

This is a repository copy of *Adaptation of the carbamoyl-phosphate synthetase (CPS) enzyme in an extremophile fish.*

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/165956/>

Version: Accepted Version

Article:

White, Lewis, Sutton, Gemma, Shechonge, Asilatu et al. (3 more authors) (Accepted: 2020) Adaptation of the carbamoyl-phosphate synthetase (CPS) enzyme in an extremophile fish. Royal Society Open Science. ISSN 2054-5703 (In Press)

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

ROYAL SOCIETY OPEN SCIENCE

Adaptation of the carbamoyl-phosphate synthetase (CPS) enzyme in an extremophile fish

| | |
|-------------------------------|---|
| Journal: | <i>Royal Society Open Science</i> |
| Manuscript ID | RSOS-201200.R1 |
| Article Type: | Research |
| Date Submitted by the Author: | n/a |
| Complete List of Authors: | White, Lewis; University of York, Department of Biology Sutton, Gemma; University of York, Department of Biology Shechonge, Asilatu ; Tanzania Fisheries Research Institute Day, Julia; University College London, Biology Dasmahapatra, Kanchon; University of York, Department of Biology Pownall, Mary; University of York, Department of Biology |
| Subject: | evolution < BIOLOGY, developmental biology < BIOLOGY, biochemistry < CROSS-DISCIPLINARY SCIENCES |
| Keywords: | Urea, Carbamoyl-phosphate synthetase, Alcolapia, extremophile, Fish |
| Subject Category: | Organismal and Evolutionary Biology |
| | |

SCHOLARONE™
Manuscripts

Author-supplied statements

Relevant information will appear here if provided.

Ethics

Does your article include research that required ethical approval or permits?:

Yes

Statement (if applicable):

Research was approved by the University of York AWERB and under the Home Office licence for Dr ME Pownall (POF 245945).

Data

It is a condition of publication that data, code and materials supporting your paper are made publicly available. Does your paper present new data?:

Yes

Statement (if applicable):

Sequence data has been made available on NCBI (accession numbers: MT119353, MT119354) and in the supplementary files.

Conflict of interest

I/We declare we have no competing interests

Statement (if applicable):

CUST_STATE_CONFLICT :No data available.

Authors' contributions

This paper has multiple authors and our individual contributions were as below

Statement (if applicable):

Experiments were designed by LJW and MEP, work was carried out by LJW and GS and the manuscript written and edited by all authors. LJW, JD and AS collected the fish from Lake Natron.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 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

1 **Adaptation of the carbamoyl-phosphate synthetase (CPS) enzyme in an extremophile fish**

2 Lewis J. White¹, Gemma Sutton¹, Asilatu Shechonge², Julia J. Day³, Kanchon K. Dasmahapatra¹, Mary E. Pownall¹

6 ¹Biology Department, University of York, York, YO10 5DD, United Kingdom

7 ²Tanzania Fisheries Research Institute, PO BOX 98 Kyela, Mbeya, Tanzania

8 ³Department of Genetics, Evolution and Environment, University College London, Darwin Building, Gower Street, London, WC1E 6BT, United Kingdom

11 Author for communication: ljw569@york.ac.uk

13 **ABSTRACT**

14 **Tetrapods and fish have adapted distinct carbamoyl-phosphate synthase (CPS) enzymes to initiate the**
15 **ornithine urea cycle during the detoxification of nitrogenous wastes. We report evidence that in the**
16 **ureotelic subgenus of extremophile fish *Oreochromis Alcolapia*, CPS III has undergone convergent**
17 **evolution and adapted its substrate affinity to ammonia which is typical of terrestrial vertebrate CPS I.**
18 **Unusually, unlike in other vertebrates, the expression of CPS III in *Alcolapia* is localised to the skeletal**
19 **muscle and is activated in the myogenic lineage during early embryonic development with expression**
20 **remaining in mature fish. We propose that adaptation in *Alcolapia* included both convergent**
21 **evolution of CPS function to that of terrestrial vertebrates, as well as changes in development**
22 **mechanisms redirecting CPS III gene expression to the skeletal muscle.**

24 **INTRODUCTION**

25 In living organisms, protein metabolism results in the production of nitrogenous wastes which need to
26 be excreted. Most teleosts are ammonotelic, excreting their toxic nitrogenous waste as ammonia across
27 gill tissue by diffusion. As an adaptation to living on land, amphibians and mammals are ureotelic, using
28 liver and kidney tissues to convert waste ammonia into the less toxic and more water soluble urea,
29 which is then excreted in urine. Other terrestrial animals such as insects, birds and reptiles are uricolytic
30 and convert nitrogenous waste into uric acid, which is eliminated as a paste; a process which requires
31 more energy but wastes less water (1).

32 While most adult fish are ammonotelic, the larval stages of some teleosts excrete nitrogenous waste as
33 both ammonia and urea before their gills are fully developed (2). Additionally, some adult fish species
34 such as the gulf toad fish (*Opsanus beta*; (3)) and the African catfish (*Clarias gariepinus*; (4)) also excrete
35 a proportion of their nitrogenous waste as urea. This is usually in response to changes in aquatic

1
2
3 36 conditions, such as high alkalinity. It has been shown experimentally that high external pH prevents
4 37 diffusion of ammonia across gill tissue (5, 6). Unusually, the cichlid fish species in the subgenus *Alcolapia*
5 38 (described by some authors as a genus but shown to nest within the genus *Oreochromis*; (7) inhabit the
6 39 highly alkaline soda lakes of Natron (Tanzania) and Magadi (Kenya), are reported to be 100% ureotelic
7 40 (8, 9).

11
12 41 Once part of a single paleo-lake, Orolonga (10), Lakes Natron and Magadi are one of the most extreme
13 42 environments supporting fish life, with water temperatures up to 42.8 °C, pH ~10.5, fluctuating
14 43 dissolved oxygen levels, and salt concentrations above 20 parts per thousand (11). *Alcolapia* is the only
15 44 group of fish to survive in these lakes, forming a recent adaptive radiation including the four species:
16 45 *Alcolapia grahami* (Lake Magadi) and *A. latilabris*, *A. ndlalani* and *A. alcalica* (Lake Natron) (11, 12). The
17 46 harsh environment of the soda lakes presents certain physiological challenges that *Alcolapia* have
18 47 evolved to overcome, including the basic need to excrete nitrogenous waste. While other species are
19 48 able to excrete urea in response to extreme conditions, none do so to the level of *Alcolapia* (13, 14), and
20 49 unlike facultative ureotelic species, the adaptation of urea production and excretion in *Alcolapia* is
21 50 considered fixed (15). Moreover, the heightened metabolic rate in *Alcolapia*, a by-product from living in
22 51 such an extreme environment (8, 16), requires an efficient method of detoxification.

26 52 *Alcolapia* and ureotelic tetrapods (including humans) detoxify ammonia using the ornithine urea cycle
27 53 (OUC) where the mitochondrial enzyme carbamoyl-phosphate synthetase (CPS) is essential for the first
28 54 and rate limiting step of urea production (17). This enzyme, together with the accessory enzyme
29 55 glutamine synthase, provide an important switch regulating the balance between ammonia removal for
30 56 detoxification and maintaining a source of ammonia for the biosynthesis of amino acids (18). CPS has
31 57 evolved into two biochemically distinct proteins: in terrestrial vertebrates CPS I uses ammonia as its
32 58 preferential nitrogen donor, while in teleosts CPS III accepts glutamine to produce urea during larval
33 59 stages (reviewed Zimmer et al 2017). While CPS I / III are mitochondrial enzymes and part of the urea
34 60 cycle, CPS II is present in the cytosol catalyzing the synthesis of carbamoyl phosphate for pyrimidine
35 61 nucleotide biosynthesis. *CPS I / III* are syntenic, representing orthologous genes; their somewhat
36 62 confusing nomenclature is based on the distinct biochemical properties of their proteins. *CPS I / III* genes
37 63 from different vertebrate species clade together, separate from *CPS II* (supplementary figure). For
38 64 simplicity, we will continue to refer to fish, glutamine binding CPS as CPS III and tetrapod, ammonia
39 65 binding CPS as CPS I. The teleost CPS III binds glutamine in the glutamine amidotransferase (GAT)
40 66 domain using two amino acid residues (19), subsequently the nitrogen source provided by the amide
41 67 group is catalysed by a conserved catalytic triad; Cys-His-Glu (20). In terrestrial vertebrates CPS I lacks a
42 68 complete catalytic triad and can only generate carbamoyl-phosphate in the presence of free ammonia

1
2
3 69 (21). This change in function from glutamine binding CPS III to ammonia binding CPS I is believed to have
4
5 70 evolved in the stem lineage of living tetrapods, first appearing in ancestral amphibians (21).
6

7 71 In tetrapods and most fish, the OUC enzymes are largely localised to the liver (22), the main urogenic
8
9 72 organ (23). *Alcolapia* are different, and the primary site for urea production in these extremophile fishes
10
11 73 is the skeletal muscle (24). Notably, glutamine synthase activity is reportedly absent in *Alcolapia* muscle
12
13 74 tissue. The kinetic properties of CPS III in *Alcolapia* therefore differ from that of other teleosts in that it
14
15 75 preferentially uses ammonia as its primary substrate, having maximal enzymatic rates above that of
16
17 76 binding glutamine (although it is still capable of doing so) as opposed to in other species where use of
18
19 77 ammonia yields enzymatic rates of around 10% to that of glutamine (24). These rates are similar to
20
21 78 ureotelic terrestrial species, where CPS I preferentially binds ammonia and is incapable of using
22
23 79 glutamine (20).

24 80 Here, we report the amino acid sequence of two *Alcolapia* species (*A. alcalica* and *A. grahami*) that
25
26 81 reveals a change in CPS III substrate binding site. In addition, we show that the expression of *Alcolapia*
27
28 82 CPS III in skeletal muscle arises early in embryonic development where transcripts are restricted to the
29
30 83 somites, the source of skeletal muscle in all vertebrates, and migrating myogenic precursors. We discuss
31
32 84 changes to the structure of functional domains and modular gene enhancers that likely underpin
33
34 85 evolutionary changes in *Alcolapia* CPS III substrate binding and the redirection of gene expression from
35
36 86 the hepatogenic to myogenic lineage (25). Our findings point to adaptation in *Alcolapia* including both
37
38 87 convergent evolution of CPS function to that of terrestrial vertebrates, as well as changes in
39
40 88 development mechanisms redirecting CPS III gene expression to the skeletal muscle.

39 89 **Methods**

41 90 ***Experimental animals***

42
43 91 Fieldwork to Lake Natron, Tanzania, was conducted during June and July of 2017 to collect live
44
45 92 specimens of the three endemic species in an attempt to produce stable breeding populations of these
46
47 93 fishes in the UK. Live fish were all collected from a single spring (site 5 (11, 26)) containing all three
48
49 94 species found in Lake Natron and identified using morphology as described in Seegers and Tichy, 1999. A
50
51 95 stand alone, recirculating aquarium was adapted to house male and female *A. alcalica* in 10 or 30 L
52
53 96 tanks at a constant temperature of 30 °C, pH 9 and salt concentration of 3800 µS at the University of
54
55 97 York.

