Phylogenetic and genomic analyses of the ribosomal oxygenases Riox1 (No66) and Riox2 (Mina53) provide new insights into their evolution

Katharina E Bräuer[1], Kevin Brockers[1], Jasmin Moneer[2], Annette Feuchtinger[3], Evi Wollscheid-Lengeling[4], Andreas Lengeling[4], Alexander Wolf[1]*

[1] Institute of Molecular Toxicology and Pharmacology, Helmholtz Zentrum München-German Research Center for Environmental Health, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany.

[2] Department of Biology II, Ludwig Maximillians University, Munich, Großhaderner Strasse 2, 82152 Planegg-Martinsried, Germany.

[3] Research Unit Analytical Pathology, Helmholtz Zentrum München, German Research Center for Environmental Health GmbH, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany.

[4] The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, United Kingdom

* Corresponding author

Bräuer KE: katharina.schubhart@helmholtz-muenchen.de

Brockers K: kevin.brockers@helmholtz-muenchen.de

Moneer J: moneer.j@gmail.com

Feuchtinger A: annette.feuchtinger@helmholtz-muenchen.de

Wollscheid-Lengeling E: Evi.Lengeling@roslin.ed.ac.uk

Lengeling A: andreas.lengeling@roslin.ed.ac.uk

Wolf A: alexander.wolf@helmholtz-muenchen.de

Abstract:

Background: Translation of specific mRNAs can be highly regulated in different cells, tissues or pathological conditions. Ribosome heterogeneity can originate from variable expression or post-translational modifications of ribosomal proteins. The ribosomal oxygenases RIOX1 (NO66) and RIOX2 (MINA53) modify ribosomal proteins via histidine hydroxylation. A similar mechanism is present in prokaryotes. Thus, ribosome hydroxylation may be a well-conserved regulatory mechanism with implications in disease and development. However, little is known about the evolutionary history of *Riox1* and *Riox2* genes and their encoded proteins across eukaryotic taxa.

Results: In this study, we have analysed *Riox1* and *Riox2* orthologues genes from 49 metazoen species and have constructed phylogenomic trees for both genes. Our genomic and phylogenetic analyses revealed that Arthropoda, Annelida, Nematoda and Mollusca lack the *Riox2* gene, although in the early phylum Cnidaria both genes, *Riox1* and *Riox2,* are present and expressed. Riox1 is an intronless single-exon-gene in several species, including humans. In contrast to *Riox2*, *Riox1* is ubiquitously present throughout the animal kingdom suggesting that *Riox1* is the phylogenetically older gene from which *Riox2* has evolved. Both proteins have maintained a unique protein architecture with conservation of active sites within the JmjC domains, a dimerization domain, and a winged helix domain. In addition, Riox1 proteins possess a unique N-terminal extension domain. Immunofluorescence analyses in Hela cells and Hydra animals identified a nucleolar localisation signal within the extended N-terminal domain of human RIOX1 and an altered subnuclear localisation for the Hydra Riox2.

Conclusions: Conserved active site residues and uniform protein domain architecture suggest a consistent enzymatic activity within the RIOX orthologues throughout evolution. However, differences in genomic architecture, like single exon genes and alterations in subnuclear localisation, as described for Hydra, point towards adaption mechanisms that may correlate with taxa- or species-specific environments. The diversification of *Riox1*/*Riox2* gene structures throughout evolution suggest, that functional requirements in expression of protein isoforms

and/or subcellular localisation of proteins may have evolved by requirements in lifestyle adaptions.

Key words:

Ribosome, ribosomal oxygenases, Fe(II) and 2-oxoglutarate dependent oxygenases, hydroxylation, JmjC, Jumonji, Mina53, NO66, single exon genes, intronless retroposed gene copies

Background

The central dogma of molecular biology has described gene expression as a straightforward process in which a gene is transcribed into mRNA followed by translation into a protein [1]. In recent years, a multitude of regulatory processes of gene expression have been identified on transcriptional and posttranscriptional levels. The mRNA itself can be subjected to regulatory events such as alternative splicing or RNA modifications [2, 3], but also mRNA translation rates or the translation selectivity of individual mRNA molecules can vary [4]. Initially the ribosome was thought to be a rather unaltered ribonucleoprotein particle responsible for translation of any incoming mRNA into a polypeptide. However, recent research has shown that ribosomes are highly specific translational machineries that underlie complex regulatory mechanisms in order to meet physiological requirements in different cell types or throughout development [4]. `Specialized´ ribosomes can be generated via various mechanisms, including changes in ribosomal protein composition or posttranslational modifications, like methylation or phosphorylation [5]. Very recently ribosomal oxygenases (ROXs) have been identified that modify ribosomal proteins via hydroxylation of amino acids [6-9]. The ROXs are a subgroup of

77 the enzyme superfamily of Fe(II) and 2-oxoglutarate (2OG) dependent oxygenases (2OG
78 oxygenases). All known human 2OG oxygenases catalyse transfer of molecular oxygen onto a
79 prime substrate, which can be either amino acids in proteins or nucleotides in DNA/RNA
80 molecules [10]. Due to the manifold set of their substrates, 2OG oxygenases have been
81 discovered to have central roles in many cellular processes including epigenetic regulation of
82 gene expression, control of transcriptional initiation and regulation of alternative splicing [11-14].
83 The ROXs, which include the nucleolar protein NO66 (official new nomenclature symbol
84 RIOX1) and the MYC-induced nuclear antigen MINA53 (new nomenclature symbol RIOX2),
85 directly modify ribosomal proteins [6, 8, 9]. NO66 / RIOX1 has been demonstrated to
86 hydroxylate histidine 216 (His216) in the ribosomal protein RPL8 (uL2), whereas MINA53 /
87 RIOX2 hydroxylates His 39 in RPL27a (uL15) [6] (Fig. 1A). 2OG-oxygenase-catalysed ribosome
88 hydroxylation is conserved from bacteria to humans. The Riox1/Riox2 counterpart in
89 prokaryotes, ycfD, hydroxylates arginine (R) 81 in Rpl16 [6] (Fig. 1B). In addition, Riox1 and
90 Riox2 share similarities in substrate binding with ycfD and exhibit similar conserved protein
91 domains [15]. Crystal structure analyses on recombinant enzymes revealed an N-terminal Jumonji
92 C (JmjC) domain, which harbours the active site and the iron-coordinating residues, characteristic
93 for all known 2OG oxygenases. A central dimerization-domain is responsible for homo-
94 oligomerization *in-vitro* and the C-terminus contains a winged-helix (WH) domain [15] (Fig. 1C).
95 Both, human RIOX1 and RIOX2 have been described to be involved in cancer cell growth.
96 *RIOX2* is a myc-target gene [16]. Its expression is upregulated in several cancers, including lung
97 and breast cancer, and knockdown of RIOX2 in A549 cells inhibited cell proliferation [17, 18].
98 Elevated RIOX2 expression has been described in non-small cell lung cancers [19] and was
99 reported to be associated with invasive colorectal cancer [20]. In addition, RIOX2 regulates
100 immune responses as a transcriptional co-repressor of the *interleukin-4* (*Il4*) encoding gene [21].
101 It has been described to polarize T helper 2 (Th2) cell responses in atopic pulmonary
102 inflammation and to have a role in parasitic worm expulsion [22]. Riox1 is involved in osteoblast
103 differentiation [23] and variations in its expression level has been reported to regulate skeletal
104 growth and bone formation in mice [24, 25].

Here we provide the first comprehensive sequence analyses of the ribosomal oxygenases Riox1 and Riox2 in different eukaryotic species. We compared the domain architecture of both proteins and their exon-intron gene structures across a wide range of metazoan species. In addition we used immunofluorescence approaches to investigate expression in human cells and in the Cnidaria *Hydra vulgaris*.

