

Doctoral Dissertation  
Academic Year 2020

Identification of anhydrobiosis related genes  
through comparative genomics in extremotolerant  
tardigrades

Graduate School of Media and Governance  
Keio University

Yuki Yoshida

## Abstract of Doctoral Dissertation Academic Year 2020

# Identification of anhydrobiosis related genes through comparative genomics in extremotolerant tardigrades

### Summary

Tardigrades, sole members of the phylum Tardigrada, are ubiquitous meiofauna with remarkable characteristics. Limno-terrestrial members of Tardigrada are capable of cryptobiosis, a latent state of life induced by extreme stress such as hypoxia (oxybiosis), hypoosmolarity (osmobiosis), hypothermia (cryobiosis), and desiccation (anhydrobiosis) at any point of their life cycle. The ametabolic state of anhydrobiosis is unique, as loss of water is far more damaging to the cell than other states. Anhydrobiosis lacks most of the conditions that are considered essential for life (*i.e.*, metabolism) and thus has been proposed as the third stage of life, a reversible stasis between life and death. The mechanisms underlying anhydrobiosis have been the subject of research in molecular biology for decades, but the lack of genomic resources has obstructed a comprehensive analysis. Here, I focused on two tardigrade species capable of diverse anhydrobiotic capabilities, *Ramazzottius varieornatus*, and *Hypsibius exemplaris*. Previous studies on the genome of *H. exemplaris* have incorrectly identified horizontally transferred genes that are now known to result from non-metazoan contamination and have also failed to resolve heterozygous regions. I obtained a 104 Mb high-quality genome encoding 19,901 genes by combining long-read sequencing with a newly established low-input DNA sequencing method from a single individual. I observed 1,422 genes regulated during *H. exemplaris* anhydrobiosis entry, of which nearly half were novel. These results extended previous studies by emphasizing the connection of tardigrade specific intrinsically disordered proteins and widely conserved anti-oxidative stress response genes with anhydrobiosis (Chapter 2). I also proposed the importance of validating candidates of horizontally transferred genes in invertebrates based on insights from the previous study (Chapter 3). I then validated the expression of these candidates during ontogeny and cross-tolerance, where I observed the expression of anhydrobiosis genes during developmental stages (Chapter 4) and identified a novel peroxidase family conserved widely in Tardigrada, emphasizing that anti-oxidative stress response is a critical component of the anhydrobiosis machinery (Chapter 5). Throughout this dissertation, I identified key factors of anhydrobiosis based on genomic and transcriptomic evidence and proposed the anti-oxidative stress response underlies the anhydrobiosis mechanism, enabling an organism to tolerate complete desiccation.

### Keywords

Tardigrade, Extremotolerant, Anhydrobiosis, Comparative genomics, Transcriptome analysis

Graduate School of Media and Governance

Keio University

Yuki Yoshida

博士論文  
2020年度（令和2年度）

## 極限環境生物クマムシの比較ゲノム解析による 乾眠関連遺伝子の同定

### 論文要旨

緩歩動物門は微小動物であるクマムシのみで形成される分類群であり、特徴的な形質を持つ。多くの陸生種は低酸素、高浸透圧、低温、そして低湿度などの極度の環境ストレスに曝露されるとクリプトビオシスと呼ばれる生命状態に移行することがどのライフステージでも可能である。特に乾眠は乾燥によって導入される無代謝状態であり、通常であれば死に直結する環境に耐えることができる。乾眠状態は代謝などの生命に必須と考えられる要素を示さないことから、生と死の間に存在する可逆的に移行可能な第3の生命状態として提唱されている。近年、この乾眠の分子機構は分子生物学の対象とされつつあったものの、ゲノム情報の不足によりその調査は困難を極めていた。本研究で私は乾眠能力の異なるヨコヅナクマムシ (*Ramazzottius varieornatus*) とエグゼンプラリスヤマクマムシ (*Hypsibius exemplaris*) に着目した。これまでのエグゼンプラリスヤマクマムシのゲノムは未解決のヘテロ接合領域や後生動物以外のコンタミネーションによる水平伝播遺伝子の予測ミスなどの問題が明らかとなっていた。本研究ではこれらの問題を解決するとともに乾眠の分子機構に関わる因子を明らかにすることを目的とした。これまでに私は一匹のクマムシから得られる超微量 DNA からゲノムシーケンスを行う技術の確立に携わり、この技術に加えてロングリードシーケンス技術をエグゼンプラリスヤマクマムシに適用することで 19,901 個の遺伝子をコードする約 104 Mb の高品質なゲノムを取得し、クマムシの比較ゲノム解析の基盤を整備した。このゲノムに基づいた乾眠前後のトランスクリプトーム解析により、約半数の機能未知遺伝子を含む 1,422 遺伝子が乾眠移行時に誘導されることを明らかにし、抗酸化経路やクマムシ特有の天然変性タンパクがエグゼンプラリスヤマクマムシでも誘導されていることを示したことで乾眠への関与をさらに強く示唆した (第2章)。また、この比較解析の過程で得られた知見を元に、微小無脊椎動物における水平伝播遺伝子候補の検証の重要性を示した (第3章)。次に、胚発生期および交叉耐性のトランスクリプトーム解析を実施し、発生期間中も乾眠遺伝子が発現しているのを示したとともに (第4章)、乾眠移行中と共通して誘導される緩歩動物のみで広く保存されている新規の抗酸化タンパクファミリーを同定し、乾眠の分子機構において抗酸化機能が中核を担う可能性を示唆した (第5章)。本学位論文ではゲノムおよびトランスクリプトームによってクマムシの生理に関わる重要な因子を同定し、真核生物において水を失ってもなお生きながらえる乾眠は抗酸化を中心としながらも極めて複雑な分子メカニズムに支えられていることを明らかにした。

### キーワード

クマムシ, 極限環境耐性, 乾眠, 比較ゲノム解析, トランスクリプトーム解析

慶應義塾大学大学院  
政策・メディア研究科

吉田祐貴

Keio University



**Copyright ©2021  
Yuki Yoshida  
All rights reserved**

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Water is essential for cellular integrity . . . . .	2
1.2	Anhydrobiosis, a latent state of life . . . . .	2
1.3	Tardigrades as a model organism for anhydrobiosis research . . . . .	6
1.4	Mechanisms of tardigrade anhydrobiosis . . . . .	8
1.5	Omics for analyzing anhydrobiosis . . . . .	14
1.6	Organization of this dissertation . . . . .	17
<b>2</b>	<b>Comparative genomics of two tardigrades</b>	<b>19</b>
2.1	Introduction . . . . .	20
2.2	Materials and methods . . . . .	23
2.2.1	Tardigrade culture and sampling . . . . .	23
2.2.2	Genome and transcriptome sequencing . . . . .	24
2.2.3	Genome assembly . . . . .	25
2.2.4	Gene finding . . . . .	26
2.2.5	HGT identification . . . . .	27
2.2.6	Anhydrobiosis analyses . . . . .	28
2.2.7	Protein family analyses and comparative genomics . . . . .	28
2.2.8	Phylogenomics . . . . .	29
2.2.9	Databasing, software usage and data manipulation . . . . .	30
2.3	Results . . . . .	30
2.3.1	The genome of <i>H. exemplaris</i> . . . . .	30
2.3.2	Comparisons with <i>R. varieornatus</i> . . . . .	35
2.3.3	Horizontal gene transfer in tardigrades . . . . .	38
2.3.4	The genomics of anhydrobiosis in tardigrades . . . . .	41
2.3.5	Phylogenetic relationships of Tardigrada . . . . .	46
2.3.6	Rare genomic changes and tardigrade relationships . . . . .	49
2.4	Discussion . . . . .	54
2.4.1	A robust estimate of the <i>H. exemplaris</i> genome . . . . .	54
2.4.2	HGT in tardigrades: <i>H. exemplaris</i> has a normal metazoan genome . . . . .	55
2.4.3	Contrasting modes of anhydrobiosis in tardigrades . . . . .	56
2.4.4	The position of tardigrades in the metazoa . . . . .	57
<b>3</b>	<b>Horizontal gene transfer in metazoa: examples and methods</b>	<b>60</b>
3.1	Introduction . . . . .	61
3.2	HGT in extremotolerant tardigrades . . . . .	65
3.3	Lessons from the tardigrades . . . . .	67
3.4	High levels of HGT in bdelloid rotifers . . . . .	70
3.5	Bioinformatic approaches to the prediction of HGT . . . . .	74

3.6	All that glitters is not gold: HGTs as hypotheses . . . . .	80
<b>4</b>	<b>Comparison of the transcriptomes of two tardigrades with different hatching coordination</b>	<b>82</b>
4.1	Introduction . . . . .	83
4.2	Materials and methods . . . . .	84
4.2.1	Tardigrade rearing and life history analysis . . . . .	84
4.2.2	Transcriptome sequencing and data preparation . . . . .	84
4.2.3	Informatics analysis . . . . .	85
4.2.4	Chemical exposure . . . . .	86
4.2.5	Statistics . . . . .	87
4.3	Results . . . . .	87
4.4	Discussion . . . . .	94
<b>5</b>	<b>A novel Mn-dependent peroxidase contributes to tardigrade anhydrobiosis</b>	<b>98</b>
5.1	Introduction . . . . .	99
5.2	Materials and methods . . . . .	100
5.2.1	Tardigrade culturing conditions and UVC exposure . . . . .	100
5.2.2	Quantification of movement in irradiated samples . . . . .	100
5.2.3	RNA extraction and library preparation . . . . .	100
5.2.4	Informatics analysis . . . . .	101
5.2.5	Transmission electron microscope observation . . . . .	102
5.2.6	Sample preparation of recombinant proteins . . . . .	102
5.2.7	Crystallization, X-ray data collection and structure determination . . . . .	103
5.2.8	Measurement of enzymatic activity of the g12777 protein . . . . .	104
5.2.9	Isothermal titration calorimetry . . . . .	104
5.2.10	Expression of GFP-tagged g12777 in HEK293 . . . . .	104
5.2.11	Validation of cellular tolerance against oxidative stress . . . . .	105
5.3	Results and discussions . . . . .	106
5.3.1	Transcriptome sequencing of UVC exposed <i>R. varieornatus</i> . . . . .	106
5.3.2	Identification of the g12777 gene family as a novel stress responsive gene family . . . . .	109
5.3.3	g12777 has a highly conserved Mn <sup>2+</sup> binding site . . . . .	109
5.3.4	g12777 is a Mn <sup>2+</sup> dependent peroxidase . . . . .	112
<b>6</b>	<b>Final remarks</b>	<b>115</b>
6.1	Establishing a high-quality genome of microscopic invertebrates . . . . .	116
6.2	A model of the anhydrobiosis machinery derived from the genome analysis . . . . .	117
6.3	Future prospects . . . . .	119
6.4	Conclusions . . . . .	121
	<b>References</b>	<b>122</b>

<b>Achievements</b>	<b>147</b>
<b>Supplementary Files</b>	<b>155</b>
Supplementary Tables . . . . .	155
Supplementary Figures . . . . .	173
Supplementary Data . . . . .	195

## List of Figures

1.1	Active and anhydrobiotic states of <i>Macrobotus shonaicus</i> . . . . .	5
1.2	Currently identified mechanisms of anhydrobiosis. . . . .	12
1.3	Conservation patterns of tardigrade specific anhydrobiosis genes. . . . .	13
1.4	The machinery of anhydrobiosis proposed prior to genomics. . . . .	16
1.5	Organization of this dissertation. . . . .	18
2.1	The genomes of <i>H. exemplaris</i> and <i>R. varieornatus</i> . . . . .	37
2.2	Horizontal gene transfer in <i>H. exemplaris</i> . . . . .	40
2.3	The genomics of anhydrobiosis in tardigrades. . . . .	45
2.4	Phylogeny of Ecdysozoa. . . . .	48
2.5	The position of tardigrada in Ecdysozoa. . . . .	53
3.1	Meiofaunal animals analysed for HGT. . . . .	64
3.2	HGT proportions in metazoan genomes. . . . .	69
4.1	Comparison of hatching timing between <i>H. exemplaris</i> and <i>R. varieornatus</i> . . . . .	88
4.2	SOM clustering of gene expression profiles during development. . . . .	90
4.3	Gene expression patterns for molting pathways during development. . . . .	91
4.4	High concentration Fenoxycarb exposure inhibited hatching in <i>H. exemplaris</i> . . . . .	93
5.1	Identification of a novel stress responsive gene family conserved within Tardigrada phylum. . . . .	108
5.2	g12777 protein has a SOD-like -sandwich fold and binds to Mn <sup>2+</sup> ion. . . . .	111
5.3	g12777 is a Golgi apparatus localizing protein that enhances anti-oxidative stress. . . . .	114

## List of Tables

2.1	Metrics of <i>H. exemplaris</i> genome assemblies. . . . .	32
2.2	Comparison of the genomes of <i>H. exemplaris</i> and <i>R. varieornatus</i> . . . . .	34
2.3	Gene family births that support different relationships of Tardigrada. . . . .	52
3.1	Methods for detection of HGT. . . . .	77

## List of Abbreviations

ADP	adenosine-5'-diphosphate
AMNP	anhydrobiosis-related mn-dependent peroxidase
AMPK	5' adenosine monophosphate-activated protein kinase
API	application-programming interface
ATM	ataxia-telangiectasia mutated
ATP	adenosine-5'-triphosphate
ATR	ataxia-telangiectasia and rad3-related protein
BAC	bacterial artificial chromosome
BER	base excision repair
BLAST	basic local alignment search tool
BSA	bovine serum albumin
BSC1	bypass of stop codon protein 1
BUSCO	benchmarking universal single-copy orthologs
BWA	burrow-wheeler aligner
CAHS	cytosolic abundant heat soluble
CaM	calmodulin
CASP	caspase
CATE	catalase
CBF	c-repeat-binding factor
CDS	coding sequence
CEGMA	core eukaryotic genes mapping approach
CHK	checkpoint kinase
CHS	consensus hit support
CNG	cyclic nucleotide-gated channel
CRISPR	clustered regularly interspaced short palindromic repeats
CSP	cold shock protein
DAPI	4',6-diamidino-2-phenylindole
DEG	differentially expressed gene
DNA-Seq	dna sequencing
DSBs	double strand breaks
Dsup	damage suppressor
EcR	ecdysone receptor
EGF	epidermal growth factor
EST	expressed sequence tag
EtAHS	<i>Echiniscus testudo</i> abundant heat soluble
FDR	false discovery rate
FOXO	forkhead box o
GFP	green fluorescent protein
GO	gene ontology
GPCR	g-protein-coupled receptor
GST	glutathione s-transferase
GTP	guanosine-5'-triphosphate
GTR+G	general time reversible model with gamma distribution of rates model
GTR-CAT+G	gtr plus rate categories model
HGT	horizontal gene transfer
HEK293	human embryonic kidney 293
HR	homologous recombination
HSF	heat shock factor
HSP	heat shock protein
IDP	intrinsically disordered proteins
INSDC	international nucleotide sequence database collaboration
IPR	interpro domain identifier
IUP	intrinsically unstructured proteins

KAAS	kegg automatic annotation server
Kr-h1	krüppel homolog-1
KEGG	kyoto encyclopedia of genes and genomes
LDL	low-density lipoprotein
LEA	late embryogenesis abundant
LIL	lea-island-located
MAHS	mitochondrial abundant heat soluble
MCL	markov cluster algorithm
MEM	minimum essential media
miRNA	micro rna
miRNA-Seq	micro rna sequencing
MMR	mismatch repair
mTOR	mechanistic target of rapamycin
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NER	nucleotide excision repair
NF-Y	nuclear transcription factor y
NHEJ	nonhomologous end joining
NUMTs	nuclear-mitochondrial transfers
PacBio	pacific biosciences
PARP	poly (adp-ribose) polymerase
PBS	phosphate-buffered saline
PF	pfam identifier
PHD	plant homeodomain
PIMT	protein l-isoaspartyl methyltransferase
piRNA	pipi-interacting rna
PP1	protein phosphatase I
PP2A	protein phosphatase 2a
RAxML	randomized accelerated maximum likelihood
RFP	red fluorescent protein
RH	relative humidity
RNAi	rna interference
RNA-Seq	rna sequencing
ROS	reactive oxygen species
rRNA	ribosomal rna
RvLEAM	late embryogenesis abundant protein mitochondrial
SAHS	secretory abundant heat soluble
SEM	scanning electron microscope
SIRT1	sirtuin 1
SOD	superoxide dismutase
SOM	self-organizing maps
SRA	sequence read archive
TALEN	transcription activator-like effector nuclease
TDP	tardigrade specific disordered protein
TEM	transmission electron microscope
TPM	transcript(s) per million
TPS	trehalose-6-phosphatase synthase
tRNA	transfer rna
TSC1/2	tuberous sclerosis 1/2
UDP	uridine diphosphate
UGT	udp-glucuronosyltransferase
USP	ultraspiracle
UVB	ultraviolet b
UVC	ultraviolet c
VHL	von hippel-lindau tumor suppressor

---

# Acknowledgments

*If I have seen further it is by standing on the shoulders of y<sup>e</sup> Giants.*

Issac Newton, *Letter to Robert Hooke*, 1675.