56 98 ***Expression of CPS III in adult tissues***

1
2
3 99 Reverse transcription polymerase chain reaction (RT-PCR) was used to determine the presence of CPS III
4
5 100 in different tissues (gill, muscle, liver, brain) of three different adult *A. alcalica*. RNA was extracted from
6
7 101 dissected tissues with TriReagent (Sigma-Aldrich) to the manufacturers' guidelines. For cDNA synthesis,
8
9 102 1µg of total RNA was reverse transcribed with random hexamers (Thermo Scientific) and superscript IV
10
11 103 (Invitrogen). PCR was performed on 2µl of the above cDNA with Promega PCR master mix and 0.5mM of
12
13 104 each primer (Forward: CAGTGGGAGGTCAGATTGC, Reverse: CTCACAGCGAAGCACAGGG). Gel
14
15 105 electrophoresis of the PCR products determined the presence or absence of CPS III RNA.

16 106 ***In situ hybridisation***

17
18 107 For the production of antisense probes, complementary to the mRNA of CPS III to use in *in situ*
19
20 108 hybridisation, the above 399bp PCR product was ligated into PGem-tEasy and transformed into the *E. coli*
21
22 109 strain DH5α. This was linearised and *in vitro* run off transcription was used to incorporate a DIG labelled
23
24 110 UTP analogue. To determine the temporal expression of these proteins in *A. alcalica*, embryos were
25
26 111 collected at different stages of development (2, 4 and 7 days post fertilisation [between 15 and 20 for
27
28 112 each stage]), fixed for 1 hour in MEMFA (0.1 M MOPS pH 7.4, 2 mM EGTA, 1 mM MgSO₄, 3.7%
29
30 113 formaldehyde) at room temperature and stored at -20°C in 100% methanol. For *in situ* hybridisation,
31
32 114 embryos were rehydrated and treated with 10 µg/mg proteinase K room temperature. After post
33
34 115 fixation and a 2 hour pre-hybridisation, embryos were hybridised with the probe at 68°C in in
35
36 116 hybridisation buffer (50% formamide (Ambion), 1mg/ml total yeast RNA, 5×SSC, 100µg/ml heparin, 1×
37
38 117 denharts, 0.1% Tween-20, 0.1% CHAPS, 10mM EDTA. Embryos were extensively washed at 68°C in 2×SSC
39
40 118 +0.1% Tween-20, 0.2×SSC +0.1% Tween-20 and maleic acid buffer (MAB; 100mM maleic acid, 150mM
41
42 119 NaCl, 0.1% Tween-20, pH7.8). This was replaced with pre-incubation buffer (4× MAB, 10% BMB, 20%
43
44 120 heat-treated lamb serum) for 2 hours. Embryos were incubated overnight (rolling at 4°C) with fresh pre-
45
46 121 incubation buffer and 1/2000 dilution of anti-DIG coupled with alkaline phosphatase (AP) (Roche). These
47
48 122 were then visualised by application of BM purple until staining had occurred.

49 123 ***Sequence analysis of CPS III***

50
51 124 cDNA was produced from the RNA extracted from whole embryos using the above method for *A. alcalica*
52
53 125 and *A. grahmi*. Multiple primer pairs (supplementary table 1) were used to amplify fragments of CPS III
54
55 126 from the cDNA via PCR and the products sent for sequencing. The coding region of CPS III was then
56
57 127 constructed using multiple alignments against the CPS I and III from other species. The amino acid
58
59 128 sequence was then examined for potential changes which could predict the functional differences seen
60
61 129 in *Alcolapia*. Phylogenetic analysis was also used to confirm the *Alcolapia* genes analysed here are CPS III
62
63 130 (supplementary figure 1). To determine potential changes in promoter region, a 3500bp section of
64
65 131 genome (accession number NCBI: MW014910) upstream of the transcriptional site start of CPS from *A.*

1
2
3 132 *alcalica* (unpublished genome), *Oreochromis niloticus* (Nile tilapia), *Xenopus tropicalis* (western clawed
4 133 frog) and *Danio rerio* (zebrafish), genomes were accessed on Ensembl, aligned and examined for binding
5 134 sites specific to the muscle transcription factor MyoD1 (E-boxes) which preferentially binds paired E-
6 135 boxes in the enhancer regions of myogenic genes with the consensus motif CAG(G/C)TG, as well as E-
7 136 boxes more broadly (CANNTG). The published genomes of *O. niloticus*, *D. rerio* and *X. tropicalis* were
8 137 accessed using Ensembl whereas the *Alcolapia* genome was constructed from whole genome
9 138 sequences.

15 139 **Results**

18 140 ***CPS III expression is activated early in the skeletal muscle lineage in A. alcalica***

20 141 Analysis of gene expression of CPS III in dissected tissues of three adult *A. alcalica* shows that transcripts
21 142 were only detected in adult muscle (**Figure 1A**). *In situ* hybridisation methods on *A. alcalica* embryos at
22 143 different stages were carried out to investigate whether this restricted muscle expression was
23 144 established during development (**Figure 1B-F**). Blue colouration indicates hybridisation of the
24 145 complementary RNA probe and shows strongest expression in the developing somites along the body
25 146 axis (**arrows**). Expression was also detected in migratory muscle precursors (MMP; **arrowheads**), which
26 147 go on to form the body wall and limb musculature, and in the developing pectoral fin buds (**white**
27 148 **arrows**). All regions of the embryo that show expression of CPS III are in the muscle lineage indicating
28 149 that in *A. alcalica* CPS III expression is restricted to muscle tissues in both adults and the developing
29 150 embryo.

31 151 Many muscle specific genes are activated during development by the muscle specific transcription
32 152 factor, MyoD. The promoter region of CPS III (3.5kb upstream of the transcriptional start site) in *A.*
33 153 *alcalica* was compared to that in *O. niloticus*, *X. tropicalis* and *D. rerio* (**Figure 2**). Examination of this
34 154 region revealed a putative paired E-box MyoD binding site 940 to 970 bases
35 155 (CAGGTGACTGTGATTATATAGTTCACAGGTG) upstream of the transcriptional start site of CPS III only in
36 156 *Alcolapia* species. Intriguingly, while no pair of MyoD E-boxes were found in the upstream region of any
37 157 other species examined, *O. niloticus* does have a single MyoD E-box motif in the same region upstream
38 158 of CPS III, and within 19 bases of this is a CAGGTT motif which a single point mutation would convert
39 159 into a pair of E-boxes (CAGGTGACTGTGATTATATAGTTCACAGGTT). This suggests that it is possible that
40 160 MyoD could bind and activate transcription of *CPS III* in the muscle of *Alcolapia* species, but not in the
41 161 closely related *O. niloticus*.

57 162 ***Convergent evolution in adaptive function of CPS III***

1
2
3 163 Sequence analysis of *A. alcalica* and *A. grahami* CPS III revealed a discrepancy in the catalytic triad
4 compared to the published sequence for CPS III in *A. grahami* (accession number NCBI: AF119250). The
5 164
6 165 coding region for *A. alcalica* and *A. grahami* was cloned and sequenced (accession numbers NCBI:
7
8 166 MT119353, MT119354). Our data confirmed the error in the published sequence of *A. grahami* CPS III
9
10 167 and shows *Alcolapia* species maintain a catalytic triad essential for catalysing the breakdown of
11 168 glutamine (**red boxes in Figure 3**). However, similar to terrestrial vertebrate CPS I which lack either one
12
13 169 but usually both residues essential for binding glutamine for utilisation by the catalytic triad
14
15 170 (**arrowheads in Figure 3**), *Alcolapia* also lack one of these residues (**asterisks Figure 3**). This amino acid
16
17 171 sequence is consistent with a change in function permitting *Alcolapia* CPS III to bind and catalyse
18 172 ammonia directly, an activity usually restricted to terrestrial vertebrate CPS I, as elucidated by extensive
19
20 173 previous biochemical analyses (20, 21).

22 174 **Discussion**

23
24 175 While most teleosts are ammonotelic, larval fish can convert ammonia to urea for excretion and to do so
25 176 express the genes coding for the enzymes of the OUC, including CPS III (27). Later these genes are
26
27 177 silenced in most fish. In the rare cases where urea is produced in adult fish, the OUC enzymes are
28
29 178 expressed in the liver (23), however there are some reports of expression in non-hepatic tissues (28, 29).
30
31 179 We report here the expression of CPS III in the muscle of adult *A. alcalica*, which is consistent with the
32
33 180 detection of CPS III protein and enzyme activity in muscle of *A. grahami* (24). We also find conserved
34
35 181 changes to the amino acid sequence which explains the convergent evolution of *A. alcalica* and *A.*
36 182 *grahami* CPS III function with CPS I in terrestrial vertebrates. This conserved change in both *Alcolapia*
37
38 183 species suggests that the adaptations in the OUC are likely to have evolved in the ancestral species
39
40 184 inhabiting paleolake Orolongo during the period of changing aquatic conditions (over the past ten
41
42 185 thousand years) that led to the extreme conditions currently found in Lakes Natron and Magadi.

43 44 186 **Activation of CPS III in the myogenic lineage**

45
46 187 We find that the expression of CPS III is activated in somites and in migratory muscle precursors that will
47
48 188 form body wall and limb musculature (indeed expression is seen in developing limb buds). All skeletal
49
50 189 muscle in the vertebrate body is derived from the somites, and these CPS III expression patterns are
51
52 190 similar to those of muscle specific genes like myosin, actins and troponins (30-32).

53 191 Muscle specific expression of CPS III in *A. alcalica* embryos is a remarkable finding as most ureotelic
54
55 192 species convert nitrogenous waste to urea in the liver (8, 20). The expression of CPS III, the first enzyme
56
57 193 in the OUC, in muscle tissue is likely significant for supporting the high catabolism in a fish species with
58
59 194 the highest recorded metabolic rate (33). There are few reports of some OUC gene expression or
60

1
2
3 195 enzyme activity in non-hepatic tissue including muscle (28, 29, 34), nonetheless other fish species only
4
5 196 evoke the activity of the OUC when exposed to high external pH or during larval stages (13, 14, 35, 36),
6
7 197 and even then, urea production is never to the high level of activity occurring in *Alcolapia* (24). There is
8
9 198 some heterogeneity of the expression patterns of CPS III during the development of different species in
10
11 199 the teleost lineage; for example *D. rerio* has reported expression in the body (37), *Oncorhynchus mykiss*
12
13 200 (rainbow trout) shows expression in the developing body but not in hepatic tissue (38) and *C. gariepinus*
14
15 201 (African catfish) had CPS III expression detected in the dissected muscle from larvae (4). In laboratory
16
17 202 conditions, adult *A. alcalica* excrete approximately 75% of their nitrogenous waste as urea, compared to
18
19 203 only 10% in adult zebrafish (White et al., in prep). The early and sustained expression of CPS III in the
20
21 204 muscle lineage is at this point an observation unique to *Alcolapia*.

22
23 205 Skeletal muscle specific gene expression is activated in cells of the myogenic lineage by a family of bHLH
24
25 206 transcription factors, including MyoD (30). MyoD binds specifically at paired E-boxes in the enhancers of
26
27 207 myogenic genes with a preference for the consensus motif of CAG(G/C)TG (39, 40). MyoD is known to
28
29 208 require the cooperative binding at two E-boxes in close proximity, to modulate transcription of
30
31 209 myogenic genes (41). The presence of a pair of E-boxes in *Alcolapia*, upstream of a gene which has
32
33 210 switched to muscle specific expression, is suggestive that MyoD is driving expression early in
34
35 211 development. Enhancer modularity is a known mechanism for selectable variation (42) and although a
36
37 212 single MyoD binding site does not define an enhancer, MyoD is known to interact with pioneer factors
38
39 213 and histone deacetylases to open chromatin and activate gene transcription in the muscle lineage (40,
40
41 214 43). Experimental analysis to determine the activity of any regulatory sequences upstream of OUC genes
42
43 215 in different species would shed light on the significance of putative transcription factor binding sites.
44
45 216 This approach could also address another intriguing question as to the elements that drive the post-
46
47 217 larval silencing of OUC genes in most fish species (37), an area with only minimal research especially
48
49 218 when compared to the well characterised promoter region in mammalian species, for instance
50
51 219 Christoffels et al 1998. A further instance of an extremophile organism redirecting expression of a
52
53 220 hepatic enzyme to muscle tissue occurs in the crucian carp (45). Under conditions of anoxia this species
54
55 221 switches to anaerobic metabolism, producing ethanol as the end product of glycolysis (46, 47). This is
56
57 222 associated with the expression of *alcohol dehydrogenase* in muscle (48). Together with our findings, this
58
59 223 potentially reveals an example of convergent evolution whereby the muscle becomes the site for
60
224 detoxifying by-products of metabolism. Elucidating any mechanisms that may include modular
225
226 enhancers that facilitate the adaptation of gene regulation in response to changing environmental
227
228 conditions will be of significant interest.

