf2 domains (H). The MerC fusions are predicted to harbor five transmembrane segments with the C-terminal bufferin domains of the chimeras on the periplasmic side of the membrane.

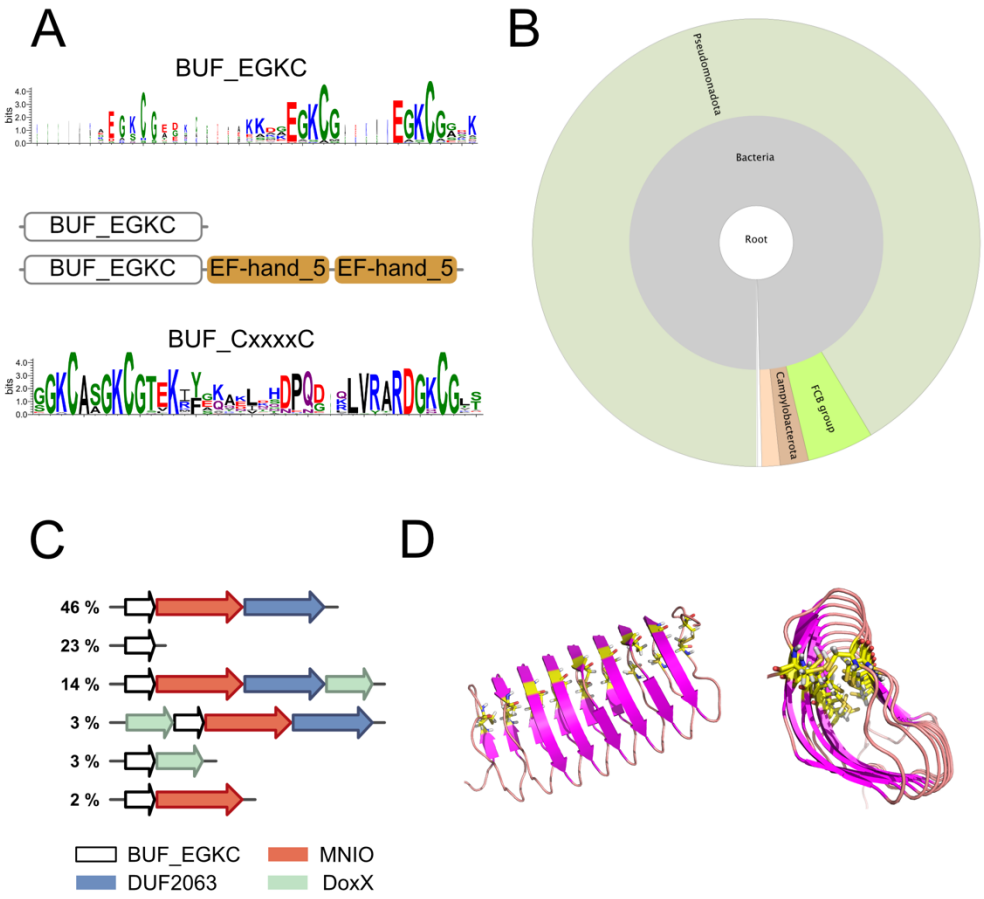

Figure 4. Features of oxazolins (BUF_EGKC) and BUF-CxxxxC proteins. A. Weblogos for the two new bufferin-like groups and schematic representation of the two major types of BUF_EGKC proteins. **B.** Taxonomic distribution of BUF_EGKC proteins. **C.** Major types of genetic organizations with their relative frequencies in the NCBI nr protein database. The open arrows represent either BUF_EGKC proteins or BUF_EGKC|EF-hand_5 chimeras. Some of the stand-alone precursor genes genuinely form single-gene transcriptional units, whereas others are flanked by genes that are not conserved to a significant extent. **D.** Structural model of a mid-size BUF_EGKC protein (WP_070991527 of *Pseudoalteromonas byunsanensis*; 195 aa). The Cys side chains (shown as sticks) that might be the targets of PTMs are predicted to point toward the inside of the helix. Note that a proportion of BUF_EGKC proteins are predicted to be unstructured, in particular those that form the N-terminal domains of the BUF_EGKC|EF-hand_5 chimeras.