Results

Evolutionary sequence and protein domain architecture conservation of the ribosomal oxygenases Riox1 and Riox2

A sequence alignment of human RIOX2 (MINA53, 465 amino acids (aa)) and human RIOX1 (NO66, 641 aa) proteins revealed 23.2% identical and 16.4% similar amino acids (Fig. 2A). Both proteins exhibit the same protein domain structure, including a JmjC domain, a dimerization domain and a winged-helix (WH) domain [15]. The RIOX1 protein has an N-terminal extension of 184 amino acids, which is absent in the RIOX2 protein (Fig. 2A). RIOX1 and RIOX2 are nuclear proteins with a strong accumulation in nucleoli [26, 27] (Fig. 2B,C). To identify *Riox1* and *Riox2* orthologs in other eukaryotic species we started searching the genomes of the model organisms *Mus musculus*, *Gallus gallus*, *Danio rerio*, *Caenorhabditis elegans* and *Drosophila melanogaster*. An Ensembl genome browser search with the human RIOX1 protein sequence (Ensembl: ENSG00000170468; Uniprot: Q9H6W3) revealed orthologous proteins in *M. musculus*, *G. gallus*, *D. rerio*, *C. elegans* and *D. melanogaster*. The *M. musculus* Riox1 protein sequence (Ensembl: ENSMUSG00000046791; Uniprot: Q9JJF3) showed 75,2% identity to the human sequence, with most sequence variation in the N-terminal extension domain (Supplementary Fig. 1). The *G. gallus* Riox1 protein (Ensembl: ENSGALG00000020454; Uniprot: Q5ZMM1) showed 49,3% identity (Supplementary Fig. 2), the *D. rerio* Riox1 43% (Ensembl: ENSDARG00000067838; Uniprot: A3KP59) (Supplementary Fig. 3), the *D.*

melanogaster Riox1 33,1% (Ensembl: FBgn0266570; Uniprot: E2QD64) (Supplementary Fig. 4) and the *C. elegans* Riox1 protein (jmjc-1, Ensembl: WBGene00020902; Uniprot: O01658) 28,7% aa sequence identity (Supplementary Fig. 5).

A search for Riox2 sequences in the same set of organisms revealed no Riox2 / Mina53 orthologes in *C. elegans* and *D. melanogaster*. Searches with human RIOX2 (Ensembl: ENSG00000170854; Uniprot: Q8IUF8) or *D. rerio* Riox2 (Ensembl: ENSDARG00000036359; Uniprot: F1R7K2) sequence respectively revealed no homologous sequences, but detected the above described Riox1 / No66 sequences for *C. elegans* and *D. melanogaster* as the sequences with highest homology (summarized in Fig. 2D).

Individual alignments of Riox2 / Mina53 sequences with the human RIOX2 showed 76,3% identical amino acids for the *M. musculus* Riox2 sequence (Ensembl: ENSMUSG00000022724; Uniprot: Q8CD15) (Supplementary Fig. 6), 70,8% for the *G. gallus* Riox2 (Ensembl: ENSGALG00000039302; Uniprot: E1C6P1) (Supplementary Fig. 7) and 57,2% for the *D. rerio* Riox2 (Ensembl: ENSDARG00000036359; Uniprot: F1R7K2) (Supplementary Fig. 8).

Genome sequence searches of additional invertebrate species from the Arthropoda, Annelida, Nematoda, and Mollusca clades did also not identify any putative *Riox2* orthologous genes (results discussed below), but the early metazoan animal *Hydra vulgaris* of the phylum Cnidaria exhibited both, Riox1 and Riox2 sequences (Fig. 3). That prompted us to analyse the expression of the hydra orthologues in more detail.

Characterization and cellular localization of hydra Riox1 and Riox2 proteins

The fresh-water polyp *Hydra* is a pre-bilaterian animal of the phylum Cnidaria [28]. Database searches of the *Hydra vulgaris* genome [29] predicted two sequences encoding for *Hydra Riox2* (HyRiox2) and *Hydra Riox1* (HyRiox1). Amplification with corresponding primers, cloning and subsequent sequencing identified the predicted HyRiox2 (551aa) and HyRiox1 (628aa) protein sequences encoded in *Hydra* cDNAs (Fig. 3A, B). An alignment of these *Hydra* protein

sequences with the human RIOX1 and RIOX2 proteins revealed 35,0% and 33,7% aa sequence identity, respectively (Supplementary Figures 9 and 10). A sequence comparison of the two identified *Hydra* ROXs showed, that HyRiox2 and HyRiox1 share 19,7% identical amino acids (Fig. 3C). ROXs proteins are characterized by a DSBH fold (JmjC domain) that harbours the active site, including residues for co-factor (Fe(II)) and co-substrate (2OG) binding [10]. Based on the crystal structures of human RIOX1 and RIOX2 proteins [15], both *Hydra* sequences exhibit the characteristic Fe(II) binding motif HxD/E…H (HyRiox2: H178xD180…H240; HyRiox1: H330xD332…H395) and also a conserved 2OG C5-carboxylate-interacting lysine residue (HyMina53: K193; HyNo66: K345) (Fig. 3C; Supplementary Fig. 9-10). Both *Hydra* homologs have also the characteristic ROX protein domain architecture with JmjC, dimerization, and winged helix domains. In addition, the *Hydra* Riox1 homolog HyRiox1 exhibits an N-terminal extension domain proximal to the JmjC domain. Interestingly, the HyRiox2 contains two extended sequence segments in its dimerization domain and one extended sequence stretch in the winged helix domain. No corresponding sequences of these extensions are detectable in the human RIOX2 (MINA53) protein. The sequence extension in the HyRiox2 winged helix domain comprises a stretch of 64 additional residues, rich in the charged amino acids lysine (K), aspartate (D) and glutamate (E) (42 of the 64aa) (Supplementary Fig. 9).

To analyse the subcellular localization of HyRiox2 and HyRiox1, we expressed GFP-tagged full-length proteins in *Hydra* cells of intact animals after transfection with a particle gun [30]. Confocal imaging of ectopically expressed GFP-tagged HyRiox1 confirmed the nuclear localization of the protein with a prominent accumulation in nucleoli (Fig. 3D,E). Transfection of HyRiox2 constructs revealed also expression in the nucleus, however in contrast to the human RIOX2, HyRiox2 did not localise in nucleoli (Fig. 3F,G).

Sequence analyses of Riox1 and Riox2 proteins from different species

The ROXs domain topology of homologous Riox1 (No66) proteins is similar in all analysed species. The lengths of their JmjC domains (NO66: 236-242aa), dimerization domains (NO66:

86-93aa) and winged helix domains (NO66: 129-133aa) are comparable in the analysed organisms, whereas the lengths of the N-terminal extension varies from 87 in *D. rerio* to 286 amino acids in *C. elegans* (Fig. 4). An alignment of these Riox1 protein sequences showed strong conservation of the Fe(II) and 2OG-binding residues in the catalytic site of these enzymes (Fig. 4). The sequence homology implies a conserved function of Riox1 (No66) across species. Structure analysis of human RIOX1 co-crystallised with an Rpl8 peptide identified interactions of Arg297, Tyr328 and Ser421 with the substrate peptide [15]. These residues are conserved in all Riox1 proteins across the analysed species (Fig. 4). Similarly, in Riox2 (Mina53) (Fig. 5), four primary amides (Asn101, Gln136, Gln139, Asn165) and Ser257 in human RIOX2 have been identified to interact with its Rpl27a substrate peptide. These residues and the Fe(II) and 2OG-binding sites are all conserved in Riox2 / Mina53 proteins across species throughout evolution (Fig. 5). Alignment of Riox2 sequences from human, mouse, chicken, zebra fish and *Hydra* revealed additional amino acid stretches in the zebra fish and *Hydra* proteins, which are not present in the other species. The zebra fish Riox2 displayed an additional 30aa stretch between the JmjC and the dimerization domain (Fig. 5). The *Hydra* sequence has an extended dimerization domain and an additional 64aa stretch within its winged helix domain (Fig. 5). We identified these aa-insertions also in the protein encoded by the cloned *Hydra* cDNAs (see above).

The N-terminal extension domain in Riox1

Despite the characteristic domain structure of the ROXs, an obvious difference between Riox2 and Riox1 is the N-terminal extension present in the Riox1 sequences. This N-terminal extension domain has no assigned function yet, but the N-terminus of Riox1 (aa 1-31) has been predicted to possess a nuclear localization signal [27]. We further investigated this hypothesis using green fluorescent protein (GFP) fusion reporter experiments. Full-length human RIOX1 tagged with GFP localized predominantly in nucleoli of transfected HeLa cells (Fig. 6A). In contrast, a GFP-tagged deletion mutant lacking aa 1-31 displayed an exclusive cytoplasmatic localisation (Fig. 6B). When we fused the N-terminus of human RIOX1 (aa 1-45) to GFP, the fusion protein

localised to the nucleus and accumulated in nucleoli as well (Fig. 6C). We used two different bioinformatic tools (NLSmapper & NLStradamus) [31, 32] to predict nuclear localization signals (NLS) in Riox1 of several species. In each species, the NLS motive was identified with NLSmapper and NLStradamus tools within the N-terminal extension of the analysed Riox1 proteins (Fig. 6D). In contrast, the mechanism of how Riox2 is directed into the nucleus and/or nucleolus is to our knowledge not known yet. Both NLS prediction tools did not identify corresponding NLSs in all analysed Riox2 protein sequences (data not shown).