227 ***Convergent evolution of adaptive CPS III function***

1
2
3 228 CPS proteins catalyse the production of carbamoyl-phosphate as a first step in nitrogen detoxification by
4
5 229 accepting either glutamine or ammonia as a nitrogen donor (17). Teleost CPS III binds glutamine: the
6
7 230 nitrogen source provided by the amide group of glutamine is catalyzed by the conserved catalytic triad
8
9 231 Cys-His-Glu in the glutamine amidotransferase (GAT) domain in the amino terminal part of CPS (20). In
10
11 232 terrestrial vertebrates, CPS I lacks the catalytic cysteine residue and only generates carbamoyl-
12
13 233 phosphate in the presence of free ammonia (21). Although CPS in *Alcolapia* shares most sequence
14
15 234 identity with fish CPS III (**Figure 3**), its ammonia binding activity is more similar in function to terrestrial
16
17 235 vertebrate CPS I (20, 24). This adaptation to preferentially bind ammonia over glutamine supports
18
19 236 efficient waste management in a fish with an exceptionally high metabolic rate (33). CPS I in terrestrial
20
21 237 vertebrates have amino acid changes in the catalytic triad which explains their binding ammonia over
22
23 238 glutamine; a reduction in glutamine binding capacity drives the use of ammonia (21). Here we show that
24
25 239 *Alcolapia* maintain the catalytic triad, but (similar to mouse and human) lack one of the two residues
26
27 240 required for efficient glutamine binding, weakening its affinity to glutamine, and driving the use of
28
29 241 ammonia as a primary substrate.

30
31 242 The interesting observation that bullfrog (*Rana catesbeiana*) CPS I retains the catalytic triad, but lacks
32
33 243 the two additional conserved amino acids required for glutamine binding, has led to the suggestion that
34
35 244 the change from preferential glutamine to ammonia binding originally evolved in the early tetrapod
36
37 245 lineage (21). A further frog species, the tree frog *Litoria caerulea*, retains its catalytic triad and only one of
38
39 246 the two residues required for glutamine binding has been altered, weakening its affinity for glutamine
40
41 247 and allowing for direct catabolism of ammonia (49). Much the same as in *Alcolapia*, *L. caerulea* CPS I is
42
43 248 still capable of using glutamine to some extent which lends further support to the notion that the
44
45 249 evolutionary transition from CPS III to CPS I occurring in amphibians and the early tetrapod lineage. The
46
47 250 changes in protein sequence of *Alcolapia* CPS III represents a convergent evolution in this extremophile
48
49 251 fish species, with acquired changes in functionally important domains which likely also evolved in early
50
51 252 terrestrial vertebrate CPS I.

52 253 **Conclusions**

53
54 254 *Alcolapia* have acquired multiple adaptations that allow continued excretion of nitrogenous waste in a
55
56 255 high pH environment. Among these is the novel expression of CPS in skeletal muscle, as well as
57
58 256 acquisition of mutations that change its function. Sequence evidence indicates that like terrestrial
59
60 257 vertebrates, and unique among fish, *Alcolapia* CPS III is capable of binding and catalysing the breakdown
61
62 258 of ammonia to carbamoyl-phosphate; a convergent evolution of CPS function. The mechanism by which
63
64 259 the novel and unique expression of CPS in muscle evolved is likely a function of enhancer regions of *A.*
65
66 260 *alcalica* and *A. grahami* that result in its regulation by muscle regulatory factors to direct CPS expression

261 in the myogenic lineage during embryonic development. Environmentally driven adaptations have
 262 resulted in changes in both the expression and activity of CPS III in *Alcolapia* that underpin its ability to
 263 turnover nitrogenous waste in a challenging environment while maintaining a high metabolic rate.

264 **Acknowledgments**

265 Experiments were designed by LJW and MEP, work was carried out by LJW and GS and the manuscript
 266 written and edited by LJW, GS, AS, JD, KKD & MEP. LJW, JD and AS collected the fish from Lake Natron.
 267 The authors are grateful to Antonia Ford and George Turner for their help in collecting the fish from
 268 Lake Natron (COSTECH permit no. 2017-259-NA-2011-182) and their continued support in this project.

269 **Funding**

270 This work was supported by the BBSRC as a studentship to LJW (BB/M011151/1). Additional support for
 271 fieldwork was provided by the Fisheries Society of the British Isles (small research grant) and the
 272 Genetics Society (Heredity fieldwork grant).

273 **Figure 1:** Expression analysis of carbamoyl-phosphate synthetase III (CPS III) from adult tissues and
 274 developing embryos of *Alcolapia alcalica*. A) Reverse transcriptase PCR and gel electrophoresis showing
 275 the muscle specific expression of CPS III, EF1 α shown as normalisation control. B-F) Lateral (C and E) and
 276 dorsal (B, D, and F) views of *in situ* hybridisation for CPS III in developing *A. alcalica* embryos at different
 277 stages (number of days post fertilisation [dpf] indicated). The blue colour indicates the detection of
 278 mRNA. The black/brown is endogenous pigment apparent in the retina and the chromatophores. Black
 279 arrows show somites, arrowheads indicate region of migrating muscle progenitors (MMP), white arrows
 280 show facial muscle (FM) and white arrowheads indicate developing pectoral fin bud bud (PFB). Black
 281 dots around the yolk and on the body are chromatophores (pigment cells).

282 **Figure 2:** Sequence comparison of a 3500bp region upstream of exon 1 of CPS III. Presence of generic E-
 283 box sites with the consensus motif of CANNTG are annotated by black bars and MyoD preferential E-box
 284 motifs of CAG(G/C)TG are represented by blue bars. Insert shows sequence of *Alcolapia alcalica*
 285 potential paired E-box, MyoD enhancer (Blue nucleotides) compared to *Oreochromis niloticus* which has
 286 a single MyoD E-box upstream of a 'presumptive' MyoD E-box (Red nucleotides).

287 **Figure 3:** Multiple amino acid alignment of residues 278 to 397 (aligned to *Alcolapia alcalica*) of
 288 carbamoyl-phosphate synthetase I and III from a number of tetrapod and teleost species respectively.
 289 Conserved amino acids are shaded in black, amino acids in the catalytic triad are boxed in red and
 290 arrowheads indicate residues vital for glutamine utilisation of the catalytic triad. The blue asterisk
 291 indicates the divergent glutamine binding residue in *Alcolapia* species that likely results in a functional
 292 change (inability to bind glutamine).

293 **References**

- 294 1. Wright PA. Nitrogen excretion: three end products, many physiological roles. *Journal of*
 295 *Experimental Biology*. 1995;198(2):273-81.
- 296 2. Zimmer AM, Wright PA, Wood CM. Ammonia and urea handling by early life stages of fishes.
 297 *Journal of Experimental Biology*. 2017;220(21):3843-55.

- 1
2
3 298 3. Barimo JF, Steele SL, Wright PA, Walsh PJ. Dogmas and controversies in the handling of
4
5 299 nitrogenous wastes: ureotely and ammonia tolerance in early life stages of the gulf toadfish, *Opsanus*
6
7 300 *beta*. Journal of Experimental Biology. 2004;207(12):2011-20.
- 8
9 301 4. Terjesen BF, Chadwick TD, Verreth JA, Rønnestad I, Wright PA. Pathways for urea production
10
11 302 during early life of an air-breathing teleost, the African catfish *Clarias gariepinus* Burchell. Journal of
12
13 303 Experimental Biology. 2001;204(12):2155-65.
- 14
15 304 5. Handy R, Poxton M. Nitrogen pollution in mariculture: toxicity and excretion of nitrogenous
16
17 305 compounds by marine fish. Reviews in Fish Biology and Fisheries. 1993;3(3):205-41.
- 18
19 306 6. Wright PA, Wood C. An analysis of branchial ammonia excretion in the freshwater rainbow
20
21 307 trout: effects of environmental pH change and sodium uptake blockade. Journal of Experimental
22
23 308 Biology. 1985;114(1):329-53.
- 24
25 309 7. Ford AGP, Bullen TR, Pang L, Genner MJ, Bills R, Flouri T, et al. Molecular phylogeny of
26
27 310 *Oreochromis* (Cichlidae: Oreochromini) reveals mito-nuclear discordance and multiple colonisation of
28
29 311 adverse aquatic environments. Molecular Phylogenetics and Evolution. 2019;136:215-26.
- 30
31 312 8. Randall D, Wood C, Perry S, Bergman H, Maloiy G, Mommsen T, et al. Urea excretion as a
32
33 313 strategy for survival in a fish living in a very alkaline environment. Nature. 1989;337:165-6.
- 34
35 314 9. Wilson PJ, Wood CM, Walsh PJ, Bergman AN, Bergman HL, Laurent P, et al. Discordance
36
37 315 between genetic structure and morphological, ecological, and physiological adaptation in Lake Magadi
38
39 316 tilapia. Physiological and Biochemical Zoology. 2004;77(4):537-55.
- 40
41 317 10. Roberts N, Taieb M, Barker P, Damnati B, Icole M, Williamson D. Timing of the Younger Dryas
42
43 318 event in East Africa from lake-level changes. Nature. 1993;366(6451):146-8.
- 44
45 319 11. Ford AGP, Dasmahapatra KK, Rüber L, Gharbi K, Cezard T, Day JJ. High levels of interspecific gene
46
47 320 flow in an endemic cichlid fish adaptive radiation from an extreme lake environment. Molecular Ecology.
48
49 321 2015;24(13):3421-40.
- 50
51 322 12. Seegers L, Tichy H. The *Oreochromis alcalicus* flock (Teleostei: Cichlidae) from lakes Natron and
52
53 323 Magadi, Tanzania and Kenya, with descriptions of two new species. Ichthyological Exploration of
54
55 324 Freshwaters. 1999;10:97-146.
- 56
57 325 13. Walsh P, Danulat E, Mommsen T. Variation in urea excretion in the gulf toadfish *Opsanus beta*.
58
59 326 Marine Biology. 1990;106(3):323-8.
- 60
327 14. Wright PA, Iwama GK, Wood CM. Ammonia and urea excretion in Lahontan cutthroat trout
328
329 328 (*Oncorhynchus clarki henshawi*) adapted to the highly alkaline Pyramid Lake (pH 9.4). Journal of
Experimental Biology. 1993;175(1):153-72.

