



HAL
open science

Phylogenetics and biochemistry elucidate the evolutionary link between l-malate and l-lactate dehydrogenases and disclose an intermediate group of sequences with mix functional properties

Céline Brochier-Armanet, Dominique Madern

► To cite this version:

Céline Brochier-Armanet, Dominique Madern. Phylogenetics and biochemistry elucidate the evolutionary link between l-malate and l-lactate dehydrogenases and disclose an intermediate group of sequences with mix functional properties. *Biochimie*, 2021, 191, pp.140-153. 10.1016/j.biochi.2021.08.004 . hal-03358405

HAL Id: hal-03358405

<https://hal.univ-grenoble-alpes.fr/hal-03358405>

Submitted on 29 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Journal Pre-proof

Phylogenetics and biochemistry elucidate the evolutionary link between L-malate and L-lactate dehydrogenases and disclose an intermediate group of sequences with mix functional properties

Céline Brochier-Armanet, Dominique Madern



PII: S0300-9084(21)00196-6

DOI: <https://doi.org/10.1016/j.biochi.2021.08.004>

Reference: BIOCHI 6175

To appear in: *Biochimie*

Received Date: 17 April 2021

Revised Date: 19 July 2021

Accepted Date: 15 August 2021

Please cite this article as: Cé. Brochier-Armanet, D. Madern, Phylogenetics and biochemistry elucidate the evolutionary link between L-malate and L-lactate dehydrogenases and disclose an intermediate group of sequences with mix functional properties, *Biochimie* (2021), doi: <https://doi.org/10.1016/j.biochi.2021.08.004>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Published by Elsevier B.V.

1 **Phylogenetics and biochemistry elucidate the evolutionary link between L-malate**
2 **and L-lactate dehydrogenases and disclose an intermediate group of sequences with**
3 **mix functional properties**

4
5 Céline-Brochier-Armanet^{1,*} and Dominique Madern^{2,*}
6

7 **Affiliations.**

8 1. Université de Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie
9 Évolutive UMR 5558, F-69622 Villeurbanne, France.

10 2. Univ. Grenoble Alpes, CEA, CNRS, IBS, 38000 Grenoble, France

11 * Corresponding authors

12 **Abstract.**

13 The NAD(P)-dependent malate dehydrogenases (MDH) (EC 1.1.1.37) and NAD-dependent
14 lactate dehydrogenases (LDH) (EC. 1.1.1.27) form a large super-family that has been
15 characterized in organisms belonging to the three Domains of Life. MDH catalyses the
16 reversible conversion of the oxaloacetate into malate, while LDH operates at the late stage of
17 glycolysis by converting pyruvate into lactate. Phylogenetic studies proposed that the
18 LDH / MDH superfamily encompasses five main groups of enzymes. Here, starting from
19 16,052 reference proteomes, we reinvestigated the relationship between MDH and LDH. We
20 showed that the LDH / MDH superfamily encompasses three main families: MDH1, MDH2,
21 and a large family encompassing MDH3, LDH, and L-2-hydroxyisocaproate dehydrogenases
22 (HicDH) sequences. An in-depth analysis of the phylogeny of the MDH3 / LDH / HicDH family
23 and of the nature of three important amino acids located within the catalytic site and involved
24 in binding and substrate discrimination, revealed a large group of sequences displaying
25 unexpected combinations of amino acids at these three critical positions. This group
26 branched in-between MDH3 and LDH sequences. The functional characterization of several
27 enzymes from this intermediate group disclosed a mix of functional properties, indicating that
28 the MDH3 / LDH / HicDH family is much more diverse than previously thought, and blurred
29 the frontier between MDH3 and LDH enzymes. Present-days enzymes of the intermediate
30 group are a valuable material to study the evolutionary steps that led to functional diversity
31 and emergence of allosteric regulation within the LDH / MDH superfamily.

32
33 **Keywords.**

34 Malate dehydrogenase, lactate dehydrogenase, molecular evolution, allosteric regulation,
35 *Archaea*, neofunctionalization.

36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70

Declarations of interest: none

Author Agreement: all authors have seen and approved the final version of the manuscript being submitted.

Abbreviations

aLRT: approximate likelihood ratio test

BV: Bootstrap value

FBP: fructose 1,6-bisphosphate

HGT: Horizontal gene transfer

HMM: Hidden Markov Model

LDH: L-Lactate dehydrogenases

LG model: Le and Gascuel model

MDH: Malate dehydrogenase

ML: Maximum Likelihood

NADH: nicotinamide adenine dinucleotide

NADPH: nicotinamide adenine dinucleotide phosphate

OAA: oxaloacetate

PYR: pyruvate

PVC superphylum: *Planctomycetes-Verrucomicrobia-Chlamydiae* superphylum

SH: Shimodaira-Hasegawa

Highlights

Phylogenetic analyses disentangle the relationships between malate dehydrogenases and lactate dehydrogenases.

The study reveals an intermediate group of enzymes that reflects an early and step-wise functional divergence between malate dehydrogenases and lactate dehydrogenases.

Neofunctionalization and subfunctionalization contribute to evolution of malate dehydrogenases and lactate dehydrogenases.

The work suggests that present-day enzymes such as in *Tolumonas auensis* are descendant of an ancient group of enzymes in which allostery evolved.

71 **1. Introduction.**

72 Central metabolism is a key chemical process that provides energy and biochemical
73 precursors for growing and multiplication of cells. Even if the total metabolic capacity is
74 mostly similar in *Archaea* and *Bacteria*, variations are observed across lineages and even at
75 the strain level [1-3]. These variations reflect the diversity of species, lifestyles, and
76 adaptations to different environments. Understanding functions and roles of enzymes
77 involved in metabolism is not only of academic importance but also of a medical interest. For
78 instance, in human, metabolic diseases are responsible of cancer and the selective inhibition
79 of metabolic enzymes that favour growing of tumour cells in tumour microenvironments is a
80 promising therapeutic strategy [4-6]. Amongst these enzymes, L-lactate dehydrogenases
81 (LDH, EC 1.1.1.27) are one of the main targets for the development of such inhibitors [7].

82 LDH are cytosolic tetrameric enzymes involved in the anaerobic metabolism of glucose when
83 oxygen is absent or in limited supply. They reversibly transform pyruvate (PYR) into lactate
84 using NADH as coenzyme (see [8] and references therein). LDHs are found in eukaryotes
85 and bacteria [9] and can be distinguished by their capacity to regulate (or not) their activity
86 owing allosteric control. Most of the bacterial LDH display sigmoidal kinetics for pyruvate
87 (homotropic activation) and are allosterically activated by fructose 1,6-bisphosphate (FBP)
88 (heterotropic activation) [9, 10]. According to chemical similarities regarding substrate and
89 catalytic mechanism, a connection between LDH and malate dehydrogenases (MDH,
90 EC 1.1.1.37) was established [11]. Indeed, both enzymes catalyse the reversible conversion
91 of 2-hydroxyacids to the corresponding 2-ketoacids. In bacteria, MDH is involved in the
92 tricarboxylic acid cycle (TCA) and oxidize oxaloacetate (OAA) into malate using either NADH
93 or NADPH as coenzyme. In eukaryotes, several forms of MDH coexist, with distinct enzymes
94 being associated to specific cell compartments. For example, in human, the mitochondrial
95 enzyme operates in the TCA cycle, whereas the cytosolic MDH is connected to membrane
96 transporters and contributes to balance the redox state of cells via the indirect transfer of
97 reducing equivalents between the cytosol and the mitochondria [12]. In *Plantae* and in *Fungi*,
98 additional MDH are present in peroxisomes and chloroplasts [13-15].

99 The hypothesis of an evolutionary link between LDH and MDH has been reinforced when the
100 first crystallographic structures have been resolved [16]. At that time, MDH were classified
101 into two main groups of dimeric enzymes (hereafter referred as MDH1 and MDH2) [17, 18].
102 MDH1 contains eukaryotic cytosolic and various bacterial enzymes, while MDH2
103 encompasses eukaryotic glyoxysomal and mitochondrial enzymes and some bacterial
104 proteins [17]. Later, a third group of MDH (hereafter referred as MDH3 also frequently
105 mentioned as LDH-like MDH) has been described (see [17] and references therein). At the
106 sequence level, MDH3 are more similar to LDH, suggesting a closer evolutionary link [17].

107 Consistently, it has been shown that the MDH3 of the crenarchaeon *Ignicoccus islandicus*
108 that recognizes OAA as main substrate has a low activity using PYR [19]. While MDH1 and
109 MDH2 are dimeric, MDH3 display an LDH-like tetrameric organisation. However, some
110 dimeric MDH3 exist as exemplified by the enzyme from *Archaeoglobus fulgidus* [20]. MDH3
111 are mainly found in bacteria, in some archaea, but not in eukaryotes, to the exception of
112 apicomplexan [21, 22]. Recently, L-2-hydroxyisocaproate dehydrogenases (hereafter
113 referred as to HicDH) have been described and are phylogenetically related to LDH and
114 MDH3 [17, 22]. HicDH are tetrameric NADH-dependent oxidoreductases that catalyse the
115 stereospecific oxidation of aliphatic branched (S)-2-hydroxycarboxylic acids [23].

116 Twenty years ago, an evolutionary scenario describing the functional divergence between
117 LDH and MDH proposed that an ancient gene duplication of an MDH gene of unknown
118 oligomeric state led to the group of tetrameric MDH (i.e. MDH3) and to the group of dimeric
119 MDH, that evolved later toward MDH1 and MDH2. This ancient duplication event did not
120 impact the function of the resulting paralogues, i.e. the production of malate from OAA. The
121 scenario proposed also that a duplication of the tetrameric MDH3 occurred. One of the two
122 resulting paralogues conserved the ancestral functionality (i.e. the use of OAA as substrate),
123 whereas the second paralogue evolved toward canonical LDH [17]. The hypothesis that LDH
124 derive from MDH was reinforced by recent studies showing that the LDH enzymes of
125 *Plasmodium*-related species and of *Cryptosporidium* derived independently from ancient
126 MDHs [21, 22, 24]. ~~Noticed that~~ Noticeably a single case of LDH evolution from cytosolic
127 dimeric MDH1 has been reported in *Trichomonas vaginalis* [25, 26]. Altogether, four
128 independent functional shifts from MDH to LDH activity have been documented [17, 22, 25].
129 Since these early studies, the number of sequences belonging to the LDH / MDH superfamily
130 has dramatically increased (up to 56,000 according to InterPro), providing a valuable material
131 to reinvestigate the evolutionary history of this superfamily. Concomitantly, comparative
132 biochemical investigations of both wild-type LDH and MDH enzymes, and site-directed
133 mutagenesis experiments have produced a large body of kinetics and ligand-binding
134 information [19, 22, 27-36]. When analysed in light of crystallographic structures, this data
135 allows to describe the catalytic mechanisms of LDH and MDH (summarized in
136 Supplementary Figure S1). The chemical reaction proceeds by transferring a hydride ion
137 from their coenzyme to the C2 carbon of their respective substrates (i.e. PYR or OAA,
138 respectively). The universally conserved R171 of LDH and MDH coordinates the carboxylate
139 moiety of PYR and OAA. For the sake of clarity, residues numbering is accordingly to the
140 normalized structural numbering based on LDH structure of *Squalus ancahthias* (PDB code
141 1LDM) [37]. All these studies also revealed the major role of three amino acid residues with

142 respect to the discrimination between PYR and OAA. These three residues are located within
143 the catalytic vacuole of LDH and MDH at positions 102, 199, and 246.

144 The selection between PYR and OAA is mostly due to the nature of amino acid at position
145 102. In MDH, the arginine at this position helps to anchor the second carboxylate extremity of
146 OAA, while in LDH, glutamine, a polar residue, accommodates the uncharged extremity of
147 PYR. The amino acid at position 246 also contributes to discriminate between the two
148 substrates. In MDH, the small lateral chain of A246 accommodates the methylene
149 carboxylate group of OAA [32], while in LDH, the larger side chain of T246 is considered as
150 unfavourable for OAA-binding because of steric hindrance [27]. In LDH and MDH, charge
151 neutrality of the catalytic site was reported as favourable for efficient catalysis [27]. In MDH,
152 the double negative charge of OAA is screened by R102 and R171, once the catalytically
153 competent complex is formed, ensuring the neutrality within the catalytic site [32, 38]. In
154 LDH, the single negative charge of the D199 lateral chain and that of PYR ensure the
155 neutrality of the catalytic site. The presence of D199 (or E199) is therefore considered as a
156 one of the specific signatures for LDH, while the presence of a neutral amino acid at the
157 equivalent position (e.g. M199) is associated to MDH. According to these ~~these~~ studies, the
158 nature of the amino acids at these three positions could be used to distinguish LDH and
159 MDH. Yet, these assumptions are mostly based on the analysis of a few enzymes and do not
160 embrace the huge diversity of the LDH / MDH3 family. For instance, the observation of M199
161 in MDH relied on a comparison between LDH and (dimeric) MDH1 and MD2, when MDH3
162 were unknown [27]. Furthermore, several recent reports suggested that the situation could
163 be more complex. For instance, MDH3 from *I. islandicus*, which displays R102, A199, and
164 T246 as sequence signature, is able to recognize both OAA and PYR as substrate [19]. In
165 addition, the recognition of PYR by LDH in apicomplexa differs from the one existing in
166 canonical LDH. In fact, in these eukaryotes the functional shift from MDH toward LDH activity
167 is due to a six amino acids insertion in the mobile loop that covers the catalytic site upon
168 substrate binding [39, 40]. The insertion induces a structural reorganisation that dramatically
169 changes the location of the residue at position 102, which consequently protrudes outside
170 the catalytic vacuole and lost its critical role of discriminating residue. A detailed evolutionary
171 scenario by which apicomplexa LDH could have acquired the capacity to recognize PYR has
172 been recently proposed [22].

173 Here, we present a large scale phylogenomic analysis of the LDH / MDH superfamily. The
174 inferred phylogenies reveal a complex evolutionary history, heavily impacted by horizontal
175 gene transfer (HGT) across and within the three Domains of Life. The mapping of the
176 discriminating amino acids at positions 102, 199, and 246 discloses a large group of
177 LDH / MDH3 sequences harbouring unexpected combinations of amino acids. This group of

178 sequences occupied an intermediate position between *stricto sensu* MDH3 and LDH
179 enzymes. Functional characterizations reveal that members of this group are very diverse
180 and harbour original enzymatic properties, combining LDH and MDH features, as exemplified
181 by the enzyme from *Tolumonas auensis*. The present work shows that the LDH / MDH3
182 family is more diverse than previously expected and that the emergence of *stricto sensu* LDH
183 occurred from this intermediate group by a three-step acquisition and fixation of critical
184 residues in the active site.