Comparative genomic analysis of *Riox1* and *Riox2* across the animal kingdom

Analyses of the human *RIOX1* and *RIOX2* genes using the Ensembl genome browser portal revealed 10 exons for human *RIOX2*. The gene encodes 9 different transcripts and comprises 9 coding exons and alternatively used non-coding 5'-UTR and 3-UTR exons. The largest transcript, RIOX2-001 (ENST00000333396.11), encodes a protein of 465 aa and encompasses a locus of 30,639 base pairs (bp) on human chromosome 3 (Fig 7A). Surprisingly, in contrast the human *RIOX1* gene encodes one transcript RIOX1-003 (ENST00000304061.7) and is an intronless gene, spanning 2,428 bp with 5'- and 3'-UTRs on chromosome 14 (Fig. 7B). Mapping of the characteristic ROX protein domains (JmjC, dimerization, winged helix) onto the exon-intron gene structure of *RIOX2* revealed that the JmjC domain is encoded by exons 2-4 and parts of exon 5. *RIOX2* exons 5, 6 and parts of exon 7 encode the dimerization domain, whereas the winged helix domain is encoded by exons 7, 8, 9 and parts of 10 (Fig. 7A). When we compared genomic structures of *Riox1* and *Riox2* orthologues throughout evolution we found that *Riox2* comprises ten (nine coding and one non-coding 3'-UTR) exons in human, mouse and chicken or 9 coding exons in *Xenopus*, 11 exons in zebra fish and 9 exons in *Hydra* (Fig. 7C).

To gain additional insights into the evolution of the *Riox1* and *Riox2* genes we included additional metazoan species that represent key taxa from the animal kingdom and constructed phylogenetic

237 trees for both genes. Protein sequences encoded by *Riox1* and *Riox2* orthologues were identified
238 in the Ensembl and EnsemblMetazoa genomic resources using the human RIOX1 and RIOX2
239 protein reference sequences in Protein BLAST (blastp) search routines or the Ensembl evidence-
240 based annotation of orthologues from pre-existing whole-genome pairwise alignments which are
241 available on both Ensembl portals. Obtained protein sequences were validated as being true
242 orthologs using the following criteria: (i) Independent blastp searches with the *Homo sapiens*
243 RIOX1/RIOX2 and *Hydra vulgaris* Riox1/Riox2 reference sequences mapped to a single locus in
244 the respective genome with an E-value $<1x10^{-40}$; (ii) Multiple blastp sequence hits (>2) where
245 obtained with the human and *Hydra* query sequence in the same locus; and (iii) The extracted
246 homologues, full-length protein sequence followed the characteristic domain structure of ROX
247 proteins with a JmjC, dimerization and winged helix domain present. As an additional criterion
248 for correct assignment of *Riox1* orthologs, we also included the presence of the above described
249 N-terminal extension domain in the candidate sequence. All Riox1 orthologs sequences
250 downloaded from the Ensembl portals matched these criteria with the exception of the *Trichoplax*
251 *adhaerens* sequence, which contained a JmjC domain but had neither a dimerization nor a winged
252 helix domains. Thus, we could not classify this *Trichoplax adhaerens* JmjC protein as an *Riox1* or
253 *Riox2* orthologues gene, however, we included the sequence in the multiple sequence alignments
254 to root the generated phylogenetic trees to a basic metazoan species for which we had sufficient
255 high quality sequence information available on the EnsemblMetazoa portal. The protein
256 phylogenetic analysis of *Riox1* orthologs together with the architectures of their genes as
257 annotated in Ensembl reveals that intronless *ROX1/Riox1* genes have evolved independently at
258 least two times in the animal kingdom. Intronless *RIOX1/Riox1* genes are present in the
259 mammalian lineage (humans, mice, and rats), in the marsupial opossum *Monodelphis domestica*,
260 and in chicken as the representative species of the avian lineage (Fig. 8). Fish, amphibians and
261 other chordates all possess multi-exon genes with up to 15 exons (*Latimeria chalumnae*) and
262 differently sized and structured non-coding 5'- and 3'-UTR exons (Fig. 8). Within the clade
263 Insecta at least another origination of an intronless *Riox1* gene must have taken place. We
264 identified intronless loci in *Drosophila melanogaster, Drosophila ananassae,* and 7 other

Drosophila species that encoded intact, full-length open reading frames (ORFs) for Riox1 proteins within a size range of 653 to 907 aa (Fig. 8, and data not shown). Importantly, these intronless *Riox1* genes are not clustered within the Insecta clade. Other *Drosophila* species like *D. erecta*, *D. mojavensis*, and *D. sechellia* have *Riox1* genes that consist of 2 exons and phylogenetically related flies such as *Anopheles gambiae*, *Culex quinquefasciatus*, *Aedes aegypti*, and *Belgica antarctica* also possess *Riox1* genes with two exons. Other representative species of the Insecta clade such as *Tribolium castaneum* and *Atta cephalotes* have multi-exon *Riox1* genes, however, in contrast to the majority of other analysed metazoan species the number of exons in these insects does not exceed a total number of 6.

Protein sequence alignments [33] of Riox1 and Riox2 JmjC domains obtained from 49 species and the *E. coli* ROX protein YcfD and construction of phylogenetic trees using maximum likelihood and bootstrap analysis (data not shown) revealed congruent tree topology. Our phylogenetic study shows a strong node support for a Riox2-Jmjc domain branch clearly separating this from Riox1-Jmjc domains (Fig. 9). Within the Riox2 branch only closely related taxa (e.g. *Ciona*) or taxonomic higher ranks (e.g. Chordata) possess high node support making an assumption about the diversification process and possible implied functions of Riox2 in different taxa difficult. The strong support for the branching node Riox2 *Hydra vulgaris*/ecycfD *Escherichia coli* to the remaining Riox2 possessing species suggests an early invention of Riox2, which became lost from early Bilateria species to Deuterostomia where it is present. Interestingly, Riox2 showed the closest sequence relatedness to the nematode Riox1 analogs, demonstrating that *Riox2* has evolved in the chordata lineage from an ancestral *Riox1* gene (Fig. 9). Again strong support for Riox1 Jmjc domains relatedness is given for closely related taxa (e.g. *D. rerio* + *O. latipes*, *B. mori* + *H. melpomene*) or higher taxa ranks (e.g. Mammalia, Chordata, Mollusca). Weak node support (e.g. *H. robusta* + Chordata, *S. purpuratus* + *L. salmonis*/Insecta) indicates that additional research is needed to fully understand the evolution of this gene and its functional diversification.

Discussion

Although both proteins, Riox1 and Riox2 are present in *Hydra vulgaris*, our phylogenetic analyses revealed the absence of Riox2 in all other investigated invertebrate lineages with the only exception of the small phylum Priapulida represented by the marine worm *Priapulus caudatus*. All analysed Chordata and Echinodermata, and Hyperoartia possess a *Riox2* gene (Fig 9). This suggests that at the basis of the Chordata, which are represented by the Tunicata *Ciona intestinalis* and *Ciona savignyi*, Riox2 was first evolved from a common ancestral Riox gene. In contrast, Riox1/No66 is ubiquitously present throughout the animal kingdom. This suggests that *Riox1* is the phylogenetically older gene from which *Riox2* evolved, likely two times independently in early Cnidaria (*H. vulgaris*) and Priapulida (*P. caudatus*). The strong sequence homology of the *H. vulgaris* Riox2 JmjC domain to the JmjC domain of the *E. coli* ycfD protein (Fig. 9) suggests that Riox2 has a ribosomal oxygenase function like its prokaryotic analog [6].