- 1
2
3 330 15. Wood CM, Wilson P, Bergman HL, Bergman AN, Laurent P, Otiang'a-Owiti G, et al. Obligatory
4 331 urea production and the cost of living in the Magadi tilapia revealed by acclimation to reduced salinity
5 332 and alkalinity. *Physiological Biochemical Zoology*. 2002;75(2):111-22.
- 6
7
8 333 16. Kavembe GD, Franchini P, Irisarri I, Machado-Schiaffino G, Meyer A. Genomics of adaptation to
9 334 multiple concurrent stresses: insights from comparative transcriptomics of a Cichlid fish from one of
10 335 earth's most extreme environments, the Hypersaline Soda Lake Magadi in Kenya, East Africa. *Journal of*
11 336 *Molecular Evolution*. 2015;81(3-4):90-109.
- 12
13
14 337 17. Rubio V, Cervera J. The carbamoyl-phosphate synthase family and carbamate kinase: structure-
15 338 function studies. Portland Press Ltd.; 1995.
- 16
17
18 339 18. de Cima S, Polo LM, Díez-Fernández C, Martínez AI, Cervera J, Fita I, et al. Structure of human
19 340 carbamoyl phosphate synthetase: deciphering the on/off switch of human ureagenesis. *Scientific*
20 341 *Reports*. 2015;5:16950.
- 21
22
23 342 19. Hong J, Salo WL, Lusty CJ, Anderson PM. Carbamyl phosphate synthetase III, an evolutionary
24 343 intermediate in the transition between glutamine-dependent and ammonia-dependent carbamyl
25 344 phosphate synthetases. *Molecular Biology*. 1994;243:131-40.
- 26
27
28 345 20. Loong A, Chng Y, Chew S, Wong W, Ip Y. Molecular characterization and mRNA expression of
29 346 carbamoyl phosphate synthetase III in the liver of the African lungfish, *Protopterus annectens*, during
30 347 aestivation or exposure to ammonia. *Journal of Comparative Physiology B*. 2012;182(3):367-79.
- 31
32
33 348 21. Saeed-Kothe A, Powers-Lee SG. Specificity determining residues in ammonia- and glutamine-
34 349 dependent carbamoyl phosphate synthetases. *Journal of Biological Chemistry*. 2002;277(9):7231-8.
- 35
36
37 350 22. Ip AY, Chew SF. Ammonia production, excretion, toxicity, and defense in fish: a review. *Frontiers*
38 351 *in Physiology*. 2010;1:134.
- 39
40
41 352 23. Mommsen TP, Walsh PJ. Evolution of urea synthesis in vertebrates: the piscine connection.
42 353 *Science*. 1989;243(4887):72-5.
- 43
44
45 354 24. Lindley TE, Scheiderer CL, Walsh PJ, Wood CM, Bergman HL, Bergman AL, et al. Muscle as the
46 355 primary site of urea cycle enzyme activity in an alkaline lake-adapted tilapia, *Oreochromis alcalicus*
47 356 *grahami*. *Journal of Biological Chemistry*. 1999;274(42):29858-61.
- 48
49
50 357 25. Rebeiz M, Tsiantis M. Enhancer evolution and the origins of morphological novelty. *Current*
51 358 *Opinion in Genetics & Development*. 2017;45:115-23.
- 52
53
54 359 26. Ford AGP, Rüber L, Newton J, Dasmahapatra KK, Balarin JD, Bruun K, et al. Niche divergence
55 360 facilitated by fine-scale ecological partitioning in a recent cichlid fish adaptive radiation. *Evolution*.
56 361 2016;70(12):2718-35.
- 57
58
59
60

- 1
2
3 362 27. Zimmer AM, Wood CM. Physiological and molecular ontogeny of branchial and extra-branchial
4 363 urea excretion in posthatch rainbow trout (*Oncorhynchus mykiss*). American Journal of Physiology-
5 364 Regulatory, Integrative and Comparative Physiology. 2016;310(3):R305-R12.
6
7
8 365 28. Korte JJ, Salo WL, Cabrera VM, Wright PA, Felskie AK, Anderson PM. Expression of carbamoyl-
9 366 phosphate synthetase III mRNA during the early stages of development and in muscle of adult rainbow
10 367 trout (*Oncorhynchus mykiss*). Journal of Biological Chemistry. 1997;272(10):6270-7.
11
12
13 368 29. Banerjee B, Koner D, Hasan R, Saha N. Molecular characterization and ornithine-urea cycle
14 369 genes expression in air-breathing magur catfish (*Clarias magur*) during exposure to high external
15 370 ammonia. Genomics. 2020;112:2247-60.
16
17
18 371 30. Pownall ME, Gustafsson MK, Emerson Jr CP. Myogenic regulatory factors and the specification of
19 372 muscle progenitors in vertebrate embryos. Annual review of Cell and Developmental Biology.
20 373 2002;18(1):747-83.
21
22
23 374 31. Talbot JC, Teets EM, Ratnayake D, Duy PQ, Currie PD, Amacher SL. Muscle precursor cell
24 375 movements in zebrafish are dynamic and require Six family genes. Development.
25 376 2019;146(10):dev171421.
26
27
28 377 32. Bucher EA, Maisonpierre P, Konieczny S, Emerson C. Expression of the troponin complex genes:
29 378 transcriptional coactivation during myoblast differentiation and independent control in heart and
30 379 skeletal muscles. Molecular and Cellular Biology. 1988;8(10):4134-42.
31
32
33 380 33. Wood CM, Brix KV, De Boeck G, Bergman HL, Bianchini A, Bianchini LF, et al. Mammalian
34 381 metabolic rates in the hottest fish on earth. Scientific Reports. 2016;6:26990.
35
36
37 382 34. Kong H, Edberg DD, Korte JJ, Salo WL, Wright PA, Anderson PM. Nitrogen excretion and
38 383 expression of carbamoyl-phosphate synthetase III activity and mRNA in extrahepatic tissues of
39 384 largemouth bass (*Micropterus salmoides*). Archives of Biochemistry and Biophysics. 1998;350(2):157-68.
40
41
42 385 35. Randall DJ, Tsui T. Ammonia toxicity in fish. Marine Pollution Bulletin. 2002;45(1-12):17-23.
43
44
45 386 36. Mommsen TP, Walsh PJ. Urea synthesis in fishes: evolutionary and biochemical perspectives:
46 387 Elsevier; 1991. 137-63 p.
47
48
49 388 37. LeMoine CM, Walsh PJ. Ontogeny of ornithine-urea cycle gene expression in zebrafish (*Danio*
50 389 *rerio*). American Journal of Physiology-Regulatory, Integrative and Comparative Physiology.
51 390 2013;304(11):R991-R1000.
52
53
54 391 38. Steele SL, Chadwick TD, Wright PA. Ammonia detoxification and localization of urea cycle
55 392 enzyme activity in embryos of the rainbow trout (*Oncorhynchus mykiss*) in relation to early tolerance to
56 393 high environmental ammonia levels. Journal of Experimental Biology. 2001;204(12):2145-54.
57
58
59 394 39. Fong AP, Yao Z, Zhong JW, Cao Y, Ruzzo WL, Gentleman RC, et al. Genetic and epigenetic
60 395 determinants of neurogenesis and myogenesis. Developmental Cell. 2012;22(4):721-35.

- 1
2
3 396 40. Conerly ML, Yao Z, Zhong JW, Groudine M, Tapscott SJ. Distinct activities of Myf5 and MyoD
4 indicate separate roles in skeletal muscle lineage specification and differentiation. *Developmental Cell*.
5 397
6 398 2016;36(4):375-85.
- 8 399 41. Weintraub H, Davis R, Lockshon D, Lassar A. MyoD binds cooperatively to two sites in a target
9 enhancer sequence: occupancy of two sites is required for activation. *Proceedings of the National*
10 400
11 401 *Academy of Sciences*. 1990;87(15):5623-7.
- 13 402 42. Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jónsson B, et al. Genetic and
14 developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*.
15 403
16 404 2004;428(6984):717.
- 18 405 43. Berkes CA, Bergstrom DA, Penn BH, Seaver KJ, Knoepfler PS, Tapscott SJ. Pbx marks genes for
19 activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential.
20 406
21 407 *Molecular Cell*. 2004;14(4):465-77.
- 23 408 44. Christoffels VM, Grange T, Kaestner KH, Cole TJ, Darlington GJ, Croniger CM, et al.
24 Glucocorticoid receptor, C/EBP, HNF3, and protein kinase A coordinately activate the glucocorticoid
25 409
26 410 response unit of the carbamoylphosphate synthetase I gene. *Molecular and Cellular Biology*.
27
28 411 1998;18(11):6305-15.
- 30 412 45. Fagernes CE, Stensløkken K-O, Røhr ÅK, Berenbrink M, Ellefsen S, Nilsson GE. Extreme anoxia
31 tolerance in crucian carp and goldfish through neofunctionalization of duplicated genes creating a new
32 413
33 414 ethanol-producing pyruvate decarboxylase pathway. *Scientific Reports*. 2017;7(1):7884.
- 35 415 46. Johnston IA, Bernard LM. Utilization of the ethanol pathway in carp following exposure to
36 416
37 417 anoxia. *Journal of Experimental Biology*. 1983;104(1):73-8.
- 38 417 47. Dhillon RS, Mandic M, Yao L, Cao Z-D, Fu S-J, Brauner CJ, et al. Ethanol metabolism varies with
39 418
40 418 hypoxia tolerance in ten cyprinid species. *Journal of Comparative Physiology B*. 2018;188(2):283-93.
- 42 419 48. Nilsson GE. A comparative study of aldehyde dehydrogenase and alcohol dehydrogenase
43 420
44 420 activities in crucian carp and three other vertebrates: apparent adaptations to ethanol production.
45 421
46 421 *Journal of Comparative Physiology B*. 1988;158(4):479-85.
- 47 422 49. Ip YK, Loong AM, Chng YR, Hiong KC, Chew SF. Hepatic carbamoyl phosphate synthetase (CPS) I
48 423
49 423 and urea contents in the hydrid tree frog, *Litoria caerulea*: transition from CPS III to CPS I. *Journal of*
50 424
51 424 *Comparative Physiology B*. 2012;182(8):1081-94.

425

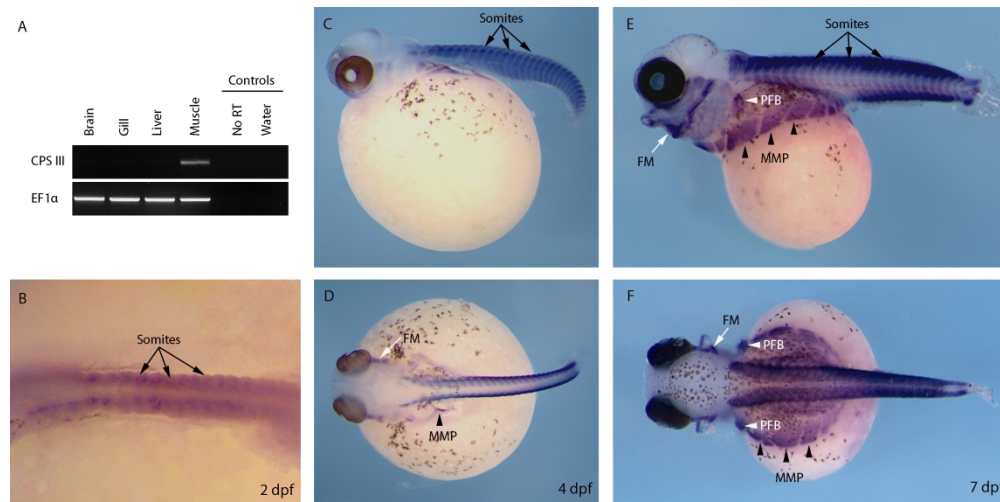


Figure 1: Expression analysis of carbamoyl-phosphate synthetase III (CPS III) from adult tissues and developing embryos of *Alcolapia alcalica*. A) Reverse transcriptase PCR and gel electrophoresis showing the muscle specific expression of CPS III, EF1 α shown as normalisation control. B-F) Lateral (C and E) and dorsal (B, D, and F) views of in situ hybridisation for CPS III in developing *A. alcalica* embryos at different stages (number of days post fertilisation [dpf] indicated). The blue colour indicates the detection of mRNA. The black/brown is endogenous pigment apparent in the retina and the chromatophores. Black arrows show somites, arrowheads indicate region of migrating muscle progenitors (MMP), white arrows show facial muscle (FM) and white arrowheads indicate developing pectoral fin bud (PFB). Black dots around the yolk and on the body are chromatophores (pigment cells).