185 **2. Materials and methods.**

186

187 **2.1. Identification of the LDH / MDH superfamily members**

188 A local protein sequence database gathering the 16,052 reference proteome sequences from
189 UniProt (<https://www.uniprot.org/proteomes/>) was built. More precisely, this corresponded to
190 433 archaea, 8,593 bacteria, 1,142 eukaryotes, and 5,884 viruses proteomes
191 (Supplementary Table S1A). The local database was queried with HMMER 3.1b2 [41] using
192 the two Hidden Markov Models (HMM) profiles from the PFAM v32.0 database that target the
193 N-terminal and the C-terminal parts of sequences belonging to the LDH / MDH superfamily:
194 the lactate / malate dehydrogenase NAD binding domain (Ldh_1_N, PF00056) and the
195 lactate / malate dehydrogenase alpha / beta C-terminal domain (Ldh_1_C, PF02866) [42].
196 This revealed 12,983 sequences containing both domains (*Evalue* threshold = 10^{-2})
197 (Supplementary Table S2). The length of the retrieved sequences ranged from 113 to 2,428
198 amino acids positions (average length = 331 amino acid positions).

199

200 **2.2. Phylogenetic analysis of the lactate / malate dehydrogenase superfamily**

201 The 12,493 sequences longer than 230 and shorter than 430 amino acid positions were
202 aligned with MAFFT v7.453 [43] using the --auto and --reorder options. The resulting multiple
203 alignment was used to infer a phylogenetic tree using FASTTREE v2.1.9 [44] with the Le and
204 Gascuel (LG) evolutionary model [45], and the -gamma and -cat = 4 options. Branch
205 supports were estimated using the Shimodaira-Hasegawa (SH) test implemented in
206 FASTTREE.

207

208 **2.3. Phylogenetic analysis of LDH / MDH3 family**

209 To avoid taxonomic redundancy, a sampling of 266 eukaryotic, 269 archaeal, and 1,737
210 bacterial proteomes was performed by keeping randomly at least one representative strain
211 per genus (Supplementary Table S1B). These 2,272 proteomes contained 1,635
212 LDH / MDH3 protein sequences longer than 230 and shorter than 430 amino acid positions.
213 These sequences were aligned with MAFFT v7.453 with the L-INSI option that allows the

214 construction of accurate multiple alignments [43]. The resulting alignment was trimmed with
215 BMGE v1.2 [46] with the BLOSUM30 substitution matrix and used to build a Maximum
216 Likelihood (ML) tree. The phylogeny was inferred with the IQ-TREE program v1.6.12 [47].
217 The LG model with a gamma distribution (G4, 4 categories, estimated alpha parameter) was
218 identified by ModelFinder [48] as the most suitable for the tree reconstruction. Branch
219 supports were estimated by the ultra-fast bootstrap procedure implemented in IQ-TREE
220 (1,000 replicates).

221

222 **2.4. Tree drawing, residue mapping, and heatmap**

223 Tree figures were drawn with iTOLv4 [49]. Heatmaps were built using ClustVis [50].

224

225 **2.5. Residue numbering**

226 The numbering of residues is defined with respect to the first structure of an LDH [37]. This
227 common numbering scheme allows one to recognize easily important active site residues,
228 such as R109, D168, R171 and H195, which are strictly conserved in all MDH and LDH.

229

230 **2.6. Protein expression**

231 Sixteen genes belonging to the LDH / MDH3 family have been chosen for functional
232 investigations. Seven genes code for proteins from mesophilic strains (set 1):
233 *Methanosarcina mazei* DSM 3647 (UniProt ID Q8PVJ7) *Planctopirus limnophila* strain
234 ATCC 43296 (UniProt ID D5S XK9), *T. auensis* TA4 (UniProt ID C4LFJ3), *Selenomonas*
235 *ruminantium* (UniProt ID Q9EVR0), *Methyloprofundus sedimenti* (UniProt ID A0A1V8M393),
236 *Cellulosilyticum lentocellum* strain ATCC 49066 (UniProt ID F2JLM3), and *Teredinibacter*
237 *turnerae* strain ATCC 39867 (UniProt ID C5BM91), while nine others genes code for
238 enzymes from (hyper)thermophilic strains (set 2): *Aquifex aeolicus* strain VF5 (UniProt ID
239 O67655), *Thermaerobacter marianensis* strain ATCC 700841 (UniProt ID E6SLT2),
240 *Thermosinus carboxydivorans* Nor1 (UniProt ID A1HSK3), *Ignicoccus hospitalis* strain KIN4/I
241 (UniProt ID A8ABY7), *Pyrobaculum aerophilum* strain ATCC 51768 (UniProt ID Q8ZVB2),
242 *Methanopyrus kandleri* strain AV19 (UniProt ID Q8TWG5), *Nitrososphaera viennensis* EN76
243 (UniProt ID A0A060HG74), *Ca. 'Bathyarchaeota archaeon'* RBG_13_46_16b (UniProt ID
244 A0A1F5CVF0), and *Pyrolobus fumarii* 1A (UniProt ID G0EHD0).

245 In addition, two genes coding for enzyme previously characterized, the HicDH from the
246 mesophilic *Weissella confusa* LBAE_C39_2 (UniProt ID H1X598) and the MDH3 from the
247 hyperthermophilic *Methanocaldococcus jannaschii* ATCC 43067 (UniProt ID Q60176), used
248 as tetramer control, were cloned and once purified their products were used as tetrameric
249 markers for calibration in Size exclusion chromatography and SEC-MALLS (see below).

250 The eighteen genes have been synthesized by the Genecust company. The sequences were
251 subjected to codon optimization for over expression in *Escherichia coli*. In order to facilitate
252 the purification using affinity chromatography, an extension of six histidine was fused to the
253 C-terminus of mesophilic enzymes (set 1), while the products of other genes (set 2) were
254 synthesized without any extension. All the enzymes were sub-cloned into the pet20a vector
255 between the *NdeI* and *BamHI* restriction sites. The constructs were transformed into
256 BL21(DE3) plysS cells and selected on LB-agar plate containing 100 ug ml⁻¹ ampicilin. A
257 single colony was cultured overnight at 37°C in 100 ml LB medium containing 100 ug ml⁻¹
258 ampicilin. 25 ml of this culture were transferred into 1L of LB medium containing 100 ug ml⁻¹
259 ampicilin. The cells were grown at 37°C until an OD600 of 0.6 was reached.

260 At this step, the overexpression protocol differs between set 1 and set 2.

261 For set1, the temperature was quickly lowered to 4°C and the cells were kept on ice for three
262 hours, then IPTG was added at a final concentration of 0.2 mM to induce expression, which
263 was continued overnight at 20°C. For set 2, IPTG was added to 0.5 mM and the cell
264 cultivation was continued for four hours at 37°C. For both sets of enzymes, bacterial cells
265 were harvested by centrifugation. The pellets were suspended in 50 mM Tris-HCl pH7, 50
266 mM NaCl (Buffer A) and frozen at -20°C for storage.

267

268 **2.7. Standard enzymatic assays.**

269 MDH activity was assessed by measuring the initial rates of OAA reduction (NADH or
270 NADPH oxidation) at 340 nm in a thermostated spectrophotometer from JASCO. The
271 standard assay mixture contained 50 mM Tris-HCl pH 7.2, 50 mM NaCl, 0.3 mM coenzyme
272 and the enzyme in a final volume of 1 ml. The reaction was initiated by addition of OAA at a
273 final concentration of 0.3 mM.

274 LDH activity was assessed in the same manner. The standard assay mixture contained 50
275 mM MES pH 6.2, 50 mM NaCl, 0.3 mM coenzyme and the enzyme. The reaction was
276 initiated by addition of PYR. Two concentrations of 1 and 10 mM were tested. The MDH and
277 LDH assays were carried out at 40°C or 70°C with enzymes considered as mesophilic or
278 thermophilic, respectively. One unit of MDH or LDH activity corresponds to the amount of
279 enzyme that catalyzes the oxidation of 1 micromole of NADH per min.

280

281 **2.8. Protein purification**

282 For set 1, after thawing of the cells, DNase at 5ug ml⁻¹ was added to the suspension. The
283 cell suspension was sonicated with a Branson sonicator at 25% intensity during three cycles
284 of 1min. The extract was centrifuged at 18,000g for 30 min at 4°C. The supernatant was
285 filtered on a 0,45 um (amicon), It was passed through a 5mL HisTrap Q HP column (GE
286 Healthcare Life Science). The column was washed with 20 mL of 50 mM Tris-HCl pH7, 200

287 mM NaCl and by 20 mL of 20 mM imidazole pH 7, 200 mM NaCl. Then, the protein under
288 study was eluted with 300 mM imidazole pH 7, 50 mM NaCl. The protein was concentrated
289 and dialyzed in Buffer A with Amicon Ultra-30 concentrators. The protein was then loaded on
290 an anion exchange UnoQ (BioRad) and eluted with a linear gradient from 0 to 1M NaCl
291 buffered by 50 mM Tris-HCl pH7. The active fractions were pooled, concentrated and loaded
292 on a size exclusion chromatography column (SEC) equilibrated with buffer A. A Superose 12
293 (GE Healthcare Life Science) or an Enrich 650 (Biorad) column were used. Fractions
294 containing the enzymes under study were concentrated and stored at -80°C in buffer A with
295 20% glycerol.

296 For set 2, the crude extract preparation was as described with set 1. Due to the expected
297 thermophilic properties of these enzymes, the soluble fraction was incubated at 75°C for 30
298 minutes and centrifuged at 18,000g for 30 minutes. The clear supernatant was loaded on a Q
299 sepharose (2x5cm) and eluted using a linear gradient of gradient from 0 to 1M NaCl buffered
300 by 50 mM Tris-HCl pH7. The end of the purification procedure using SEC and the storage
301 were as with set 1. In the case of the *M. kandleri* sequence, the SEC column was
302 equilibrated in a 0.4 M NaCl, Tris-HCl 50 mM pH7.

303 Each recombinant MDH3 and LDH purification yielded 10 mg of pure enzyme that was used
304 for characterization.

305

306 **2.9. Determination of native molecular masses of the purified enzymes.**

307 Size Exclusion Chromatography was carried out with a flow rate of 0.5 mL.min⁻¹ on an Enrich
308 650 column (Biorad). Calibration of the column was performed with the gel filtration standard
309 from Biorad. Direct comparison of the elution profiles was achieved using the Compare mode
310 of the BioLogic FPLC operating system (Biorad).

311

312 **2.10. Size Exclusion Chromatography - Multi Angle Laser Light scattering (SEC- 313 MALLS).**

314 SEC combined with online detection by MALLS and refractometry (RI) was used to measure
315 the absolute molecular mass of proteins in solution. The SEC run was performed using an
316 ENrich™ SEC650 10x300 gel-filtration column (Biorad) equilibrated with a buffer composed
317 of 50 mM Tris-HCl pH7.2 and 50 mM NaCl. Separation was performed at room temperature
318 and 50 µl of protein sample, concentrated at ~5 mg ml⁻¹, was injected with a constant flow
319 rate of 0.5 ml⁻¹ min⁻¹. Online MALLS detection was performed with a DAWN-HELEOS II
320 detector (Wyatt Technology Corp.) using a laser emitting at 690 nm. Protein concentration
321 was determined by measuring the differential refractive index online using an Optilab T-rEX
322 detector (Wyatt Technology Corp.) with a refractive index increment dn/dc of 0.185 ml⁻¹ g⁻¹.

323 Weight-averaged molecular weight (Mw) determination was done with the ASTRA6 software
324 (Wyatt Technologies) and curve was represented with GraphPad Prism.

325

326 **2.11. Analytical ultra-centrifugation.**

327 Ultra-centrifugation experiments were conducted in an XLI analytical ultracentrifuge
328 (Beckman, Palo Alto, CA) using an ANTi-50 rotor, using double channel Epon centerpieces
329 (Beckman, Palo Alto, CA) of 12 mm optical path length equipped with sapphire windows, with
330 the reference channel being typically filled with the solvent of the sample. Acquisitions were
331 done at 20°C and at 42,000 rpm (130,000g), overnight, using absorbance (280 nm) and
332 interference detection. Data processing and analysis was done using the program SEDFIT
333 [51], and GUSSE [52] using standard equations and protocols [53].

334

335 **2.12. Kinetics parameters.**

336 For each experiment assay, initial rates of NADH or NADPH oxidation were recorded at
337 various concentrations of OAA and PYR. The concentration of NADH and NADPH was
338 0.2mM. The data were analyzed using Graph Pad Prism V6 using the Michaelis-Menten or
339 Allosteric sigmoidal option. Because of the difficulties to determine the exact Km values
340 gained for kinetics at very low substrate concentration, the values reported here correspond
341 to an apparent Km.

342

343 **3. Results.**

344

345 **3.1. The LDH / MDH superfamily has a complex evolutionary history**

346 A survey of 16,052 reference proteomes covering the three Domains of Life and Viruses (i.e.
347 433 archaea, 8,593 bacteria, 1,142 eukaryotes, and 5,884 viruses) identified 12,983
348 LDH / MDH homologues with an average length of 331 amino acids positions in 7,575
349 proteomes (47.2%) (Supplementary Tables S1 and S2). More precisely, 290 homologues
350 were detected in *Archaea*, 8,010 in *Bacteria*, 4,683 in *Eucarya*, while none were present in
351 viruses (Supplementary Tables S1A and S2). Among them, 490 (3.78%) have unexpected
352 length (less than 230 or more than 430 amino acid positions), whereas the 12,493 other
353 sequences have a size compatible with LDH and MDH enzymes: 288 in *Archaea*: 7,978 in
354 *Bacteria*, and 4,227 in *Eucarya*. The phylogenetic analysis of these sequences recovered
355 three major clusters, corresponding to dimeric MDH1 and MDH2 enzymes (2,792
356 sequences, SH support = 0.991, and 2,571 sequences, SH support = 0.936, respectively),
357 and a large clade corresponding to tetrameric MDH3, HicDH, and LDH enzymes (7,130
358 sequences, SH support = 0.892) (Figure 1). These three families displayed very different
359 taxonomic distributions (Figure 2A-D). In fact, MDH1 are mainly present in *Bacteria*, MDH2 in

360 *Eucarya*, while LDH / MDH3 are broadly distributed in the three Domains of Life. As a
361 consequence, the three Domains of Life harbour different profiles (Figure 2E-H): LDH / MDH3
362 prevail in *Archaea* and *Bacteria*, MDH2 are abundant in *Eucarya*, whereas MDH1 are
363 majority in *Bacteria*.