I first would like to acknowledge my supervisors, Professor Masaru Tomita, Associate Professor Kazuharu Arakawa, Project Lecturer Nobuaki Kono, and Professor Akio Kanai. Professor Tomita offered me the first invitation to the inspiring world of biology. Without Professor Tomita's intuition at the Summer Bio College event in 2011, I would not have been able to meet the splendid members of the scientific community, submerge in the ocean of knowledge, and reveal the unknown. Associate Professor Arakawa has been my supervisor since I started my bachelor's research and has graciously been my advisor for the long years. His knowledge and philosophy go deeper than what I can see and still is an unfathomed crevasse. The philosophical basis of what I stand on today has is based on his guidance. Project Lecturer Kono has been someone I can look up to for not only scientific discussions but also for things that I needed comments on. He has encouraged me on many occasions and helped me throughout the years. Lastly, Professor Kanai has been my supervisor on my experimental side; in many of the experimental assays conducted in this thesis, I have sought his opinions. The studies conducted here all have deep guidance from these four supreme scientists, and I feel great gratitude to each and every one of them.

Secondly, I would like to thank the members of the Arakawa lab: Dr. Daiki D. Horikawa, Dr. Masayuki Fujiwara, Dr. Keizo Takasuka, Dr. Koyuki Kondo, Dr. James Flemming, Yuki Takai, Nozomi Abe Shoji, Nahoko Ishii, Yuki Onozawa, Kaori Igarashi, Kayoko Kurimoto, Masami Shimura and Makiko Anzai. Current and former student members have also supported my studies; Yuta Yoshida, Shiori Komine,

Dr. Soh Ishiguro, Kotone Itaya, Taiyo Miyahara, Dr. Kayln Kawamoto Kono, Kyoko Ishino, Tomoya Nishiwaki, Kounosuke Mark Ii, Hisanori Iwai, Takahiro Masuda, Yuri Tsuchizawa, Yumi Murai, Sayuri Konno, Ryosuke Nishino, Yu Kurihara, Tomoki Takeda, Hitoshi Kawakami, Yui Yamada, Peng Yongsheng, and Sora Ishikawa. These members have been my support these years. I hope the best for every member.

I then would like to acknowledge the current members and graduates of the Masaru Tomita lab of Keio University. I enjoyed the time spent with these cheerful members as they have been friends (and sometimes foes); Senka Deno, Kaho Yajima, Masahiro Miura, Ayano Sakai, Nao Takeuchi, Yuji Yasuda, Yoshiro Kita, Megumi Tsurumaki, Daniel Yamamoto Evans, Kazuki Tanaka, Motofumi Saito, Shohei Nagata, Akane Nishigata, Tomoya Tsukimi, Fumie Nakasuka, Ryosuke Hayasaka, Airi Hosoya, Hideto Mori, Kenta Morimoto, Tsubasa Watabe, Mia Ito-Yoshikawa, Wataru Takahagi, Keita Sasaki, Yuriko Kawasaki, Gakuto Makino, Junnosuke Imai, Dr. Shahlizah Shah, Dr. Yuri Yamamoto, Dr. Chiharu Murakami-Ishii, Dr. Shinnosuke Murakami, Dr. Kentaro Hayashi, Dr. Satoshi Tamaki, Kohei Yamaga. In particular, Dr. Murakami, Dr. Ishii, Hideto Mori and Yuriko Kawasaki have been delightful friends whom I could depend on.

My gratitude extends to out of the lab members to the office staff members Akiko Sakai, Masumi Honma, Kaori Kano, Yoko Tsuchiya, Maki Oike, Satomi Yokoi, Aya Saito, and collaborators Dr. Kenta Sugiura, Hiroki Minato, Dr. Takahiro Yamada, and my great neighbor Satoru Ariga. Dr. Kenta Sugiura has been a splendid colleague studying these facinating creatures. I hope one day that we can collaborate once more.

Lastly, but not least, I would like to thank my family, Yuichiro, Hiroko, Yuta, and Mayu Yoshida. Without them, I would not have been able to do the works written here in this dissertation.

---

---

# **Chapter 1**

---

## Introduction

## 1.1 Water is essential for cellular integrity

*αλλα θαλης μεν ο της τριαντης αρχηγος φιλοσοφιας νδωρ φησιν ειναι*<sup>1</sup>

Aristotle, *Metaphysics*, 983β, 17-21

Thales of Miletus, recognized as the first philosopher by Aristotle, regarded water as the matter from which nature originated. Although this point of view is no longer held, water is considered essential for life on earth. Water comprises up to approximately 70% of the cell, making it the most abundant molecule in our bodies (Sheng and Huggins, 1979). It serves as a solvent for cellular molecular reactions (Franks, 2000), which are themselves a “definition of life” (Koshland, 2002). So synonymous is water with life that the existence of water vapor in the atmosphere of exoplanets has been recognized as “*the first clue ... that a planet may be habitable*” (Lammer *et al.*, 2009, p. 236).

Living organisms have developed countermeasures to compensate for the loss of water molecules in the cell, as extreme dehydration would lead to death in most cases. Terrestrial and limno-terrestrial organisms constantly face desiccation stress, either by loss of internal water or desiccation of the surrounding environment. When these organisms face environmental desiccation, large organisms, such as terrestrial vertebrates, are capable of moving to a nearby humid environment where they can ingest water. However, microscopic organisms lack such freedom of mobility, and thus require a means of tolerating such conditions. Anhydrobiosis, a latent state of life induced by desiccation, is one of the more unusual responses to extreme water loss.

## 1.2 Anhydrobiosis, a latent state of life

*Den 2. September was het mede seer warm en droog weder, en ik nam des mergens de klok ontrent negen ure eenige van die stoffe uyt de loode goot, die nu ingedroogt sijnde, geen halve rugge van een Mes dikte konde uyt maken, die de voorledene nagt op myn comptoir hadde gelegen, en ik dede deselve in een tuba, die de dikte van een swane schagt hadde, en ik goot op deselve een weynig Regen water, uyt myn steene regen-bak, daar in wel levende dierkens waren, dog uyt kleynder soort bestaande, en ik vermengde soo aanstonts de geseyde stoffe met het water, op dat de stoffe die als vast aan een was, van een soude scheidt, om of'er levende dierkens in die drooge stoffe mogten beslooten leggen, die in 't korte te voorschijn*

<sup>1</sup>“but Thales, the founder of this school of philosophy, says the permanent entity is water.. Translation from [https://www.loebclassics.com/view/aristotle-metaphysics/1933/pb\\_LCL271.19.xml](https://www.loebclassics.com/view/aristotle-metaphysics/1933/pb_LCL271.19.xml)

*souden komen, hoe wel ik geen gedagten hadde, dat in soo een uyt gedroogde stoffe eenig levend schepsel soude wesen.* <sup>2</sup>

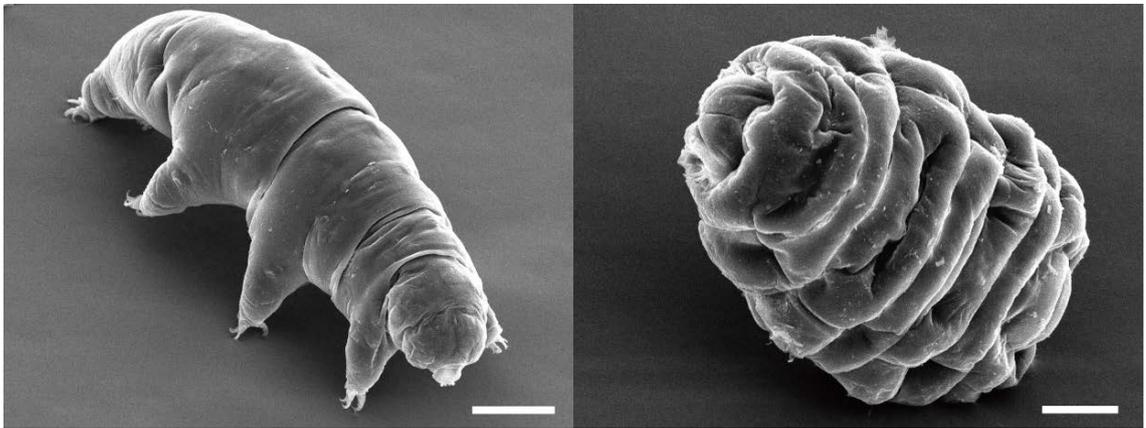
van Leeuwenhoek, A. *Letter to Hendrik van Bleyswijk*, 1702.

The first observation of an animal anhydrobiote was reported by Antonie van Leeuwenhoek, “the father of microbiology”, in 1702 (van Leeuwenhoek, 1702). Leeuwenhoek rehydrated dust collected from a roof gutter and observed the recovery of animals (probably bdelloid rotifers) within the hour. His work was followed by Needham and Corti (Corti, 1774; Needham, 1743), who observed anhydrobiosis in nematodes and tardigrades (Figure 1.1). This phenomenon was first termed anabiosis by Preyer (Preyer, 1891), but was refined to anhydrobiosis by Giard (Giard, 1894). Currently, anhydrobiotic animals are found in various invertebrate lineages, including *Artemia* (Clegg 2001; Clegg 2005), chironomids (Cornette *et al.*, 2017; Hinton, 1960), rotifers (van Leeuwenhoek, 1702), nematodes (Needham, 1743), and tardigrades (Corti, 1774; Keilin, 1959). Tolerance of extreme environments was first observed by Spallanzani (Spallanzani, 1776), whose report inspired several later studies. In particular, Doyère (Doyère, 1842) and Rahm (Rahm, 1921) performed extensive studies in tardigrades.

One major discussion point between Doyère and Pouchet was whether metabolism occurred in these desiccated animals (Broca, 1860). Other research provided evidence for both possibilities: Lance and Baumann suggested that metabolism may persist at very low levels (Baumann, 1922; Lance, 1896), while Rahm and Pigon argued that metabolism must completely cease, because anhydrobiotic organisms could survive sub-zero temperatures (Rahm 1921; Rahm 1926) and because carbon dioxide production was not detected in dried organisms (Pigon and Weglarska, 1955). This discussion was resolved by Keilin (Keilin, 1959), who concluded that Doyère’s hypothesis was correct, *i.e.*, that metabolism comes to a halt during anhydrobiosis. Furthermore, he coined the term cryptobiosis, which includes cryobiosis (induced by low temperature), osmobiosis (osmotic stress), anoxybiosis (low oxygen levels), and finally anhydrobiosis (desiccation). He stated cryptobiosis to be “*the state of an organism when it shows no visible signs of life and when its metabolic activity becomes hardly measurable or comes reversibly to a standstill*” (Keilin, 1959, p. 166). As previously stated, metabolism is an indispensable property of life, and therefore Clegg has proposed anhydrobiosis to be a third state of life, between “life” and “death” (Clegg, 2001).

<sup>2</sup>“*The following day the sky was again very hot and dry, and, about nine in the morning, I took some of the sediment which had been in the leaden gutter, which was then quite dried, and no thicker than half the back of a knife; it had also lain from the preceding evening in my study; this I put into a glass tube, about the thickness of a swan’s quill, and poured on it a small quantity of rain water taken out of my stone cistern, in which water were swimming some of the before mentioned Animalcules of the smaller sort; having poured in this water I mixed it up with the dry sediment or matter put into the tube, and which seemed very hard and compact, in order to dissolve the same; that thus, if there were still any living Animalcules in it, they might issue forth; though I confess I never thought that there could be any living creature in a substance so dried as this was.*” Translation from van Leeuwenhoek A, “On certain animalcules found in the sediment in gutters of the roofs of houses. Letter 144.”, p209, [https://www.dbnl.org/tekst/leeu027alle14\\_01/leeu027alle14\\_01\\_0010.php#b0233](https://www.dbnl.org/tekst/leeu027alle14_01/leeu027alle14_01_0010.php#b0233).

The rise of molecular biology during the 20th century has led to an understanding of how desiccation affects cells. The cell comprises several components, of which the cellular membrane is most affected by the loss of water. The lipid bilayer is sustained by hydrophobic interactions that are diminished by water depletion, causing cell lysis or, at least, leakage of internal components. Even if the membrane is sustained by some protective compound (Crowe *et al.*, 1987), the loss of cellular water molecules increases the concentration of intracellular solutes, resulting in osmotic stress. Perturbation of mitochondrial function also causes an increase in levels of reactive oxygen species (ROS), which in turn results in oxidative damage to various cell components. In particular, damage by oxidative stress in proteins are particularly important in the case of proteins (Daly *et al.*, 2007). Oxidation can damage a protein in a number of ways, some of which damage can be repaired or ameliorated. An example is protein carbonylation, which has been observed in tardigrades undergoing anhydrobiosis (Daly, 2012; Kuzmic *et al.*, 2018). DNA damage has also been observed in anhydrobiotic individuals (Gusev *et al.*, 2010; Hespels *et al.*, 2014; Neumann *et al.*, 2009; Rebecchi *et al.*, 2009). The damage that occurs during desiccation is hypothesized to be the result of ROS activity (Savic *et al.*, 2015). Thus, oxidative stress is considered the major cause of cellular damage during desiccation. Anhydrobiotes have evolved many mechanisms or adapted conserved mechanisms to protect against such damage.



**Figure 1.1: Active and anhydrobiotic states of *Macrobiotus shonaicus*.**

Scale bars indicate 50  $\mu\text{m}$  and 20  $\mu\text{m}$  for active (left) and tun (right), respectively. SEM images taken by Kenta Sugiura.

### 1.3 Tardigrades as a model organism for anhydrobiosis research

*Seltsam ist dieses Thierchen, weil der ganze Bau seines Körpers ausserordentlich und seltsam ist, und weil es in seiner äusserlichen Gestalt, dem ersten Anblicke nach, die grösste Aehnlichkeit mit einem Bäre im Kleinen hat. Das hat mich auch bewogen, ihm den Namen des kleinen Wasserbärs zu geben*<sup>3</sup>

Goeze, J. A. E. *Beobachtung: Über den kleinen Wasserbären*. 1773.