267x132mm (300 x 300 DPI)

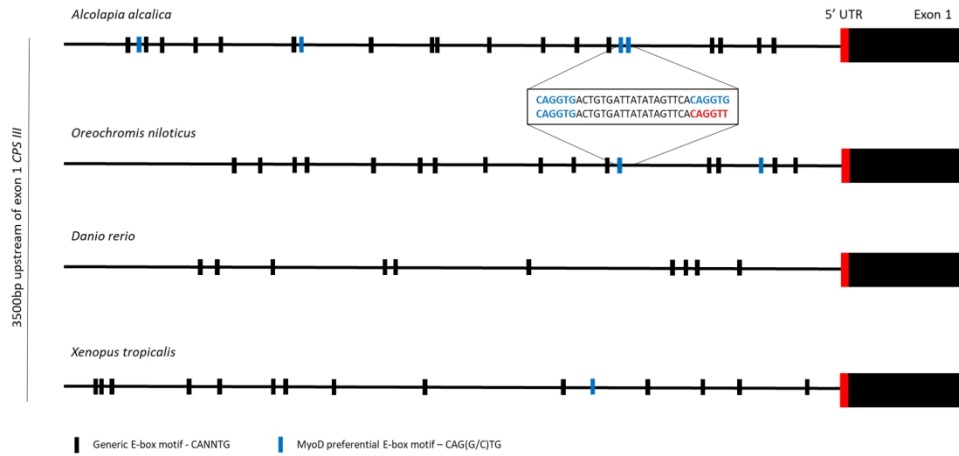


Figure 2: Sequence comparison of a 3500bp region upstream of exon 1 of CPS III. Presence of generic E-box sites with the consensus motif of CANNTG are annotated by black bars and MyoD preferential E-box motifs of CAG(G/C)TG are represented by blue bars. Insert shows sequence of *Alcolapia alcalica* potential paired E-box, MyoD enhancer (Blue nucleotides) compared to *Oreochromis niloticus* which has a single MyoD E-box upstream of a 'presumptive' MyoD E-box (Red nucleotides).

321x150mm (150 x 150 DPI)

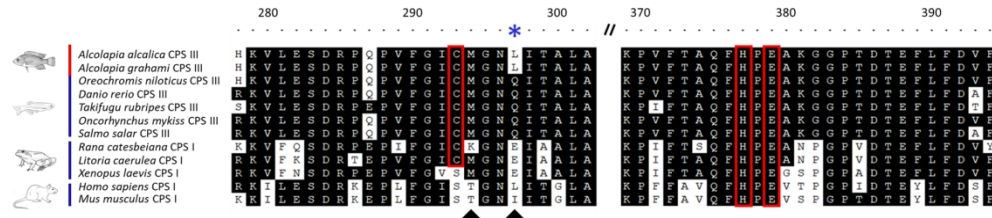


Figure 3: Multiple amino acid alignment of residues 278 to 397 (aligned to *Alcolapia alcalica*) of carbamoyl-phosphate synthetase I and III from a number of tetrapod and teleost species respectively. Conserved amino acids are shaded in black, amino acids in the catalytic triad are boxed in red and arrowheads indicate residues vital for glutamine utilisation of the catalytic triad. The blue asterisk indicates the divergent glutamine binding residue in *Alcolapia* species that likely results in a functional change (inability to bind glutamine).

321x70mm (150 x 150 DPI)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 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

1 **Adaptation of the carbamoyl-phosphate synthetase (CPS) enzyme in an extremophile fish**

2 Lewis J. White¹, Gemma Sutton¹, Asilatu Shechonge², Julia J. Day³, Kanchon K. Dasmahapatra¹, Mary E. Pownall¹

6 ¹Biology Department, University of York, York, YO10 5DD, United Kingdom

7 ²Tanzania Fisheries Research Institute, PO BOX 98 Kyela, Mbeya, Tanzania

8 ³Department of Genetics, Evolution and Environment, University College London, Darwin Building, Gower Street, London, WC1E 6BT, United Kingdom

11 Author for communication: ljw569@york.ac.uk

13 **ABSTRACT**

14 **Tetrapods and fish have adapted distinct carbamoyl-phosphate synthase (CPS) enzymes to initiate the**
15 **ornithine urea cycle during the detoxification of nitrogenous wastes. We report evidence that in the**
16 **ureotelic subgenus of extremophile fish *Oreochromis Alcolapia*, CPS III has undergone convergent**
17 **evolution and adapted its substrate affinity to ammonia which is typical of terrestrial vertebrate CPS I.**
18 **Unusually, unlike in other vertebrates, the expression of CPS III in *Alcolapia* is localised to the skeletal**
19 **muscle and is activated in the myogenic lineage during early embryonic development with expression**
20 **remaining in mature fish. We propose that adaptation in *Alcolapia* included both convergent**
21 **evolution of CPS function to that of terrestrial vertebrates, as well as changes in development**
22 **mechanisms redirecting CPS III gene expression to the skeletal muscle.**

24 **INTRODUCTION**

25 In living organisms, protein metabolism results in the production of nitrogenous wastes which need to
26 be excreted. Most teleosts are ammonotelic, excreting their toxic nitrogenous waste as ammonia across
27 gill tissue by diffusion. As an adaptation to living on land, amphibians and mammals are ureotelic, using
28 liver and kidney tissues to convert waste ammonia into the less toxic and more water soluble urea,
29 which is then excreted in urine. Other terrestrial animals such as insects, birds and reptiles are uricotelic
30 and convert nitrogenous waste into uric acid, which is eliminated as a paste; a process which requires
31 more energy but wastes less water (1).

32 While most adult fish are ammonotelic, the larval stages of some teleosts excrete nitrogenous waste as
33 both ammonia and urea before their gills are fully developed (2). Additionally, some adult fish species
34 such as the gulf toad fish (*Opsanus beta*; (3)) and the African catfish (*Clarias gariepinus*; (4)) also excrete
35 a proportion of their nitrogenous waste as urea. This is usually in response to changes in aquatic

1
2
3 36 conditions, such as high alkalinity. It has been shown experimentally that high external pH prevents
4 37 diffusion of ammonia across gill tissue (5, 6). Unusually, the cichlid fish species in the subgenus *Alcolapia*
5 38 (described by some authors as a genus but shown to nest within the genus *Oreochromis*; (7)) inhabit the
6
7
8 39 highly alkaline soda lakes of Natron (Tanzania) and Magadi (Kenya), are reported to be 100% ureotelic
9
10 40 (8, 9).

11
12 41 Once part of a single paleo-lake, Orolonga (10), Lakes Natron and Magadi are one of the most extreme
13 42 environments supporting fish life, with water temperatures up to 42.8 °C, pH ~10.5, fluctuating
14 43 dissolved oxygen levels, and salt concentrations above 20 parts per thousand (11). *Alcolapia* is the only
15 44 group of fish to survive in these lakes, forming a recent adaptive radiation including the four species:
16 45 *Alcolapia grahami* (Lake Magadi) and *A. latilabris*, *A. ndlalani* and *A. alcalica* (Lake Natron) (11, 12). The
17 46 harsh environment of the soda lakes presents certain physiological challenges that *Alcolapia* have
18 47 evolved to overcome, including the basic need to excrete nitrogenous waste. While other species are
19 48 able to excrete urea in response to extreme conditions, none do so to the level of *Alcolapia* (13, 14), and
20 49 unlike facultative ureotelic species, the adaptation of urea production and excretion in *Alcolapia* is
21 50 considered fixed (15). Moreover, the heightened metabolic rate in *Alcolapia*, a by-product from living in
22 51 such an extreme environment (8, 16), requires an efficient method of detoxification.

23
24
25
26
27
28
29
30
31 52 *Alcolapia* and ureotelic tetrapods (including humans) detoxify ammonia using the ornithine urea cycle
32 53 (OUC) where the mitochondrial enzyme carbamoyl-phosphate synthetase (CPS) is essential for the first
33 54 and rate limiting step of urea production (17). This enzyme, together with the accessory enzyme
34 55 glutamine synthase, provide an important switch regulating the balance between ammonia removal for
35 56 detoxification and maintaining a source of ammonia for the biosynthesis of amino acids (18). CPS has
36 57 evolved into two biochemically distinct proteins: in terrestrial vertebrates CPS I uses ammonia as its
37 58 preferential nitrogen donor, while in teleosts CPS III accepts glutamine to produce urea during larval
38 59 stages (reviewed Zimmer et al 2017). While CPS I / III are mitochondrial enzymes and part of the urea
39 60 cycle, CPS II is present in the cytosol catalyzing the synthesis of carbamoyl phosphate for pyrimidine
40 61 nucleotide biosynthesis. CPS I / III are syntenic, representing orthologous genes; their somewhat
41 62 confusing nomenclature is based on the distinct biochemical properties of their proteins. CPS I / III genes
42 63 from different vertebrate species clade together, separate from CPS II (supplementary figure). For
43 64 simplicity, we will continue to refer to fish, glutamine binding CPS as CPS III and tetrapod, ammonia
44 65 binding CPS as CPS I. -The teleost CPS III binds glutamine in the glutamine amidotransferase (GAT)
45 66 domain using two amino acid residues (19), subsequently the nitrogen source provided by the amide
46 67 group is catalysed by a conserved catalytic triad; Cys-His-Glu (20). In Tterrestrial vertebrates CPS I lacks a
47 68 complete catalytic triad and can only generate carbamoyl-phosphate in the presence of free ammonia
48 69 (21). This change in function from glutamine binding CPS III to ammonia binding CPS I is believed to have

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
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

70 ~~occurred in the evolution of amphibian species~~ evolved in the stem lineage of living tetrapods, first
71 appearing in ancestral amphibians (21).

72 In tetrapods and most fish, the OUC enzymes are largely localised to the liver (22), the main urogenic
73 organ (23). *Alcolapia* are different, and the primary site for urea production in these extremophile fishes
74 is the skeletal muscle (24). Notably, glutamine synthase activity is reportedly absent in *Alcolapia* muscle
75 tissue. The kinetic properties of CPS III in *Alcolapia* therefore differ from that of other teleosts in that it
76 preferentially uses ammonia as its primary substrate, having maximal enzymatic rates above that of
77 binding glutamine (although it is still capable of doing so) as opposed to in other species where use of
78 ammonia yields enzymatic rates of around 10% to that of glutamine (24). These rates are similar to
79 ureotelic terrestrial species, where CPS I preferentially binds ammonia and is incapable of using
80 glutamine (20).

81 Here, we report the amino acid sequence of two *Alcolapia* species (*A. alcalica* and *A. grahami*) that
82 reveals a change in CPS III substrate binding site. In addition, we show that the expression of *Alcolapia*
83 CPS III in skeletal muscle arises early in embryonic development where transcripts are restricted to the
84 somites, the source of skeletal muscle in all vertebrates, and migrating myogenic precursors. We discuss
85 changes to the structure of functional domains and modular gene enhancers that likely underpin
86 evolutionary changes in *Alcolapia* CPS III substrate binding and the redirection of gene expression from
87 the hepatogenic to myogenic lineage (25). Our findings point to adaptation in *Alcolapia* including both
88 convergent evolution of CPS function to that of terrestrial vertebrates, as well as changes in
89 development mechanisms redirecting CPS III gene expression to the skeletal muscle.