364 The phylogenetic analysis of the LDH / MDH superfamily showed major inconsistencies with
365 the current systematics. In fact, bacterial, archaeal, and eukaryotes sequences are mixed in
366 the tree, indicating that inter-domain HGT occurred (Figure 1). Furthermore, the taxonomic
367 distribution of MDH1, MDH2, MDH3, and LDH is highly variable depending on the considered
368 lineages (Figure 3). For instance, regarding the *Planctomycetes*, *Verrucomicrobia*, and
369 *Chlamydiae* (PVC) superphylum, MDH1 dominates in *Chlamydiae* and *Verrucomicrobia*,
370 while MDH3 is majority in *Planctomycetes*. Worth to note, while both cytosolic MDH1 and
371 mitochondrial MDH2 were likely inherited in eukaryotes from their last common ancestor,
372 these enzymes are not universal in present-day lineages, suggesting that secondary and
373 independent losses occurred during the diversification of *Eucarya*. Yet, all eukaryotic
374 lineages encode at least one of these two MDH, to the exception of most apicomplexan that
375 have neither of them, and instead harbour a MDH3 likely acquired from a bacterial donor
376 (Figures 3A and 4). This suggests that the native MDH1 and MDH2 enzymes were
377 secondary replaced by MDH3 in the Apicomplexa. Furthermore, the number of LDH / MDH is
378 highly variable even at small evolutionary scale, within species and genera (Supplementary
379 Table S1A). As an example, *Klebsiella pneumoniae* subspecies *pneumoniae* and *Klebsiella*
380 *oxytoca* harbour two homologues, while a single homologue is found in *Klebsiella aerogenes*
381 and *Klebsiella pneumoniae* ISC21, and none is present in *K. pneumoniae* strains IS39, IS43,
382 and IS46. This suggests that the evolutionary history of LDH / MDH enzymes is complex and
383 may have been also heavily impacted by gene loss in addition to HGT.

384 To go further, we performed an accurate ML phylogenetic analysis of each of the three
385 families. To limit taxonomic redundancy, we sampled the 16,052 reference proteomes by
386 keeping one or two representative strains per genus. The 2,272 retained proteomes (266
387 eukaryotes, 269 archaeal, and 1,737 bacterial) contained 602 MDH1, 561 MDH2, and 1,635
388 LDH / MDH3 sequences longer than 230 and shorter than 430 amino acids. The resulting
389 trees confirmed that the evolution of the LDH / MDH has been heavily impacted by HGT and
390 revealed puzzling patterns (Supplementary Figures S2-S4). For instance, most bacterial
391 sequences from different phyla are intermixed on the trees and do not form monophyletic
392 groups as expected if they have had vertical inheritance (Supplementary Figures S2A, S3A,
393 and S4A). A similar situation is observed for archaeal LDH / MDH3, the main protein family
394 present in *Archaea*. In fact, LDH / MDH3 sequences are patchily distributed across archaeal
395 lineages (Figure 3C), and their phylogenetic relationships are at odds with the current
396 archaeal systematics (Supplementary Figures S4C).

397

398 3.2. Eukaryotic MDH / LDH have not been acquired from archaeal ancestors or
399 through the mitochondrial endosymbiosis

400 Regarding eukaryotic sequences, one would expect a close relationship between eukaryotic
401 cytosolic MDH1 and archaeal enzymes if the former had been inherited from the archaeal
402 lineage at the origin of eukaryotes on the one hand, and a close relationship between
403 mitochondrial MDH2 and alphaproteobacterial sequences in agreement with the
404 mitochondrial endosymbiosis hypothesis on the other hand [54]. The analysis of the MDH1,
405 MDH2, and LDH/MDH3 families provided a very different picture.

406 The MDH1 family gathers mainly actinobacterial, PVC, gammaproteobacterial, and
407 betaproteobacterial sequences, whereas it is patchy distributed in a few archaea and
408 alphaproteobacteria (Figure 3). Furthermore, these sequences do not display any specific
409 link with eukaryote sequences (Supplementary Figure S2B). This precludes the possibility
410 that the cytosolic MDH1 of eukaryotes have been acquired vertically from *Archaea* or through
411 the mitochondrial endosymbiosis. In fact, MDH1 have been likely acquired by eukaryotes via
412 HGT from a non-alphaproteobacterial bacterial donor. Because MDH1 sequences from
413 eukaryotes are not specifically linked to any specific bacterial lineage, the identity of the
414 bacterial donor lineage cannot be determined.

415 Regarding MDH2, this family encompasses very few bacterial sequences, mainly from
416 gammaproteobacteria, and none are present in *Alphaproteobacteria* or *Archaea* (Figures 1
417 and 3, and Supplementary Figure S2). This suggests either a gammaproteobacterial origin of
418 eukaryotic mitochondrial MDH2 or an HGT event from eukaryotes to gammaproteobacteria
419 followed by HGT among gammaproteobacterial lineages. Altogether, an alphaproteobacterial
420 or an archaeal origin of eukaryotic MDH1 and MDH2 is excluded. Puzzlingly, while most
421 alphaproteobacterial proteomes contain LDH / MDH3 sequences, they are unrelated to
422 eukaryotic sequences (Figure 3 and Supplementary Figure S4). This means that if
423 eukaryotes have acquired an alphaproteobacterial LDH / MDH3 coding gene through the
424 mitochondrial endosymbiosis, it has been lost, possibly after the acquisition of a MDH2 from
425 bacterial, as proposed early on by Madern (2002).

426 Eukaryotic proteomes from some alveolata, fungi, metazoa, and viridiplantae contain
427 LDH / MDH3 sequences (Figure 3). These sequences are intermixed with prokaryotic
428 sequences reflecting again secondary and independent acquisitions by HGT from different
429 donors (Figure 1 and Supplementary Figure S4). In the case of apicomplexa, a subsequent
430 neofunctionalization event allowed their MDH3 to acquire the capacity to converts PYR to
431 lactate [22].

432

433 **3.3. Mapping of residues involved in substrate recognition sheds a new light on the** 434 **LDH / MDH3 family**

435 The LDH / MDH3 family encompasses LDH, MDH, and the enigmatic HicDH enzymes. The
436 unrooted phylogeny of this large family does not clearly distinguish subfamilies that would
437 correspond to these three enzymes (Supplementary Figure S4). Previous experimental
438 works have shown that substrate discrimination depends strongly on the conformation of the
439 substrate-binding site and thus on the nature of the residues present, with canonical MDH3
440 harbouring R102, A246 or S246, while canonical LDH Q102 and T246 at the homologous
441 sites. In addition, canonical LDH have an acidic residue (aspartate or glutamate) at position
442 199. We mapped these residues on the LDH / MDH3 phylogeny, as well as the phylogenetic
443 position of characterized enzymes (Figure 4A). As expected, the nature of these three
444 residues involved in lactate or malate production and experimental data were consistent (see
445 filled / empty blue triangles as well as filled / empty grey triangles on Figure 4A).
446 Consistently, two large clades corresponding to LDH and MDH3 can be delineated (clades
447 with blue and pink branches on Figure 4A, respectively), and will be hereafter referred as to
448 LDH and MDH3 *stricto sensu*. These two clades are homogenous regarding the nature of the
449 residue at position 102, to the exception of apicomplexan LDH sequences that are nested
450 within MDH3 *stricto sensu* sequences. This is not surprising because these eukaryotes LDH
451 result of a recent neofunctionalization event of a MDH3 enzyme involving structural
452 changes of the catalytic site. As a consequence, the R102 residue has lost its role in
453 substrate discrimination (see above). Another exception concerns a small group of
454 sequences nested within *stricto sensu* LDH that display residues typical of MDH3 enzymes
455 at position 102 and 246. LDH *stricto sensu* sequences are mainly present in bacteria, some
456 eukaryotes (i.e. in some metazoa, viridiplantae, fungi, and in a few protists), and in a few
457 archaea, while MDH3 sequences are mainly present in bacteria (Figure 4A).

458 Surprisingly, a large group of sequences branch in-between *stricto sensu* LDH and MDH3
459 sequences (Figure 4). Sequences from this intermediate group harbour different
460 combinations of amino acids at positions 102, 199, and 246 (Figures 4A and 4B). In fact,
461 most of these sequences harbour only one residue involved in PYR recognition (T246),
462 associated in some cases to residues specifically associated to OAA recognition (i.e. R102)
463 or to PYR recognition (i.e. D199 or E199). They encompass most archaeal sequences and a
464 mix of bacterial sequences from different phyla (Figure 4B). Notice that LDH enzymes from
465 *Thermotogales* are nested within the intermediate group and not within the *stricto sensu*
466 LDH, although they display their three amino acids signature (Figure 4B). According to the
467 literature, this intermediate group contain enzymes using OAA and PYR as substrate, as well
468 as HicDH (empty and filled grey triangles on Figure 4B, respectively). Yet, functional data are
469 scarce and do not encompass the whole diversity of the group, especially regarding the

470 nature of the amino acids observed at the three important positions for substrate
471 discrimination.

472

473 **3.4. Functional characterization reveals new unexpected enzymatic properties.**

474 To go further, we characterized enzymes occupying key positions in the LDH / MDH3 family
475 and/or harbouring various and original amino acid combinations at critical positions 102, 199,
476 and 246 with the aim to increase the description of the LDH / MDH3 family, and especially
477 the intermediate group (Table 1). More precisely we selected three enzymes belonging to
478 group of MDH3 *stricto sensu*: *A. aeolicus* strain VF5, *T. marianensis* strain ATCC 700841,
479 and *M. mazei* DSM 3647, to enrich the functional information for this very large cluster
480 (Figure 4A). We selected also *P. limnophila* ATCC 43296 as representative of the enigmatic
481 small subgroup of sequences nested within LDH *sensu stricto* that harbor amino acids
482 associated to MDH3 (i.e. R102, M199, and A246) instead of the three residues associated to
483 PYR recognition (i.e. Q102, T246 and D/E199) (Figure 4A). Regarding the intermediate
484 group, functional data are available in the literature for a few members of subgroups C, E, G,
485 H, I, J, and one enzyme located in-between subgroups H and J (empty and filled grey
486 triangles on Figure 4B and Supplementary Table S3). In this study, we investigated the
487 properties of twelve additional enzymes from the intermediate group: *N. viennensis* and *Ca.*
488 'Bathyarchaeota archaeon' RBG_13_46_16b (subgroup A), *P. aerophilum* strain (subgroup
489 B), *T. carboxydivorans* Nor1 (a lonely sequence located in-between subgroups E and F),
490 *T. auensis* (subgroup F), *M. kandleri* (subgroup I), *I. hospitalis* (subgroup J), *P. fumarii*
491 (subgroup J), *S. ruminantium*, *C. lentocellum*, *T. turnerae*, and *M. sedimenti* (subgroup L)
492 (Table 1).

493 To determine whether the products of the sixteen selected genes, corresponding to three
494 MDH3 *stricto sensu*, one LDH *stricto sensu*, and twelve sequences from the intermediate
495 group, could be assigned as MDH or LDH, the overexpressed soluble proteins were purified,
496 their oligomeric state was determined, and their capacity to use OAA or PYR was tested.
497 Among the selected targets, six from the intermediate group (i.e. the two enzymes from
498 subgroup A and some of subgroups J and L) did not fold properly after overexpression in
499 *Escherichia coli*.

500

501 **3.4.1. Determination of the oligomeric state**

502 Elution profiles of the properly folded enzymes showed a major peak on SEC column
503 suggesting it contains a single oligomeric species. The *M. jannaschii* MDH3 (subgroup I) and
504 *W. confusa* HicDH (subgroup P), for which the crystal structures indicate that they are
505 tetramer were used as control experiments [55, 56]. Both enzymes display an elution volume
506 around 13.5 mL (see I and J panels on Supplementary Figure S5). To the exception of the

507 *T. auensis* enzyme, which has a higher elution volume of 14,5 mL suggesting it behaves as a
508 dimer, the other enzymes display elution volume similar to the control experiments indicating
509 that they are tetramers. Because the determination of molecular weight of a protein by SEC
510 is sensitive to several artefacts [57], eight measurements were completed by SEC MALLS
511 analysis (Supplementary Figure S6). The eight enzymes analysed (including the
512 *M. jannaschii* MDH3 as control) display molecular weights in agreement with the tetrameric
513 state. The specific case of *T. auensis* enzyme, which was disclosed as dimer is presented
514 below. Altogether, these data confirm that the tetrameric association is likely dominant within
515 the MDH3 / LDH clade.

516

517 **3.4.2. Determination of the functional activity**

518 Then, proteins under study were assessed for their capacity to recognize OAA and PYR with
519 NADH or NADPH as coenzyme. To the exception of *S. ruminantium*, which uses only PYR
520 and should be considered as an LDH, the other enzymes strictly use OAA MDH. Saturation
521 curves are presented and maximal specific activity have been determined (Table 1 and
522 Supplementary Figure S7). The enzymes shown to used OAA harbor profiles expected for
523 Michaelian enzymes with either a hyperbolic profile or a profile showing an inhibition by
524 increasing concentration of substrate. Activity values are in the typical data ranges reported
525 for MDH [58] to the exception of *P. limnophila* and *T. auensis*, which have maximal specific
526 activity below 100 U. mg⁻¹.

527 As expected, activity measurements confirm that enzymes from *A. aeolicus*, *T. marianensis*,
528 and *M. mazei*, belonging to the clade of *stricto sensu* MDH3, are functional tetrameric MDH,
529 which use OAA as substrate and NADH as coenzyme, an observation in agreement with the
530 presence of R102, M199, and A246, and with previous measurements on bacterial
531 sequences from the MDH3 clade [59-61]. A malate activity, albeit with a low specific activity,
532 is also observed for the *P. limnophila* enzyme. The corresponding sequence is nested within
533 *stricto sensu* LDH but harbor, R102, M199, and A246 like most MDH3 - and not the three
534 LDH residues associated to PYR recognition. This strongly suggests that a shift from lactate
535 toward malate production occurred, likely as a consequence of the replacement of the three
536 residues favoring PYR recognition by amino acid favoring OAA recognition. To our
537 knowledge, this is the first report of a functional shift from LDH toward MDH activity.