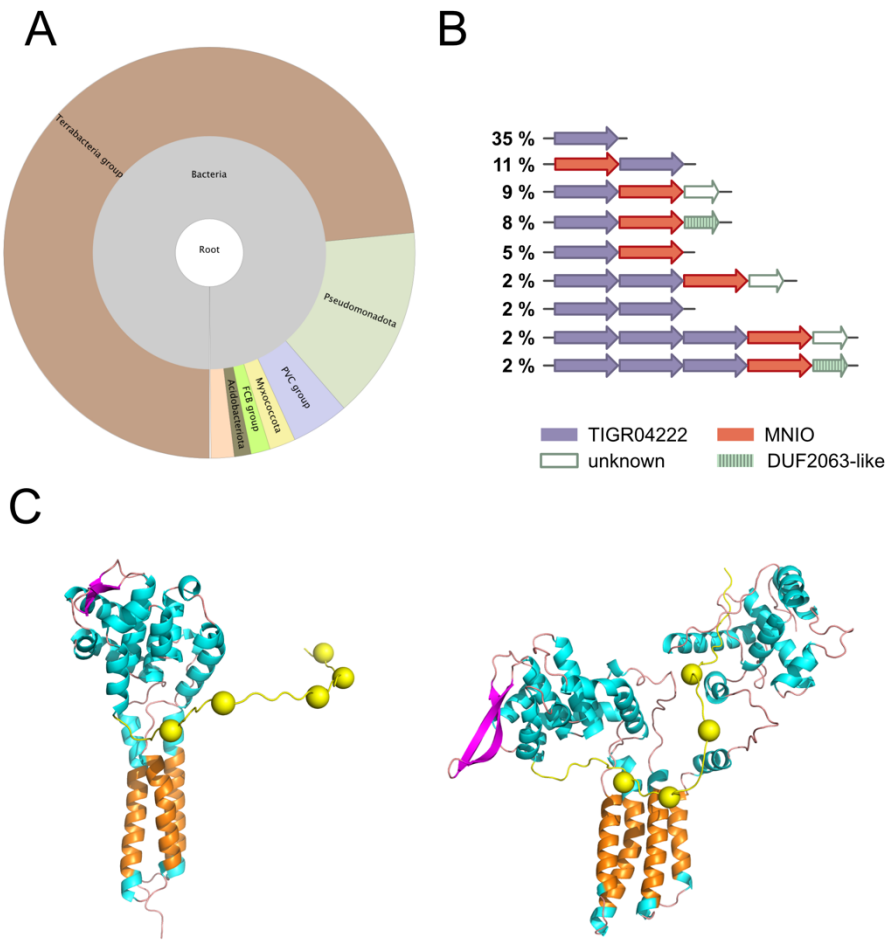

Figure 5. Features of TIGR04222 proteins. A. Taxonomic distribution. **B.** Major types of genetic organizations with their relative frequencies in the NCBI nr protein database. See text and Fig. 6 for the 'DUF2063-like' coding genes. **C.** AlphaFold2 models of selected TIGR04222 proteins (WP_179806563 of *Micromonospora purpureochromogenes*, 300 residues, and ADB18956 of *Pirellula staley*i DSM 6068, 516 residues). The TM segments are in orange, and the Cys residues are shown as yellow balls. The C-terminal Cys-rich domains are predicted to be exposed to the periplasm.