90 **Methods**

91 ***Experimental animals***

92 Fieldwork to Lake Natron, Tanzania, was conducted during June and July of 2017 to collect live
93 specimens of the three endemic species in an attempt to produce stable breeding populations of these
94 fishes in the UK. Live fish were all collected from a single spring (site 5 (11, 26)) containing all three
95 species found in Lake Natron and identified using morphology as described in Seegers and Tichy, 1999. A
96 ~~n-aquatics habitat~~, stand alone, recirculating aquarium was adapted to house male and female *A.*
97 *alcalica* in 10 or 30 L tanks at a constant temperature of 30 °C, pH 9 and salt concentration of 3800 µS
98 at the University of York.

99 ***Expression of CPS III in adult tissues***

1
2
3 100 Reverse transcription polymerase chain reaction (RT-PCR) was used to determine the presence of CPS III
4
5 101 in different tissues (gill, muscle, liver, brain) of three different adult *A. alcalica*. RNA was extracted from
6
7 102 dissected tissues with TriReagent (Sigma-Aldrich) to the ~~manufactures~~manufacturers' guidelines. For
8
9 103 cDNA synthesis, 1µg of total RNA was reverse transcribed with random hexamers (Thermo Scientific)
10
11 104 and superscript IV (Invitrogen). PCR was performed on 2µl of the above cDNA with Promega PCR master
12
13 105 mix and 0.5mM of each primer (Forward: CAGTGGGAGGTCAGATTGC, Reverse:
14
15 106 CTCACAGCGAAGCACAGGG). Gel electrophoresis of the PCR products determined the presence or
16
17 107 absence of CPS III RNA.

108 ***In situ hybridisation***

109 For the production of antisense probes, complementary to the mRNA of CPS III to use in *in situ*
110 hybridisation, ~~these above 399bp~~ PCR products ~~were~~was ligated into PGem-tEasy and transformed into
111 the *E. coli* strain DH5α. ~~Transformations were plated onto LB agar plates containing ampicillin (100~~
112 ~~µg/ml) and grown overnight at 37°C. A single colony was cultured in 3ml of LB medium at 37°C overnight~~
113 ~~at 180rpm and Qiagen miniprep kits used to extract the cultured plasmids. Plasmids were sequenced to~~
114 ~~determine orientation which in turn indicates which restriction enzyme and polymerase pair to use to~~
115 ~~produce the antisense RNA probe. This was linearised and~~ in vitro run off transcription was used to
116 incorporate a DIG labelled UTP analogue. To determine the temporal expression of these proteins in *A.*
117 *alcalica*, embryos were collected at different stages of development (2, 4 and 7 days post fertilisation
118 [between 15 and 20 for each stage]), fixed for 1 hour in MEMFA (0.1 M MOPS pH 7.4, 2 mM EGTA, 1 mM
119 MgSO₄, 3.7% formaldehyde) at room temperature and stored at -20°C in 100% methanol. For *in situ*
120 hybridisation, embryos were rehydrated and treated with 10 µg/mg proteinase K room temperature.
121 After post fixation and a 2 hour pre-hybridisation, embryos were hybridised with the probe at 68°C in in
122 hybridisation buffer (50% formamide (Ambion), 1mg/ml total yeast RNA, 5×SSC, 100µg/ml heparin, 1×
123 denharts, 0.1% Tween-20, 0.1% CHAPS, 10mM EDTA~~50% formamide~~. Embryos were extensively washed
124 at 68°C in 2×SSC +0.1% Tween-20, 0.2×SSC +0.1% Tween-20 and maleic acid buffer (MAB; 100mM maleic
125 acid, 150mM NaCl, 0.1% Tween-20, pH7.8). This was replaced with pre-incubation buffer (4× MAB, 10%
126 BMB, 20% heat-treated lamb serum) for 2 hours. Embryos were incubated overnight (rolling at 4°C) with
127 fresh pre-incubation buffer and 1/2000 dilution of anti-DIG coupled with alkaline phosphatase (AP)
128 (Roche). After extensive washing, hybridisation was visualised by applying anti-DIG coupled with alkaline
129 phosphatase followed by BM purple. BM purple is a substrate for alkaline phosphatase. These were then
130 visualised by application of BM purple until staining had occurred.

131 ***Sequence analysis of CPS III***

1
2
3 132 cDNA was produced from the RNA extracted from whole embryos using the above method for *A. alcalica*
4
5 133 and *A. grahmi*. Multiple primer pairs ([supplementary table 1](#)) were used to amplify fragments of CPS III
6
7 134 from the cDNA via PCR and the products sent for sequencing. The coding region of CPS III was then
8
9 135 constructed using multiple alignments against the CPS I and III from other species. The amino acid
10
11 136 sequence was then examined for potential changes which could predict the functional differences seen
12
13 137 in *Alcolapia*. [Phylogenetic analysis was also used to confirm the *Alcolapia* genes analysed here are CPS III](#)
14
15 138 [\(supplementary figure 1\)](#). To determine potential changes in promoter region, a 3500bp section of
16
17 139 genome ([accession number NCBI: MW014910](#)) upstream of the transcriptional site start of CPS from *A.*
18
19 140 *alcalica* (unpublished genome), *Oreochromis niloticus* (Nile tilapia), *Xenopus tropicalis* (western clawed
20
21 141 frog) and *Danio rerio* (zebrafish), genomes were accessed on Ensembl, ~~was~~ aligned and examined for
22
23 142 binding sites specific to the muscle transcription factor MyoD (E-boxes) which preferentially binds
24
25 143 paired E-boxes in the enhancer regions of myogenic genes with the consensus motif CAG(G/C)TG, as
26
27 144 well as E-boxes more broadly (CANNTG). The published genomes of *O. niloticus*, *D. rerio* and *X. tropicalis*
28
29 145 were accessed using Ensembl whereas the *Alcolapia* genome was constructed from whole genome
30
31 146 sequences.

147 Results

148 ***CPS III expression is activated early in the skeletal muscle lineage in A. alcalica***

149 Analysis of gene expression of CPS III in dissected tissues of three adult *A. alcalica* shows that transcripts
150 were only detected in adult muscle (**Figure 1A**). *In situ* hybridisation methods on *A. alcalica* embryos at
151 different stages were carried out to investigate whether this restricted muscle expression was
152 established during development (**Figure 1B-F**). Blue colouration indicates hybridisation of the
153 complementary RNA probe and shows strongest expression in the developing somites along the body
154 axis (**arrows**). Expression was also detected in migratory muscle precursors (MMP; **arrowheads**), which
155 go on to form the body wall and limb musculature, and in the developing ~~limb~~ [pectoral fin](#) buds (**white**
156 **arrows**). All regions of the embryo that show expression of CPS III are in the muscle lineage indicating
157 that in *A. alcalica* CPS III expression is restricted to muscle tissues in both adults and the developing
158 embryo.

159 Many muscle specific genes are activated during development by the muscle specific transcription
160 factor, MyoD. The promoter region of CPS III (3.5kb upstream of the transcriptional start site) in *A.*
161 *alcalica* was compared to that in *O. niloticus*, *X. tropicalis* and *D. rerio* (**Figure 2**). Examination of this
162 region revealed a [putative](#) paired E-box MyoD binding site 940 to 970 bases
163 ([CAGGTGACTGTGATTATATAGTTACAGGTG](#)) upstream of the transcriptional start site of CPS III only in
164 *Alcolapia* species. Intriguingly, while no pair of MyoD E-boxes were found in the upstream region of any

1
2
3 165 other species examined, *O. niloticus* does have a single MyoD E-box motif in the same region upstream
4 166 of CPS III, and within 19 bases of this is a CAGGTT motif which a single point mutation would convert
5 167 into a pair of E-boxes (CAGGTGACTGTGATTATATAGTTACAGGTT). This suggests that it is possible that
6 168 MyoD could bind and activate transcription of *CPS III* in the muscle of *Alcolapia* species, but not in the
7 169 closely related *O. niloticus*.

12 170 ***Convergent evolution in adaptive function of CPS III***

14 171 Sequence analysis of *A. alcalica* and *A. grahami* CPS III revealed a discrepancy in the catalytic triad
15 172 compared to the published sequence for CPS III in *A. grahami* (accession number [NCBI: AF119250](#)). The
16 173 coding region for *A. alcalica* and *A. grahami* was cloned and sequenced (accession numbers [NCBI:](#)
17 174 [MT119353](#), [MT119354](#)). Our data confirmed the error in the published sequence of *A. grahami* CPS III
18 175 and shows *Alcolapia* species maintain a catalytic triad essential for catalysing the breakdown of
19 176 glutamine (**red boxes in Figure 3**). However, similar to terrestrial vertebrate CPS I which lack either one
20 177 but usually both residues essential for binding glutamine for utilisation by the catalytic triad
21 178 (**arrowheads in Figure 3**), *Alcolapia* also lack one of these residues (**asterisks Figure 3**). This amino acid
22 179 sequence is consistent with a change in function permitting *Alcolapia* CPS III to bind and catalyse
23 180 ammonia directly, an activity usually restricted to terrestrial vertebrate CPS I, as elucidated by extensive
24 181 previous biochemical analyses (20, 21).

33 182 **Discussion**

34 183 While most teleosts are ammonotelic, larval fish can convert ammonia to urea for excretion and to do so
35 184 express the genes coding for the enzymes of the OUC, including CPS III (27). Later these genes are
36 185 silenced in most fish. In the rare cases where urea is produced in adult fish, the OUC enzymes are
37 186 expressed in the liver (23), however there are some reports of expression in non-hepatic tissues (28, 29).
38 187 We report here the expression of CPS III in the muscle of adult *A. alcalica*, which is consistent with the
39 188 detection of CPS III protein and enzyme activity in muscle of *A. grahami* (24). We also find conserved
40 189 changes to the amino acid sequence which explains the convergent evolution of *A. alcalica* and *A.*
41 190 *grahami* CPS III function with CPS I in terrestrial vertebrates. This conserved change in both *Alcolapia*
42 191 species suggests that the adaptations in the OUC are likely to have evolved in the ancestral species
43 192 inhabiting paleolake Orolongo during the period of worsening-changing aquatic conditions (over the
44 193 past ten thousand years) that led to the extreme conditions currently found in Lakes Natron and
45 194 Magadi.

57 195 ***Activation of CPS III in the myogenic lineage***

1
2
3 196 We find that the expression of CPS III is activated in somites and in migratory muscle precursors that will
4
5 197 form body wall and limb musculature (indeed expression is seen in developing limb buds). All skeletal
6
7 198 muscle in the vertebrate body is derived from the somites, and these CPS III expression patterns are
8
9 199 similar to those of muscle specific genes like myosin, actins and troponins (30-32).