538 Within the intermediate group, subgroups B and I harbor R102, M199, and A246 (or S246),
539 found in MDH3 *stricto sensu* enzymes. Consistently, *P. aerophilum* and *Aeropyrum pernix*
540 enzymes (subgroup B), and and *M. kandleri* enzymes (subgroup I) use OAA as substrate
541 (Table 1 and Supplementary Table S3). However, while the two enzymes of subgroup B use
542 only NADH as coenzymes, the enzymes of *M. jannaschii* (subgroup I) and
543 *Methanothermobacter thermoautotrophicus* (subgroup H) use both NADH and NADPH with a

544 preference for the latter [62, 63]. Activity measurements using the *M. kandleri* enzyme
545 (subgroup I), confirms it is a NADPH-dependent MDH. The strict preference for NADH in
546 dehydrogenase relies on the presence of an aspartate at the position 52 (Supplementary
547 Figure S8) in the coenzyme binding site that prevents, owing to electrostatic repulsion, the
548 binding of the additional phosphate in NADPH [64]. In *Methanothermobacter*
549 *thermoautotrophicus*, *M. jannaschii*, and *M. kandleri* sequences a glycine (neutral residue) is
550 present at position 52, allowing accommodating the NADPH. Most sequences LDH / MDH3
551 family harbour D52. The presence of alternative residues (mainly glycine or serine) at
552 position 52 is restricted to sequences of subgroups H, I, and K, a few members of subgroup
553 L (not shown). This suggests that the use of NADPH is rather rare and that this feature
554 emerged recently and independently during the diversification of the LDH / MDH3 family.
555 Interestingly, sequences from subgroups C, E, H, J, and the two sequences located in-
556 between E and F subgroups harbour R102 typical of MDH3 *sensu stricto* and T246 specific
557 of LDH enzymes, while various residues are observed at position 199 (e.g. A199, Q199, or
558 M199). Accordingly, we wondered the impact of this mix combination of MDH and LDH
559 signatures on the activity of these enzymes. Previous reports indicate that *A. fulgidus*
560 (subgroup C), *Haloarcula marismortui* (subgroup E), as well as *Metallosphaera sedula* and
561 *I. islandicus* (both from subgroup J) have an NADH-dependent MDH activity with high activity
562 for OAA. Our data indicate that this is also the case for *T. carboxydivorans*, an enzyme that
563 branch in-between E and F subgroups and that shows an MDH activity with a low affinity for
564 OAA (i.e. K_m values higher than 0.5 mM) (Table 1). Accordingly, it would be tempting to
565 conclude that the presence of R102 has a stronger influence than T246 on substrate
566 recognition, and that subgroup A and O enzymes could have MDH activity. However, in a
567 recent study, the NADH-dependent MDH3 from *I. islandicus* was shown to use OAA (with
568 high affinity) but also PYR in the absence of OAA to produce lactate yet with very low affinity
569 and efficiency [19]. The structure of this enzyme has revealed a slight deformation of the
570 catalytic site compared to other enzymes harbouring the same combination of amino acids at
571 positions 102, 199, and 246, and changes of the relative orientation between subunits that
572 make the tetrameric scaffold. It has been proposed that these features are responsible for
573 the dual activity [19]. The situation is likely more complex, as since we show here that the
574 enzyme from *I. hospitalis*, a close relative of *I. islandicus*, strictly recognizes OAA
575 (Supplementary Figure S7). Both enzymes share the same three amino acids signature
576 (R102, A199, T246), but diverge by 55 amino acid replacements. These changes may impact
577 the local topology and dynamics of their catalytic site, making them more or less efficient in
578 substrate differentiation. This illustrates that functional differences may exist even at small
579 evolutionary scale. This indicates that even if amino signatures are good predictors of

580 substrate recognition and enzymatic activity, they are not sufficient, especially in the case of
581 enzymes with dual activities.

582 Puzzlingly, most of the sequences of the intermediate group harbour neither arginine nor
583 glutamine at amino acid position 102, whereas all of them harbour either a threonine at
584 position 246 as LDH or a serine at the equivalent position as MDH3. This is surprising given
585 the important role of the position 102 in substrate differentiation, but it opens the possibility to
586 investigate the contribution of amino acid at position 246. More precisely, most enzymes
587 harbour a threonine at position 246 (sequences belonging in-between subgroups H and I,
588 and in subgroups L, N, and P), while the presence of S246 is restricted to sequences from
589 the subgroup F. Functional information on those sequences are rare. Regarding sequences
590 harboring T246, available data rely on the characterization of the HicDH of *W. confusa* [65]
591 and *Mycoplasma agalactiae* PG2 [66]. Both enzymes display a PYR-dependent LDH activity,
592 even if the HicDH recognizes preferentially branched oxoacids and to a lesser extent PYR.
593 The comparison between structures of *W. confusa* HicDH and some LDH showed that it is
594 the consequence of hydrophobic changes in the substrate binding pocket, which direct
595 substrate specificity towards large substrate [56]. Here we have performed the
596 characterization of a representative of the large subgroup L, the enzyme of *S. ruminantium*.
597 Our data indicates that this enzyme is an LDH with a low affinity for PYR. This is consistent
598 of the indirect assignation of the *S. ruminantium* enzyme as a LDH because of its capacity to
599 complement the anaerobic growth deficiency of an *E. coli* mutant [67]. We noticed that in this
600 protein, the three amino acids signature is I102, A199 and T246. In order to establish firmly
601 its function, the gene product was overexpressed and purified. The protein is a tetramer as
602 analyzed with SEC MALLS (Table 1 and Supplementary Figure S6). It uses only NADH as
603 coenzyme and recognizes PYR with a sigmoid profile (Table 1 and Supplementary Figure
604 S7). The affinity for PYR is 2 mM and FBP does not exert any activation. The Hill' coefficient
605 was 2.5. OAA was not recognized. The characterization demonstrates that the enzyme of
606 *S. ruminantium* is therefore an LDH of low affinity for PYR that displays homotropic
607 activation. The characterization of the *S. ruminantium*, *Mycoplasma agalactiae*, and
608 *W. confusa* enzymes suggests that the presence of a threonine at amino acid position 246
609 associated to the absence of R102, could be sufficient to abolish the recognition of OAA and
610 to induce an affinity for PYR. However, additional data is needed to test this hypothesis.

611

612 **3.5. The puzzling enzyme from *Tolomonas auensis***

613 The case of *T. auensis* within subgroup F is of special interest because it harbours a P at
614 position 102, while other sequences of subgroup F have a glycine, an alanine or a serine.
615 Amino acid at position 102 is part of the mobile loop that covers the catalytic site upon
616 substrate binding. Because of its intrinsic structure, a proline residue decreases the local

617 flexibility of the protein backbone [68]. Consequently, it is likely that P102 could strongly
618 influence the substrate differentiation capacity of the *T. auensis* enzyme, even if amino acid
619 at position M199 and S246 correspond to those generally encountered in *sensu stricto*
620 MDH3. The inspection of the *T. auensis* enzyme primary sequence showed also a noticeable
621 exception with respect to the putative coenzyme utilization. While, all LDH and most MDH3
622 harbor D52 in agreement with their capacity to use NADH [64] (see above), *T. auensis* and
623 other members of subgroup F display an asparagine at the equivalent position, suggesting
624 these enzymes preferentially uses NADPH. Because of this atypical combination of amino
625 acids, the enzyme of *T. auensis* was therefore purified and characterized.

626 Surprisingly, the elution profile using SEC at the end of the purification revealed that the
627 enzyme elutes at 14.5 mL (Supplementary Figure S5). Such a value suggests that the
628 enzyme does not behave as a tetramer (see above). The oligomeric state of the enzyme was
629 further analysed using SEC-MALLS and by analytical ultra-centrifugation. Its theoretical
630 molecular weight is 138 kDa for a tetrameric association. The protein elutes as a single peak
631 on SEC with an elution volume of 13.8 ml. The experimental weight-averaged molecular
632 mass was 56.4 ± 5 kDa (Figure 5A). Such a value is consistent with a dimeric species
633 instead of a tetramer. The *T. auensis* enzyme oligomeric state was also analyzed using ultra-
634 centrifugation. The sedimentation coefficient $S_{20,w}$ profile shows a major peak of $4.67 \pm$
635 0.2 S with a frictional coefficient ratio of 1.29 that confirms it is a globular dimer (Figure 5B).
636 Previous studies using ultra-centrifugation showed tetrameric MDH3 have $S_{20,w}$ values in the
637 range 7.0-7.2 S [19, 69].

638 The experimental data show that the enzyme is able to reduce both OAA and PYR with
639 NADPH (0.3 mM), but not with NADH. This dual capacity of substrate recognition strongly
640 suggests a phenomenon of promiscuous enzymatic activity that could be due to the absence
641 of canonical discriminating residue Q or R at position 102. The *W. conf* HicDH with no
642 canonical (OAA/PYR) substrate discriminating residue at position 102 was shown to be able
643 to use 2-Oxoisocaproate 2-oxocaproate, 2-phenylpyruvate in addition to PYR (Feil, et al.
644 1994). None of these compounds were recognized by the *T. aue* enzyme.

645 In the presence of OAA, the hyperbolic profile for activity of the enzyme is consistent with an
646 MDH profile. It has a maximal activity of $19 \text{ U} \cdot \text{mg}^{-1}$ and a K_m value for OAA of 8.5 mM (Figure
647 5C). Such a K_m value demonstrates a very low affinity for OAA, in contrast to most of *stricto*
648 *sensu* MDH3 for which K_m for OAA is below 0.3 mM [19, 58]. The capacity of OAA
649 recognition by the enzyme of *T. auensis* is consistent with the presence of a serine at
650 position 246, as in MDH3 enzymes. Unexpectedly, this enzyme catalyzes also PYR oxidation
651 using NADPH with a sigmoid profile as in the case of homotropic activation of allosteric
652 NADH-dependent LDH [10]. The enzyme has a maximal specific activity of $94 \text{ u} \cdot \text{mg}^{-1}$, a Hill'
653 coefficient of 1.9 and a $K_{1/2}$ (Km) value for PYR of 45 mM indicating a cooperative

654 substrate binding and a low affinity for this substrate (Figure 5D). It is assumed that the
655 concentration of free OAA within cells of living organisms is very low, because it is
656 continuously recycled in the TCA cycle [70]. Therefore, the functionality of the *T. auensis*
657 enzyme as MDH is crippled by its high K_m value for OAA. In the proteome of *T. auensis*,
658 another protein (C4LAG8 UniProt) is assigned by sequence homology to a MDH1,
659 suggesting it acts as the “true” MDH.

660 With *stricto sensu* allosteric LDH, the maximal catalytically efficiency is achieved in the
661 presence of the allosteric effector fructose 1, 6 bis-phosphate (FBP), which induces a strong
662 favorable shift of affinity for PYR toward sub millimolar value [10]. The effect of FBP was
663 assessed on the *T. auensis* enzyme. It did not induce any significant activation effect. Recall
664 that in LDH, FBP binding occurs at interfaces between the dimer of dimers, which makes the
665 final tetrameric assembly [30]. Because the *T. auensis* protein is a dimer, the condition to
666 create an FBP-binding site is not achieved. As a conclusion of this functional investigation,
667 *T. auensis* enzyme, displays a new unexpected behavior. It corresponds to the first
668 description of a dimeric NADPH-dependent LDH of low efficiency that solely displays
669 homotropic activation.

670

671 4. Discussion

672 Gene duplication is considered as one of the main evolutionary forces that favors the
673 diversification of protein function [71]. A duplication event can lead to (i) the amplification of
674 the preexisting function due to multiple copies, (ii) the specialization through the
675 subfunctionalization of one of the two resulting paralogues, or (iii) the emergence of a novel
676 function owing to the accumulation of key mutations via the neofunctionalization of one of the
677 two paralogues [72]. In the super family of LDH / MDH, it was suggested that the emergence
678 of canonical LDH resulted from the duplication of an ancestral tetrameric MDH. After the
679 duplication, one of the paralogues kept the ancestral function and is at the origin of the
680 MDH3 subfamily, while the second paralogue acquired the capacity to transform PYR into
681 lactate [17].

682 Our in-depth investigation of the huge LDH / MDH superfamily shows that the family of
683 LDH / MDH3 could be split into three main groups accordingly to the combination of the three
684 amino acids involved in substrate binding and differentiation. They correspond to (i) the
685 *stricto sensu* tetrameric MDH3 that display the combination R102, A/S246, and M199 and
686 recognize OAA with a high affinity, (ii) an intermediate group gathering very diverse enzymes
687 with respect to the nature of amino acids found at positions 102, 199, and 246, oligomeric
688 states, coenzyme use, and substrate specificities and affinities, and (iii) the *stricto sensu*
689 tetrameric LDH that mostly harbor Q102, D/E199, and T246, and that use PYR as substrate
690 (Supplementary Table S3).

691 Our analysis indicates also that the evolutionary history of the LDH / MDH3 family was
692 heavily impacted by HGT, including HGT across the three Domains of Life. Furthermore,
693 there is no or very few taxonomic redundancies between LDH and MDH3 (Figure 4 and
694 Supplementary Figure S4), as expected if these two subfamilies derived from a duplication
695 event. These results challenge previous hypotheses to explain the emergence of canonical
696 LDH and MDH3 from a duplication event [17]. A more likely scenario would be that during the
697 diversification of MDH3, a subgroup of sequences, which is at the origin of the intermediate
698 group, has progressively acquired the capacity to use PYR instead of OAA. A series of
699 substitution events affecting important residues for the enzymatic activity, relaxed the strong
700 affinity for OAA, allowing accommodating for other substrates and led to the creation of
701 reservoir of enzymes of various oligomeric states, with different affinities and substrate
702 differentiation capacities (Figure 6). In particular, the replacement of the alanine or the serine
703 by a threonine at position 246, would have been a permissive event for the future evolution
704 toward LDH. Indeed, T246 is considered as unfavorable for OAA binding [73, 74]. The
705 conservation of R102 in archaeal proteins belonging to the intermediate group, could explain
706 why these enzymes are always functional MDH even if they display for most of them T246.
707 This observation confirms that R102 is the most efficient amino acid to define MDH
708 functionality.

709 The role of subfunctionalization that is related to functional promiscuity cannot be ignored as
710 playing a role in the evolution of LDH and MDH. Indeed, a previous work showed that the
711 archaeal *I. islandicus* MDH belonging in the intermediate group has a promiscuous LDH low
712 activity [19]. Noticed this sequence harbors threonine at position 246. The present work
713 revealed that another enzyme from the intermediate group, C4LFJ3 from *T. auensis* (with a
714 P102, M199, and S246 signature) could recognize both OAA and PYR, the substrates typical
715 of MDH and LDH, with low affinity. The MDH from *T. carboxydivorans* (with a R102, M199
716 and T246 signature) has a low affinity for OAA, suggesting that the decrease of affinity for
717 the main substrate of MDH was very likely one of the steps that contributed to favor the
718 emergence of LDH. Eventually, such a phenomenon could have been amplified so that the
719 MDH function is fully abolished. This likely happened in the subgroups of sequences that do
720 not display an R at position 102. The characterization of the enzyme from *S. ruminantium*
721 demonstrates it is an LDH with low PYR affinity and which no longer recognize OAA. Noticed
722 that the subgroup G, that encompasses sequences harboring the three residues found in
723 *sensu stricto* LDH and the experimentally characterized LDH enzyme from *Thermotoga*
724 *maritima* [75], corresponds likely to LDH enzymes.