Comparison of the genomic architecture manifested a common exon/intron structure for *Riox2* genes throughout evolution. In all analysed organisms, the *Riox2* gene consists of 9-10 exons. In contrast, *Riox1* genes possess strikingly different gene structures throughout the animal kingdom. All Mammalia and Aves (birds) have intronless single exon *Riox1* genes (Fig. 8). Insecta have *Riox1* genes with few exons (2 to 6) with some species of *Drosophila* possessing also single exon genes (Fig. 8).

Intronless genes usually evolve through retroposition or retroduplication of an ancient intron-containing gene in which a new copy of the gene is generated by reverse transcription of a spliced mRNA from a parental gene which is then inserted as a novel locus into a new location somewhere in the genome [34, 35]. The presence of intronless *Riox1* genes in some *Drosophila* species and the occurrence of *Riox1* genes in other insects that just contain a few exons (2-6) suggest two possible scenarios of gene evolution in the Insecta lineage. Either the *Drosophila* clade has acquired an intronless *Riox1* gene at the branch point of divergence from other fly species and this intronless gene copy has been maintained in particular *Drosophila* species or a

single exon *Riox1* gene was evolved by retroduplication at the branch point of all Insecta and introns were then independently acquired in different fly subclades with the exception of the *Drosophila* clade, in which intronless *Riox1* genes have been maintained in some but not all *Drosophila* species. In the latter scenario, an evolutionary selection pressure must have preserved the maintenance of those intronless genes in particular *Drosophila* species. Reasons for such a preservation of intronless *Riox1* loci throughout evolution are unknown but as discussed below different explanations might be feasible.

Approximately 9% of all genes in the human genome are intronless [36]. Intronless genes are typically expressed at a lower level compared to intron-containing genes and evolve through retroduplication events in which mRNA transcripts are reverse-transcribed into DNA and inserted into genomes [34, 35, 37]. Such retroposition events are an important mechanism for the origination and evolution of new genes from ancestral, parental source genes [35, 38]. New retropositioned genes may eventually evolve new functions through recruitment of promoter and/or enhancer elements from other regions in the genome and may thereby acquire new spatiotemporal expression patterns, which are different from the ancestral source gene [38]. This can then result in novel tissue- or cell-specific functions of retroposed intronless genes [34]. Interestingly, *Riox1* single exon genes are mainly present in taxa with terrestrial lifestyles. Taxa representing animals with predominantly aquatic lifestyles have multiexonic genes with up to 18 exons (Fig. 8). The reason for this divergence in *Riox1* gene architectures is unknown but it may be related to oxygen content and physiological requirements in these divers environments. Riox proteins are members of the JmjC protein superfamily and as such their enzymatic dioxygenase activity highly depends on the concentration and availability of molecular oxygen [10]. Multiexonic genes tend to have substantially broader expression patterns than single exon genes and alternative splicing allows evolution and emergence of multiple protein isoforms that are functionally distinct by having differential expression patterns and/or biochemical properties [38]. Thus, multiexonic genes and their capability to express multiple Riox1 isoforms with different functions might suit aquatic organism, that are dependent on dissolved oxygen for respiraton

346 better to adapt to such environments. In contrast, intronless genes have generally a narrower
347 expression pattern and are usually also expressed at lower levels compared to multiexonic genes.
348 With human and mouse *Riox1* genes we have undertaken a preliminary *in silico* expression
349 analysis with publically available microarray datasets and expression atlases of primary cells and
350 tissues using the network analysis tool BioLayout / Miru [39, 40]. Searches for *Riox1* co-
351 expression networks did not identify defined clusters of co-expressed genes, probably due to the
352 low level and ubiquitous expression pattern of *Riox1* in most analysed cells and tissues (data not
353 shown). However, many intronless genes are known to be associated with stress or immune
354 response induced signalling pathways, which require fast induction of gene and protein
355 expression. Here, the circumvention of complex primary mRNA processing steps such as splicing
356 may insure faster response rates of expression induction. Examples of such intronless genes are
357 members of the heat shock 70 gene family [41], interferon-coding genes [42] and genes encoding
358 G protein-coupled receptors [43]. Future research may identify conditions under which *Riox1*
359 expression is specifically induced and these might be related to stress inducing conditions. This
360 hypothesis is supported by the finding that the closely sequence related paralog of *Riox1* and
361 *Riox2*, *Jmjc-1* of *Caenorhabditis elegans*, is involved in the regulation of an evolutionary
362 conserved stress-response network [44].

363 Our protein sequence analyses of the ribosomal oxygenases Riox1 (No66) and Riox2 (Mina53) in
364 several eukaryotic organisms revealed a conserved protein domain architecture, which initially
365 has been identified by X-ray crystallography in human RIOX2 (Mina53), RIOX1 (No66) and the
366 prokaryotic yfcD. An *N*-terminal JmjC domain harbours the iron-coordinating active site residues
367 [15]. The dimerization domain encourages homo-oligomerization at least *in vitro* with
368 recombinant protein (dimer/tetramer) [15, 45], and the *C*-terminal winged helix (WH) domain has
369 no assigned function yet. Generally the WH domain represents a helix-turn-helix motif, which has
370 been shown to participate in protein-DNA or protein-protein interactions [46]. However, the
371 lengths of these three domains reside within only a small range of variation, although some
372 organisms exhibit some specific extended sequence elements. As genome sequence coverage and

qualities of genome assembly still varies among the analysed organisms, some of these extended sequence elements might represent protein sequence annotation mistakes generated by Ensembl's automatic genebuild pipeline. In the cnidarian *Hydra vulgaris* the Riox2 sequence encodes two extended amino acid stretches with no matches in the human sequence. Cloning of the respective *Hydra* Riox2 from cDNA confirmed the occurrence of those sequence elements, one of which (64 `additional´ amino acids in the WH domain) consists of mainly charged amino acids (K, D, E). Interestingly, cellular localisation of the *Hydra* Riox2 is also slightly different to the human protein. While human RIOX2 shows nucleoplasmic localisation with distinct accumulation in nucleoli (Fig. 2) [26, 27], the *Hydra* orthologue is a nuclear protein but seem to relieve nucleoli (Fig. 3). Human RIOX2 catalyses histidine hydroxylation in the ribosomal protein Rpl27A [6]. Crystal structure analyses of human RIOX2 in complex with the Rpl27A substrate-peptide discovered four primary amides (Asn101, Gln136, Gln139, Asn165) and Ser257 to be essential for substrate-enzyme interaction [15]. In *Hydra* Riox2 these residues are conserved (Fig. 5) and also the iron coordinating residues HxD…H in the active site imply a fully functional enzyme. Further detailed molecular analyses would be necessary to unravel the enzymatic activity of *Hydra* Riox2. The original described activity of the human RIOX2 protein was demethylation of histone 3 lysine 9 tri-methylated (H3K9me3) residues [47], however, the evidence for this assignment is controversial [48]. Whether or not Riox2 might have dual functionality as a ribosomal oxygenase and/or as a histone demethylase needs to be separately investigated in each species, which possesses an orthologous gene. Such studies should be further supported with structural approaches, such as those described for another *Hydra* 2OG oxygenase, Jmjd6 [49-51]. It is likely that throughout evolution Riox2/Mina53 proteins have acquired different or additional substrate targets and hydroxylation activities.

In *Hydra* both proteins, Riox2 and Riox1, are expressed. Riox1 in *Hydra* and all other analysed organisms displayed an *N*-terminal extension of variable length with unknown function, which discriminates Riox1 and Riox2 sequences. Initial investigations identified a nuclear localisation signal in aa1-30 of the human RIOX1 [27]. For Riox1 in several other organisms NLS sequences

have been also predicted in the N-terminal regions (Fig. 6). Our deletion mutants of human RIOX1 (aa32-641) confirmed cytoplasmic localisation upon lack of proposed NLS (Fig. 6). For the first time our GFP fusion experiments unravelled that aa1-45 of human RIOX1 are actually sufficient to target GFP into nucleoli, which resembles the nuclear localization pattern of full-length RIOX1. Generally nucleolar localisation signals (NoLS) are not very well characterized, but are known to be sometimes part of nuclear localisation signals (NLS) and composed of Arg and Lys residues [52].