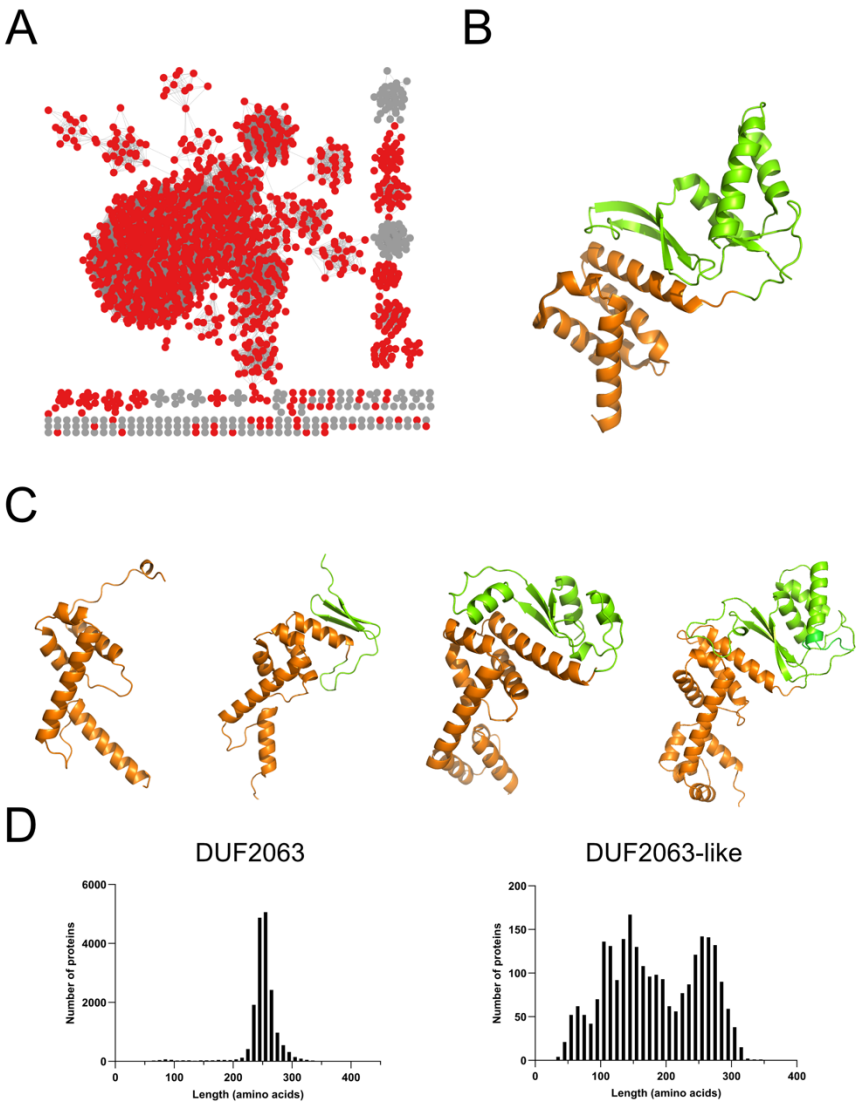

Figure 6. Characterization of new MNIO partner proteins. A. Representative node network of the unknown proteins encoded by genes that follow the MNIO-coding genes (E value of 50, sequence alignment score threshold of 5). The proteins found to be structurally related to DUF2063 proteins but that do not respond to the Pfam DUF2063 signature ('DUF2063-like') are colored red. **B**. X-ray structure of the *NGO1945* DUF2063-RRE partner protein of *Neisseria* (pdb #3dee). The DUF2063 and RRE domains are in orange and green, respectively. **C**. Alphafold2 models of four proteins found in the major sequence cluster shown in A, with the same color code as in B. From left to right: WP_033442786 of *Saccharothrix* sp. NRRL B-16314 (UNIPROT A0A7W9M5Z5, 115 residues); WP_091237309 of *Micromonospora matsumotoense* (UNIPROT A0A246RJP9, 158 residues); WP_151034108 of *Cellvibrio* sp. KY-GH-1 (UNIPROT A0A5J6PC27, 215 residues); WP_006975235 of *Plesiocystis pacifica* (UNIPROT A6GEU5, 276 residues). **D.** Size distribution of the DUF2063- and DUF2063-like proteins. The majority length of 250-260 residues in the left-hand graph corresponds to a fusion between a *bona fide* DUF2063 domain and a complete RRE-fold type domain. In the right-hand graph, the shorter proteins that are heterogeneous in length most likely do not contain full-length RRE-type domains.