725 Previous studies focused the LDH / MDH superfamily indicated that LDH functionality
726 evolved independently four times from MDH enzymes [17, 21, 22, 25]. Through the
727 characterization of the *P. limnophila* enzyme, we disclose the first case of conversion of LDH

728 toward MDH. However, it occurred very rarely during the functional diversification of the
729 LDH / MDH superfamily. By revealing new cases of functional changes from MDH toward
730 LDH, the present work demonstrates that this phenomenon is not as rare as previously
731 thought. These functional changes are consistent with an enzyme evolution hypothesis,
732 which proposes that functional promiscuity and conformational diversity are strong drivers of
733 protein evolution [76]. Accordingly, the emergence of *stricto sensu* LDH could have resulted
734 from the neofunctionalization of an NADH-dependent enzyme, with a T246 signature, from
735 the intermediate group with a low (or abolished) affinity for OAA and/or a low affinity for PYR.
736 The fixation of an acidic amino acid (D or E) at position 199 promoting charge neutrality
737 within the catalytic site and the acquisition of Q102, for which the neutral chain interacts with
738 the neutral PYR methyl group, were the last steps to favor an efficient PYR recognition as
739 observed in *stricto sensu* LDH. Altogether, our results confirmed that the action of mutations
740 relies strongly on epistasis, i.e. the amplitude of the effect depends on the sequence of the
741 rest of the protein acids [77]. This may explain why attempts to transform tetrameric MDH3
742 into LDH on the basis of unique R102 to Q mutation have encountered a limited success
743 [28]. With respect to the role of the three positions analyzed in the present work, an efficient
744 transformation would have been achieved likely, if the M199, ensuring the charge neutrality
745 within the catalytic site, was replaced by an acidic amino acid at position 199. Finally, until
746 the present work, the allosteric behavior was considered as constitutive of the tetrameric
747 state in LDH, while a unique case of hidden allosteric capacity was reported for a MDH3
748 enzyme [19]. The characterization of the enzyme from *T. auensis* demonstrates that, in the
749 clade MDH3 / LDH, the minimal oligomeric state which is able to support homotropic
750 activation is a dimer.

751

752 5. Conclusion

753 Our phylogenetic and experimental investigations disclose an unexpected diversity of
754 LDH / MDH3 family and shed a new light on the evolution of this important family of
755 enzymes. In particular, we were able to characterize representative enzymes of most of the
756 subgroups that branch in-between *stricto sensu* LDH and MDH3, yet this is far from covering
757 the whole diversity of these atypical enzymes. Biochemical, structural and dynamical
758 investigations of members of the intermediate group are thus of primary importance to gain
759 insights into the evolution of allostery and substrate recognition. Our data allow us to propose
760 an evolutionary scenario for the emergence of LDH from MDH enzymes involving punctual
761 and permissive substitutions at key positions slightly modified and relaxed the properties of
762 MDH enzymes. In one of these intermediate lineages, the acquisition of a few critical
763 residues, including a glutamine at position 102, hampered the recognition of PYR, a
764 signature of *stricto sensu* LDH.

765

766 Acknowledgements

767 We thank Drs Eric Girard, Fabio Sterpone, Jacques Coves, Caroline Barette, Caroline Mas,
768 and Laura Eme for stimulating discussions and critical reading of the manuscript. This work
769 used the platforms of the Grenoble Instruct center (ISBG, UMS 3518 CNRS-CEA-UJF-
770 EMBL) within the Grenoble Partnership for Structural Biology (PSB). This work also used the
771 computing facilities of the CC LBBE/PRABI.

772

773 Funding

774 This work has been supported by the Agence Nationale de la Recherche (ANR-16-CE11-
775 0011-02).

Journal Pre-proof

776 **Table 1. Summary of the experimental characterization of new members of the LDH / MDH3 family.**

Strain (Taxonomy) Accession number – key residues	Phylogenetic position	Key residues	Expected function / Real function	Coenzyme	OAA		PYR	
					Maximal specific activity (V _{max} , U mg ⁻¹)	K _m (mM)	Maximal specific activity (V _{max} , U mg ⁻¹)	K _m (mM)
<i>Methanosarcina mazei</i> (Archaea, Methanosarcinales) Q8PVJ7	MDH3 <i>stricto sensu</i>	R102, M199, A246	MDH / MDH	NADH NADPH	301 -	0.26 -	- -	- -
<i>Thermaerobacter marianensis</i> (Bacteria, Firmicutes) E6SLT2	MDH3 <i>stricto sensu</i>	R102, M199, A246	MDH / MDH	NADH NADPH	402 -	0.11 -	- -	- -
<i>Aquifex aeolicus</i> (Bacteria, Aquificales) O67655	MDH3 <i>stricto sensu</i>	R102, M199, A246	MDH / MDH	NADH NADPH	510 -	- -	- -	- -
<i>Planctopirus limnophila</i> (Bacteria, Planctomycetes) D5SXX9	LDH <i>stricto sensu</i>	R102, M199, A246	MDH / MDH	NADH NADPH	76 -	0.05 -	- -	- -
<i>Pyrobaculum aerophilum</i> (Archaea, Thermoproteales) Q8ZVB2	IG – subgroup B	R102, M199, T246	MDH / MDH	NADH NADPH	254 -	0.10 -	- -	- -
<i>Thermosinus carboxydivorans</i> (Bacteria, Firmicutes) A1HSK3	IG – in-between E and F	R102, M199, T246	MDH / MDH	NADH NADPH	368 -	2.50 -	- -	- -
<i>Tolumonas auensis</i> (Bacteria, Gammaproteobacteria) C4LFJ3	IG – subgroup F	P102, M199, S246	? / LDH	NADH NADPH	- 19	- 8.50	- 94	- 45
<i>Methanopyrus kandleri</i> (Archaea, Methanopyrales) Q8TWG5	IG – subgroup I	R102, M199, S246	MDH / MDH	NADH NADPH	nd 123	- 0.10	- -	- -
<i>Ignicoccus hospitalis</i> (Archaea, Desulfurococcales) A8ABY7	IG – subgroup J	R102, A199, T246	MDH / MDH	NADH NADPH	253 -	0.18 -	- -	- -
<i>Selenomonas ruminantium</i> (Bacteria, Firmicutes) Q9EVR0	IG – subgroup L	I102, A199, T246	? / LDH	NADH NADPH	- -	- -	130 -	2.0 -

777 IG = intermediate group, nd = not determined, “-“ = no significant activity detected.

778

779

780 **Figure legends**

781

782 **Figure 1. Phylogeny of LDH / MDH superfamily.**

783 The tree encompasses 12,493 sequences longer than 230 and shorter than 430 amino acid
 784 positions. The three main families are represented with different colours: MDH1 in brown
 785 (aLRT support = 0.936), dimeric MDH2 in purple (aLRT support = 0.991), and MDH3 and
 786 LDH (aLRT support = 0.892). Sequences from *Archaea* are represented in red, *Bacteria* in
 787 green, and *Eucarya* in orange.

788

789 **Figure 2. LDH / MDH superfamily distribution in the three Domains of Life.**

790 Taxonomic distribution of LDH / MDH superfamily across the three Domains of Life: **(A)**
 791 MDH1, **(B)** MDH2, **(C)** MDH3 / LDH / HicDH, and **(D)** all sequences. Distribution of MDH1,
 792 MDH2, MDH3 / LDH / HicDH in *Archaea* **(E)**, *Bacteria* **(F)**, *Eucarya* **(G)**, and in the three
 793 Domains of Life **(H)**.

794

795 **Figure 3. Taxonomic distribution of MDH1, MDH2, LDH / MDH3 family sequences.**

796 The colour scale indicates the proportion of proteomes containing at least one sequence of
 797 the family (white = not present, dark orange = present in all members).

798

799 **Figure 4. Unrooted maximum likelihood phylogeny of the LDH / MDH3 family.**

800 **(A)** Full tree showing the three main groups of LDH / MDH3 enzymes: the LDH *stricto sensu*
 801 (blue branches) that harbour the three amino acids allowing the recognition and binding of
 802 PYR (Q102, D199/E199, and T246), the MDH3 *stricto sensu* (pink branches) that harbour
 803 the two amino acids allowing the recognition and binding of OAA (R102, and SA46/S246),
 804 and the intermediate group (black branches) that harbour various combination of these
 805 amino acids.

806 **(B)** Zoom on the intermediate group. LDH and MDH3 *stricto sensu* have been collapsed.

807 The tree was inferred with the 1,635 sequences present in 2,272 representative proteomes.

808 In both trees, sequences from *Archaea* are represented in red, *Bacteria* in green, and
 809 *Eucarya* in orange. The scale bar represents the average number of substitutions per site.

810 Circles at branch correspond to ultrafast bootstrap values. For clarity, values lower than 90%
 811 were omitted. Critical residues involved in substrate recognition and functional characterized

812 enzymes are mapped with dark blue triangles. The most internal circle indicates a Q (filled
813 triangles) or a R (empty triangles) at position 102, corresponding to lactate or malate
814 recognition, respectively. The second circle highlights a T (filled triangles) or a S/A (empty
815 triangles) at position 246, corresponding to lactate or malate recognition, respectively. The
816 filled triangles on third circle design an acidic residue (D or E) at position 199 that ensure
817 charge neutrality within the catalytic site. Light grey triangles on the fourth circle design the
818 position of experimentally characterized LDH (filled triangles) and MDH3 (empty triangles)
819 from previous studies (Supplementary Table S3). Black triangles on the fifth circle indicate
820 the position of experimentally characterized LDH (filled triangles) and MDH3 (empty
821 triangles) in this study.

822

823 **Figure 5. Properties of *Tolomonas auensis* enzyme.**

824 **(A)** Oligomeric state determination of the *T. auensis* enzyme using SEC-MALLS analysis. The
825 chromatogram shows the elution profile monitored by excess refractive index (left ordinate
826 axis) and the molecular weight as dashed line (right ordinate axis) derived from MALLS and
827 refractometry measurements. The estimated average molecular weight is indicated on the
828 graph. **(B)** Sedimentation profile monitored by absorbance at 280 nm. **(C)** and **(D)** Enzymatic
829 activity profiles of *T. auensis* enzyme using Oxaloacetate or Pyruvate respectively.
830 Measurements were done in the presence of the indicated concentrations of substrates with
831 NADPH as coenzyme.

832

833 **Figure 6. Scheme of the functional evolution within the MDH/LDH superfamily.**

834 The different groups of enzymes defined on the basis of three keys residues are indicated.
835 The pink arrow shows the variation of oxaloacetate (OAA) affinity between the various MDHs
836 groups: from high (solid line) to low (dashed line) values. The blue arrow illustrates the
837 selection of LDH functionality due to pyruvate usage. The variation of the three important
838 residues involved in substrate recognition are indicated. Only two of the various combination
839 that can be observed within the intermediate group are shown.

840

841

842

843 **Author contributions.**

844 - conception of the work: C.B-A & D.M

845 - collection of data: C.B-A & D.M

846 - analysis of data: C.B-A & D.M

847 - writing of manuscript: C.B-A & D.M

848

849 **Author Agreement.**

850 All authors have seen and approved the final version of the manuscript being submitted.

851

852

853

854 **References**855 [1] B.J. Baker, V. De Anda, K.W. Seitz, N. Dombrowski, A.E. Santoro, K.G. Lloyd, Diversity, ecology
856 and evolution of Archaea, *Nat Microbiol*, 5 (2020) 887-900.857 [2] J.L. Macalady, T.L. Hamilton, C.L. Grettenberger, D.S. Jones, L.E. Tsao, W.D. Burgos, Energy,
858 ecology and the distribution of microbial life, *Philosophical transactions of the Royal Society of*
859 *London. Series B, Biological sciences*, 368 (2013) 20120383.860 [3] A. Kreimer, E. Borenstein, U. Gophna, E. Ruppin, The evolution of modularity in bacterial metabolic
861 networks, *Proceedings of the National Academy of Sciences of the United States of America*, 105
862 (2008) 6976-6981.863 [4] K.O. Alfarouk, D. Verduzco, C. Rauch, A.K. Muddathir, H.H. Adil, G.O. Elhassan, M.E. Ibrahim, J.
864 David Polo Orozco, R.A. Cardone, S.J. Reshkin, S. Harguindey, Glycolysis, tumor metabolism, cancer
865 growth and dissemination. A new pH-based etiopathogenic perspective and therapeutic approach to
866 an old cancer question, *Oncoscience*, 1 (2014) 777-802.867 [5] U.E. Martinez-Outschoorn, M. Peiris-Pages, R.G. Pestell, F. Sotgia, M.P. Lisanti, Cancer
868 metabolism: a therapeutic perspective, *Nat Rev Clin Oncol*, 14 (2017) 11-31.869 [6] H. Pelicano, D.S. Martin, R.H. Xu, P. Huang, Glycolysis inhibition for anticancer treatment,
870 *Oncogene*, 25 (2006) 4633-4646.871 [7] S.L. Zhang, Y. He, K.Y. Tam, Targeting cancer metabolism to develop human lactate
872 dehydrogenase (hLDH)5 inhibitors, *Drug Discov Today*, 23 (2018) 1407-1415.873 [8] M. Adeva-Andany, M. Lopez-Ojen, R. Funcasta-Calderon, E. Ameneiros-Rodriguez, C. Donapetry-
874 Garcia, M. Vila-Altesor, J. Rodriguez-Seijas, Comprehensive review on lactate metabolism in human
875 health, *Mitochondrion*, 17 (2014) 76-100.876 [9] E.I. Garvie, Bacterial lactate dehydrogenases, *Microbiological reviews*, 44 (1980) 106-139.877 [10] H. Taguchi, The Simple and Unique Allosteric Machinery of *Thermus caldophilus* Lactate
878 Dehydrogenase : Structure-Function Relationship in Bacterial Allosteric LDHs, *Advances in*
879 *experimental medicine and biology*, 925 (2017) 117-145.880 [11] L.J. Banaszak, R.A. Bradshaw, Malate Dehydrogenases, in: P.D. Boyer (Ed.) *The Enzymes*,
881 *Academic Press, New-York*, 1975, pp. 369-396.882 [12] Y.M. Go, D.P. Jones, Redox compartmentalization in eukaryotic cells, *Biochim Biophys Acta*,
883 1780 (2008) 1273-1290.884 [13] P. Minarik, N. Tomaskova, M. Kollarova, M. Antalík, Malate dehydrogenases--structure and
885 function, *Gen Physiol Biophys*, 21 (2002) 257-265.886 [14] J. Selinski, R. Scheibe, Malate valves: old shuttles with new perspectives, *Plant Biol (Stuttg)*, 21
887 *Suppl 1* (2019) 21-30.888 [15] C. Gietl, Partitioning of malate dehydrogenase isoenzymes into glyoxysomes, mitochondria, and
889 chloroplasts, *Plant Physiol*, 100 (1992) 557-559.890 [16] J.J. Birktoft, R.T. Fernley, R.A. Bradshaw, L.J. Banaszak, Amino acid sequence homology among
891 the 2-hydroxy acid dehydrogenases: mitochondrial and cytoplasmic malate dehydrogenases form a
892 homologous system with lactate dehydrogenase, *Proceedings of the National Academy of Sciences of*
893 *the United States of America*, 79 (1982) 6166-6170.