Conclusions

In conclusion, our phylogenetic and genomic analysis of Riox1 and Riox2 has revealed the maintenance of a unique protein architecture with conservation of active enzymatic sites throughout evolution in the whole animal kingdom. This strongly suggests that both orthologues have a consistent enzymatic function as Fe(II) and 2OG-dependent dioxygenases with likely ribosomal protein hydroxylation as their main function. However, at the genomic level both orthologs show diversifications in the evolution of their gene architectures and presence or absence of the *Riox2* gene in different taxa. Many higher vertebrates and certain fly species (e.g. *Drosophila*) possess an intronless *Riox1* orthologue and *Riox2* is absent in most invertebrates. This suggests that Riox1 is the evolutionary older JmjC-domain containing protein with ribosomal oxidase function. The more complex gene structure of *Riox1* with multiple exons and introns in lower, marine metazoans suggests a different, perhaps, more complex regulation of protein expression in these organisms. If this hypothesis is true that complex *Riox*1 gene structures and expression regulation correlates with adaptation to different environments needs to be studied in the relevant species in the future. Differences in expression of protein isoforms and/or subcellular localisations of Riox orthologues in different species, as shown here for *Hydra*, maybe explained by different functional requirements and evolutionary lifestyle adaptation of different taxa.

Methods

Molecular cloning

Human full-length and truncated RIOX1 and RIOX2 sequences were sub-cloned into the mammalian expression plasmid pEGFP-N1 (Clontech). *Hydra Riox1* and *Riox2* were amplified from cDNA (*Hydra vulgaris*, total RNA extracted from whole animals, primer: hyNO66_NheF 5'-CAGGCTAGCATGAATAACAACAAAGTATCAGC-3', hyNO66_XmaR 5'-GACCCGGGTGTATGGACCAATGGAACC-3' for *Riox1*, and hyMina53_NheF CAGGCTAGCATGGTGAAACGCAAAGGTTC, hyMina53_XmaR GGCCCGGGTTTGATTTCAATCAAATCATCAC, for *Riox2*, respectively) and sub-cloned into the *Hydra* eGFP expression plasmid pHotG [30] by using the *Nhe*1 and *Xma*1 restriction sites.

Cell culture and transfection

HeLa cells (ATCC; CCL-2) were cultured in Dulbecco´s modified Eagle´s medium (DMEM) supplemented with 10% fetal calf serum (FCS), penicillin (100 Uml^{-1}) and streptomycin (100 μgml^{-1}) at 37°C, 5% CO_2. Cells were transfected with expression constructs using Lipofectamine 2000 (Invitrogen) according to manufacturer´s instructions.

Hydra culture

Hydra vulgaris strain Basel [53] were held in mass culture in hydra medium (0,1mM KCL, 1mM NaCl, 0,1mM $MgSO_4$, 1mM Tris, 1mM $CaCl_2$), at a constant temperature of 18°C and were fed regularly with freshly hatched *Artemia nauplii*.

Transfection of Hydra cells

2,4mg Gold particles (1,0μm, BioRad) were coated with 10μg plasmid DNA according to instructions of manufacturer. They were introduced into the Hydra cells with a Helios gene gun system (BioRad) as described [30].

Fixation and mounting of *Hydra*

Animals were relaxed in 2% urethane in hydra medium and fixed with 4% paraformaldehyde (in PBS) at room temperature for one hour. After three washes with PBS, they were counterstained for DNA with DAPI (Sigma, 1μg/ml) and mounted on slides with Vectashield mounting medium (Axxora).

Confocal imaging of *Hydra*

Light optical sections were acquired with a Leica TCS SP5-2 confocal laser-scanning microscope. Fluorochromes were visualised with the 405 laser with an excitation wavelength of 405nm and emission filters 413 to 443nm for DAPI. The argon laser with excitation wavelength of 488nm and emission filters 496 to 537nm was used for GFP. Image resolution was 512x512 pixel. To obtain an improved signal-to-noise ratio, each section image was averaged from three successive scans.

Immunostaining and microscopy of Hela cells

HeLa cells were grown to 50-70% confluence on 18mm diameter coverslips. 24 hours post-transfection cells were fixed with 4% paraformaldehyde (10 min). GFP-expressing cells were stained with 1μg/ml DAPI and slides mounted in Vectashield. For antibody staining cells were permeabilized after fixation with 1% Triton X-100 in PBS and subsequently kept in blocking solution for an hour (10%FCS, 0.2% Tween-20 in PBS). Primary antibodies: α-Riox1 (Mina53, ab169154, Abcam), α-Riox1 (NO66, ab113975, Abcam) and α-UBF (sc-9131, Santa Cruz).

Secondary antibodies: Alexa Fluor 488 chicken anti-mouse (A21200, Thermo Fisher), Alexa Fluor 594 donkey anti-rabbit (A21206, Thermo Fisher). Slides were imaged with a fluorescence microscope Carl Zeiss LSM 510 META.

Ensemble database searches

Orthologes of *Riox1* and *Riox2* genes were identified using Ensembl (http://www.ensembl.org/info/about/species.html) and EnsemblMetazoa (http://metazoa.ensembl.org/species.html) portals by employing pblast searches. As described in the result section, the genome of each selected species was queried with the human or *Hydra vulgaris* RIOX1/Riox1 (NP_078920.2 and XP_002157896.3) and ROX2/Riox2 (NP_694822.2 and XP_002167270) protein reference sequences, respectively. For phylogenetic analyses of orthologues proteins only sequences were included in which both, the human and hydra pblast queries, matched to a single locus of the selected species and the identified protein displayed the characteristic domain architecture of Riox1 and Riox2 proteins as described below in the results section. Species were selected throughout the animal kingdom to represent main taxonomic classes, where available with at least two species per class depending on genome sequence coverage and quality of gene structure annotations. *Hydra vulgaris* protein sequences and gene annotations for *Riox1* and *Riox2* were obtained from https://metazome.jgi.doe.gov/pz/portal.html.

Multiple-sequence alignments

FASTA-formatted amino-acid sequences were aligned using the MAFFT 7 tool (http://www.ebi.ac.uk/Tools/msa/mafft/) provided by the European Bioinformatics Institute (EBI) using the ClustalW algorithm [54]. The resulting multiple amino acid sequence alignment was used to generate phylogenetic trees.

Construction of phylogenetic trees

To analyse the phylogenetic relationship of 41 Riox1 proteins from different species and to infer the evolutionary relationship of Ribosomal oxygenases, the JmjC domain sequences of Riox1 (No66) and Riox2 (Mina53) from 49 species and of YcfD (ecycfD), the ROX protein from *Escherichia coli* [6], were subjected to a maximum likelihood analysis using the online phylogenetic tool W-IQ-TREE (Version 1.5.4 at http://iqtree.cibiv.univie.ac.at) [55]. In the IQ-Tree webserver the 'Substitution model' and the default 'Auto' settings were selected to determine the best-fit substitution model followed by tree construction. Within the 'Branch Support Analysis' the default settings of an ultrafast bootstrap analysis with 1000 replicates was used, with maximum number of iterations set at 1000 and a minimum correlation coefficient of 0.99 and for the 'Single branch tests' the SH-aLRT branch test with 1000 replicates was selected [56].

Bioinformatic prediction of nuclear localisation sites

Nuclear localisation signal searches have been performed with NLS Mapper (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi) [32] and NLStradamus (http://www.moseslab.csb.utoronto.ca/NLStradamus/) [31].

List of abbreviations

ROX: ribosomal oxygenase, 2OG: 2-oxoglutarate, His: histidine, JmjC: Jumonji C, WH: winged-helix, Th2: T helper 2, K: lysine, D: aspartate, E: glutamate, GFP: green fluorescent protein, NLS: nuclear localisation signal, NoLS: nucleolar localisation signal, UTR: untranslated region, ORF: open reading frame

Declarations

Ethics approval and consent to participate: Not applicable.

Consent to publish: Not applicable.

Availability of data and materials: The datasets analysed during the current study are available in the Ensembl database (http://www.ensembl.org/info/about/species.html), the EnsemblMetazoa database (http://metazoa.ensembl.org/species.html) or the Hydra Genome Database (https://metazome.jgi.doe.gov/pz/portal.html). The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests: The authors declare no competing or financial interests.