Tardigrades were originally described in 1773 by Goeze by the name of “Wasserbärs” - water bears in German (Goeze, 1773, p. 368). A year later, Corti first observed the recovery of tardigrades from anhydrobiosis (Corti, 1774), followed by Spallanzani who described the induction of anhydrobiosis in these organisms (Spallanzani, 1777). Spallanzani also coined the name Tardigrada from the Latin “*tardigrado*” (slow walker), which is also used as the name of the phylum to which tardigrades belong (Spallanzani, 1776). The phylogeny of Tardigrada within Metazoa has long been discussed, but has yet to be resolved (Borner *et al.*, 2014; Campbell *et al.*, 2011; Dunn *et al.*, 2008; Edgecombe, 2010; Nielsen, 2013). Ramazzotti established the phylum Tardigrada (Ramazzotti and Maucci, 1983), of which over 1,330 species are currently members (Degma *et al.*, 2020). Three orders exist within Tardigrada: approximately two-thirds of the phylum are classified as Eutardigrada, the remaining as Heterotardigrada. A third order of tardigrades, Mesotardigrada, has been proposed (Rahm, 1937), but is considered *nomen dubium* due to difficulties locating this species following the destruction of the holotype (Grothman *et al.*, 2017).

Tardigrades can be found in various aquatic and terrestrial environments (Nelson, 2002). Previous studies suggest that tardigrades originated in the ocean and acquired anhydrobiotic capability during terrestrialization (Renaud-Mornant, 1982). Individuals would be exposed to a variety of fluctuating stresses in the tidal zone (May, 1953), particularly water stress, and therefore would be under severe selective pressure to develop appropriate mechanisms of cellular protection. While many species in the two classes Parachaela and Apochela (Eutardigrada) can undergo anhydrobiosis, studies have also identified non-anhydrobiotic species, including *Isohypsibius myrops* (Ito *et al.*, 2016), *Dactylobiotus parthenogeneticus* (Guidetti *et al.*, 2006), *Thulinus ruffoi* (Sobczyk *et al.*, 2015), *Halobiotus crispae* (Guidetti *et al.*, 2006). Currently, only species within Eutardigrada can be reared under laboratory conditions, and most are presumed or have proven to be anhydrobiotic, *i.e.*, *Milnesium* (Sugiura *et al.*, 2020; Suzuki, 2003), *Ramazzottius* (Horikawa *et al.*, 2008), *Hypsibius* (Gabriel *et al.*, 2007; Gąsiorek *et al.*, 2018), *Isohypsibius* (Altiero and Rebecchi, 2001; Ito

<sup>3</sup>“This little animal is strange because the whole structure of its body is extraordinary and strange, and because in its outward form, at first sight, it has the greatest resemblance to a small bear. That also moved me to give him the name of the little water bear.”

*et al.*, 2016), *Acutuncus* (Tsujiimoto *et al.*, 2015), *Mesobiotus* (Guidetti *et al.*, 2019), *Macrobiotus* (Altiero and Rebecchi, 2001; Stec *et al.*, 2018; Sugiura *et al.*, 2019), *Paramacrobiotus* (Schill, 2013), *Diphascion* (Altiero and Rebecchi, 2001); thus, various molecular studies have focused on these species. In contrast, many species of Heterotardigrada inhabit the ocean; therefore, these species are hypothesized to be non-anhydrobiotic. Anhydrobiotic capabilities have been observed in several terrestrial, limno-terrestrial, and tidal species (Wright, 1989a), such as *Echiniscoides sigismundi* (Marcus, 1929), *Echiniscus testudo* (Doyère, 1842), *Cornechiniscus lobatus* (Schill *et al.*, 2007). Other species have been collected from terrestrial moss (Nelson, 2002) and are also presumed to be anhydrobiotic. These observations suggest the possibility of convergent evolution; anhydrobiosis may have been acquired independently by Heterotardigrada and Eutardigrada.

An important feature of tardigrades is that these organisms can enter anhydrobiosis at any point in their life cycle (Horikawa *et al.*, 2008; Schill and Fritz, 2008), which differs from other well-known anhydrobiotic invertebrates, textit*e.*, only the larval stages of *Polypedilum vanderplanki*, dauer stages of *Caenorhabditis elegans* (Erkut *et al.*, 2011; Hinton, 1960), and resting eggs of monogonont rotifers (Ricci, 1998) are desiccation tolerant. On the other hand, many Nematoda species and bdelloid rotifers are capable of anhydrobiosis at non-dauer stages (Barrett, 1982; Ricci and M., 1997; Shannon *et al.*, 2005). Most of these lineages are centered in a small phylogenetic clade (excluding bdelloid rotifers, for which most of the class is anhydrobiotic), and thus the acquisition of anhydrobiosis at the order level seen in tardigrades is characteristic. Such observations also suggest adaptation through the whole life cycle to enable survival within their variable habitats. Interestingly, anhydrobiosis capabilities differ greatly even within Tardigrada (Wright, 1989a), possibly according to the environment each species inhabits. For example, *Ramazzottius varieornatus* (Bertolani and Kinchin, 1993; Horikawa *et al.*, 2008), *Milnesium* sp. (described as *Milnesium tardigradum* in the original article) (Doyère, 1842; Horikawa and Higashi, 2004), and *Echiniscus testudo* (Doyère, 1842; Murai *et al.*, 2020) are capable of immediate desiccation. In contrast, *Hypsibius exemplaris* (Gąsiorek *et al.*, 2018) (formerly described as *Hypsibius dujardini* Doyère 1840) requires preconditioning for 1 – 2 days (Kondo *et al.*, 2015). These observations suggest the mechanisms of anhydrobiosis may differ between these two groups, either due to genome composition or regulation of vital components. Interestingly, the desiccation-sensitive *I. myrops* Marcus 1944 and *T. ruffoi* Bertolani 1981 are phylogenetically close relative of *R. varieornatus* and *H. exemplaris* (Bertolani *et al.*, 2014; Ito *et al.*, 2016; Sobczyk *et al.*, 2015). This suggests anhydrobiosis-related genes may be lost or not expressed in these species.

A famous characteristic of anhydrobiotic species is the ability to tolerate extreme environments, usually in the dry state. Studies in the late 1700s tested hydrated and anhydrobiotic specimens for their tolerance of various stresses and observed survival after exposure to high (approximately 100°C) and low (−196°C) temperatures (Becquerel 1950; Horikawa *et al.* 2008; Hengherr *et al.* 2009a; Hengherr *et al.* 2009b; Hengherr

*et al.* 2010), high pressure (7.5 GPa) (Horikawa *et al.*, 2009; Ono *et al.*, 2016; Seki and Toyoshima, 1998), organic solvents (1-butanol and 1-hexanol) (Ramløv and Westh, 2001), ammonia (Sobczyk *et al.*, 2015), salinity (100-1,000 mOsmkg<sup>-1</sup>) (Halberg *et al.*, 2009), copper ions (0.5–2 µg/L) (Hygum *et al.*, 2017), hydrogen peroxide (65 mM) (Bonifacio *et al.*, 2012), up to 30-60 kJ/m<sup>2</sup> ultraviolet B (UVB) and 20 kJ/m<sup>2</sup> ultraviolet C (UVC) radiation (Altiero *et al.*, 2011; Horikawa *et al.*, 2013; May *et al.*, 1964), over 5,000 Gy gamma radiation (Beltran-Pardo *et al.*, 2013; 2015; Horikawa *et al.*, 2006; 2008; 2012; Jönsson *et al.*, 2005; 2013; 2016; Jönsson and Wojcik, 2017; May *et al.*, 1964; Nilsson *et al.*, 2010), and the vacuum of space (Persson *et al.*, 2011). In particular, tardigrades are capable of tolerating UVC and gamma radiation even in the active state, suggesting that these species have enhanced repair systems (Jönsson, 2007). Such harsh environments do not exist on planet Earth, prompting the hypothesis that this remarkable hardiness is the result of cross-tolerance (Gade *et al.*, 2020), *i.e.*, using anhydrobiosis mechanisms to tolerate other stresses (Jönsson, 2003). Exposure to ionizing radiation causes direct damage to DNA, but also causes oxidative stress through the ionization of water molecules. The anti-oxidative stress mechanisms that enable anhydrobiosis are probably therefore an important factor in radiation tolerance (and other extremotolerances). Thus, a comparison of stress responses would be an ideal method to understand the mechanisms of anhydrobiosis.

Together, these features make tardigrades useful model organisms for understanding how a multicellular eukaryote with highly differentiated tissues tolerates complete desiccation.

## 1.4 Mechanisms of tardigrade anhydrobiosis

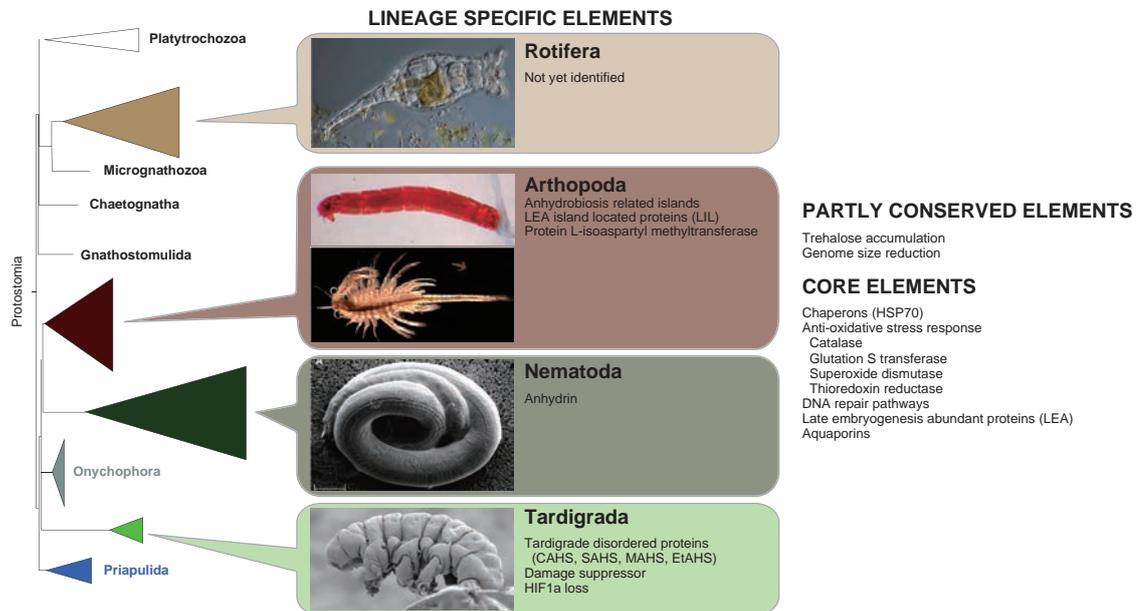
When tardigrades begin to dehydrate, they contract in the axial direction forming a “tun” shape. Tun formation has been proposed to be an active response; studies immersing active tardigrade specimens in ethanol – thus killing them – resulted in loss of tun formation during desiccation (Wright, 1989b). Furthermore, studies have suggested that the muscle system is involved in this contraction (Hygum *et al.*, 2016). Recent studies have successfully induced anhydrobiosis without tun formation, showing that the tun state is not necessary for anhydrobiosis (Hygum *et al.*, 2016), but these experiments used conditions rare in nature; thus, tun formation should be considered a normal feature of tardigrade anhydrobiosis. Ultrastructural studies have shown minimal structural changes in a tun compared to active tardigrades (Czerneková *et al.*, 2017; Richaud *et al.*, 2020; Schmitt, 1973; Walz, 1979; Wharton and Barrett, 1985). This suggests that tun formation may be a major contribution to anhydrobiosis and additional biochemical or molecular adaptations are responsible for the preservation of cellular structures. In contrast, changes in mitochondrial morphology have been observed in individuals recovering from the dry state, but not during entry into anhydrobiosis (Richaud *et al.*, 2020; Wharton and Barrett, 1985), suggesting that mitochondria are damaged at some point during anhydrobiosis, but not during the early stages.

Turning from morphological aspects, we find that evidence for the molecular mechanisms of anhydrobiosis has accumulated in the last 20 years using genetic information obtained in model organisms (Figure 1.2). An important factor of anhydrobiosis in some species is the disaccharide trehalose (Erkut *et al.*, 2011; Hengherr *et al.*, 2008a; Jönsson and Persson, 2010; Sakurai *et al.*, 2008), which has been proposed to protect cellular components by vitrification inside the cell. Accumulation of trehalose to high levels – up to 20-30% of dry weight (Erkut *et al.*, 2013; Watanabe *et al.*, 2002) – has been observed in various lineages (*e.g.*, chironomids, brine shrimp cysts, nematodes). However, such accumulation requires time, which reduces the number of environments such organisms can inhabit. Indeed, accumulation of such sugars at levels observed in nematodes and arthropods have not been observed in tardigrades and bdelloid rotifers (Arakawa, 2013; Jönsson and Persson, 2010; Lapinski and Tunnacliffe, 2003; Westh and Ramløv, 1991). The tardigrades *Milnesium* sp. seems to lack the trehalose synthesis pathway (Hengherr *et al.*, 2008b). This suggests that there may be an analogous or different mechanism that provides physical cellular protection. One candidate is the renowned intrinsically unstructured proteins (IUPs). Late embryogenesis abundant (LEA) proteins, initially found to be highly expressed in plant seeds (Dure *et al.*, 1981), have been suggested to play several protective roles (Czernik *et al.*, 2020; Marunde *et al.*, 2013; Moore *et al.*, 2016). Recent studies have suggested that several LEA proteins condense into membrane-less organelles by liquid-liquid phase separation, which protects cellular molecules and cellular integrity (Belott *et al.*, 2020). LEA orthologs have been found in various anhydrobiotic organisms, suggesting a common role in anhydrobiosis (Fu *et al.*, 2020; Hatanaka *et al.*, 2015; Tanaka *et al.*, 2015). Other well-known proteins that may have a role in anhydrobiosis include those highly conserved in eukaryotes: heat shock proteins (HSPs) (Jönsson and Schill, 2007; Ramløv and Westh, 2001; Reuner *et al.*, 2010; Schill *et al.*, 2004), cold shock proteins (CSPs) (Kamilari *et al.*, 2019), anti-oxidative stress proteins (superoxide dismutase, SOD; glutathione S-transferase, GST; thioredoxin; catalase) (Rizzo *et al.*, 2010), aquaporins (Grohme *et al.*, 2013; Kikawada *et al.*, 2008), and DNA repair enzymes such as RAD51 (Beltran-Pardo *et al.*, 2013; Carrero *et al.*, 2019; Förster *et al.*, 2012). Additionally, several proteins are reported to participate in the regulation of anhydrobiosis genes. In the anhydrobiotic midge *P. vanderplanki*, nuclear transcription factor Y (NF-Y) is suggested to be the master regulator of anhydrobiotic regulation (Yamada *et al.*, 2020), while heat shock factor (HSF) is implicated in arthropod anhydrobiosis (Mazin *et al.*, 2018; Tan and MacRae, 2018). In contrast, the tardigrade employs the protein phosphatase 1 (PP1)/ protein phosphatase 2A (PP2A) phosphorylation pathway combined with 5' adenosine monophosphate-activated protein kinase (AMPK) for stress signaling (Kondo *et al.* 2015; Kondo *et al.* 2019; Kondo *et al.* 2020). These findings suggest that different lineages may utilize different signaling pathways to control anhydrobiosis. Furthermore, metabolome analysis shows the accumulation of the osmoprotectant glycerophosphocholine during *R. varieornatus* anhydrobiosis (Arakawa, 2013), suggesting a rapid response to desiccation stress.