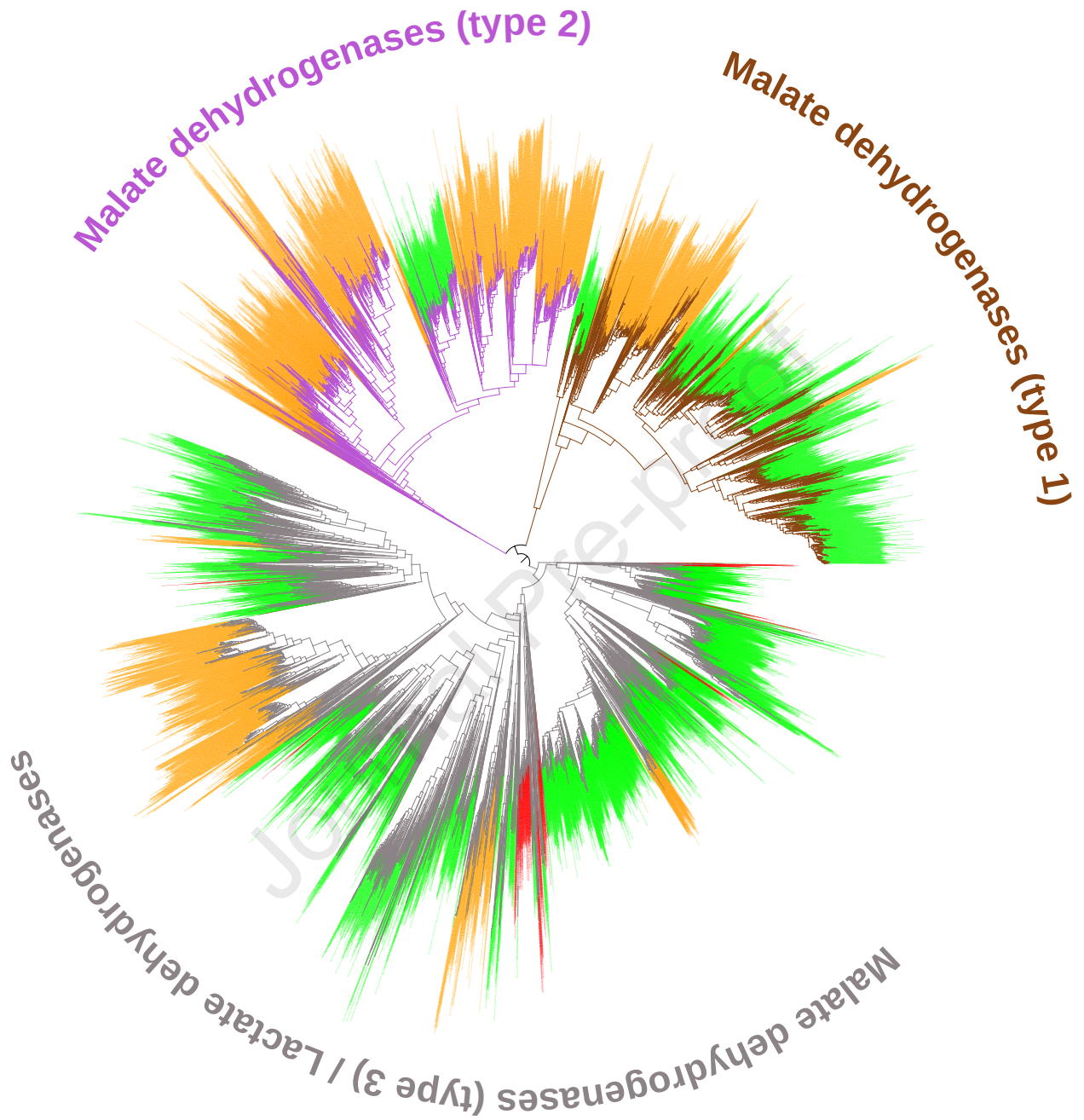
- 894 [17] D. Madern, Molecular evolution within the L-malate and L-lactate dehydrogenase super-family,
895 Journal of molecular evolution, 54 (2002) 825-840.
- 896 [18] C.R. Goward, D.J. Nicholls, Malate dehydrogenase: a model for structure, evolution, and
897 catalysis, Protein Sci, 3 (1994) 1883-1888.
- 898 [19] J. Roche, E. Girard, C. Mas, D. Madern, The archaeal LDH-like malate dehydrogenase from
899 Ignicoccus islandicus displays dual substrate recognition, hidden allostery and a non-canonical
900 tetrameric oligomeric organization, J Struct Biol, 208 (2019) 7-17.
- 901 [20] A. Irimia, F.M. Vellieux, D. Madern, G. Zaccai, A. Karshikoff, G. Tibbelin, R. Ladenstein, T. Lien,
902 N.K. Birkeland, The 2.9Å resolution crystal structure of malate dehydrogenase from Archaeoglobus
903 fulgidus: mechanisms of oligomerisation and thermal stabilisation, Journal of molecular biology, 335
904 (2004) 343-356.
- 905 [21] D. Madern, X. Cai, M.S. Abrahamsen, G. Zhu, Evolution of Cryptosporidium parvum lactate
906 dehydrogenase from malate dehydrogenase by a very recent event of gene duplication, Molecular
907 biology and evolution, 21 (2004) 489-497.
- 908 [22] J.I. Boucher, J.R. Jacobowitz, B.C. Beckett, S. Classen, D.L. Theobald, An atomic-resolution view
909 of neofunctionalization in the evolution of apicomplexan lactate dehydrogenases, eLife, 3 (2014).
- 910 [23] H. Schütte, W. Hummel, M.R. Kula, L-2-hydroxyisocaproate dehydrogenase—A new enzyme
911 from Lactobacillus confusus for the stereospecific reduction of 2-ketocarboxylic acids, Appl Microbiol
912 Biotechnol, 19 (1984) 167-176.
- 913 [24] G. Zhu, J.S. Keithly, Alpha-proteobacterial relationship of apicomplexan lactate and malate
914 dehydrogenases, The Journal of eukaryotic microbiology, 49 (2002) 255-261.
- 915 [25] P.A. Steindel, E.H. Chen, J.D. Wirth, D.L. Theobald, Gradual neofunctionalization in the
916 convergent evolution of trichomonad lactate and malate dehydrogenases, Protein Sci, 25 (2016) 1319-
917 1331.
- 918 [26] G. Wu, A. Fiser, B. ter Kuile, A. Sali, M. Muller, Convergent evolution of *Trichomonas vaginalis*
919 lactate dehydrogenase from malate dehydrogenase, Proceedings of the National Academy of
920 Sciences of the United States of America, 96 (1999) 6285-6290.
- 921 [27] H.M. Wilks, K.W. Hart, R. Feeney, C.R. Dunn, H. Muirhead, W.N. Chia, D.A. Barstow, T. Atkinson,
922 A.R. Clarke, J.J. Holbrook, A specific, highly active malate dehydrogenase by redesign of a lactate
923 dehydrogenase framework, Science, 242 (1988) 1541-1544.
- 924 [28] F. Cendrin, J. Chroboczek, G. Zaccai, H. Eisenberg, M. Mevarech, Cloning, sequencing, and
925 expression in Escherichia coli of the gene coding for malate dehydrogenase of the extremely
926 halophilic archaeobacterium Haloarcula marismortui, Biochemistry, 32 (1993) 4308-4313.
- 927 [29] W.E. Boernke, C.S. Millard, P.W. Stevens, S.N. Kakar, F.J. Stevens, M.I. Donnelly, Stringency of
928 substrate specificity of Escherichia coli malate dehydrogenase, Archives of biochemistry and
929 biophysics, 322 (1995) 43-52.
- 930 [30] T. Ohta, K. Yokota, T. Minowa, S. Iwata, Mechanism of allosteric transition of bacterial L-lactate
931 dehydrogenase, Faraday discussions, (1992) 153-162.
- 932 [31] M.J. Reddish, H.L. Peng, H. Deng, K.S. Panwar, R. Callender, R.B. Dyer, Direct evidence of
933 catalytic heterogeneity in lactate dehydrogenase by temperature jump infrared spectroscopy, J Phys
934 Chem B, 118 (2014) 10854-10862.
- 935 [32] J.M. Gonzalez, R. Marti-Arbona, J.C.H. Chen, B. Broom-Peltz, C.J. Unkefer, Conformational
936 changes on substrate binding revealed by structures of Methylobacterium extorquens malate
937 dehydrogenase, Acta Crystallogr F Struct Biol Commun, 74 (2018) 610-616.
- 938 [33] M. Katava, M. Marchi, D. Madern, M. Sztucki, M. Maccarini, F. Sterpone, Temperature Unmasks
939 Allosteric Propensity in a Thermophilic Malate Dehydrogenase via Dewetting and Collapse, J Phys
940 Chem B, 124 (2020) 1001-1008.
- 941 [34] A.R. Clarke, D.B. Wigley, W.N. Chia, D. Barstow, T. Atkinson, J.J. Holbrook, Site-directed
942 mutagenesis reveals role of mobile arginine residue in lactate dehydrogenase catalysis, Nature, 324
943 (1986) 699-702.
- 944 [35] A. Cortes, D.C. Emery, D.J. Halsall, R.M. Jackson, A.R. Clarke, J.J. Holbrook, Charge balance in
945 the alpha-hydroxyacid dehydrogenase vacuole: an acid test, Protein Sci, 1 (1992) 892-901.
- 946 [36] T.J. Nobbs, A. Cortes, J.L. Gelpi, J.J. Holbrook, T. Atkinson, M.D. Scawen, D.J. Nicholls,
947 Contribution of a buried aspartate residue towards the catalytic efficiency and structural stability of
948 Bacillus stearothermophilus lactate dehydrogenase, Biochem J, 300 (Pt 2) (1994) 491-499.
- 949 [37] W. Eventoff, M.G. Rossmann, S.S. Taylor, H.J. Torff, H. Meyer, W. Keil, H.H. Kiltz, Structural
950 adaptations of lactate dehydrogenase isozymes, Proceedings of the National Academy of Sciences of
951 the United States of America, 74 (1977) 2677-2681.
- 952 [38] A.D. Chapman, A. Cortes, T.R. Dafforn, A.R. Clarke, R.L. Brady, Structural basis of substrate
953 specificity in malate dehydrogenases: crystal structure of a ternary complex of porcine cytoplasmic