Funding: This work was supported by the institute strategic grant funding from the Biotechnology and Biological Sciences Research Council (BBSRC, UK; to AL and EWL) and the Deutsche Forschungsgemeinschaft (DFG BO1748/12-1 to JM). Funding bodies had no impact on the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors´contributions: AW designed the study. KEB, AL and AW performed the genomic analyses. EWL performed the phylogenetic analyses. KB produced the deletion constructs. KEB, KB and AF analysed proteins in human cells. KEB cloned the Hydra sequences. KEB and JM performed the experiments in Hydra. AL and AW wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements: The authors thank Prof Angelika Böttger and Prof Charles N David for support and advice.

Figure legends

Figure 1: The ribosomal oxygenases (ROXs) are a subgroup of Fe(II) and 2OG-dependent oxygenases that modify the ribosome and are present in pro- and eukaryotes. (A) The human ROXs RIOX2 / MINA53 and RIOX1 / NO66 hydroxylate histidine residues in the ribosomal proteins

552 Rpl27a and Rpl8, respectively, whereas (B) the *E.coli* ycfD protein hydroxlyates an arginine in Rpl16
553 [6]. Protein sequence and crystal structure analyses confirmed a similar protein-domain architecture
554 for the three proteins [15]. (C) They consist of a JmjC-domain (red), a dimerization domain (brown)
555 for homo-oligomerization and a winged-helix domain (blue) amino acid triad HxD…H that
556 coordinates the iron and is essential for catalytic activity is indicated in green (C).

557

558 **Figure 2: Clustal omega sequence alignment [33] of human RIOX2 and RIOX1 amino acid**
559 **sequences.** (A) Protein domains are indicated in red (JmjC), brown (dimerization) and blue (winged
560 helix). The RIOX1-specific N-terminal extension is marked in grey. The Fe(II) and 2OG-binding
561 residues are highlighted in green or purple respectively. Both, human RIOX2 and RIOX1 localise to
562 nucleoli. Immunofluorescent staining of endogenous RIOX2 (α-Mina53) in HeLa cells showed co-
563 localisation with (B) the nucleolar marker protein upstream-binding factor 1 (α-UBF) and (C) with
564 RIOX1 (α-NO66). Scalebar: 10 μm. (D) BLAST searches in genomes of several species revealed
565 Riox1 to be present in all tested species from *Hydra* to human. In contrast Riox2 seemed to be absent
566 in *C.elegans* and *D.melanogaster*.

567

568 **Figure 3: Riox1 and Riox2 expression in Hydra.** (A, B) PCR-Amplification of *HyRiox2* and
569 *HyRiox1* from *Hydra* cDNA resulted in two individual bands. Subsequent cloning and sequencing
570 confirmed presence of *HyRiox2* and *HyRiox1*. (C) Alignment of both *Hydra* sequences, with
571 highlighted iron-binding motif (HxD…H) (green) and predicted lysine-residue for 2OG binding
572 (blue). (D, E) Expression of GFP-tagged HyRiox1 in Hydra animals displayed nuclear localisation
573 and strong accumulation in nucleoli. (F, G) HyRiox2 is also localised in nuclei, but an accumulation
574 in nucleoli is not detectable. DNA-stain: DAPI. Scalebar: 10 μm.

575

576 **Figure 4: Riox1 protein alignment.** Clustal omega alignment of Riox1 protein sequences from
577 *C.elegans*, *D.melanogaster*, *H.vulgaris*, zebrafish, chicken, mouse and human. The protein domains
578 JmjC (red), dimerization (brown) and winged-helix (blue) are indicated based on the human sequence

579 [15]. Lengths of the individual N-terminal extension domains are indicated (grey). The prospective
580 iron-binding motif HxD…H (green) and the 2OG C5-carboxylate-binding residue (K, purple) are
581 conserved in all species. Crystal structure analysis of human RIOX1 with substrate Rpl8 identified
582 R297, Y328 and S421 residues of human RIOX1 involved in Rpl8 peptide binding [15] (red).
583
584 **Figure 5: Riox2 protein alignment.** Clustal omega alignment of Riox2 protein sequences from
585 *H.vulgaris*, zebrafish, chicken, mouse and human. The protein domains JmjC (red), dimerization
586 (brown) and winged-helix (blue) are indicated based on the human sequence [15]. The prospective
587 iron-binding motif HxD…H (green) and the 2OG C5-carboxylate-binding residue (K, purple) are
588 conserved in all species. Crystal structure analysis of human RIOX2 with substrate Rpl27a identified
589 the amino acids N101, Q136, Q139, N165 and S257 of human RIOX2 involved in Rpl27a peptide
590 binding [15] (red).
591
592 **Figure 6: Nuclear localisation of Riox1 and Riox2.** The N-terminal extension domain of human
593 RIOX1 has been shown to harbour the nuclear localisation signal (NLS) [27]. (A) Expression of GFP-
594 tagged full-length RIOX1 in Hela cells resulted mainly in nucleolar accumulation. (B) A RIOX1
595 deletion mutant lacking amino acids 1-31 localised to the cytoplasm. (C) Fusion of amino acids 1-45
596 of human RIOX1 resulted in nuclear GFP localisation with strong accumulation in the nucleoli. DNA
597 stain: DAPI. Scalebar: 5μm. (D) Predictions of NLS in Riox1 of other species with either NLSmapper
598 (blue) or NLStradamus (pink) identified NLS in the N-terminal extension domains of Riox1, whereas
599 for Riox2 no NLS were predicted.
600
601 **Figure 7: Riox1 and Riox2 gene structures.** (A) The human *RIOX2* gene exhibits ten exons
602 distributed on chromosome 3. The exon lengths are indicated. The RIOX2 protein domains (JmjC,
603 dimerization, winged helix) are mapped on the gene. (B) The human *RIOX1* gene is a single exon
604 gene of 1962 bases on chromosome 14. The RIOX1 protein domains (JmjC, dimerization, winged
605 helix) are mapped on the gene. (C) Analysis of the genomic structures of *Riox2* and *Riox1* genes in *H.*
606 *sapiens*, *M. musculus*, *G. gallus*, *X. laevis*, *D. rerio*, *C. elegans*, *D. melanogaster* and *H. vulgaris* with

607 the number of exons are given in the table. *C. elegans* and *D .melanogaster* lack a *Riox1* gene. The
608 *Riox2* genes of human (Hs), mouse (Mm) and chicken (Gg) exhibit one non-coding exon (5´) and nine
609 coding exons.
610
611 **Figure 8: Phylogenetic tree of Riox1 (No66) orthologues proteins and exon-intron architecture**
612 **of the corresponding protein encoding genes.** The tree was inferred through a maximum-likelihood
613 analysis of 41 representative species (IQ-TREE). The tree shown is a consensus tree with SH-like
614 aLRT and ultrafast bootstrap (UFboot) values (numbers in parentheses SH-aLRT support (%) /
615 ultrafast bootstrap support (%)) given as branch support values. Good branch support is confirmed
616 with SH-aLRT >= 80% and UFboot >= 95%. The Tree is unrooted although the outgroup taxon
617 'Trichoplax' is drawn at root. The scale bar indicates 0.64 substitutions per site. Blue boxes and lines
618 on the right show the gene architectures of the corresponding genes with exons and introns,
619 respectively. Filled boxes represent protein-coding exons, empty boxes represent non-coding 5'- and
620 3'-UTR exons. Numbers in parentheses indicate total number of exons. Red rectangles encircle single
621 exon, intronless genes present in three different taxa (Mammalia, Aves, and Insecta, grey background
622 shading). Stars indicate species for which completed (non-fragmented) gene architecture annotations
623 are yet not available.
624
625 **Figure 9: Phylogenetic relationship of Riox1 (No66) and Riox2 (Mina53) JmjC domain**
626 **sequences in Metazoa.** Riox1- and Riox2-JmjC domain sequences from species used in this study
627 were extracted from full-length protein sequences, aligned using ClustalW and maximum-likelihood
628 analysis used for tree construction (IQ-TREE). The tree shown is a consensus tree with SH-like aLRT
629 and ultrafast bootstrap (UFboot) values (numbers in parentheses SH-aLRT support (%) / ultrafast
630 bootstrap support (%)) given as branch support values. Good branch support is confirmed with SH-
631 aLRT >= 80% and UFboot >= 95%. The Tree is unrooted although the outgroup taxon 'Trichoplax' is
632 drawn at root. The scale bar indicates 1.00 substitutions per site. The JmjC-domain of the *E. coli* ycfD
633 ribosomal oxygenase (ecycfD) was included in the alignment to analyse its phylogenetic relationship
634 to metazoan Riox1 and Riox2 proteins (indicated in red). Riox2-JmjC domain branch is highlighted

with grey background shading to show its separate branch node relationship to Riox1-JmjC domains. Note, Riox2 (Mina53) is also present in *Hydra vulgaris* and *Priapulus caudatus* which both possess *Riox1* and *Riox2* orthologous genes as invertebrates. Species with a Riox1 gene, but which lack a Riox2 gene are highlighted in red.