Tardigrades have acquired several novel IUPs that are highly expressed during anhydrobiosis (Figure 1.3), suggesting they contribute to desiccation tolerance. Recent studies have used the term “Tardigrade-specific intrinsically disordered proteins” (TDPs, Boothby *et al.* 2017) to describe these families (Janis *et al.*, 2018; Jönsson, 2019; Richaud *et al.*, 2020), but considering the broad definition of “disorderedness”, this is not sufficient to categorize these unstructured proteins. These proteins localize to specific cellular compartments: cytoplasmic abundant heat soluble (CAHS) protein in the cytoplasm; secretory abundant heat soluble (SAHS) protein to the extracellular matrix (Yamaguchi *et al.*, 2012); mitochondrial abundant heat soluble (MAHS) protein, and *Ramazzottius varieornatus* late embryogenesis abundant, mitochondrial (RvLEAM) protein, both to the mitochondria (Tanaka *et al.*, 2015). Additionally, a CAHS structural analog in Echiniscidae (*Echiniscus testudo* abundant heat soluble; EtAHS) has also been identified (Murai *et al.*, 2020). The CAHS protein has been suggested to contribute to cellular protection (Boothby *et al.*, 2017; Piszkiwicz *et al.*, 2019), but the evidence indicates that the protection provided does not surpass that of bovine serum albumin (BSA) and ubiquitin. The “glass transition” hypothesis proposed by Boothby *et al.* (2017) as a mechanism of protection by CAHS proteins is not widely accepted (Arakawa and Numata, 2021), and thus the function of these proteins remains unresolved. However, MAHS protein grants tolerance of osmotic stress to human cell lines (Tanaka *et al.*, 2015), implying that it contributes to anhydrobiosis. Another novel protein identified in tardigrades is the damage suppressor protein (Dsup), localizing to the nucleus (Hashimoto *et al.*, 2016). This protein is perhaps the most interesting protein found in tardigrades, as its expression in cells and plants confers increased X-ray tolerance (Hashimoto *et al.*, 2016; Kirke *et al.*, 2020). Computational analysis has suggested that this DNA-binding protein scavenges ROS to protect DNA (Minguez-Toral *et al.*, 2020). Nevertheless, exactly how these proteins contribute to anhydrobiosis remains to be elucidated. Additionally, these proteins are conserved only in Eutardigrada or a specific lineage (Bemm *et al.*, 2017; Chavez *et al.*, 2019; Hashimoto and Kunieda, 2017; Kamilari *et al.*, 2019), and thus a phylum-wide anhydrobiosis mechanism has not yet been identified.

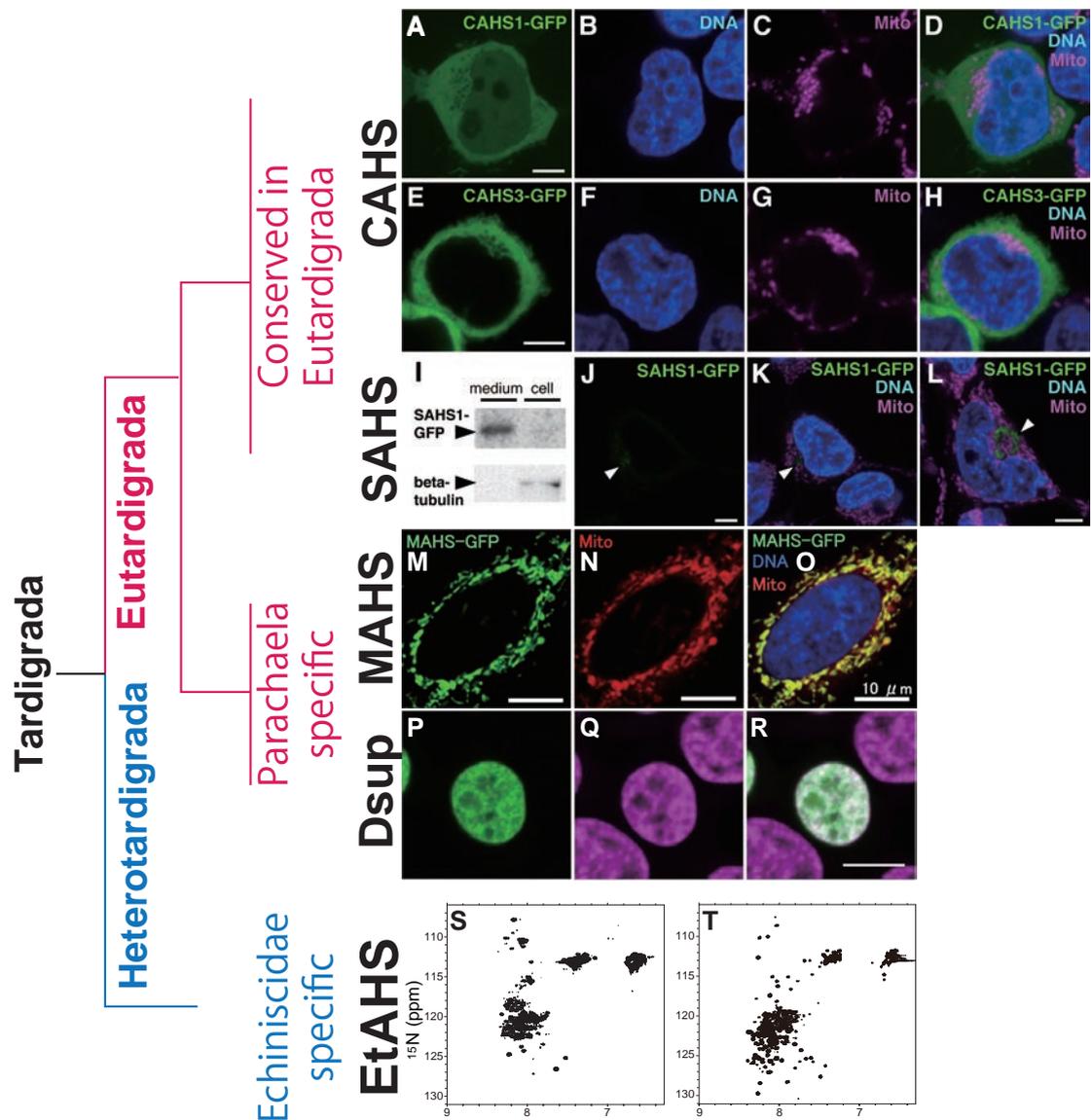
Together, these findings suggest that anhydrobiotic animals employ species-specific proteins as well as highly conserved factors to enable anhydrobiosis. However, the current catalogue of anhydrobiosis genes is not sufficient to *understand* anhydrobiosis. Many of the currently proposed anhydrobiosis proteins are conserved widely within eukaryotes, but why then are other species not capable of anhydrobiosis? Furthermore, CAHS genes alone do not confer anhydrobiosis (Boothby *et al.*, 2017; Yamaguchi *et al.*, 2012) and, although MAHS and Dsup do confer greater tolerance of osmotic stress or X-irradiation (Hashimoto and Kunieda, 2017; Tanaka *et al.*, 2015), this is not at the levels typical of anhydrobiosis. These proteins have been identified through omics-based analysis, and thus a *de novo* screening approach based on genomics would be an effective means of identifying other such genes. Additionally, previous studies have employed gain-of-function methods in human cell lines or other well-characterized organisms to attempt to demonstrate

how such genes/proteins effect anhydrobiosis (Hashimoto *et al.*, 2016; Kirke *et al.*, 2020; Tanaka *et al.*, 2015; Yamaguchi *et al.*, 2012). This approach was necessary because few genetic tools, such as gene knockout (*i.e.*, CRISPR/Cas9, TALEN, *etc.*), are not available in non-model organims like tardigrades. RNA interference (RNAi) is possible in *H. exemplaris* (Tenlen *et al.*, 2013), but knockdown experiments involving multi-copy genes with relatively high sequence similarity would prove to be a problem.



**Figure 1.2: Currently identified mechanisms of anhydrobiosis.**

Genes implicated to contribute to anhydrobiosis in various invertebrates. Images and phylogenetic trees were replicated from the following: Phylogenetic tree (Laumer *et al.*, 2019), *R. varieornatus* (Higashiyama and Arakawa), *Artemia* (Boothby and Pielak, 2017), *P. vanderplanki* (Watanabe *et al.*, 2004), *Panagrolaimus superbus* (Shannon *et al.*, 2005), *Rotaria macrura* (Yoshida *et al.*, 2019b).



**Figure 1.3: Conservation patterns of tardigrade specific anhydrobiosis genes.**

The cellular localization of tardigrade specific anhydrobiosis genes are indicated along their conservation patterns within Tardigrada. CAHS and SAHS within Eutardigrada, MAHS and Dsup are conserved within Parachaela, and EtAHS in Echiniscidae. [A,B,C,D] CAHS1 [E,F,G,H] CAHS2 [I,J,K,L] SAHS [M,N,O] MAHS [P,Q,R] Dsup [S,T] EtAHS. [A,E,J,M,P] GFP fluorescence of cells expressing the corresponding GFP-tagged protein [B,F,Q] Hoechst 33342 staining [C,G,N] MitoTracker Red CMXRos [D,H,K,L,O,R] Merged images of multiple channels. [I] Anti-GFP western blot of the SAHS expressing cells and its culture medium [S,T] 1H-15N HSQC spectra of [S] EtAHS A, [T] EtAHS B. The images were replicated from the following: CAHS and SAHS (Yamaguchi *et al.*, 2012), MAHS (Tanaka *et al.*, 2015), Dsup (Hashimoto *et al.*, 2016), EtAHS (Murai *et al.*, 2020).

## 1.5 Omics for analyzing anhydrobiosis

Biological research has accelerated massively following the establishment of sequencing technologies, represented by the completion of the human draft genome by the Human Genome Project in 2003. The genome sequence provides a comprehensive catalog of the genes that define the corresponding organism. Sequencing technologies have taken a further leap by high-throughput sequencing, using so-called next-generation sequencers. Illumina technologies now enable human genome sequencing at \$1,000, substantially cheaper than the three billion dollars used in the original Human Genome Project. Additionally, long-read sequencing technologies established by Pacific Biosciences (PacBio) and Oxford Nanopore Technologies have enabled the acquisition of high-quality genomes at low prices and low labor cost compared to conventional techniques requiring bacterial artificial chromosomes (BACs) or fosmids. These technologies have been applied to non-model organisms, allowing researchers to access genetic information to understand biological phenomena in these organisms.

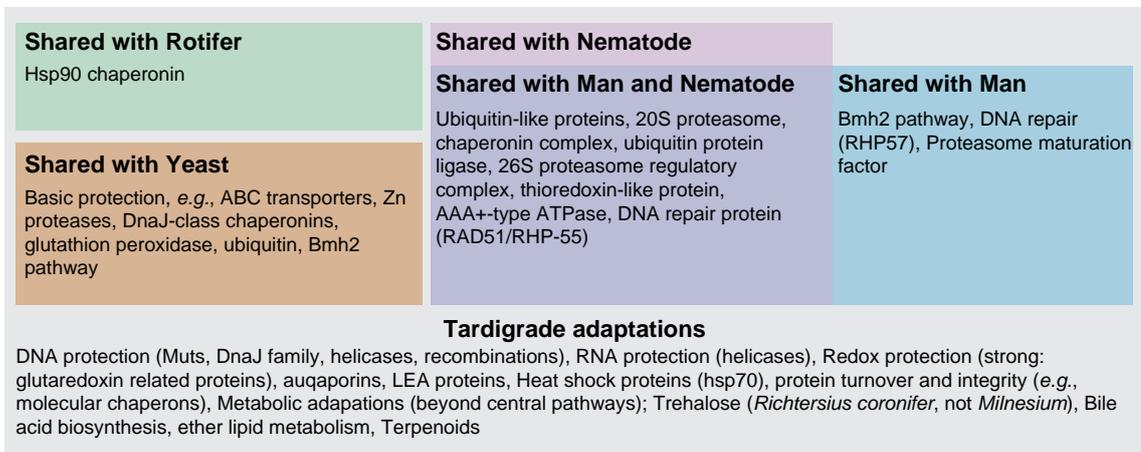
Prior to the rise of high-throughput sequencing, the majority of molecular studies were focused on the orthologs of genes that had been found to contribute to anhydrobiosis in other organisms (Grohme *et al.*, 2013; Jönsson and Schill, 2007; Reuner *et al.*, 2010; Schill *et al.*, 2004). Expressed sequence tag (EST) analysis followed these studies (Förster *et al.*, 2012; Mali *et al.*, 2010), identifying genes conserved in other well-analyzed species, *i.e.*, *C. elegans* and *Drosophila melanogaster*. This led to the proposal of the first model of the molecular machinery of anhydrobiosis (Figure 1.4). The establishment of high-throughput transcriptome sequencing (Wang *et al.*, 2014), shotgun proteomics (Schokraie *et al.* 2010; Schokraie *et al.* 2012), and metabolomics (Arakawa, 2013; Beisser *et al.*, 2012), allowed for more comprehensive and quantitative analysis of anhydrobiosis genes.

The first genome sequence in Tardigrada was that of *H. exemplaris* (Boothby *et al.*, 2015), closely followed by that of *R. varieornatus* (Hashimoto *et al.*, 2016). Hashimoto *et al.* (2016) combined fosmid clones sequenced by Sanger sequencing and Illumina reads to obtain a high-quality 56 Mbp genome of *R. varieornatus*. Comparison of genomic content with other metazoans identified gene duplication in anhydrobiosis genes and gene loss in stress-signaling pathways, including hypoxia inducible factor 1 (*HIF1*), *p53*, ataxia-telangiectasia mutated (*ATM*) genes and the mechanistic target of rapamycin (mTOR) pathway. Furthermore, this study identified the *Dsup* gene, a novel DNA-binding protein that enables nearly two-fold better survival of X-irradiation in human cell lines. This study emphasized the importance of genomic data for identifying components of the anhydrobiosis machinery, involving gene loss, gene duplication, and acquisition of novel genes. In contrast, Boothby *et al.* (2015) combined Illumina Moleculo long reads and short-insert mate-paired libraries, resulting in the 213 Mbp *H. exemplaris* genome. This study reported that the *H. exemplaris* genome has incorporated numerous horizontal gene transfers (HGTs),

comprising up to 17% of the genome. However, the genome size of *H. exemplaris* has been estimated to be 70-100 Mbp (Gabriel *et al.*, 2007; Koutsovoulos *et al.*, 2016), implying the existence of non-*Hypsibius* sequences. Indeed, several groups that independently sequenced the *H. exemplaris* genome or reanalyzed the released data observed many non-metazoan contaminants (Arakawa, 2016; Bemm *et al.*, 2016; Delmont and Eren, 2016; Koutsovoulos *et al.*, 2016). Furthermore, independent genome assemblies (Arakawa, 2016; Koutsovoulos *et al.*, 2016) also observed heterozygosity, resulting in a 130 Mbp genome. Thus, these genomes, including the contaminant-free versions, could not be subjected to comparative genomics to identify anhydrobiosis-related loci.

This genome contamination catastrophe indicated an important problem in invertebrate genomics, *i.e.*, the removal of contaminants. Due to the size of a tardigrade, one can obtain only tens of picograms of genomic DNA from a single individual. The sequencing projects of both the above species used large numbers of tardigrade specimens (*e.g.*, ~15,000 individuals for *R. varieornatus*). However, such large samples increase the possibility of non-tardigrade contamination, as seen in the *H. exemplaris* genome. Additionally, preparing such numbers of individuals is only possible for species that can be grown in large quantities in the laboratory, making the study of specimens collected from the wild impossible. Therefore, enabling contaminant-free genome sequencing from a small number of specimens has become a necessity for tardigrade genomics. Arakawa and others have established a DNA and RNA sequencing method from a single individual (Arakawa *et al.*, 2016; Arakawa, 2016; Yoshida *et al.*, 2018), which would be applicable for such analysis.

In summary, a high-quality genome of *H. exemplaris* is needed to resolve the controversy surrounding this species and to identify genomic loci contributing to anhydrobiosis.