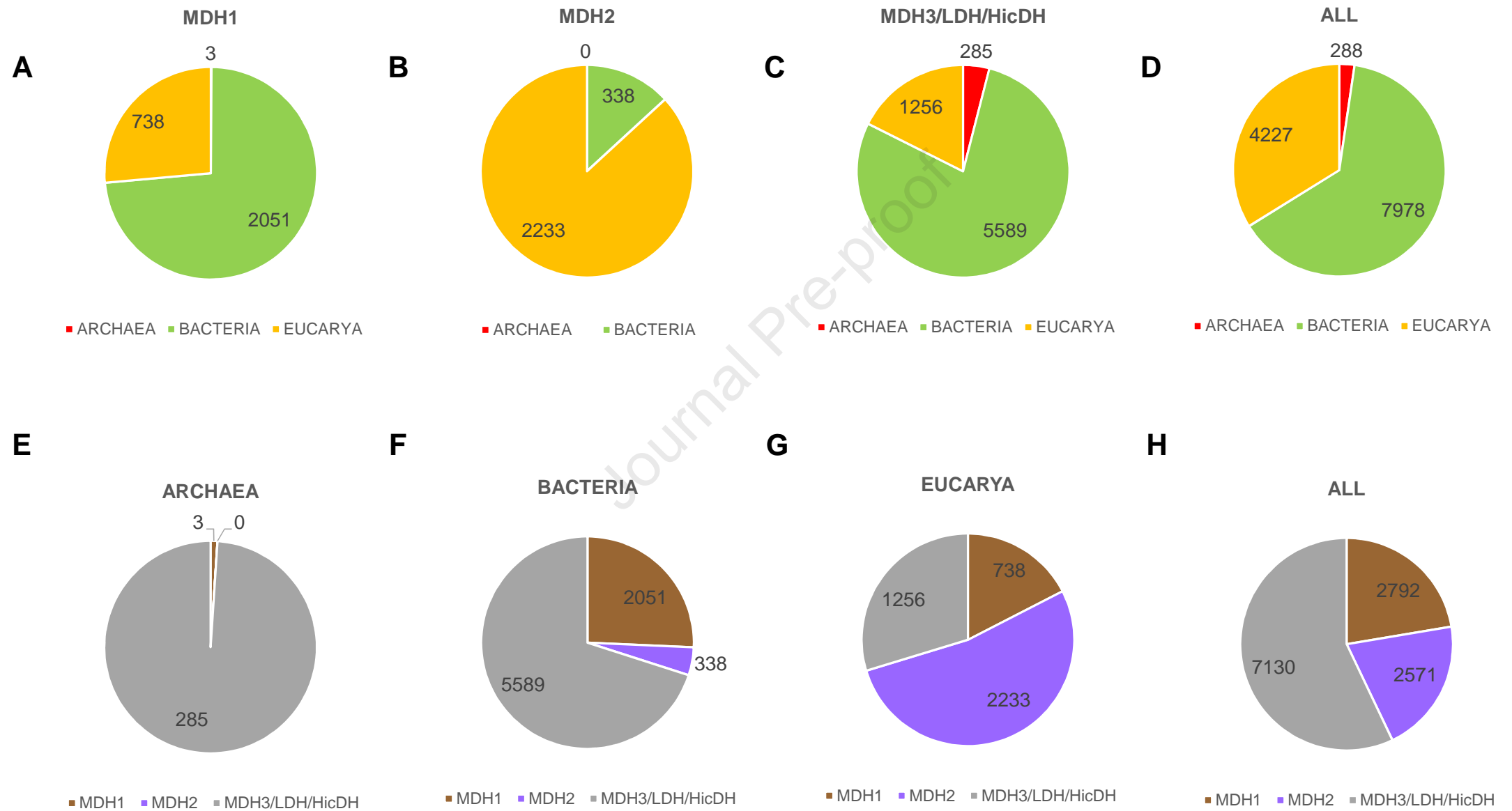
- 954 malate dehydrogenase, alpha-ketomalonate and tetrahydroNAD, *Journal of molecular biology*, 285
955 (1999) 703-712.
- 956 [39] V.J. Winter, A. Cameron, R. Tranter, R.B. Sessions, R.L. Brady, Crystal structure of *Plasmodium*
957 *berghei* lactate dehydrogenase indicates the unique structural differences of these enzymes are
958 shared across the *Plasmodium* genus, *Mol Biochem Parasitol*, 131 (2003) 1-10.
- 959 [40] S. Lunev, S. Butzloff, A.R. Romero, M. Linzke, F.A. Batista, K.A. Meissner, I.B. Muller, A. Adawy,
960 C. Wrenger, M.R. Groves, Oligomeric interfaces as a tool in drug discovery: Specific interference with
961 activity of malate dehydrogenase of *Plasmodium falciparum* in vitro, *PloS one*, 13 (2018) e0195011.
- 962 [41] S.R. Eddy, Accelerated Profile HMM Searches, *PLoS computational biology*, 7 (2011) e1002195.
- 963 [42] S. El-Gebali, J. Mistry, A. Bateman, S.R. Eddy, A. Luciani, S.C. Potter, M. Qureshi, L.J.
964 Richardson, G.A. Salazar, A. Smart, E.L.L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S.C.E.
965 Tosatto, R.D. Finn, The Pfam protein families database in 2019, *Nucleic acids research*, 47 (2019)
966 D427-D432.
- 967 [43] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements
968 in performance and usability, *Molecular biology and evolution*, 30 (2013) 772-780.
- 969 [44] M.N. Price, P.S. Dehal, A.P. Arkin, FastTree 2--approximately maximum-likelihood trees for large
970 alignments, *PloS one*, 5 (2010) e9490.
- 971 [45] S.Q. Le, O. Gascuel, An improved general amino acid replacement matrix, *Molecular biology and*
972 *evolution*, 25 (2008) 1307-1320.
- 973 [46] A. Criscuolo, S. Gribaldo, BMGE (Block Mapping and Gathering with Entropy): a new software for
974 selection of phylogenetic informative regions from multiple sequence alignments, *BMC evolutionary*
975 *biology*, 10 (2010) 210.
- 976 [47] L.T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh, IQ-TREE: a fast and effective stochastic
977 algorithm for estimating maximum-likelihood phylogenies, *Molecular biology and evolution*, 32 (2015)
978 268-274.
- 979 [48] S. Kalyaanamoorthy, B.Q. Minh, T.K.F. Wong, A. von Haeseler, L.S. Jermin, ModelFinder: fast
980 model selection for accurate phylogenetic estimates, *Nature methods*, 14 (2017) 587-589.
- 981 [49] I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v4: recent updates and new developments,
982 *Nucleic acids research*, 47 (2019) W256-W259.
- 983 [50] T. Metsalu, J. Vilo, ClustVis: a web tool for visualizing clustering of multivariate data using
984 Principal Component Analysis and heatmap, *Nucleic acids research*, 43 (2015) W566-570.
- 985 [51] P. Schuck, Size-distribution analysis of macromolecules by sedimentation velocity
986 ultracentrifugation and lamm equation modeling, *Biophys J*, 78 (2000) 1606-1619.
- 987 [52] C.A. Brautigam, Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation
988 Data, *Methods in enzymology*, 562 (2015) 109-133.
- 989 [53] A. Le Roy, K. Wang, B. Schaack, P. Schuck, C. Breyton, C. Ebel, AUC and Small-Angle
990 Scattering for Membrane Proteins, *Methods in enzymology*, 562 (2015) 257-286.
- 991 [54] J.B. Dacks, M.C. Field, R. Buick, L. Eme, S. Gribaldo, A.J. Roger, C. Brochier-Armanet, D.P.
992 Devos, The changing view of eukaryogenesis - fossils, cells, lineages and how they all come together,
993 *J Cell Sci*, 129 (2016) 3695-3703.
- 994 [55] B.I. Lee, C. Chang, S.J. Cho, S.H. Eom, K.K. Kim, Y.G. Yu, S.W. Suh, Crystal structure of the
995 MJ0490 gene product of the hyperthermophilic archaeobacterium *Methanococcus jannaschii*, a novel
996 member of the lactate/malate family of dehydrogenases, *Journal of molecular biology*, 307 (2001)
997 1351-1362.
- 998 [56] K. Niefind, H.J. Hecht, D. Schomburg, Crystal structure of L-2-hydroxyisocaproate dehydrogenase
999 from *Lactobacillus confusus* at 2.2 Å resolution. An example of strong asymmetry between subunits,
1000 *Journal of molecular biology*, 251 (1995) 256-281.
- 1001 [57] D. Some, H. Amartely, A. Tsadok, M. Lebendiker, Characterization of Proteins by Size-Exclusion
1002 Chromatography Coupled to Multi-Angle Light Scattering (SEC-MALS), *J Vis Exp*, (2019).
- 1003 [58] L.J. Yennaco, Y. Hu, J.F. Holden, Characterization of malate dehydrogenase from the
1004 hyperthermophilic archaeon *Pyrobaculum islandicum*, *Extremophiles : life under extreme conditions*,
1005 11 (2007) 741-746.
- 1006 [59] A.K. Rolstad, E. Howland, R. Sirevag, Malate dehydrogenase from the thermophilic green
1007 bacterium *Chloroflexus aurantiacus*: purification, molecular weight, amino acid composition, and
1008 partial amino acid sequence, *Journal of bacteriology*, 170 (1988) 2947-2953.
- 1009 [60] C. Charnock, U.H. Refseth, R. Sirevag, Malate dehydrogenase from *Chlorobium vibrioforme*,
1010 *Chlorobium tepidum*, and *Heliobacterium gestii*: purification, characterization, and investigation of
1011 dinucleotide binding by dehydrogenases by use of empirical methods of protein sequence analysis,
1012 *Journal of bacteriology*, 174 (1992) 1307-1313.

- 1013 [61] T. Oikawa, N. Yamamoto, K. Shimoke, S. Uesato, T. Ikeuchi, T. Fujioka, Purification,
1014 characterization, and overexpression of psychrophilic and thermolabile malate dehydrogenase of a
1015 novel antarctic psychrotolerant, *Flavobacterium frigidimaris* KUC-1, *Biosci Biotechnol Biochem*, 69
1016 (2005) 2146-2154.
- 1017 [62] D. Madern, The putative L-lactate dehydrogenase from *Methanococcus jannaschii* is an NADPH-
1018 dependent L-malate dehydrogenase, *Molecular microbiology*, 37 (2000) 1515-1520.
- 1019 [63] H. Thompson, A. Tersteegen, R.K. Thauer, R. Hedderich, Two malate dehydrogenases in
1020 *Methanobacterium thermoautotrophicum*, *Archives of microbiology*, 170 (1998) 38-42.
- 1021 [64] M. Nishiyama, J.J. Birktoft, T. Beppu, Alteration of coenzyme specificity of malate dehydrogenase
1022 from *Thermus flavus* by site-directed mutagenesis, *The Journal of biological chemistry*, 268 (1993)
1023 4656-4660.
- 1024 [65] I.K. Feil, H.P. Lerch, D. Schomburg, Deletion variants of L-hydroxyisocaproate dehydrogenase.
1025 Probing substrate specificity, *European journal of biochemistry / FEBS*, 223 (1994) 857-863.
- 1026 [66] Y. Masukagami, K.A. Tivendale, G.F. Browning, F.M. Sansom, Analysis of the *Mycoplasma bovis*
1027 lactate dehydrogenase reveals typical enzymatic activity despite the presence of an atypical catalytic
1028 site motif, *Microbiology (Reading)*, 164 (2018) 186-193.
- 1029 [67] J.D. Evans, S.A. Martin, Cloning of the L-lactate dehydrogenase gene from the ruminal bacterium
1030 *Selenomonas ruminantium* HD4, *Curr Microbiol*, 44 (2002) 155-160.
- 1031 [68] B.K. Ho, E.A. Coutsiaris, C. Seok, K.A. Dill, The flexibility in the proline ring couples to the protein
1032 backbone, *Protein Sci*, 14 (2005) 1011-1018.
- 1033 [69] N. Coquelle, R. Talon, D.H. Juers, E. Girard, R. Kahn, D. Madern, Gradual adaptive changes of a
1034 protein facing high salt concentrations, *Journal of molecular biology*, 404 (2010) 493-505.
- 1035 [70] R. Steuer, A.N. Nesi, A.R. Fernie, T. Gross, B. Blasius, J. Selbig, From structure to dynamics of
1036 metabolic pathways: application to the plant mitochondrial TCA cycle, *Bioinformatics*, 23 (2007) 1378-
1037 1385.
- 1038 [71] H. Innan, F. Kondrashov, The evolution of gene duplications: classifying and distinguishing
1039 between models, *Nat Rev Genet*, 11 (2010) 97-108.
- 1040 [72] S. Ohno, *Evolution by gene duplication*, Springer –Verlag, New-York, 1970.
- 1041 [73] B. Binay, R.B. Sessions, N.G. Karaguler, A double mutant of highly purified *Geobacillus*
1042 *stearothermophilus* lactate dehydrogenase recognises l-mandelic acid as a substrate, *Enzyme Microb*
1043 *Technol*, 52 (2013) 393-399.
- 1044 [74] D. Bur, T. Clarke, J.D. Friesen, M. Gold, K.W. Hart, J.J. Holbrook, J.B. Jones, M.A. Luyten, H.M.
1045 Wilks, On the effect on specificity of Thr246----Gly mutation in L-lactate dehydrogenase of *Bacillus*
1046 *stearothermophilus*, *Biochemical and biophysical research communications*, 161 (1989) 59-63.
- 1047 [75] T. Dams, R. Ostendorp, M. Ott, K. Rutkat, R. Jaenicke, Tetrameric and octameric lactate
1048 dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima*. Structure and stability of
1049 the two active forms, *European journal of biochemistry / FEBS*, 240 (1996) 274-279.
- 1050 [76] L.C. James, D.S. Tawfik, Conformational diversity and protein evolution--a 60-year-old hypothesis
1051 revisited, *Trends in biochemical sciences*, 28 (2003) 361-368.
- 1052 [77] A. Wellner, M. Raitsev Gurevich, D.S. Tawfik, Mechanisms of protein sequence divergence and
1053 incompatibility, *PLoS genetics*, 9 (2013) e1003665.

1054

Tree scale: 10





EUCARYA

LDH/MDH3/HicDH

MDH2

MDH1

Alveolata (61)
 Amoebozoa (8)
 Apusozoa (1)
 Choanoflagellata (2)
 Cryptophyta (1)
 Diplomonadida (3)
 Fungi (573)
 Haptophyceae (2)
 Heterolobosea (1)
 Ichthyosporia (2)
 Kinetoplastida (19)
 Metazoa (327)
 Nucleariidae (1)
 Parabasalia (2)
 Rhizaria (2)
 Rhodophyta (2)
 Stramenopiles (28)
 Viridiplantae (107)
 TOTAL_EUCARYA (1142)

BACTERIA

LDH/MDH3/HicDH

MDH2

MDH1

Acidobacteria (23)
 Actinobacteria (1230)
 Aquificae (17)
 Armatimonadetes (8)
 Bacteroidetes (809)
 Bacteriota (5)
 Calditerricae (3)
 Calditerricae (2)
 Candidate divisions (1088)
 Chlorobi (16)
 Chloroflexi (168)
 Chrysiogenetes (1)
 Coprothermobacteriota (1)
 Cyanobacteria (165)
 Deferribacteres (6)
 Deinococcus-Thermus (28)
 Dictyoglomi (2)
 Elusimicrobia (15)
 Fibrobacteres (11)
 Firmicutes (1478)
 Fusobacteria (20)
 Gemmatimonadetes (4)
 Haloplasmatales (1)
 Ignaribacteriae (32)
 Kiritimatiellaota (3)
 Lentisphaerae (11)
 Microgenomates group (34)
 Nitrospirae/Tectomicrobia (7)
 Nitrospirae (18)
 Parcubacteria group (161)
 Proteobacteria Alphaproteobacteria (1028)
 Proteobacteria Betaproteobacteria (498)
 Proteobacteria Deltaproteobacteria (145)
 Proteobacteria Epsilonproteobacteria (92)
 Proteobacteria Gammaproteobacteria (1007)
 Proteobacteria other (69)
 PVC Chlamydiae (23)
 PVC Planctomycetes (56)
 PVC Verrucomicrobia (38)
 Rhodothermaeota (3)
 Spirochaetes (35)
 Synergistetes (16)
 Tenericutes (103)
 Thermobaculum (1)
 Thermodesulfobacteria (5)
 Thermotogae (20)
 Unclassified bacteria (67)
 TOTAL_BACTERIA (8593)