Additional files

Additional file 1: Protein sequence alignment (Clustal Omega) [33] of RIOX1 (*H.sapiens*) and Riox1 (*M.musculus*). The proposed iron-binding motif (H340, D342, H405) and the 2OG–interacting lysine residue (K355) for the human sequence [15] are indicated in green or blue respectively.

Additional file 2: Protein sequence alignment (Clustal Omega) [33] of RIOX1 (*H.sapiens*) and Riox1 (*G.gallus*). The proposed iron-binding motif (H340, D342, H405) and the 2OG–interacting lysine residue (K355) for the human sequence [15] are indicated in green or blue respectively.

Additional file 3: Protein sequence alignment (Clustal Omega) [33] of RIOX1 (*H.sapiens*) and Riox1 (*D.rerio*). The proposed iron-binding motif (H340, D342, H405) and the 2OG–interacting lysine residue (K355) for the human sequence [15] are indicated in green or blue respectively.

Additional file 4: Protein sequence alignment (Clustal Omega) [33] of RIOX1 (*H.sapiens*) and Riox1 (*D.melanogaster*). The proposed iron-binding motif (H340, D342, H405) and the 2OG–interacting lysine residue (K355) for the human sequence [15] are indicated in green or blue respectively.

Additional file 5: Protein sequence alignment (Clustal Omega) [33] of RIOX1 (*H.sapiens*) and Riox1 (*C.elegans*). The proposed iron-binding motif (H340, D342, H405) and the 2OG–interacting lysine residue (K355) for the human sequence [15] are indicated in green or blue respectively.

Additional file 6: Protein sequence alignment (Clustal Omega) [33] of RIOX2 (*H.sapiens*) and Riox2 (*M.musculus*). The proposed iron-binding motif (H179, D181, H240) and the 2OG–interacting lysine residue (K194) for the human sequence [15] are indicated in green or blue respectively.

Additional file 7: Protein sequence alignment (Clustal Omega) [33] of RIOX2 (*H.sapiens*) and Riox2 (*G.gallus*). The proposed iron-binding motif (H179, D181, H240) and the 2OG–interacting lysine residue (K194) for the human sequence [15] are indicated in green or blue respectively.

Additional file 8: Protein sequence alignment (Clustal Omega) [33] of RIOX2 (*H.sapiens*) and Riox2 (*D.rerio*). The proposed iron-binding motif (H179, D181, H240) and the 2OG–interacting lysine residue (K194) for the human sequence [15] are indicated in green or blue respectively.

Additional file 9: Nucleotide sequence (open reading frame) and corresponding amino acid sequence of Riox2 (*H.vulgaris*). Protein sequence alignment (Clustal Omega) [33] of RIOX2 (*H.sapiens*) and Riox2 (*H.vulgaris*). The proposed iron-binding motif (H179, D181, H240) and the 2OG–interacting lysine residue (K194) for the human sequence [15] are indicated in green or blue respectively. An additional stretch of charged amino acids in the *Hydra* sequence is highlighted in red.

Additional file 10: Nucleotide sequence (open reading frame) and corresponding amino acid sequence of Riox1 (*H.vulgaris*). Protein sequence alignment (Clustal Omega) [33] of RIOX1 (*H.sapiens*) and Riox1 (*H.vulgaris*). The proposed iron-binding motif (H340, D342, H405) and the 2OG–interacting lysine residue (K355) for the human sequence [15] are indicated in green or blue respectively.

References

1. Crick F: **Central dogma of molecular biology**. *Nature* 1970, **227**(5258):561-563.

2. Rissland OS: **The organization and regulation of mRNA-protein complexes**. *Wiley Interdiscip Rev RNA* 2017, **8**(1).
3. Roundtree IA, He C: **RNA epigenetics--chemical messages for posttranscriptional gene regulation**. *Curr Opin Chem Biol* 2016, **30**:46-51.
4. Shi Z, Barna M: **Translating the genome in time and space: specialized ribosomes, RNA regulons, and RNA-binding proteins**. *Annu Rev Cell Dev Biol* 2015, **31**:31-54.
5. Xue S, Barna M: **Specialized ribosomes: a new frontier in gene regulation and organismal biology**. *Nature reviews Molecular cell biology* 2012, **13**(6):355-369.
6. Ge W, Wolf A, Feng T, Ho CH, Sekirnik R, Zayer A, Granatino N, Cockman ME, Loenarz C, Loik ND *et al*: **Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans**. *Nat Chem Biol* 2012, **8**(12):960-962.
7. Katz MJ, Acevedo JM, Loenarz C, Galagovsky D, Liu-Yi P, Perez-Pepe M, Thalhammer A, Sekirnik R, Ge W, Melani M *et al*: **Sudestada1, a Drosophila ribosomal prolyl-hydroxylase required for mRNA translation, cell homeostasis, and organ growth**. *Proceedings of the National Academy of Sciences of the United States of America* 2014.
8. Loenarz C, Sekirnik R, Thalhammer A, Ge W, Spivakovsky E, Mackeen MM, McDonough MA, Cockman ME, Kessler BM, Ratcliffe PJ *et al*: **Hydroxylation of the eukaryotic ribosomal decoding center affects translational accuracy**. *Proc Natl Acad Sci U S A* 2014, **111**(11):4019-4024.
9. Singleton RS, Liu-Yi P, Formenti F, Ge W, Sekirnik R, Fischer R, Adam J, Pollard PJ, Wolf A, Thalhammer A *et al*: **OGFOD1 catalyzes prolyl hydroxylation of RPS23 and is involved in translation control and stress granule formation**. *Proc Natl Acad Sci U S A* 2014, **111**(11):4031-4036.
10. Aik W, McDonough MA, Thalhammer A, Chowdhury R, Schofield CJ: **Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases**. *Curr Opin Struct Biol* 2012, **22**(6):691-700.
11. Bottger A, Islam MS, Chowdhury R, Schofield CJ, Wolf A: **The oxygenase Jmjd6-a case study in conflicting assignments**. *The Biochemical journal* 2015, **468**(2):191-202.
12. Kooistra SM, Helin K: **Molecular mechanisms and potential functions of histone demethylases**. *Nature reviews Molecular cell biology* 2012.
13. Rasmussen KD, Helin K: **Role of TET enzymes in DNA methylation, development, and cancer**. *Genes & development* 2016, **30**(7):733-750.
14. Webb JD, Coleman ML, Pugh CW: **Hypoxia, hypoxia-inducible factors (HIF), HIF hydroxylases and oxygen sensing**. *Cellular and molecular life sciences : CMLS* 2009, **66**(22):3539-3554.
15. Chowdhury R, Sekirnik R, Brissett NC, Krojer T, Ho CH, Ng SS, Clifton IJ, Ge W, Kershaw NJ, Fox GC *et al*: **Ribosomal oxygenases are structurally conserved from prokaryotes to humans**. *Nature* 2014, **510**(7505):422-426.
16. Tsuneoka M, Koda Y, Soejima M, Teye K, Kimura H: **A novel myc target gene, mina53, that is involved in cell proliferation**. *J Biol Chem* 2002, **277**(38):35450-35459.
17. Zhang Y, Lu Y, Yuan BZ, Castranova V, Shi X, Stauffer JL, Demers LM, Chen F: **The Human mineral dust-induced gene, mdig, is a cell growth regulating gene associated with lung cancer**. *Oncogene* 2005, **24**(31):4873-4882.
18. Thakur C, Lu Y, Sun J, Yu M, Chen B, Chen F: **Increased expression of mdig predicts poorer survival of the breast cancer patients**. *Gene* 2014, **535**(2):218-224.
19. Suzuki C, Takahashi K, Hayama S, Ishikawa N, Kato T, Ito T, Tsuchiya E, Nakamura Y, Daigo Y: **Identification of Myc-associated protein with JmjC domain as a novel therapeutic target oncogene for lung cancer**. *Molecular cancer therapeutics* 2007, **6**(2):542-551.
20. Nishizawa Y, Nishida N, Konno M, Kawamoto K, Asai A, Koseki J, Takahashi H, Haraguchi N, Nishimura J, Hata T *et al*: **Clinical Significance of Histone Demethylase NO66 in Invasive Colorectal Cancer**. *Annals of surgical oncology* 2016.