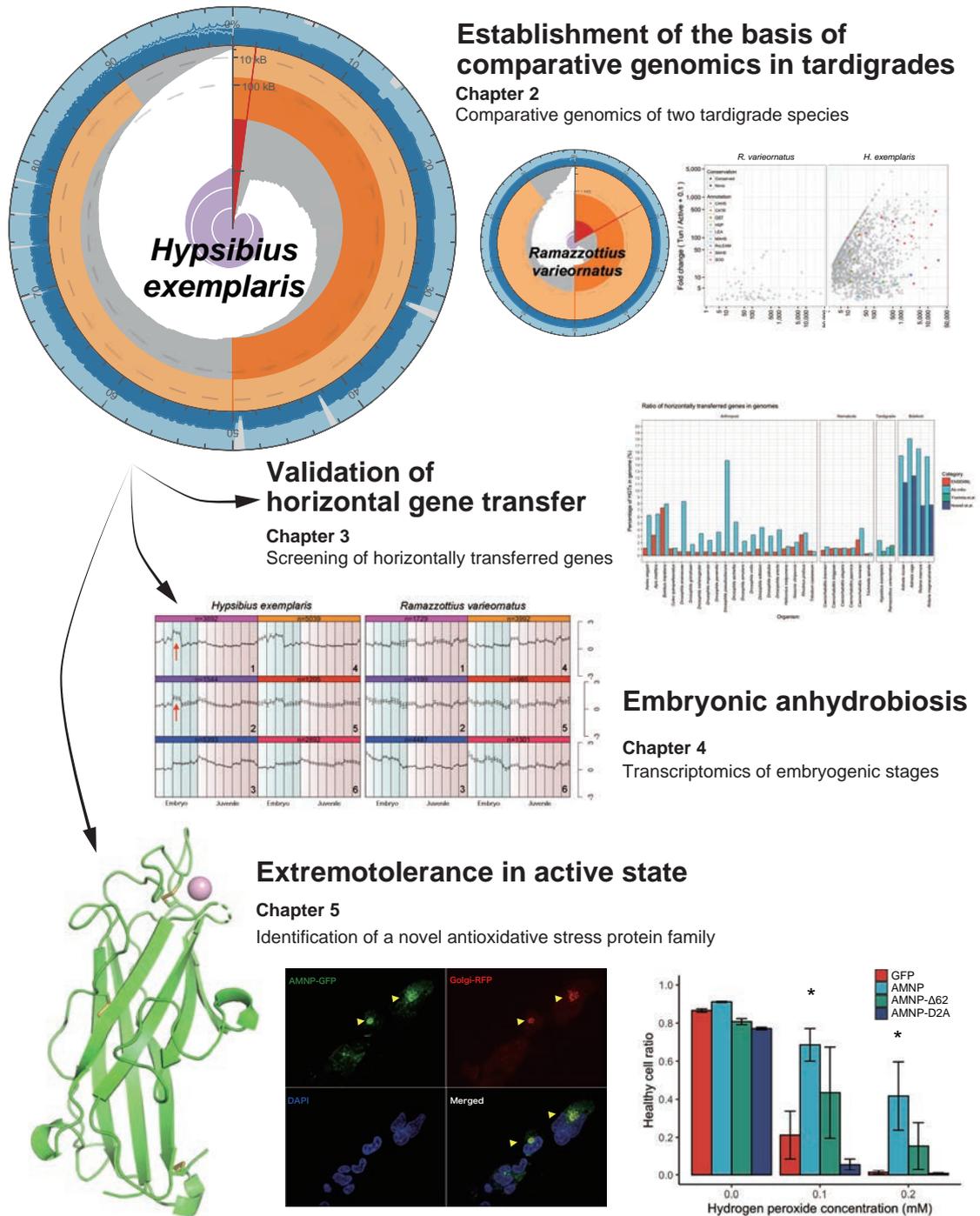
**Figure 1.4: The machinery of anhydrobiosis proposed prior to genomics.**

The model of anhydrobiosis mechanisms proposed before the application of comprehensive omics methods. Recreated from Förster *et al.* (2009).

## 1.6 Organization of this dissertation

As stated above, revealing the mechanisms that allow organisms to enter anhydrobiosis would lead to an understanding of the robustness of eukaryotic cells. A basic approach to understanding anhydrobiosis would be to compare the genetic content of anhydrobiotic and non-anhydrobiotic species. However, gathering large quantities of tardigrade specimens is challenging. Nevertheless, by obtaining comprehensive information on genetic content and its regulation, we should be able to identify novel and common anhydrobiotic factors that were previously overlooked.

In this dissertation (Figure 1.5), I provide evidence for the adaptation of tardigrade genomes to allow anhydrobiosis, obtained through omics-based analysis. I first obtained the genome sequence of *H. exemplaris* to use as a basis for comparative genomics in tardigrades (Chapter 2). Using this, I conducted comparative genomic and transcriptome analysis to identify candidate anhydrobiosis genes and to understand their regulation during entry into anhydrobiosis (Chapter 2). Additionally, because previous studies have discussed the HGT content of *H. exemplaris*, I reviewed the methods underlying HGT detection and determined that HGT levels are not elevated in tardigrades (Chapter 2 and 3). Furthermore, as *H. exemplaris* and *R. varieornatus* are both capable of anhydrobiosis in the embryonic stages, we anticipated that anhydrobiosis genes would be regulated during development. Hence, we submitted the embryos of both species to comparative transcriptomics to identify anhydrobiosis genes expressed at these stages (Chapter 4). Finally, we evaluated the expression of candidate anhydrobiosis genes in conditions other than anhydrobiosis, to obtain genes that may be involved in cross-tolerance in tardigrades (Chapter 5). Finally, I will discuss the perspectives derived from these results (Chapter 6).



**Figure 1.5: Organization of this dissertation.**

A summary of this dissertation. Chapters 2, 3, 4, 5 each are reprints of my published articles or preprints (Yoshida *et al.* 2017; Yoshida *et al.* 2019b; Yoshida *et al.* 2019a; Yoshida *et al.* 2020).

---

---

## **Chapter 2**

---

Comparative genomics of two tardigrades

## 2.1 Introduction

The superphylum Ecdysozoa emerged in the Precambrian, and ecdysozoans not only dominated the early Cambrian explosion but also are dominant (in terms of species, individuals, and biomass) today. The relationships of the 8 phyla within Ecdysozoa remain contentious, with morphological assessments, developmental analyses, and molecular phylogenetics yielding conflicting signals (Borner *et al.*, 2014; Campbell *et al.*, 2011; Dunn *et al.*, 2008). It has generally been accepted that Arthropoda, Onychophora (velvet worms), and Tardigrada (water bears or moss piglets) form a monophylum, Panarthropoda (Campbell *et al.*, 2011), and that Nematoda (roundworms) are closely allied to Nematomorpha (horsehair worms) and distinct from Panarthropoda. However, molecular phylogenies have frequently placed representatives of Tardigrada as sisters to Nematoda (Borner *et al.*, 2014; Dunn *et al.*, 2008), invalidating Panarthropoda and challenging models of the evolution of complex morphological traits such as segmentation, serially repeated lateral appendages, the triradiate pharynx, and a tripartite central nervous system (Edgecombe, 2010).

The key taxon in these disagreements is phylum Tardigrada. Nearly 1,200 species of tardigrades have been described (Degma *et al.*, 2017). All are members of the meiofauna—small animals that live in the water film and in interstices between sediment grain (Degma *et al.*, 2017). There are marine, freshwater, and terrestrial species. Many species of terrestrial tardigrades are cryptobiotic: they have the ability to survive extreme environmental challenges by entering a dormant state (Clegg, 2001). Common to these resistances is an ability to lose or exclude the bulk of body water, and anhydrobiotic tardigrades have been shown to have tolerance to high and low temperatures (including freezing), organic solvents, X- and gamma-rays, high pressure, and the vacuum of space (Altiero *et al.*, 2011; Hengherr *et al.*, 2009a; Horikawa *et al.*, 2013; Jönsson *et al.*, 2005; May *et al.*, 1964; Ono *et al.*, 2016; Persson *et al.*, 2011; Ramløv and Westh, 2001). The physiology of anhydrobiosis in tardigrades has been explored extensively, but little is currently known about its molecular bases (Beltran-Pardo *et al.*, 2013; Mali *et al.*, 2010). Many other animals have cryptobiotic abilities, including some nematodes and arthropods (Gusev *et al.*, 2014), and comparison of the mechanisms in different independent acquisitions of this trait will reveal underlying common mechanisms.

Central to the development of tractable experimental models for cryptobiosis is the generation of high-quality genomic resources. Genome assemblies of two tardigrades, *H. exemplaris* (Arakawa, 2016; Boothby *et al.*, 2015; Koutsovoulos *et al.*, 2016) and *R. varieornatus* (Hashimoto *et al.*, 2016), both in the family Hypsibiidae, have been published. *H. exemplaris* is a limnoterrestrial tardigrade that is easy to culture (Kondo *et al.*, 2015), while *R. varieornatus* is a terrestrial tardigrade and highly tolerant of environmental extremes (Horikawa *et al.*, 2008). An experimental toolkit for *H. exemplaris*, including RNA interference (RNAi) and *in situ* hybridization, is being developed (Smith *et al.*, 2016; Tenlen *et al.*, 2013). *H. exemplaris* is poorly cryptobiotic compared to *R. varieornatus*. *H. exemplaris* requires 48 h of preconditioning at

85% relative humidity (RH) and a further 24 h in 30% RH (Kondo *et al.*, 2015) to enter cryptobiosis with high survival, while *R. varieornatus* can form a tun (the cryptobiotic form) within 30 minutes at 30% RH (Horikawa *et al.*, 2008).

Several anhydrobiosis-related genes have been identified in Tardigrada. Catalases, superoxide dismutases (SODs), and glutathione reductases may protect against oxidative stress (Rizzo *et al.*, 2010), and chaperones, such as heat shock protein 70 (HSP70) (Jönsson and Schill, 2007; Reuner *et al.*, 2010; Schill *et al.*, 2004), may act to protect proteins from the denaturing effects of water loss (Beltran-Pardo *et al.*, 2013; Hengherr *et al.*, 2008b; Moore *et al.*, 2016). Additionally, several tardigrade-specific gene families have been implicated in anhydrobiosis, based on their expression patterns. Cytosolic abundant heat soluble (CAHS), secretory abundant heat soluble (SAHS), late embryogenesis abundant protein mitochondrial (LEAM), mitochondrial abundant heat soluble protein (MAHS), and damage suppressor (Dsup) gene families have been implicated in *R. varieornatus* extremotolerance (Boothby *et al.*, 2017; Hashimoto *et al.*, 2016; Tanaka *et al.*, 2015; Yamaguchi *et al.*, 2012). These gene families were named by their subcellular location or function, and expression of MAHS and Dsup in human tissue culture cell lines resulted in elevated levels of tolerance against osmotic stress and X-ray irradiation (approximately 4 Gy). Surprisingly, analyses of the *R. varieornatus* genome showed extensive gene loss in the peroxisome pathway and in stress signaling pathways, suggesting that this species is compromised in terms of reactive oxygen resistance and repair of cellular damage (Hashimoto *et al.*, 2016). While loss of these pathways would be lethal for a normal organism, loss of these resistance pathways may be associated with anhydrobiosis.

Desiccation in some taxa induces the production of anhydroprotectants, small molecules that likely replace cellular water to stabilize cellular machinery. Trehalose, a disaccharide shown to contribute to anhydrobiosis in midges (da Costa Morato Nery *et al.*, 2008; Kikawada *et al.*, 2007), nematodes (Madin and Crowe, 1975), and artemia (Clegg, 1967), is not present in the tardigrade *Milnesium tardigradum* (Hengherr *et al.*, 2008b). Coupled with the ability of *R. varieornatus* to enter anhydrobiosis rapidly (*i.e.*, without the need for extensive preparatory biosynthesis), this suggests that tardigrade anhydrobiosis does not rely on induced synthesis of protectants. Entry into anhydrobiosis in *H. exemplaris* does require active transcription during preconditioning (Kondo *et al.*, 2015), suggesting the activation of a genetic program to regulate physiology. Inhibition of PP1/2A, a positive regulator of the forkhead box O (FOXO) transcription factor that induces antioxidative stress pathways, led to high lethality in *H. exemplaris* during anhydrobiosis induction (Kondo *et al.*, 2015). As *R. varieornatus* does not require preconditioning, systems critical to anhydrobiosis in *R. varieornatus* are likely to be constitutively expressed.

An interesting artifact of anhydrobiosis is the acquisition of tolerance against other extreme stress. One of the main focus of tardigrade cross-tolerance has been radiation, due to the fact that tardigrades show high tolerance during not only the anhydrobiotic but the active state as well. Although conditions such as

food addition and radiation sources differ between studies, most studies are consistent with the conclusion that eutardigrades, as well as heterotardigrades, have LD<sub>50/48h</sub> values of 3,000 to 5,000 Gy (Bakri *et al.*, 2005; Beltran-Pardo *et al.*, 2015; Chinnasri *et al.*, 1997; Horikawa *et al.*, 2006; 2008; Jönsson *et al.*, 2005; 2016; May *et al.*, 1964; Nilsson *et al.*, 2010), suggesting that tardigrades generally have high radiation tolerance. Ionizing radiation, a form of electromagnetic wave, is known to damage cellular components, and to cause somatic mutations, protein damage, and apoptosis (Daly, 2012). In particular, gamma rays and heavy ion beams can induce devastating damage to cellular molecules. Gamma-ray irradiation at a high dosage can cause massive DNA damage, through direct and indirect actions. High tolerance in the anhydrobiotic state may originate from the lack of water or protective molecules that enable anhydrobiosis. The absence of water in the anhydrobiotic state would prevent generation of reactive oxygen species, thus indirect damage would be suppressed. Previous studies on *R. varieornatus* have identified Dsup, a protein that may contribute to DNA protection (Hashimoto *et al.*, 2016), although this protein is not conserved in other tardigrades (Bemm *et al.*, 2017). On the other hand, these mechanisms do not explain the conserved tolerance in the active state. One hypothesis is that high tolerance in the active state may be due to functional anti-oxidant and DNA repair pathways: such mechanisms would not be operative during lack of water. While most analyses on cross-tolerance in tardigrades have concentrated on zoological aspects, there are several molecular studies focused on a limited number of highly conserved genes. Additionally, it has been hypothesized in bacteria and plants that genome size contributes greatly to tolerance against gamma rays. Although the total amount of gamma-ray tolerable differs between organisms, the number of double strand breaks (DSBs) per 1 Mb per 1 Gy of gamma-ray a genome can tolerate has been estimated to be 0.002 to 0.006, conserved from bacteria to fungi, rotifers, and human cell lines (Anno *et al.*, 2003; Bakri *et al.*, 2005; Chinnasri *et al.*, 1997; Daly, 2012; Su *et al.*, 1990; Watanabe *et al.*, 2006; 2007). Such quantitative analysis has yet to be conducted in tardigrades, however, provided that anhydrobiosis itself caused DNA damage in *M. tardigradum* (Neumann *et al.*, 2009), irradiation of ionizing radiation would presumably cause DNA damage. These hypotheses suggest that there may be prominent cellular repair mechanisms in tardigrades, in which a more comprehensive analysis may be effective.