10
11 200 Muscle specific expression of CPS III in *A. alcalica* embryos is a remarkable finding as most ureotelic
12
13 201 species convert nitrogenous waste to urea in the liver (8, 20). The expression of CPS III, the first enzyme
14
15 202 in the OUC, in muscle tissue is likely significant for supporting the high catabolism in a fish species with
16
17 203 the highest recorded metabolic rate (33). There are few reports of some OUC gene expression or
18
19 204 enzyme activity in non-hepatic tissue including muscle (28, 29, 34), nonetheless other fish species only
20
21 205 evoke the activity of the OUC when exposed to high external pH or during larval stages (13, 14, 35, 36),
22
23 206 and even then, urea production is never to the high level of activity occurring in *Alcolapia* (24). There is
24
25 207 some heterogeneity of the expression patterns of CPS III during the development of different species in
26
27 208 the teleost lineage; for example; *D. rerio* has reported expression in the body (37), *Oncorhynchus mykiss*
28
29 209 (rainbow trout) shows expression in the developing body but not in hepatic tissue (38) and *C. gariepinus*
30
31 210 (African catfish) had CPS III expression detected in the dissected muscle from larvae (4). In laboratory
32
33 211 conditions, adult *A. alcalica* excrete approximately 75% of their nitrogenous waste as urea, compared to
34
35 212 only 10% in adult zebrafish (White et al., in prep). The early and sustained expression of CPS III in the
36
37 213 muscle lineage is at this point an observation unique to *Alcolapia*.

38
39 214 Skeletal muscle specific gene expression is activated in cells of the myogenic lineage by a family of bHLH
40
41 215 transcription factors, including MyoD (30). MyoD binds specifically at paired E-boxes in the enhancers of
42
43 216 myogenic genes with a preference for the consensus motif of CAG(G/C)TG (39, 40). MyoD is known to
44
45 217 require the cooperative binding at two E-boxes in close proximity, to modulate transcription of
46
47 218 myogenic genes (41). The presence of a pair of E-boxes in *Alcolapia*, upstream of a gene which has
48
49 219 switched to muscle specific expression, is suggestive that MyoD is driving expression early in
50
51 220 development. Enhancer modularity is a known mechanism for selectable variation (42) and although a
52
53 221 single MyoD binding site does not define an enhancer, MyoD is known to interact with pioneer factors
54
55 222 and histone deacetylases to open chromatin and activate gene transcription in the muscle lineage (40,
56
57 223 43). Experimental analysis to determine the activity of any regulatory sequences upstream of OUC genes
58
59 224 in different species would shed light on the significance of putative transcription factor binding sites.
60
225 This approach could also address another intriguing question as to the elements that drive the post-
226 larval silencing of OUC genes in most fish species (37), an area with only minimal research especially
227 when compared to the well characterised promoter region in mammalian species, for instance
228 Christoffels et al 1998. A further instance of an extremophile organism redirecting expression of a
229 hepatic enzyme to muscle tissue occurs in the crucian carp (45). Under conditions of anoxia this species

1
2
3 230 switches to anaerobic metabolism, producing ethanol as the end product of glycolysis (46, 47). This is
4
5 231 associated with the expression of *alcohol dehydrogenase* in muscle (48). Together with our findings, this
6
7 232 potentially reveals an example of convergent evolution whereby the muscle becomes the site for
8
9 233 detoxifying by-products of metabolism. Elucidating any mechanisms that may include modular
10
11 234 enhancers that facilitate the adaptation of gene regulation in response to changing environmental
12
13 235 conditions will be of significant interest.

14 236 ***Convergent evolution of adaptive CPS III function***

15
16 237 CPS proteins catalyse the production of carbamoyl-phosphate as a first step in nitrogen detoxification by
17
18 238 accepting either glutamine or ammonia as a nitrogen donor (17). Teleost CPS III binds glutamine: the
19
20 239 nitrogen source provided by the amide group of glutamine is catalyzed by the conserved catalytic triad
21
22 240 Cys-His-Glu in the glutamine amidotransferase (GAT) domain in the amino terminal part of CPS (20). In
23
24 241 terrestrial vertebrates, CPS I lacks the catalytic cysteine residue and only generates carbamoyl-
25
26 242 phosphate in the presence of free ammonia (21). Although CPS in *Alcolapia* shares most sequence
27
28 243 identity with fish CPS III (**Figure 3**), its ammonia binding activity is more similar in function to terrestrial
29
30 244 vertebrate CPS I (20, 24). This adaptation to preferentially bind ammonia over glutamine supports
31
32 245 efficient waste management in a fish with an exceptionally high metabolic rate (33). CPS I in terrestrial
33
34 246 vertebrates have amino acid changes in the catalytic triad which explains their binding ammonia over
35
36 247 glutamine; a reduction in glutamine binding capacity drives the use of ammonia (21). Here we show that
37
38 248 *Alcolapia* maintain the catalytic triad, but (similar to mouse and human) lack one of the two residues
39
40 249 required for efficient glutamine binding, weakening its affinity to glutamine, and driving the use of
41
42 250 ammonia as a primary substrate.

43
44 251 The interesting observation that bullfrog (*Rana catesbeiana*) CPS I retains the catalytic triad, but lacks
45
46 252 the two additional conserved amino acids required for glutamine binding, has led to the suggestion that
47
48 253 the change from preferential glutamine to ammonia binding originally evolved in [the early tetrapod](#)
49
50 254 [lineageamphibia](#) (21). A further frog species, the tree frog *Litoria caerulea*, retains its catalytic triad and
51
52 255 only one of the two residues required for glutamine binding has been altered, weakening its affinity for
53
54 256 glutamine and allowing for direct catabolism of ammonia (49). Much the same as in *Alcolapia*, *L.*
55
56 257 *caerulea* CPS I is still capable of using glutamine to some extent which lends further support to the
57
58 258 notion that the evolutionary transition from CPS III to CPS I occurring in [the amphibians and the early](#)
59
60 259 [tetrapod](#) lineage. The changes in protein sequence of *Alcolapia* CPS III represents a convergent evolution
61
62 260 in this extremophile fish species, with acquired changes in functionally important domains which likely
63
64 261 also evolved in early terrestrial vertebrate CPS I.

65 262 ***Conclusions***

263 *Alcolapia* have acquired multiple adaptations that allow continued excretion of nitrogenous waste in a
 264 high pH environment. Among these is the novel expression of CPS in skeletal muscle, as well as
 265 acquisition of mutations that change its function. Sequence evidence indicates that like terrestrial
 266 vertebrates, and unique among fish, *Alcolapia* CPS III is capable of binding and catalysing the breakdown
 267 of ammonia to carbamoyl-phosphate; a convergent evolution of CPS function. The mechanism by which
 268 the novel and unique expression of CPS in muscle evolved is likely a function of enhancer regions of *A.*
 269 *alcalica* and *A. grahami* that result in its regulation by muscle regulatory factors to direct CPS expression
 270 in the myogenic lineage during embryonic development. Environmentally driven adaptations have
 271 resulted in changes in both the expression and activity of CPS III in *Alcolapia* that underpin its ability to
 272 turnover nitrogenous waste in a challenging environment while maintaining a high metabolic rate.

273 Acknowledgments

274 Experiments were designed by LJW and MEP, work was carried out by LJW and GS and the manuscript
 275 written and edited by LJW, GS, AS, JD, KKD & MEP. LJW, JD and AS collected the fish from Lake Natron.
 276 The authors are grateful to Antonia Ford and George Turner for their help in collecting the fish from
 277 Lake Natron (COSTECH permit no. 2017-259-NA-2011-182) and their continued support in this project.

278 Funding

279 This work was supported by the BBSRC as a studentship to LJW (BB/M011151/1). Additional support for
 280 fieldwork was provided by the Fisheries Society of the British Isles (small research grant) and the
 281 Genetics Society (Heredity fieldwork grant).

282 **Figure 1:** Expression analysis of carbamoyl-phosphate synthetase III (CPS III) from adult tissues and
 283 developing embryos of *Alcolapia alcalica*. A) Reverse transcriptase PCR and gel electrophoresis showing
 284 the muscle specific expression of CPS III, EF1 α shown as normalisation control. B-F) Lateral (C and E) and
 285 dorsal (B, D, and F) views of *in situ* hybridisation for CPS III in developing *A. alcalica* embryos at different
 286 stages (number of days post fertilisation [dpf] indicated). The blue colour indicates the detection of
 287 mRNA. The black/brown is endogenous pigment apparent in the retina and the chromatophores. Black
 288 arrows show somites, arrowheads indicate region of migrating muscle progenitors (MMP), white arrows
 289 show facial muscle (FM) and white arrowheads indicate developing ~~limb~~ pectoral fin bud (LPFB).
 290 Black dots around the yolk and on the body are chromatophores (pigment cells).

291 **Figure 2:** Sequence comparison of a 3500bp region upstream of exon 1 of CPS III. Presence of generic E-
 292 box sites with the consensus motif of CANNTG are annotated by black bars and MyoD preferential E-box
 293 motifs of CAG(G/C)TG are represented by blue bars. Insert shows sequence of *Alcolapia alcalica*
 294 potential paired E-box, MyoD enhancer (Blue nucleotides) compared to *Oreochromis niloticus* which has
 295 a single MyoD E-box upstream of a 'presumptive' MyoD E-box (Red nucleotides).

296 **Figure 3:** Multiple amino acid alignment of residues 278 to 397 (aligned to *Alcolapia alcalica*) of
 297 carbamoyl-phosphate synthetase I and III from a number of tetrapod and teleost species respectively.
 298 Conserved amino acids are shaded in black, amino acids in the catalytic triad are boxed in red and
 299 arrowheads indicate residues vital for glutamine utilisation of the catalytic triad. The blue asterisk

1
2
3 300 indicates the divergent glutamine binding residue in *Alcolapia* species that likely results in a functional
4 301 change (inability to bind glutamine).