21. Okamoto M, Van Stry M, Chung L, Koyanagi M, Sun X, Suzuki Y, Ohara O, Kitamura H, Hijikata A, Kubo M *et al*: **Mina, an Il4 repressor, controls T helper type 2 bias**. *Nature immunology* 2009, **10**(8):872-879.
22. Pillai MR, Lian S, Bix M: **Mina: a Th2 response regulator meets TGFbeta**. *Current opinion in immunology* 2014, **31**:38-43.
23. Sinha KM, Yasuda H, Coombes MM, Dent SY, de Crombrugghe B: **Regulation of the osteoblast-specific transcription factor Osterix by NO66, a Jumonji family histone demethylase**. *The EMBO journal* 2010, **29**(1):68-79.
24. Chen Q, Sinha K, Deng JM, Yasuda H, Krahe R, Behringer RR, de Crombrugghe B: **Mesenchymal Deletion of Histone Demethylase NO66 in Mice Promotes Bone Formation**. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research* 2015, **30**(9):1608-1617.
25. Chen Q, Zhang L, de Crombrugghe B, Krahe R: **Mesenchyme-specific overexpression of nucleolar protein 66 in mice inhibits skeletal growth and bone formation**. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 2015, **29**(6):2555-2565.
26. Eilbracht J, Kneissel S, Hofmann A, Schmidt-Zachmann MS: **Protein NO52--a constitutive nucleolar component sharing high sequence homologies to protein NO66**. *Eur J Cell Biol* 2005, **84**(2-3):279-294.
27. Eilbracht J, Reichenzeller M, Hergt M, Schnolzer M, Heid H, Stohr M, Franke WW, Schmidt-Zachmann MS: **NO66, a highly conserved dual location protein in the nucleolus and in a special type of synchronously replicating chromatin**. *Mol Biol Cell* 2004, **15**(4):1816-1832.
28. Leclere L, Copley RR, Momose T, Houliston E: **Hydrozoan insights in animal development and evolution**. *Current opinion in genetics & development* 2016, **39**:157-167.
29. Chapman JA, Kirkness EF, Simakov O, Hampson SE, Mitros T, Weinmaier T, Rattei T, Balasubramanian PG, Borman J, Busam D *et al*: **The dynamic genome of Hydra**. *Nature* 2010, **464**(7288):592-596.
30. Bottger A, Alexandrova O, Cikala M, Schade M, Herold M, David CN: **GFP expression in Hydra: lessons from the particle gun**. *Development genes and evolution* 2002, **212**(6):302-305.
31. Nguyen Ba AN, Pogoutse A, Provart N, Moses AM: **NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction**. *BMC Bioinformatics* 2009, **10**:202.
32. Kosugi S, Hasebe M, Tomita M, Yanagawa H: **Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs**. *Proc Natl Acad Sci U S A* 2009, **106**(25):10171-10176.
33. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J *et al*: **Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega**. *Mol Syst Biol* 2011, **7**:539.
34. Grzybowska EA: **Human intronless genes: functional groups, associated diseases, evolution, and mRNA processing in absence of splicing**. *Biochem Biophys Res Commun* 2012, **424**(1):1-6.
35. Kaessmann H: **Origins, evolution, and phenotypic impact of new genes**. *Genome Res* 2010, **20**(10):1313-1326.
36. Jorquera R, Ortiz R, Ossandon F, Cardenas JP, Sepulveda R, Gonzalez C, Holmes DS: **SinEx DB: a database for single exon coding sequences in mammalian genomes**. *Database (Oxford)* 2016, **2016**.
37. Shabalina SA, Ogurtsov AY, Spiridonov AN, Novichkov PS, Spiridonov NA, Koonin EV: **Distinct patterns of expression and evolution of intronless and intron-containing mammalian genes**. *Mol Biol Evol* 2010, **27**(8):1745-1749.

38. Carelli FN, Hayakawa T, Go Y, Imai H, Warnefors M, Kaessmann H: **The life history of retrocopies illuminates the evolution of new mammalian genes**. *Genome Res* 2016, **26**(3):301-314.
39. Theocharidis A, van Dongen S, Enright AJ, Freeman TC: **Network visualization and analysis of gene expression data using BioLayout Express(3D)**. *Nat Protoc* 2009, **4**(10):1535-1550.
40. Mabbott NA, Baillie JK, Brown H, Freeman TC, Hume DA: **An expression atlas of human primary cells: inference of gene function from coexpression networks**. *BMC Genomics* 2013, **14**:632.
41. Maquat LE, Li X: **Mammalian heat shock p70 and histone H4 transcripts, which derive from naturally intronless genes, are immune to nonsense-mediated decay**. *RNA* 2001, **7**(3):445-456.
42. Krause CD: **Intron loss in interferon genes follows a distinct set of stages, and may confer an evolutionary advantage**. *Cytokine* 2016, **83**:193-205.
43. Markovic D, Challiss RA: **Alternative splicing of G protein-coupled receptors: physiology and pathophysiology**. *Cell Mol Life Sci* 2009, **66**(20):3337-3352.
44. Kirienko NV, Fay DS: **SLR-2 and JMJC-1 regulate an evolutionarily conserved stress-response network**. *EMBO J* 2010, **29**(4):727-739.
45. Wang C, Zhang Q, Hang T, Tao Y, Ma X, Wu M, Zhang X, Zang J: **Structure of the JmjC domain-containing protein NO66 complexed with ribosomal protein Rpl8**. *Acta Crystallogr D Biol Crystallogr* 2015, **71**(Pt 9):1955-1964.
46. Teichmann M, Dumay-Odelot H, Fribourg S: **Structural and functional aspects of winged-helix domains at the core of transcription initiation complexes**. *Transcription* 2012, **3**(1):2-7.
47. Lu Y, Chang Q, Zhang Y, Beezhold K, Rojanasakul Y, Zhao H, Castranova V, Shi X, Chen F: **Lung cancer-associated JmjC domain protein mdig suppresses formation of tri-methyl lysine 9 of histone H3**. *Cell Cycle* 2009, **8**(13):2101-2109.
48. Williams ST, Walport LJ, Hopkinson RJ, Madden SK, Chowdhury R, Schofield CJ, Kawamura A: **Studies on the catalytic domains of multiple JmjC oxygenases using peptide substrates**. *Epigenetics* 2014, **9**(12):1596-1603.
49. Cikala M, Alexandrova O, David CN, Proschel M, Stiening B, Cramer P, Bottger A: **The phosphatidylserine receptor from Hydra is a nuclear protein with potential Fe(II) dependent oxygenase activity**. *BMC Cell Biol* 2004, **5**:26.
50. Mantri M, Krojer T, Bagg EA, Webby CJ, Butler DS, Kochan G, Kavanagh KL, Oppermann U, McDonough MA, Schofield CJ: **Crystal structure of the 2-oxoglutarate- and Fe(II)-dependent lysyl hydroxylase JMJD6**. *J Mol Biol* 2010, **401**(2):211-222.
51. Webby CJ, Wolf A, Gromak N, Dreger M, Kramer H, Kessler B, Nielsen ML, Schmitz C, Butler DS, Yates JR, 3rd *et al*: **Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing**. *Science* 2009, **325**(5936):90-93.
52. Emmott E, Hiscox JA: **Nucleolar targeting: the hub of the matter**. *EMBO Rep* 2009, **10**(3):231-238.
53. Hemmrich G, Anokhin B, Zacharias H, Bosch TC: **Molecular phylogenetics in Hydra, a classical model in evolutionary developmental biology**. *Mol Phylogenet Evol* 2007, **44**(1):281-290.
54. Katoh K, Standley DM: **MAFFT multiple sequence alignment software version 7: improvements in performance and usability**. *Mol Biol Evol* 2013, **30**(4):772-780.
55. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ: **W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis**. *Nucleic Acids Res* 2016, **44**(W1):W232-235.
56. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O: **New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0**. *Syst Biol* 2010, **59**(3):307-321.