*H. exemplaris* and *R. varieornatus* are relatively closely related (both are members of Hypsibiidae), and both have available genome sequences. The *R. varieornatus* genome has high contiguity and scores highly in all metrics of gene completeness (Hashimoto *et al.*, 2016). For *H. exemplaris*, three assemblies have been published. One has low contiguity (N50 length of 17 kb) and contains a high proportion of contaminating nontardigrade sequence, including approximately 40 Mb of bacterial sequence, and spans 212 Mb (Boothby *et al.*, 2015). The other two assemblies, both at approximately 130 Mb (Arakawa, 2016; Koutsovoulos *et al.*, 2016), eliminated most contamination, but contained uncollapsed haploid segments because of unrecognized heterozygosity. The initial low-quality *H. exemplaris* genome was published alongside a claim of extensive

horizontal gene transfer (HGT) from bacteria and other taxa into the tardigrade genome and a suggestion that HGT might have contributed to the evolution of cryptobiosis (Boothby *et al.*, 2015). The extensive HGT claim has been robustly challenged (Arakawa, 2016; Bemm *et al.*, 2016; Boothby and Goldstein, 2016; Delmont and Eren, 2016; Koutsovoulos *et al.*, 2016), but the debate as to the contribution of HGT to cryptobiosis remains open. The genomes of these species could be exploited for understanding the mechanisms of rapid-desiccation versus slow-desiccation strategies in tardigrades, the importance of HGT, and the resolution of the deep structure of the Ecdysozoa. However, the available genomes are not of equivalent quality.

We have generated a high-quality genome assembly for *H. exemplaris*, from an array of data including single-tardigrade sequencing (Arakawa *et al.*, 2016) and long, single-molecule reads, and using a heterozygosity-aware assembly method (Challis *et al.*, 2017; Kajitani *et al.*, 2014). Gene finding and annotation with extensive RNA sequencing (RNA-Seq) data allowed us to predict a robust gene set. While most (60%) of the genes of *H. exemplaris* had direct orthologues in an improved gene prediction for *R. varieornatus*, levels of synteny were very low. We identified an unremarkable proportion of potential HGTs. *H. exemplaris* showed losses of peroxisome and stress signaling pathways, as described in *R. varieornatus*, as well as additional unique losses. Transcriptomic analysis of anhydrobiosis entry detected higher levels of regulation in *H. exemplaris* compared to *R. varieornatus*, as predicted, including regulation of genes with anti-stress and apoptosis functions. Using single-copy orthologues, we reanalyzed the position of Tardigrada within Ecdysozoa and found strong support for a Tardigrade+Nematode clade, even when data from transcriptomes of a nematomorph, onychophorans, and other ecdysozoan phyla were included. However, rare genomic changes tended to support the traditional Panarthropoda. We discuss our findings in the context of how best to improve genomics of neglected species, the biology of anhydrobiosis, and conflicting models of ecdysozoan relationships.

## 2.2 Materials and methods

### 2.2.1 Tardigrade culture and sampling

The tardigrade *H. exemplaris* Z151 was purchased from Sciento (Manchester, United Kingdom). *H. exemplaris* Z151 and *R. varieornatus* strain YOKOZUNA-1 were cultured as previously described (Arakawa *et al.*, 2016; Horikawa *et al.*, 2008). Briefly, tardigrades were fed *Chlorella vulgaris* (Chlorella Industry) on 2% Bacto Agar (Difco) plates prepared with Volvic water, incubated at 18°C for *H. exemplaris* and 22°C for *R. varieornatus* under constant dark conditions. Culture plates were renewed approximately every 7-8 d. Anhydrobiotic adult samples were isolated on 30 µM filters (Millipore) and placed in a chamber maintained

at 85% RH for 48 h for *H. exemplaris*; 30% RH for 24 h and additional 24 h at 0% RH for slow-dried *R. varieornatus*; and 0% RH for 30 min on a 4 cm x 4 cm Kim-towel with 300  $\mu$ L of distilled water and an additional 2 h without the towel for fast-dried *R. varieornatus*. Successful anhydrobiosis was assumed when >90% of the samples prepared in the same chamber recovered after rehydration.

### 2.2.2 Genome and transcriptome sequencing

Genomic DNA for long read sequencing was extracted using MagAttract HMW DNA Kit (Qiagen) from approximately 900,000 *H. exemplaris*. DNA was purified twice with AMPure XP beads (Beckman Coulter). A 20 kb PacBio library was prepared following the manual “20 kb Template Preparation Using BluePippin Size-Selection System (15 kb Size Cutoff)” (PacBio SampleNet—Shared Protocol) using SMARTBell Template Prep Kit 1.0 (Pacific Biosciences) and was sequenced using eight SMRT Cells Pac V3 with DNA Sequencing Reagent 4.0 on a PacBio RSII System (Pacific Biosciences) at Takara Bio. Briefly, purified DNA was sheared, targeting 20 kb fragments, using a g-TUBE (Covaris). Following end-repair and ligation of SMRTbell adapters, the library was size selected using BluePippin (Sage Science) with a size cutoff of 10 kb. The size distribution of the library was assayed on TapeStation 2200 (Agilent Technologies) and quantified using the Quant-iT dsDNA BR Assay Kit (Invitrogen). MiSeq reads from a single *H. exemplaris* individual (DRR055040) are from our previous reports (Arakawa, 2016).

For mRNA-Seq to be used in genome annotation, 30 individuals were collected from each of the following conditions in 3 replicates: active and dried adults (slow dried for *R. varieornatus*), eggs (1, 2, 3, 4, 5, 6, and 7 d after laying), and juveniles (1, 2, 3, 4, 5, 6, and 7 d after hatching). Because of sample preparations, *R. varieornatus* juveniles were sampled until juvenile first day. For gene expression analysis, we sampled approximately 2–3 individuals of fast-dried *R. varieornatus*. All mRNA-Seq analyses were conducted with three replicates. Specimens were thoroughly washed with Milli-Q water on a sterile nylon mesh (Millipore) and immediately lysed in TRIzol reagent (Life Technologies) using 3 cycles of immersion in liquid nitrogen followed by 37°C incubation. Total RNA was extracted using the Direct-zol RNA kit (Zymo Research) following the manufacturer’s instructions, and RNA quality was checked using the High Sensitivity RNA ScreenTape on a TapeStation (Agilent Technologies). For library preparation, mRNA was amplified using the SMARTer Ultra Low Input RNA Kit for Sequencing v.4 (Clontech), and Illumina libraries were prepared using the KAPA HyperPlus Kit (KAPA Biosystems). Purified libraries were quantified using a Qubit Fluorometer (Life Technologies), and the size distribution was checked using the TapeStation D1000 ScreenTape (Agilent Technologies). Libraries size selected above 200 bp by manually excision from agarose were purified with a NucleoSpin Gel and PCR Clean-up Kit (Clontech). The samples were then sequenced on the Illumina NextSeq 500 in High Output Mode with a 75-cycle kit (Illumina) as single end reads, with 48 multiplexed samples per run. Adapter sequences were removed, and the sequences

were demultiplexed using the bcl2fastq v.2 software (Illumina). For active and dried adults, RNA-Seq was also conducted starting from approximately 10,000 individuals, similarly washed, but RNA extraction was conducted with TRIzol reagent (Life Technologies) followed by RNeasy Plus Mini Kit (Qiagen) purification. Library preparation and sequencing were conducted at Beijing Genomics Institute.

For miRNA-Seq, 5,000 individuals were homogenized using Biomasher II (Funakoshi), and TRIzol (Invitrogen) was used for RNA extraction; the individuals were then purified by isopropanol precipitation. Size selection of fragments of 18-30 nt using electrophoresis, preparation of the sequencing library for Illumina HiSeq 2000, and subsequent (single end) sequencing were carried out by Beijing Genomics Institute.

All sequence data were validated for quality using FastQC (Andrews, 2015).

### 2.2.3 Genome assembly

The MiSeq reads from whole genome amplified DNA were merged with Usearch (Edgar, 2010), and both merged and unmerged pairs were assembled with SPAdes (Bankevich *et al.*, 2012) as single end. The SPAdes assembly was checked for contamination with BLAST+ BLASTN (Camacho *et al.*, 2009) against the nr (O'Leary *et al.*, 2016) database, and no observable contamination was found with blobtools (Kumar *et al.*, 2013). Illumina data from Boothby *et al.* (Boothby *et al.*, 2015) were mapped to the SPAdes assembly with Bowtie2 (Langmead and Salzberg, 2012), and read pairs were retained if at least 1 of them mapped to the assembly. These reads were then assembled, scaffolded, and gap closed with Platanus (Kajitani *et al.*, 2014). The Platanus assembly was further scaffolded and gap closed using the PacBio data with PBJelly (English *et al.*, 2012).

Falcon (Challis *et al.*, 2017) assembly of PacBio data was performed on the DNAnexus platform. Using this Falcon assembly, Platanus assembly was extended using SSPACE-LongReads (Boetzer and Pirovano, 2014) and gap-filled with PBJelly (English *et al.*, 2012) with default parameters. Single-individual MiSeq reads were mapped to the assembly with BWA MEM (Li *et al.*, 2009), and all contigs with coverage < 1 or length < 1,000 bp or those corresponding to the mitochondrial genome were removed. At this stage, 1 CEGMA gene became unrecognized by CEGMA (Parra *et al.*, 2007), probably because of multiple PBJelly runs, and therefore, the contig harboring that missing CEGMA gene was corrected by Pilon (Walker *et al.*, 2014) using the single individual MiSeq reads. We also validated genomic completeness with BUSCO using the eukaryote lineage gene set.

### 2.2.4 Gene finding

mRNA-Seq data (Development, Active-tun 10k animals, Supplementary Table S1) were mapped to the genome assembly with TopHat2 (Kim *et al.*, 2013; Langmead and Salzberg, 2012) without any options. Using the mapped data from TopHat, BRAKER (Hoff *et al.*, 2016) was used with default settings to construct a GeneMark-ES (Borodovsky and Lomsadze, 2011) model and an Augustus (Keller *et al.*, 2011) gene model, which are used for *ab initio* prediction of genes. The getAnnotFasta.pl script from Augustus was used to extract coding sequences from the GFF3 file. Similarly, to construct a modified version of the *R. varieornatus* genomes annotation, we used the development and anhydrobiosis (Supplementary Table S2) RNA-Seq data for BRAKER annotation. We found that a few genes were misannotated (MAHS in both species, a CAHS orthologue in *R. varieornatus*), due to fusion with a neighboring gene, and these were manually curated. tRNA and rRNA genes were predicted with tRNAscan-SE (Lowe and Eddy, 1997) and RNAmmer (Lagesen *et al.*, 2007), respectively. BUSCO was used again to validate the completeness of the predicted gene set for both tardigrades.

The mRNA-Seq data used to predict the gene models were mapped with BWA MEM (Li *et al.*, 2009) against the predicted coding sequences, the genome, and a Trinity (v2.2.0) (Grabherr *et al.*, 2011) assembled transcriptome. We also mapped the mRNA-Seq data used for gene expression analysis (single individual *H. exemplaris* and fast/slow dry of *R. varieornatus*) of the active state and tun state. After SAM to BAM conversion and sorting with SAMtools view and sort (Li and Durbin, 2009), we used QualiMap (Okonechnikov *et al.*, 2016) for mapping quality check.

To annotate the predicted gene models, we performed similarity searches using BLAST BLASTP (Altschul *et al.*, 1997) against Swiss-Prot, TrEMBL (UniProt Consortium, 2015), and HMMER hmmsearch (Mistry *et al.*, 2013) against Pfam-A (Finn *et al.*, 2016) and Dfam (Hubley *et al.*, 2016), KAAS analysis for KEGG orthologue mapping (Moriya *et al.*, 2007), and InterProScan (Goujon *et al.*, 2010) for domain annotation. We used RepeatScout (Price *et al.*, 2005) and RepeatMasker (Smit *et al.*, 2015) for *de novo* repeat identification. To compare *H. exemplaris* gene models to those of *R. varieornatus*, we also ran BLAST BLASTP searches against the updated *R. varieornatus* proteome and TBLASTN search against the *R. varieornatus* genome and extracted bidirectional best hits with in-house Perl scripts.

For miRNA prediction, we used miRDeep (Friedlander *et al.*, 2012) to predict mature miRNA within the genome, using the mature miRNA sequences in miRBase (Kozomara and Griffiths-Jones, 2014). The predicted mature miRNA sequences were then searched against miRBase with ssearch36 (Pearson and Lipman, 1988) for annotation by retaining hits with identity >70% and a complete match of bases 1-7, 2-8, or 3-9.

### 2.2.5 HGT identification

HGT genes were identified using the HGT index approach (Boschetti *et al.*, 2012). Swiss-Prot and TrEMBL were downloaded (UniProt Consortium, 2015), and sequences with “Complete Proteome” in the Keyword were extracted. Following the method of Boschetti *et al.*, an Arthropoda-less and Nematoda-less database was constructed. These databases were searched with Diamond BLASTX (Buchfink *et al.*, 2015) using as query all coding sequences, using the longest transcript for each gene. Hits with an E-value below  $1e^{-5}$  were kept. The HGT index ( $Hu$ ) was calculated as  $Bo - Bm$ , the bit score difference between the best nonmetazoan hit ( $Bo$ ) and the best metazoan hit ( $Bm$ ), and genes with  $Hu \geq 30$  were identified as HGT candidates.

To assess if *ab initio* annotation of genomes biases the calculation of the HGT index, we calculated HGT indices for genomes in Ensembl-Metazoa (Aken *et al.*, 2017) that had corresponding Augustus (Keller *et al.*, 2011) gene models and ran *ab initio* gene prediction. We analyzed *Aedes aegypti*, *Apis mellifera*, *Bombus impatiens*, *Caenorhabditis brenneri*, *Caenorhabditis briggsae*, *Caenorhabditis elegans*, *Caenorhabditis japonica*, *Caenorhabditis remanei*, *Culex quinquefasciatus*, *Drosophila ananassae*, *Drosophila erecta*, *Drosophila grimshawi*, *Drosophila melanogaster*, *Drosophila mojavensis*, *Drosophila persimilis*, *Drosophila pseudoobscura*, *Drosophila sechellia*, *Drosophila simulans*, *Drosophila virilis*, *Drosophila willistoni*, *Drosophila yakuba*, *Heliconius melpomene*, *Nasonia vitripennis*, *Rhodnius prolixus*, *Tribolium castaneum*, and *Trichinella spiralis*. Gene predictions for each organism were conducted using autoAugPred.pl from the Augustus package with the corresponding model (Supplementary Table S3). The longest isoform sequences for all genes were extracted for both Ensembl and *ab initio* annotations, and the HGT index was calculated for each gene in all organisms. To assess if using Diamond BLASTX biases HGT index calculation, we ran BLAST BLASTX (Altschul *et al.*, 1997) searches with *H. exemplaris* and calculated the HGT index using the same pipeline.

The blast-score-based HGT index provided a first-pass estimate of whether a gene had been horizontally transferred from a nonmetazoan species. Phylogenetic trees were constructed for each of the 463 candidates (based on the HGT index) along with their best blast hits as described above (Supplementary Data S1). Protein sequences for the blast hits were aligned with the HGT candidate using MAFFT (Katoh and Standley, 2013). Randomized Accelerated Maximum Likelihood (RAxML) (Stamatakis, 2014) was used to build 450 individual trees, as 13 of the protein sets had less than 4 sequences and trees could not be built for them. HGT candidates were categorized as prokaryotes, viruses, metazoan, and nonmetazoan (specifically, eukaryotes that were nonmetazoan, such as fungi) based on the monophyletic clades that they were placed in. Any that could not be classified monophyletically were classified as “complex”: these were split into complex non-HGT (where the complexity was within a metazoan radiation) or complex HGT (where HGT was affirmed but affinities remained unclear) (Supplementary Data S2). OrthoFinder (Emms and Kelly,

2015) with default BLAST+ BLASTP search settings and an inflation parameter of 1.5 was used to identify orthogroups containing *H. exemplaris* and *R. varieornatus* protein-coding genes. These orthogroups were used to identify the *R. varieornatus* HGT homologues of *H. exemplaris* HGT candidates. HGT candidates were classified as having high gene expression levels if they had an average gene expression greater than the overall average gene expression level of 1 TPM.