5
6 302 **References**

- 7
8 303 1. Wright PA. Nitrogen excretion: three end products, many physiological roles. *Journal of*
9 304 *Experimental Biology*. 1995;198(2):273-81.
- 10 305 2. Zimmer AM, Wright PA, Wood CM. Ammonia and urea handling by early life stages of fishes.
11 306 *Journal of Experimental Biology*. 2017;220(21):3843-55.
- 12 307 3. Barimo JF, Steele SL, Wright PA, Walsh PJ. Dogmas and controversies in the handling of
13 308 nitrogenous wastes: ureotely and ammonia tolerance in early life stages of the gulf toadfish, *Opsanus*
14 309 *beta*. *Journal of Experimental Biology*. 2004;207(12):2011-20.
- 15 310 4. Terjesen BF, Chadwick TD, Verreth JA, Rønnestad I, Wright PA. Pathways for urea production
16 311 during early life of an air-breathing teleost, the African catfish *Clarias gariepinus* Burchell. *Journal of*
17 312 *Experimental Biology*. 2001;204(12):2155-65.
- 18 313 5. Handy R, Poxton M. Nitrogen pollution in mariculture: toxicity and excretion of nitrogenous
19 314 compounds by marine fish. *Reviews in Fish Biology and Fisheries*. 1993;3(3):205-41.
- 20 315 6. Wright PA, Wood C. An analysis of branchial ammonia excretion in the freshwater rainbow
21 316 trout: effects of environmental pH change and sodium uptake blockade. *Journal of Experimental*
22 317 *Biology*. 1985;114(1):329-53.
- 23 318 7. Ford AGP, Bullen TR, Pang L, Genner MJ, Bills R, Flouri T, et al. Molecular phylogeny of
24 319 *Oreochromis* (Cichlidae: Oreochromini) reveals mito-nuclear discordance and multiple colonisation of
25 320 adverse aquatic environments. *Molecular Phylogenetics and Evolution*. 2019;136:215-26.
- 26 321 8. Randall D, Wood C, Perry S, Bergman H, Maloij G, Mommsen T, et al. Urea excretion as a
27 322 strategy for survival in a fish living in a very alkaline environment. *Nature*. 1989;337:165-6.
- 28 323 9. Wilson PJ, Wood CM, Walsh PJ, Bergman AN, Bergman HL, Laurent P, et al. Discordance
29 324 between genetic structure and morphological, ecological, and physiological adaptation in Lake Magadi
30 325 tilapia. *Physiological and Biochemical Zoology*. 2004;77(4):537-55.
- 31 326 10. Roberts N, Taieb M, Barker P, Damnati B, Icole M, Williamson D. Timing of the Younger Dryas
32 327 event in East Africa from lake-level changes. *Nature*. 1993;366(6451):146-8.
- 33 328 11. Ford AGP, Dasmahapatra KK, Rüber L, Gharbi K, Cezard T, Day JJ. High levels of interspecific gene
34 329 flow in an endemic cichlid fish adaptive radiation from an extreme lake environment. *Molecular Ecology*.
35 330 2015;24(13):3421-40.
- 36 331 12. Seegers L, Tichy H. The *Oreochromis alcalicus* flock (Teleostei: Cichlidae) from lakes Natron and
37 332 Magadi, Tanzania and Kenya, with descriptions of two new species. *Ichthyological Exploration of*
38 333 *Freshwaters*. 1999;10:97-146.
- 39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 334 13. Walsh P, Danulat E, Mommsen T. Variation in urea excretion in the gulf toadfish *Opsanus beta*.
4 335 Marine Biology. 1990;106(3):323-8.
5
6 336 14. Wright PA, Iwama GK, Wood CM. Ammonia and urea excretion in Lahontan cutthroat trout
7 337 (*Oncorhynchus clarki henshawi*) adapted to the highly alkaline Pyramid Lake (pH 9.4). Journal of
8 338 Experimental Biology. 1993;175(1):153-72.
9
10 339 15. Wood CM, Wilson P, Bergman HL, Bergman AN, Laurent P, Otiang'a-Owiti G, et al. Obligatory
11 340 urea production and the cost of living in the Magadi tilapia revealed by acclimation to reduced salinity
12 341 and alkalinity. Physiological Biochemical Zoology. 2002;75(2):111-22.
13
14 342 16. Kavembe GD, Franchini P, Irisarri I, Machado-Schiaffino G, Meyer A. Genomics of adaptation to
15 343 multiple concurrent stresses: insights from comparative transcriptomics of a Cichlid fish from one of
16 344 earth's most extreme environments, the Hypersaline Soda Lake Magadi in Kenya, East Africa. Journal of
17 345 Molecular Evolution. 2015;81(3-4):90-109.
18
19 346 17. Rubio V, Cervera J. The carbamoyl-phosphate synthase family and carbamate kinase: structure-
20 347 function studies. Portland Press Ltd.; 1995.
21
22 348 18. de Cima S, Polo LM, Díez-Fernández C, Martínez AI, Cervera J, Fita I, et al. Structure of human
23 349 carbamoyl phosphate synthetase: deciphering the on/off switch of human ureagenesis. Scientific
24 350 Reports. 2015;5:16950.
25
26 351 19. Hong J, Salo WL, Lusty CJ, Anderson PM. Carbamyl phosphate synthetase III, an evolutionary
27 352 intermediate in the transition between glutamine-dependent and ammonia-dependent carbamyl
28 353 phosphate synthetases. Molecular Biology. 1994;243:131-40.
29
30 354 20. Loong A, Chng Y, Chew S, Wong W, Ip Y. Molecular characterization and mRNA expression of
31 355 carbamoyl phosphate synthetase III in the liver of the African lungfish, *Protopterus annectens*, during
32 356 aestivation or exposure to ammonia. Journal of Comparative Physiology B. 2012;182(3):367-79.
33
34 357 21. Saeed-Kothe A, Powers-Lee SG. Specificity determining residues in ammonia- and glutamine-
35 358 dependent carbamoyl phosphate synthetases. Journal of Biological Chemistry. 2002;277(9):7231-8.
36
37 359 22. Ip AY, Chew SF. Ammonia production, excretion, toxicity, and defense in fish: a review. Frontiers
38 360 in Physiology. 2010;1:134.
39
40 361 23. Mommsen TP, Walsh PJ. Evolution of urea synthesis in vertebrates: the piscine connection.
41 362 Science. 1989;243(4887):72-5.
42
43 363 24. Lindley TE, Scheiderer CL, Walsh PJ, Wood CM, Bergman HL, Bergman AL, et al. Muscle as the
44 364 primary site of urea cycle enzyme activity in an alkaline lake-adapted tilapia, *Oreochromis alcalicus*
45 365 *grahami*. Journal of Biological Chemistry. 1999;274(42):29858-61.
46
47 366 25. Rebeiz M, Tsiantis M. Enhancer evolution and the origins of morphological novelty. Current
48 367 Opinion in Genetics & Development. 2017;45:115-23.
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 368 26. Ford AGP, Rüber L, Newton J, Dasmahapatra KK, Balarin JD, Bruun K, et al. Niche divergence
4 369 facilitated by fine-scale ecological partitioning in a recent cichlid fish adaptive radiation. *Evolution*.
5 370 2016;70(12):2718-35.
- 6
7
8 371 27. Zimmer AM, Wood CM. Physiological and molecular ontogeny of branchial and extra-branchial
9 372 urea excretion in posthatch rainbow trout (*Oncorhynchus mykiss*). *American Journal of Physiology-*
10 373 *Regulatory, Integrative and Comparative Physiology*. 2016;310(3):R305-R12.
- 11
12
13 374 28. Korte JJ, Salo WL, Cabrera VM, Wright PA, Felskie AK, Anderson PM. Expression of carbamoyl-
14 375 phosphate synthetase III mRNA during the early stages of development and in muscle of adult rainbow
15 376 trout (*Oncorhynchus mykiss*). *Journal of Biological Chemistry*. 1997;272(10):6270-7.
- 16
17
18 377 29. Banerjee B, Koner D, Hasan R, Saha N. Molecular characterization and ornithine-urea cycle
19 378 genes expression in air-breathing magur catfish (*Clarias magur*) during exposure to high external
20 379 ammonia. *Genomics*. 2020;112:2247-60.
- 21
22
23 380 30. Pownall ME, Gustafsson MK, Emerson Jr CP. Myogenic regulatory factors and the specification of
24 381 muscle progenitors in vertebrate embryos. *Annual review of Cell and Developmental Biology*.
25 382 2002;18(1):747-83.
- 26
27
28 383 31. Talbot JC, Teets EM, Ratnayake D, Duy PQ, Currie PD, Amacher SL. Muscle precursor cell
29 384 movements in zebrafish are dynamic and require Six family genes. *Development*.
30 385 2019;146(10):dev171421.
- 31
32
33 386 32. Bucher EA, Maisonpierre P, Konieczny S, Emerson C. Expression of the troponin complex genes:
34 387 transcriptional coactivation during myoblast differentiation and independent control in heart and
35 388 skeletal muscles. *Molecular and Cellular Biology*. 1988;8(10):4134-42.
- 36
37
38 389 33. Wood CM, Brix KV, De Boeck G, Bergman HL, Bianchini A, Bianchini LF, et al. Mammalian
39 390 metabolic rates in the hottest fish on earth. *Scientific Reports*. 2016;6:26990.
- 40
41
42 391 34. Kong H, Edberg DD, Korte JJ, Salo WL, Wright PA, Anderson PM. Nitrogen excretion and
43 392 expression of carbamoyl-phosphate synthetase III activity and mRNA in extrahepatic tissues of
44 393 largemouth bass (*Micropterus salmoides*). *Archives of Biochemistry and Biophysics*. 1998;350(2):157-68.
- 45
46
47 394 35. Randall DJ, Tsui T. Ammonia toxicity in fish. *Marine Pollution Bulletin*. 2002;45(1-12):17-23.
- 48
49
50 395 36. Mommsen TP, Walsh PJ. Urea synthesis in fishes: evolutionary and biochemical perspectives:
51 396 Elsevier; 1991. 137-63 p.
- 52
53
54 397 37. LeMoine CM, Walsh PJ. Ontogeny of ornithine-urea cycle gene expression in zebrafish (*Danio*
55 398 *rerio*). *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*.
56 399 2013;304(11):R991-R1000.
- 57
58
59
60

- 1
2
3 400 38. Steele SL, Chadwick TD, Wright PA. Ammonia detoxification and localization of urea cycle
4 401 enzyme activity in embryos of the rainbow trout (*Oncorhynchus mykiss*) in relation to early tolerance to
5 402 high environmental ammonia levels. *Journal of Experimental Biology*. 2001;204(12):2145-54.
6
7 403 39. Fong AP, Yao Z, Zhong JW, Cao Y, Ruzzo WL, Gentleman RC, et al. Genetic and epigenetic
8 404 determinants of neurogenesis and myogenesis. *Developmental Cell*. 2012;22(4):721-35.
9
10 405 40. Conerly ML, Yao Z, Zhong JW, Groudine M, Tapscott SJ. Distinct activities of Myf5 and MyoD
11 406 indicate separate roles in skeletal muscle lineage specification and differentiation. *Developmental Cell*.
12 407 2016;36(4):375-85.
13
14 408 41. Weintraub H, Davis R, Lockshon D, Lassar A. MyoD binds cooperatively to two sites in a target
15 409 enhancer sequence: occupancy of two sites is required for activation. *Proceedings of the National*
16 410 *Academy of Sciences*. 1990;87(15):5623-7.
17
18 411 42. Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jónsson B, et al. Genetic and
19 412 developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*.
20 413 2004;428(6984):717.
21
22 414 43. Berkes CA, Bergstrom DA, Penn BH, Seaver KJ, Knoepfler PS, Tapscott SJ. Pbx marks genes for
23 415 activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential.
24 416 *Molecular Cell*. 2004;14(4):465-77.
25
26 417 44. Christoffels VM, Grange T, Kaestner KH, Cole TJ, Darlington GJ, Croniger CM, et al.
27 418 Glucocorticoid receptor, C/EBP, HNF3, and protein kinase A coordinately activate the glucocorticoid
28 419 response unit of the carbamoylphosphate synthetase I gene. *Molecular and Cellular Biology*.
29 420 1998;18(11):6305-15.
30
31 421 45. Fagernes CE, Stensløkken K-O, Røhr ÅK, Berenbrink M, Ellefsen S, Nilsson GE. Extreme anoxia
32 422 tolerance in crucian carp and goldfish through neofunctionalization of duplicated genes creating a new
33 423 ethanol-producing pyruvate decarboxylase pathway. *Scientific Reports*. 2017;7(1):7884.
34
35 424 46. Johnston IA, Bernard LM. Utilization of the ethanol pathway in carp following exposure to
36 425 anoxia. *Journal of Experimental Biology*. 1983;104(1):73-8.
37
38 426 47. Dhillon RS, Mandic M, Yao L, Cao Z-D, Fu S-J, Brauner CJ, et al. Ethanol metabolism varies with
39 427 hypoxia tolerance in ten cyprinid species. *Journal of Comparative Physiology B*. 2018;188(2):283-93.
40
41 428 48. Nilsson GE. A comparative study of aldehyde dehydrogenase and alcohol dehydrogenase
42 429 activities in crucian carp and three other vertebrates: apparent adaptations to ethanol production.
43 430 *Journal of Comparative Physiology B*. 1988;158(4):479-85.
44
45 431 49. Ip YK, Loong AM, Chng YR, Hiong KC, Chew SF. Hepatic carbamoyl phosphate synthetase (CPS) I
46 432 and urea contents in the hylid tree frog, *Litoria caerulea*: transition from CPS III to CPS I. *Journal of*
47 433 *Comparative Physiology B*. 2012;182(8):1081-94.
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
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

434