ARCHAEA

LDH/MDH3/HicDH

MDH2

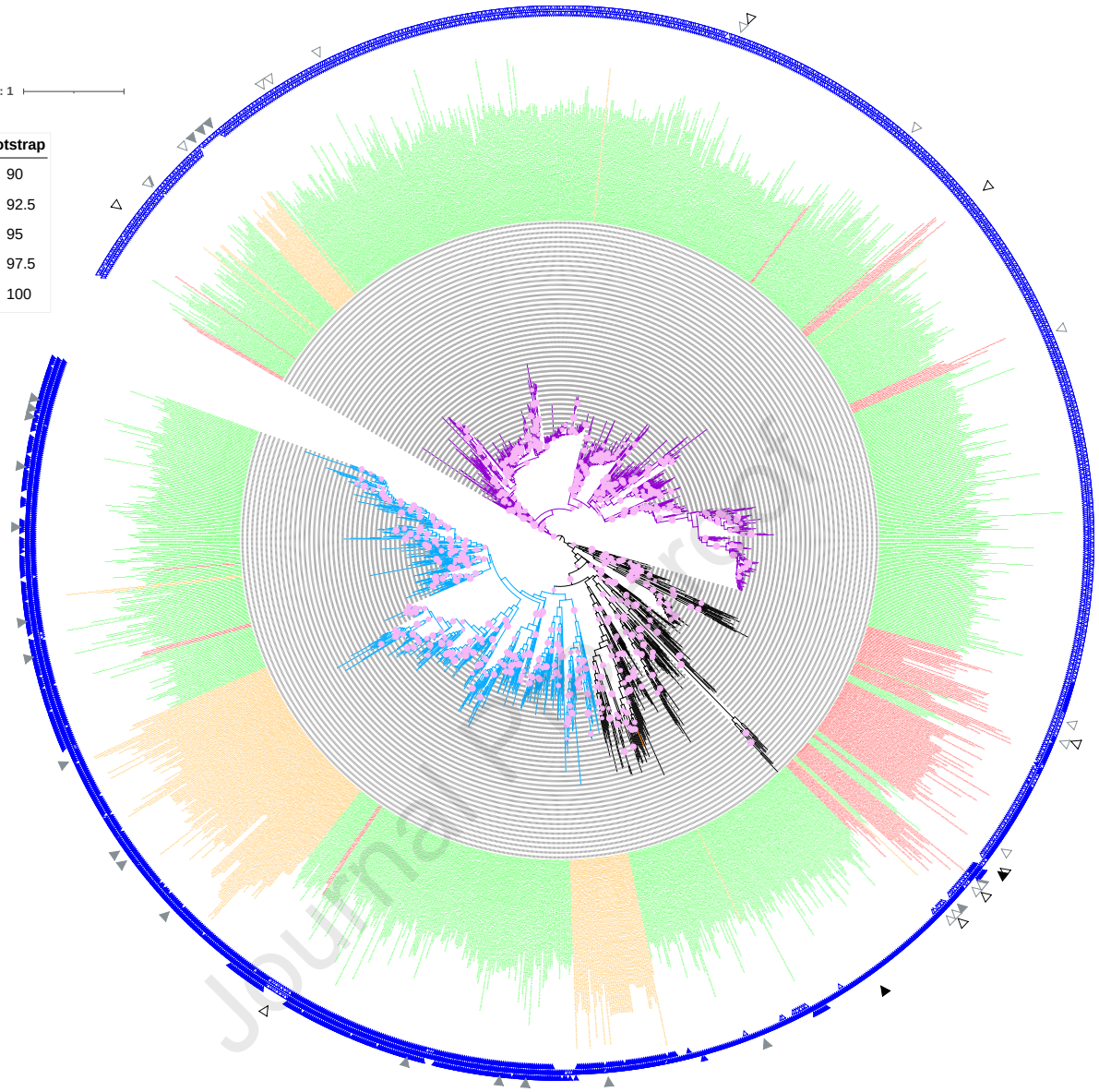
MDH1

Archaeoglobales (6)
 Asgard_group (11)
 Candidatus_Aenigmarchaeota (32)
 Candidatus_Aenigmarchaeota (5)
 Candidatus_Altarchaeales (12)
 Candidatus_Bathyarchaeota (29)
 Candidatus_Diapherotrites (7)
 Candidatus_Korarchaeota (1)
 Candidatus_Micrarchaeota (13)
 Candidatus_Nanoarchaeota (3)
 Candidatus_Pacearchaeota (4)
 Candidatus_Parvarchaeota (1)
 Candidatus_Woesearchaeota (3)
 Candidatus_Pacearchaeota (1)
 Diatorarchaea_DHVE2_group (1)
 Diatorarchaea_Marine_Group_II (1)
 Diatorarchaea_Methanomassiliococcales (6)
 Diatorarchaea_Thermoplasmatales (15)
 Hadesarchaea (5)
 Halobacteriales (97)
 Methanomada_Methanobacteriales (23)
 Methanomada_Methanococcales (6)
 Methanomada_Methanopyrales (1)
 Methanomicrobia_ANME-1_cluster (1)
 Methanomicrobia_Arc_1_group (3)
 Methanomicrobia_Methanocellales (2)
 Methanomicrobia_Methanomicrobiales (16)
 Methanomicrobia_Methanosarcinales (26)
 Methanonatronarchaeales (2)
 Nanoarchaeales (3)
 TACK_Acidilobales (3)
 TACK_Desulfurococcales (16)
 TACK_Ferriidococcales (3)
 TACK_Sulfolobales (9)
 TACK_Thaumarchaeota (18)
 TACK_Thermoproteales (12)
 Theionarchaea (2)
 Thermococcales (12)
 Unclassified_archaea (16)
 Uncultured_archaea (4)
 TOTAL_ARCHAEA (433)

Tree scale: 1

bootstrap

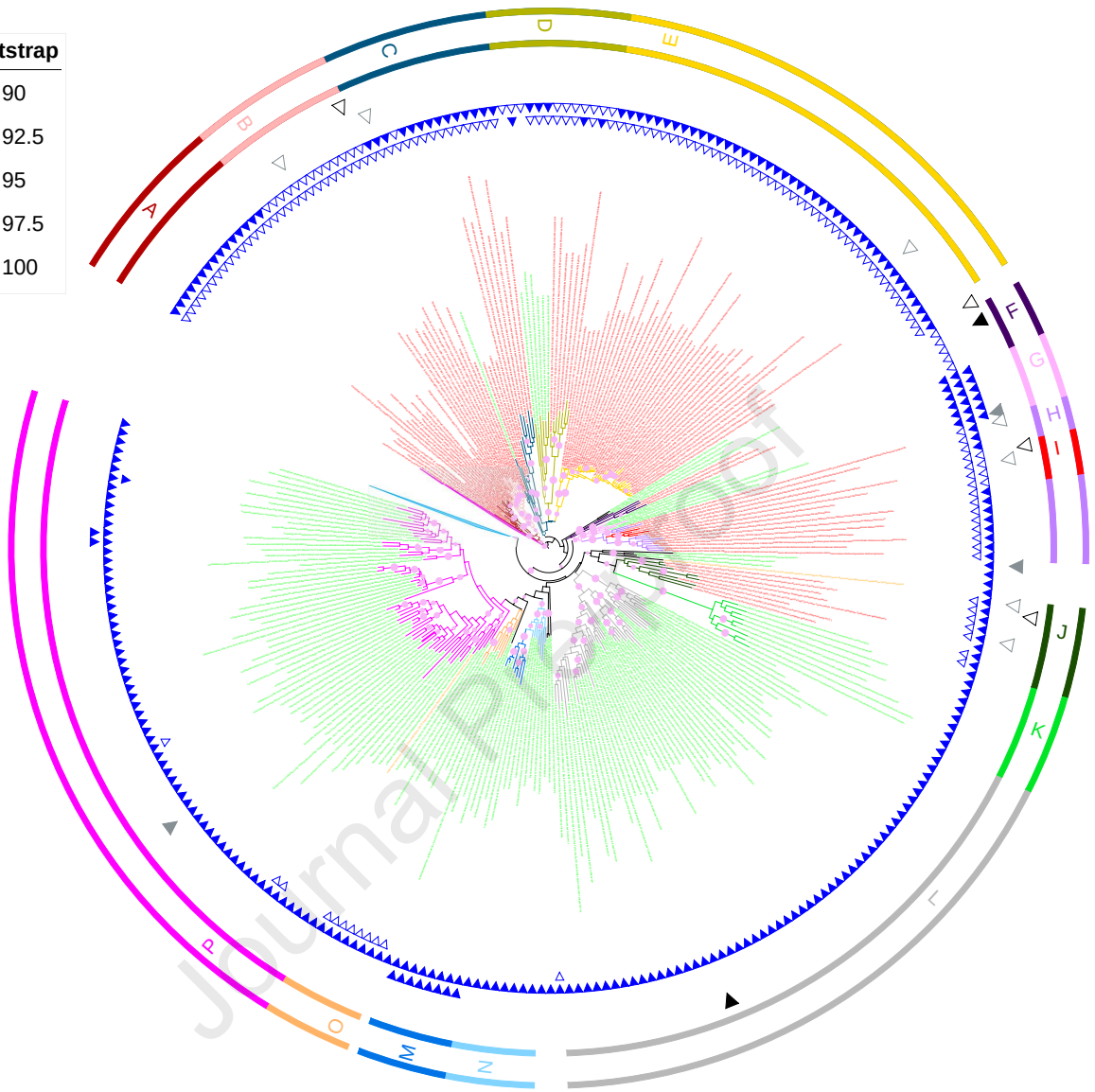
- 90
- 92.5
- 95
- 97.5
- 100

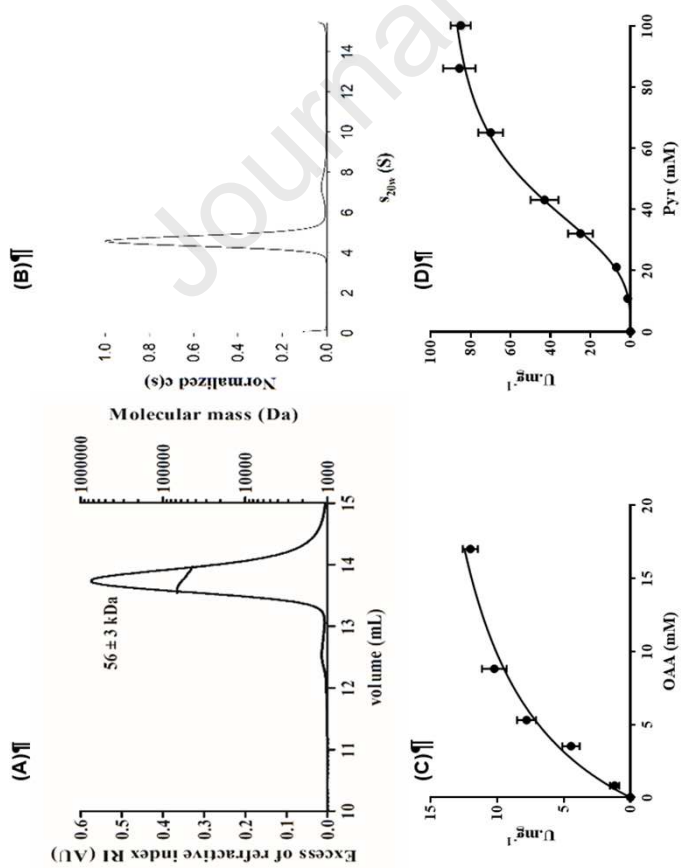


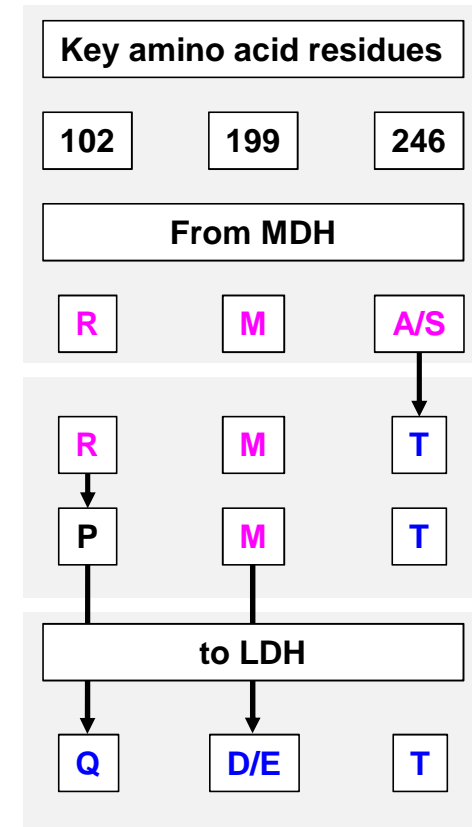
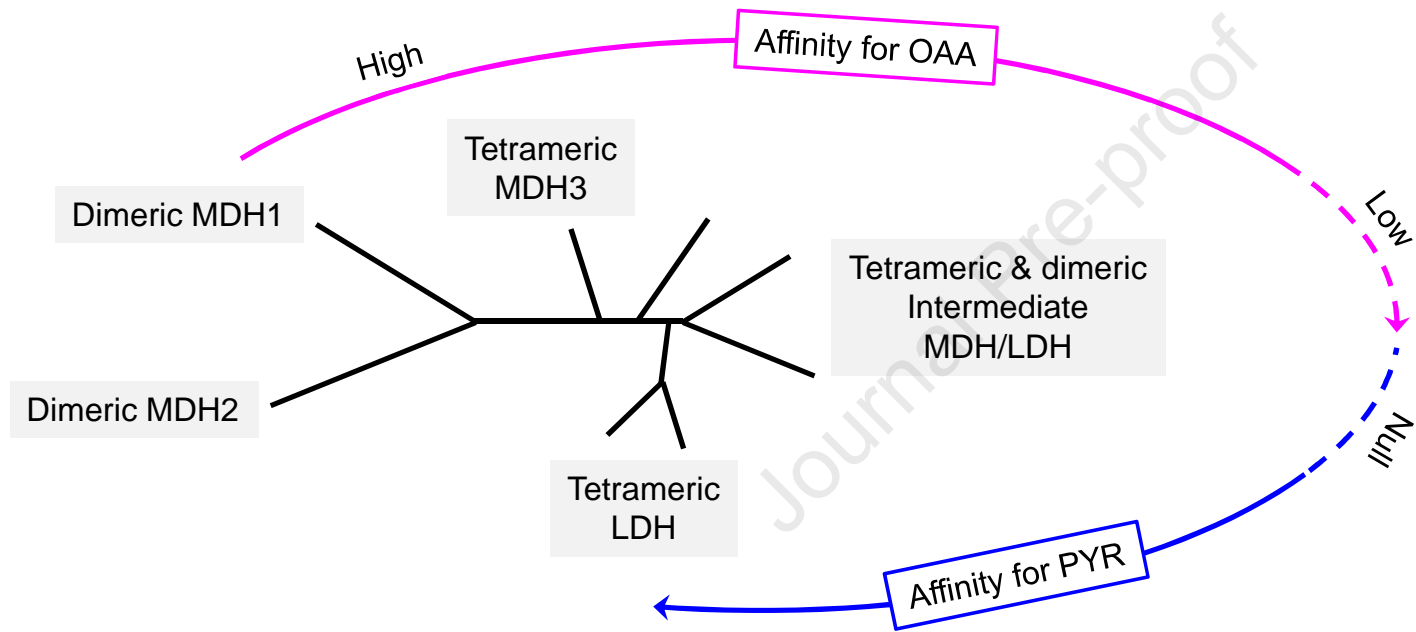
Tree scale: 1

bootstrap

- 90
- 92.5
- 95
- 97.5
- 100







Highlights

Phylogenetic analyses disentangle the relationships between malate dehydrogenases and lactate dehydrogenases.

The study reveals an intermediate group of enzymes that reflects an early and step-wise functional divergence between malate dehydrogenases and lactate dehydrogenases.

Neofunctionalization and subfunctionalization contribute to evolution of malate dehydrogenases and lactate dehydrogenases.

The work suggests that present-day enzymes such as in *Tolumonas auensis* are descendant of an ancient group of enzymes in which allostery evolved.

Journal Pre-proof

Declarations of interest: none

Journal Pre-proof