### 2.2.6 Anhydrobiosis analyses

To identify genes responsive to anhydrobiosis, we explored transcriptome (Illumina mRNA-Seq) data for both *H. exemplaris* and *R. varieornatus*. Individual mRNA-Seq data for *H. exemplaris* (Arakawa *et al.*, 2016) before and during anhydrobiosis were contrasted with new sequence data for *R. varieornatus* similarly treated. We mapped the mRNA-Seq reads to the coding sequences of the relevant species with BWA MEM (Li *et al.*, 2009), and after summarizing the read count of each gene, we used DESeq2 (Love *et al.*, 2014) for differential expression calculation, using false discovery rate (FDR) correction. Genes with a FDR below 0.05, an average expression level (in transcripts per kilobase of model per million mapped fragments; TPM) of over 1, and a fold change over 2 were defined as differentially expressed genes. Gene expression (TPM) was calculated with Kallisto (Bray *et al.*, 2016) and was parsed with custom Perl scripts. To assess if there were any differences in fold change distributions, we used R to calculate the fold change for each gene (anhydrobiotic / [active + 0.1]) and conducted a U test using the `wilcox.test()` function. We mapped the differentially expressed genes to KEGG pathway maps (Okuda *et al.*, 2008) to identify pathways that were likely to be differentially active during anhydrobiosis.

### 2.2.7 Protein family analyses and comparative genomics

For comparison with *R. varieornatus*, we first aligned the genomes of *H. exemplaris* and *R. varieornatus* with Murasaki and visualized with `gmv` (Popendorf *et al.*, 2010). The lower `tf-idf` anchor filter was set to 500. A syntenic block was observed between scaffold0001 of *H. exemplaris* and scaffold002 of *R. varieornatus*. We extracted the corresponding regions (*H. exemplaris*: scaffold0001; 363,334 – 2,100,664, *R. varieornatus*: scaffold002; 2,186,607 – 3,858,816), and conducted alignment with Mauve (Darling *et al.*, 2004). Tardigrade-specific, protection-related genes (CAHS, SAHS, MAHS, RvLEAM, and Dsup) were identified by BLASTP, subjected to phylogenetic analysis using Clustalw2 (Goujon *et al.*, 2010) and FastTree (Price *et al.*, 2010), and visualized with FigTree (Rambaut, 2016).

Single-copy orthologues between *H. exemplaris* and *R. varieornatus* were identified using the orthologous groups defined by OrthoFinder. Using the Ensembl Perl API, gene structure information (gene lengths, exon counts, and intron spans per gene) were extracted for these gene pairs. To avoid erroneous gene predic-

tions biasing observed trends, *H. exemplaris* genes that were 20% longer or 20% shorter were considered outliers.

HOX loci were identified using BLAST, and their positions on scaffolds and contigs assessed. To identify HOX loci in other genomes, genome assembly files were downloaded from Ensembl Genomes (Aken *et al.*, 2017) or Wormbase ParaSite (Howe *et al.*, 2016a;b) and formatted for local search with BLAST+ (Camacho *et al.*, 2009). Homeodomain alignments were generated using Clustal Omega (Sievers *et al.*, 2011), and phylogenies estimated with RAxML (Stamatakis, 2014) to classify individual homeodomains.

Protein predictions from genomes of Annelida (1 species), Nematoda (9), Arthropoda (15), Mollusca (1), and Priapulida (1) were retrieved from public databases (Supplementary Table S4). Proteomes were screened for isoforms (Supplementary Data S3), and the longest isoforms were clustered with the proteins of *H. exemplaris* and *R. varieornatus* using OrthoFinder 1.1.2 (Emms and Kelly, 2015) at different inflation values (Supplementary Data S4). Proteins from all proteomes were functionally annotated using InterProScan (Goujon *et al.*, 2010). OrthoFinder output was analyzed using KinFin (Laetsch and Blaxter, 2017) under 2 competing phylogenetic hypotheses: either (1) “Panarthropoda”, where Tardigrada and Arthropoda share a more recent common ancestor distinct from Nematoda, or (2) Tardigrada and Nematoda sharing a more recent common ancestor distinct from Arthropoda. (see Supplementary Data S5 for input files used in KinFin analysis). Enrichment and depletion in clusters containing proteins from Tardigrada and other taxa was tested using a 2-sided Mann-Whitney-U test of the count (if at least 2 taxa had non-zero counts), and results were deemed significant at a  $p$ -value threshold of  $p = 0.01$ .

A graph representation of the OrthoFinder clustering (at inflation value = 1.5) was generated using the `generate_network.py` script distributed with KinFin. The nodes in the graph were positioned using the ForceAtlas2 layout algorithm implemented in Gephi.

### 2.2.8 Phylogenomics

The whole-genome OrthoFinder clustering at inflation value 1.5 was mined for potential single-copy orthologues for phylogenetic analysis. Transcriptome data were retrieved for additional tardigrades (2 species), a priapulid (1), kinorhynch (2), and onychophoran (3) (Supplementary Table S5). Assembled transcripts for *Echiniscus testudo*, *Milnesium tardigradum*, *Pycnophyes kielensis*, and *Halicryptus spinulosus* were downloaded from the NCBI Transcriptome Shotgun Assembly (TSA) Database. EST sequences of *Euperipatoides kanangrensis*, *Peripatopsis sedgwicki*, and *Echinoderes horni* were downloaded from NCBI Trace Archive and assembled using CAP3 (Huang, 1999). Raw mRNA-Seq reads for *Peripatopsis capensis* were downloaded from NCBI SRA, trimmed using skewer (Jiang *et al.*, 2014), and assembled with Trinity (Grabherr *et al.*, 2011). Protein sequences were predicted from all transcriptome data using TransDecoder (Brian

and Papanicolaou, 2017), retaining a single open reading frame per transcript. Predicted proteins from these transcriptomes were used along with the genome-derived proteomes in a second OrthoFinder clustering analysis.

We identified putatively orthologous genes in the OrthoFinder clusters for the genome and the genome-plus-transcriptome datasets. For both datasets, the same pipeline was followed. Clusters with 1-to-1 orthology were retained. For clusters with median per-species membership equal to 1 and mean less than 2.5, a phylogenetic tree was inferred with RAxML (using the LG+G model)(Stamatakis, 2014). Each tree was visually inspected to identify the largest possible monophyletic clan, and in-paralogues and spuriously included sequences were removed. Individual alignments of each locus were filtered using trimal (Capella-Gutierrez *et al.*, 2009) and then concatenated into a supermatrix using fastconcat (Kuck and Longo, 2014). The supermatrices were analyzed with RAxML with 100 ML bootstraps and PhyloBayes (Lartillot and Philippe, 2004) (see Supplementary Table S5 for specific commands). Trees were summarized in FigTree.

### 2.2.9 Databasing, software usage and data manipulation

A dedicated Ensembl genome browser (version 85) (Aken *et al.*, 2017) using the EasyImport pipeline (Challis *et al.*, 2017) was constructed on <http://www.tardigrades.org>, and the *H. exemplaris* genome and annotations described in this thesis and the new gene predictions for *R. varieornatus* were imported. We used open source software tools where available, as detailed in Supplementary Table S5. Custom scripts developed for the project are uploaded to [https://github.com/abs-yy/Hypsibius\\_exemplaris\\_manuscript](https://github.com/abs-yy/Hypsibius_exemplaris_manuscript). We used G-language Genome Analysis Environment (Arakawa *et al.*, 2003; Arakawa and Tomita, 2006) for sequence manipulation.

## 2.3 Results

### 2.3.1 The genome of *H. exemplaris*

The genome size of *H. exemplaris* has been independently estimated by densitometry to be approximately 100 Mb (Gabriel *et al.*, 2007; Koutsovoulos *et al.*, 2016), but the spans of existing assemblies exceed this, because of contamination with bacterial reads and uncollapsed heterozygosity of approximately 30% to 60% of the span estimated from k-mer distributions. We generated new sequencing data (Supplementary Table S6) for *H. exemplaris*. Tardigrades, originally purchased in mixed cultures from Sciento, were cultured with a single algal food source. Illumina short reads were generated from a single, cleaned tardigrade (Arakawa *et al.*, 2016), and Pacific Biosciences (PacBio) long single-molecule reads from DNA from a bulk, cleaned tardigrade population (approximately 900,000 animals). We employed an assembly strategy that eliminated

evident bacterial contamination (Kumar *et al.*, 2013) and eliminated residual heterozygosity. Our initial *Platanus* (Kajitani *et al.*, 2014) genome assembly had a span of 99.3 Mb in 1,533 contigs, with an N50 length of 250 kb. Further scaffolding and gap filling (Boetzer and Pirovano, 2014) with PacBio reads and a Falcon (Challis *et al.*, 2017) assembly of the PacBio reads produced a 104 Mb assembly in only 1,421 scaffolds and an N50 length of 342 kb and N90 count of 343 (Table 2.1). In comparison with previous assemblies, this assembly has improved contiguity and improved coverage of complete core eukaryotic genes (Parra *et al.*, 2007; Simao *et al.*, 2015). Blobplot analysis revealed low levels of contamination (Supplementary Figure S1). Read coverage was relatively uniform throughout the genome (Supplementary Figure S2, Supplementary Table S7), with only a few short regions, likely repeats, with high coverage. We identified 29.6 Mb (28.5%) of the *H. exemplaris* genome as being repetitive (Supplementary Table S8). Simple repeats covered 5.2% of the genome, with a longest repeat unit of 8,943 bp. Seven of the 8 longest repeats were of the same repeat unit (GATGGGTTTT)<sub>n</sub>, were found exclusively at 9 scaffold ends, and may correspond to telomeric sequence (Supplementary Table S9). The other long repeat was a simple repeat of (CAGA)<sub>n</sub> and its complementary sequence (GTCT)<sub>n</sub>, and spanned 3.2 Mb (3% of the genome, longest unit 5,208 bp). We identified eighty-one 5.8S rRNA, two 18S rRNA, and three 28S rRNA loci with RNAmmer (Lagesen *et al.*, 2007). Scaffold0021 contains both 18S and 28S loci, and it is likely that multiple copies of the ribosomal RNA repeat locus have been collapsed in this scaffold, as it has very high read coverage (approximately 5,400-fold, compared to approximately 113-fold overall, suggesting approximately 48 copies). tRNAs for each amino acid were found (Supplementary Figure S3) (Lowe and Eddy, 1997). Analysis of microRNA sequencing (miRNA-Seq) data with miRDeep (Friedlander *et al.*, 2012) predicted 507 mature miRNA loci (Supplementary Data S6), of which 185 showed similarity with sequences in miRbase (Kozomara and Griffiths-Jones, 2014).

**Table 2.1: Metrics of *H. exemplaris* genome assemblies.**

Data source	This Work	Edinburgh	North Carolina
Sequencing technologies	Illumina and PacBio	Illumina	Illumina and PacBio
Genome version	nHd.3.0	nHd.2.3	tg
Scaffold number	1,421	13,202	161,75
Total scaffold length (bp)	104,155,103	134,961,902	212,302,995
Average scaffold length (bp)	73,297	10,222	13,125
Longest Scaffold Length (bp)	2,115,976	594,143	1,208,507*
Shortest Scaffold Length (bp)	1,000	500	2,002
N50 (bp) (no. scaffs in N50)	342,180 (#85)	50,531 (#701)	17,496 (#3,422)
N90 (bp) (no. scaffs in N90)	65,573 (#343)	6,194 (#3,280)	6,637 (#11,175)
CEGMA genes found (partial)	237 (240)	220 (241)	221 (235)
CEGMA gene duplication ratio	1.17 (1.23)	1.35 (1.56)	3.26 (3.53)

\* The longest scaffolds in the tg assembly are derived from bacterial contaminants.

We generated RNA-Seq data from active and anhydrobiotic (“tun” stage) tardigrades and developmental stages of *H. exemplaris* (Supplementary Table S1). Gene finding using BRAKER (Hoff *et al.*, 2016) predicted 19,901 genes, with 914 isoforms (version nHd3.0). This set of gene models had higher completeness and lower duplication BUSCO and CEGMA scores compared to those predicted with MAKER (Holt and Yandell, 2011), which uses RNA-Seq and protein evidence (Predicted proteome based BRAKER: 90.7% MAKER: 77.9%, genome based BRAKER: 86.3%, Metazoan lineage used). Minor manual editing of this gene set to break approximately 40 fused genes generated version nHd3.1. Mapping of RNA-Seq data to the predicted coding transcriptome showed an average mapping proportion of approximately 50% to 70%, but the mapping proportion was over 95% against the genome (Supplementary Table S10 and S11). A similar mapping pattern for RNA-Seq data to predicted transcriptome was also observed for *R. varieornatus*. Over 70% of the *H. exemplaris* transcripts assembled with Trinity (Grabherr *et al.*, 2011) mapped to the predicted transcriptome, and a larger proportion to the genome (Supplementary Table S12). RNA-Seq reads that are not represented in the predicted coding transcriptome likely derived from UTR regions, unspliced introns, or promiscuous transcription. We inferred functional and similarity annotations for approximately 50% of the predicted proteome (Table 2.2). 60% of the genes (12,338) had a TPM expression of over 1.

The *H. exemplaris* nHd.3.0 and nHd.3.1 genome assembly is available on a dedicated Ensembl (Aken *et al.*, 2017) server, <http://ensembl.tardigrades.org>, where it can be compared with previous assemblies of *H. exemplaris* and with the *R. varieornatus* assembly. The Ensembl database interface includes an application-programming interface (API) for scripted querying (Yates *et al.*, 2015). All data files (including supplementary data files and other analyses) are available from <http://download.tardigrades.org>, and a dedicated Basic Local Alignment Search Tool (BLAST) server is available at <http://blast.tardigrades.org>. All raw data files (excluding the RNA-Seq data for gamma ray irradiated samples) have been deposited in International Nucleotide Sequence Database Collaboration (INSDC) databases (National Center for Biotechnology Information [NCBI] and Sequence Read Archive [SRA] (Supplementary Table S1 and S2, S6), and the assembly (nHd3.1) has been submitted to NCBI under the accession ID MTYJ00000000.

**Table 2.2: Comparison of the genomes of *H. exemplaris* and *R. varieornatus*.**

Assembly	<i>H. exemplaris</i> 3.0		<i>R. varieornatus</i> 1.1		Difference	
GENOME	Mb	%	bp	%	Mb	%
<b>Total span</b>	104.16	-	55.83	-	48.33	
<b>Genic</b>	59.03	56.67%	31.94	57.21%	27.09	56.06%
exon span	25.25	24.24%	19.56	35.03%	5.69	11.78%
intron span	33.78	32.43%	12.38	22.17%	21.40	44.28%
<b>Intergenic</b>	45.13	43.33%	23.89	42.79%	21.23	43.94%
repeat	27.11	26.03%	10.11	18.12%	17.00	35.17%
GENES	# families	# genes	# families	# genes	#families	#genes
Number of genes	11,705	19,901	9,029	13,917	2,676	5,984
Number of proteins (including isoforms)		20,815		14,538		6,277
Species-specific singletons	4,364	4,364	1,995	1,995	2,369	2,369
Species-specific gene families	45	258	20	123	25	135
Shared gene families	7,296	15,279	7,014	11,799	258	3,480
Uniquely retained ancestral genes *	471	999	189	311	282	688
Genes with BLAST matches to SwissProt		8,337		6,978		
Genes with BLAST matches to TrEMBL		10,202		8,359		
Genes with InterPro domain matches		11,227		8,633		
Genes with Gene Ontology terms		7,804		6,030		
All genes	mean	median	mean	median	ratio of means	ratio of medians
Gene length (bp)	2,966	2,131	2,295	1,641	1.29	1.30
Exon span (bp)	1,269	978	1,405	1,074	0.90	0.91
Exon count (#)	5.94	4	6.02	4	0.99	1.00
Intron span (bp)	1,697	1,109	889	520	1.91	2.13
Intron count (#)	4.94	3	5.02	3	0.98	1.00
Single-copy Orthologues **	mean	median	mean	median	ratio of means	ratio of medians
Gene length (bp)	3,716	2,776	2,579	1,929	1.44	1.44
Exon span (bp)	1,615	1,278	1,581	1,253	1.02	1.02
Exon count (#)	7.64	6	6.96	6	1.10	1.00
Intron span (bp)	2,101	1,475	998	635	2.11	2.32
Intron count (#)	3,716	2,776	2,579	1,929	1.44	1.44

\* Uniquely retained ancestral genes include genes shared by only one tardigrade and at least one non-tardigrade taxon.