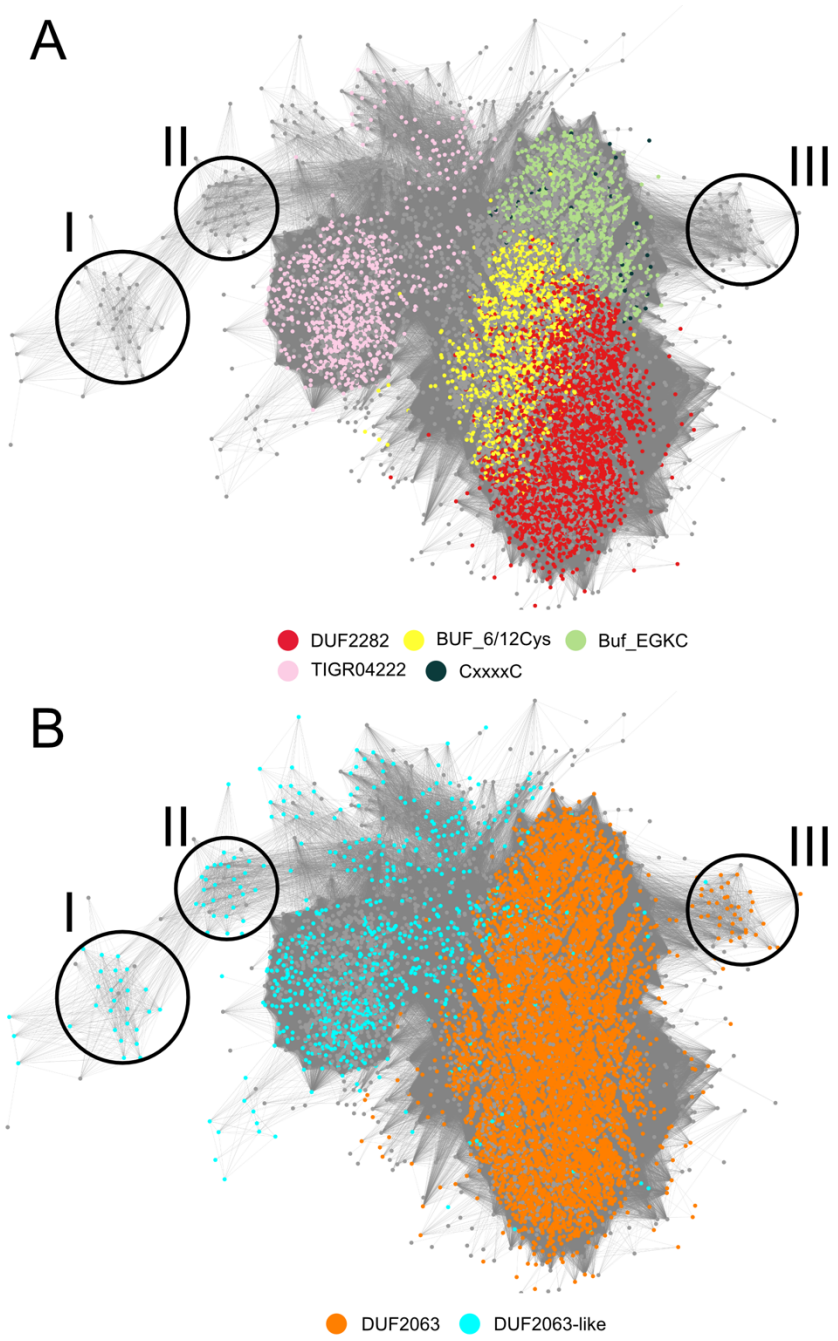

Figure 7. Subgroups of MNIO enzymes are associated with specific RiPPs and partner proteins. Members of the mega-cluster of the node network of MNIO enzymes (shown in Fig. 1B) are here colored according to their genetically linked RiPP precursors and partner proteins. In **A**, the colors of the dots represent the MNIO enzymes according to their putative precursor substrates. The pale green dots include both BUF_EGKC proteins and BUF_EGKC|EF-hand_5 chimeras. In **B**, the orange and cyan dots represent MNIO enzymes associated with *bona fide* DUF2063 partners and those associated with DUF2063-like partners, respectively. The precursors and BGCs of the MNIO enzymes of the sequence clusters numbered I to III are described in more details in Fig. S2.

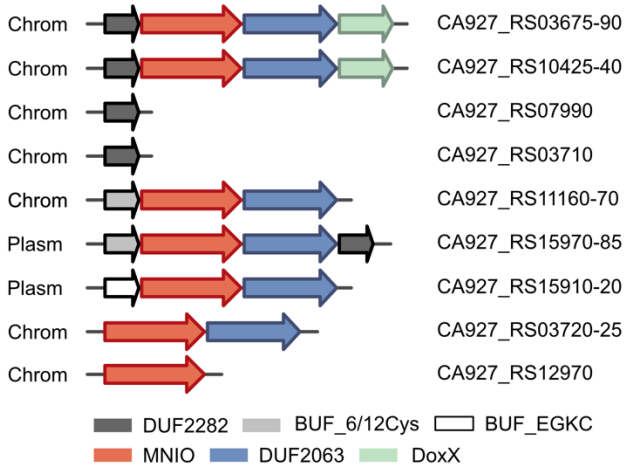

Figure 8. Bufferin-type BGCs in the genome of *Legionella pneumophila subsp fraseri*. The locus tags are provided at the right. Two BGCs are found on a mega plasmid. No genes coding for ECF sigma factors and anti-sigma proteins were found, and no other conserved genes were identified. Note that the bufferin associated with locus tag CA927_RS12970 is encoded by a truncated gene (not shown), suggesting loss of function of this BGC. The BGC of CA927_RS07990 also contains genes coding for a YceI protein (small beta barrel protein described to bind hydrophobic molecules) and a cytochrome B protein. The RiPP precursor of the CA927_RS03710 locus is in translational coupling with a guanylate cyclase domain.