\*\* Single-copy Orthologues: orthologues with CDS lengths differing by more than 20% were not considered.

### 2.3.2 Comparisons with *R. varieornatus*

We compared this high-quality assembly of *H. exemplaris* to that of *R. varieornatus* (Hashimoto *et al.*, 2016). In initial comparisons, we noted that *R. varieornatus* had many single-exon loci that had no *H. exemplaris* (or other) homologues. Reasoning that this might be a technical artifact, we updated gene models for *R. varieornatus* using BRAKER (Hoff *et al.*, 2016) with additional comprehensive RNA-Seq of developmental stages (Supplementary Table S2). The new prediction included 13,917 protein-coding genes (612 isoforms). This lower gene count compared to the original (19,521 genes) was largely due to a reduction in single-exon genes with no transcript support (from 5,626 in version 1 to 1,777 in the current annotation). Most (12,752, 90%) of the BRAKER-predicted genes were also found in the original set. Approximately 80% of the genes had TPM expression of over 1. In both species, some predicted genes may derive from transposons, as 2,474 *H. exemplaris* and 626 *R. varieornatus* proteins matched Dfam domains (Hubley *et al.*, 2016). While several of these putatively transposon-derived predictions have a Swiss-Prot (UniProt Consortium, 2015) homologue (*H. exemplaris*: 915, 37%; *R. varieornatus*: 274, 44%), most had very low expression levels.

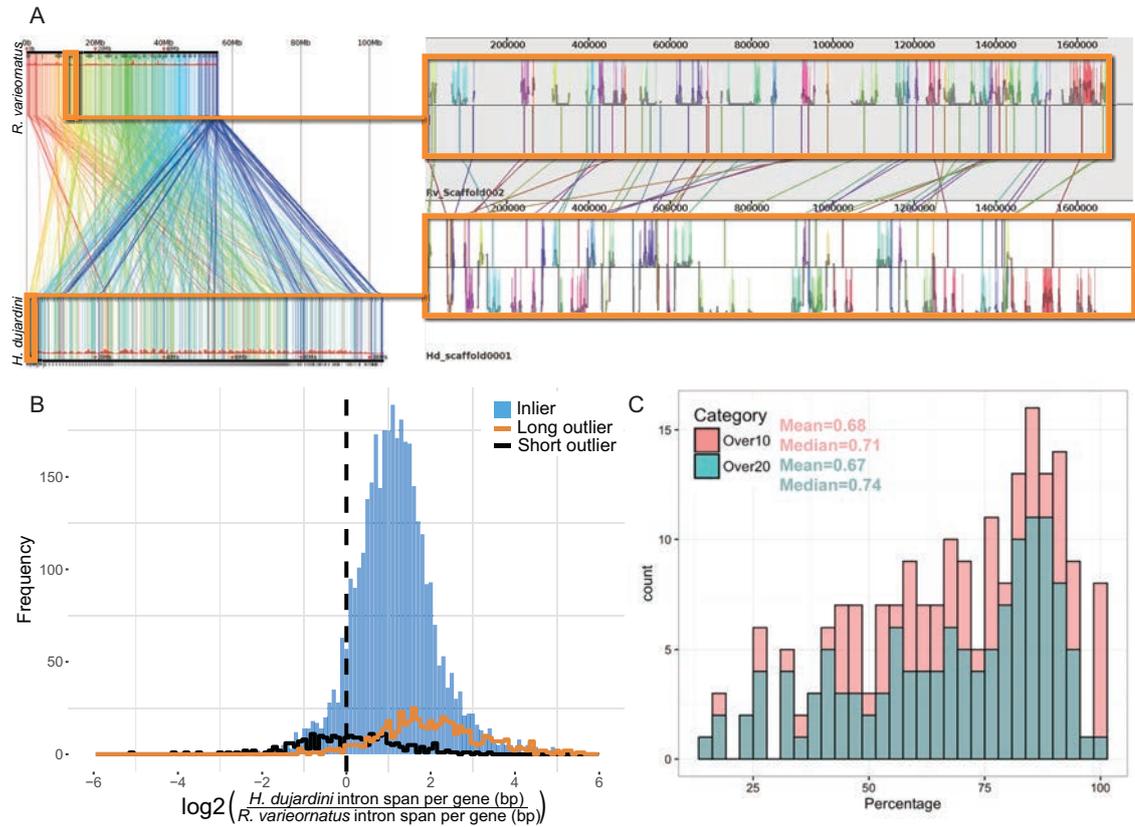
One striking difference between the 2 species was in genome size, as represented by assembly span: the *R. varieornatus* assembly had a span of 55 Mb, half that of *H. exemplaris* (Table 2.2). This difference could have arisen through whole genome duplication, segmental duplication, or more piecemeal processes of genome expansion or contraction. *H. exemplaris* had 5,984 more predicted genes than *R. varieornatus*. These spanned approximately 23 Mb and accounted for about half of the additional span. There was no difference in number of exons per gene between orthologues or in the whole predicted gene set. However, comparing orthologues, the intron span per gene in *H. exemplaris* was on average twice that in *R. varieornatus* (Figure 2.1b), and gene length (measured as start codon to stop codon in coding exons) was approximately 1.3-fold greater in *H. exemplaris* (Table 2.2, Supplementary Figure S4). There was more intergenic noncoding DNA in *H. exemplaris*, largely explained by an increase in the repeat content (28.6 Mb in *H. exemplaris* versus 11.1 Mb in *R. varieornatus*).

Whole genome alignments of *R. varieornatus* and *H. exemplaris* using Murasaki (Popendorf *et al.*, 2010) revealed a low level of synteny but evidence for conserved linkage at the genome scale, with little conservation of gene order beyond a few loci. For example, comparison of *R. varieornatus* Scaffold002 of with *H. exemplaris* scaffold0001 showed linkage, with many orthologous (genome-wide bidirectional best BLAST hit) loci across approximately 1.7 Mb of the *H. exemplaris* genome (Figure 2.1a). A high proportion of orthologues of genes located on the same scaffold in *H. exemplaris* were also in one scaffold in *R. varieornatus*, implying that intrachromosomal rearrangement may be the reason for the low level of synteny (Figure 2.1c).

We defined protein families in the *H. exemplaris* and new *R. varieornatus* predicted proteomes, along

with a selection of other ecdysozoan and other animal predicted proteomes (Supplementary Table S4), using OrthoFinder (Emms and Kelly, 2015), including predicted proteomes from fully sequenced genomes or predicted proteomes from the fully sequenced genomes and (likely partial) transcriptomes in two independent analyses. Using these protein families, we identified orthologues for phylogenetic analysis and explored patterns of gene family expansion and contraction, using KinFin (Laetsch and Blaxter, 2017). We identified 144,610 protein families in the analysis of 29 fully sequenced genome species. Of these families, 87.9% were species specific (with singletons accounting for 11.6% of amino acid span, and multi-protein clusters accounting for 1.2% of span). While only 12.1% of clusters contained members from  $\geq 2$  predicted proteomes, they accounted for the majority of the amino acid span (87.2%). *H. exemplaris* had more species-specific genes than *R. varieornatus* and had more duplicate genes in gene families shared with *R. varieornatus* (Table 2.2). *H. exemplaris* also had more genes shared with nontardigrade outgroups, suggesting loss in *R. varieornatus*. Many families had more members in tardigrades compared to other taxa, and 3 had fewer members (115 had uncorrected Mann-Whitney U-test probabilities  $< 0.01$ , but none had differential presence after Bonferroni correction). In 9 of the families with tardigrade overrepresentation, tardigrades had more than four times as many members as the average of the other species (Supplementary Data S7).

There were 1,486 clusters composed solely of proteins predicted from the two tardigrade genomes. Of those, 365 (24.56%) had a congruent domain architecture in both species, including 53 peptidase clusters, 27 kinase clusters, and 29 clusters associated with signaling function, including 18 G-protein coupled receptors (see Supplementary Data S8). While these annotations are commonly found in clade-specific families, suggesting that innovation in these classes of function is a general feature in metazoan evolution, of particular interest was innovation in the Wnt signaling pathway. Tardigrade-unique clusters included Wnt, Frizzled, and chibby proteins. Of relevance to cryptobiosis, 21 clusters with domain annotation relevant to genome repair and maintenance were synapomorphic for Tardigrada, including molecular chaperones (2), histone/chromatin maintenance proteins (11), genome repair systems (4), nucleases (2), and chromosome cohesion components (2) (see below).



**Figure 2.1: The genomes of *H. exemplaris* and *R. varieornatus*.**

**(A) Linkage conservation but limited synteny between *H. exemplaris* and *R. varieornatus*.** Whole genome alignment was performed with Murasaki (Popendorf *et al.*, 2010). The left panel shows the whole genome alignment. Similar regions are linked by a line colored following a spectrum based on the start position in *R. varieornatus*. To the right is a realignment of the initial segment of *H. exemplaris* scaffold0001 (lower), showing matches to several portions of *R. varieornatus* Scaffold0002 (above), illustrating the several inversions that must have taken place. The histograms show pairwise nucleotide sequence identity between these 2 segments. **(B) Increased intron span in *H. exemplaris*.** *H. exemplaris* genes are longer because of expanded introns. Frequency histogram of  $\log_2$  ratio of intron span per gene in 4,728 *H. exemplaris* genes compared to their orthologues in *R. varieornatus*. Outliers are defined as genes in *H. exemplaris* whose coding sequences (CDSs) are 20% longer (long outliers; orange; 576 genes) or 20% shorter (short outliers; black; 294 genes) than their orthologues in *R. varieornatus*. **(C) Gene neighborhood conservation between *H. exemplaris* and *R. varieornatus*.** To test conservation of gene neighborhoods, we asked whether genes found together in *H. exemplaris* were also found close together in *R. varieornatus*. Taking the set of genes on each long *H. exemplaris* scaffold, we identified the locations of the reciprocal best Basic Local Alignment Search Tool (BLAST) hit orthologues in *R. varieornatus* and counted the maximal proportion mapping to 1 *R. varieornatus* scaffold. *H. exemplaris* scaffolds were binned and counted by this proportion. As short scaffolds, with fewer genes, might bias this analysis, we performed analyses independently on scaffolds with >10 genes and scaffolds with >20 genes. *H. dujardini* has been redescribed as *H. exemplaris*. Replicated from Yoshida *et al.* (2017).

### 2.3.3 Horizontal gene transfer in tardigrades

HGT is an interesting but contested phenomenon in animals. Many newly sequenced genomes have been reported to have relatively high levels of HGT, and genomes subject to intense curation efforts tend to have lower HGT estimates. We performed *ab initio* gene finding on the genomes of the model species *C. elegans* and *D. melanogaster* with Augustus (Keller *et al.*, 2011) and used the HGT index approach (Boschetti *et al.*, 2012), which simply classifies loci based on the ratio of their best BLAST scores to ingroup and potential donor taxon databases, to identify candidates. Compared with their mature annotations, we found elevated proportions of putative HGTs in both species (*C. elegans*: 2.09% of all genes; *D. melanogaster*: 4.67%). We observed similarly elevated rates of putative HGT loci, as assessed by the HGT index, in gene sets generated by *ab initio* gene finding in additional arthropod and nematode genomes compared to their mature annotation (Figure 2.2a, Supplementary Table S3). Thus, the numbers of HGT events found in the genomes of *H. exemplaris* and *R. varieornatus* are likely to have been overestimated in these initial, uncured gene predictions, even after sequence contamination has been removed, as seen in the *H. exemplaris* assembly of Boothby *et al.* (Boothby and Goldstein, 2016).

Using the HGT index approach, we identified 463 genes as potential HGT candidates in *H. exemplaris* (Supplementary Data S2). Using Diamond BLASTX (Buchfink *et al.*, 2015) instead of standard BLASTX (Altschul *et al.*, 1997; Camacho *et al.*, 2009) made only a minor difference in the number of potential HGT events predicted (446 genes). We sifted the initial 463 *H. exemplaris* candidates through a series of biological filters. A true HGT locus will show affinity with its source taxon when analyzed phylogenetically (a more sensitive test than simple BLAST score ratio). Four-fifths of these loci (357) were confirmed as HGT events by RAxML (Stamatakis, 2014) analysis of aligned sequences (Figure 2.2b). For 13 candidates, there were not enough homologues found in public databases to estimate phylogenies. HGT genes are expected to be incorporated into the host genome and to persist through evolutionary time. Only 164 of the RAxML-confirmed *H. exemplaris* candidates had homologues in *R. varieornatus*, indicating phyletic perdurance (Supplementary Data S7). HGT loci will acquire gene structure and expression characteristics of their host metazoan genome. We identified expression at greater than 1 transcript per million (TPM) in any library for 338 HGT candidates. While metazoan genes usually contain spliceosomal introns, and 367 of the candidate HGT gene models included introns, we regard this a lower-quality validation criterion, as gene-finding algorithms are programmed to identify introns. Therefore, our highest-credibility current estimate for HGT into the genome of *H. exemplaris* is 133 genes (0.7% of all genes), with a less-credible set, showing conservation, expression, and/or phylogenetic validation of 357 (1.8%) and an upper bound of 463 (2.3%). This is congruent with estimates of 1.6% HGT candidates (out of 13,917 genes) for *R. varieornatus* (Hashimoto *et al.*, 2016).

The putative HGT loci tended to be clustered in the tardigrade genomes, with many gene neighbors of

HGT loci also predicted to be HGTs (Supplementary Figure S5). We found 58 clusters of HGT loci in *H. exemplaris* and 14 in *R. varieornatus* (Supplementary Data S9). The largest clusters included up to 6 genes from the same gene family and may have arisen through tandem duplication. These tandem duplication clusters included intradiol ring-cleavage dioxygenases, uridine diphosphate (UDP) glycosyltransferases, and alpha/beta fold hydrolases. Several clusters of UDP glycosyltransferases with signatures of HGT from plants were identified in the *H. exemplaris* genome, 1 of which included 6 UDP glycosyltransferases within 12 genes (loci between the genes bHd03905 and bHd03916). *H. exemplaris* had 40 UDP glycosyltransferase genes, 29 of which were classified as glucuronosyltransferase (UGT, K00699) by Kyoto Encyclopedia of Genes and Genomes (KEGG) ORTHOLOG mapping with the KEGG Automatic Annotation Server (KAAS) (Moriya *et al.*, 2007), and of these 27 were HGT candidates. While UGT can function in a number of pathways, we found that the whole ascorbate synthesis pathway, in which UGT metabolizes UDP-D-glucuronate to D-Glucuronate, has been acquired by HGT in *H. exemplaris*. *R. varieornatus* has only acquired L-gulonolactone oxidase (Supplementary Figure S6). Gluconolactonase and L-gulonolactone oxidase were consistently expressed at low levels (approximately 10 – 30 TPM), but L-ascorbate degradation enzyme L-ascorbate oxidase was not expressed (TPM < 1).

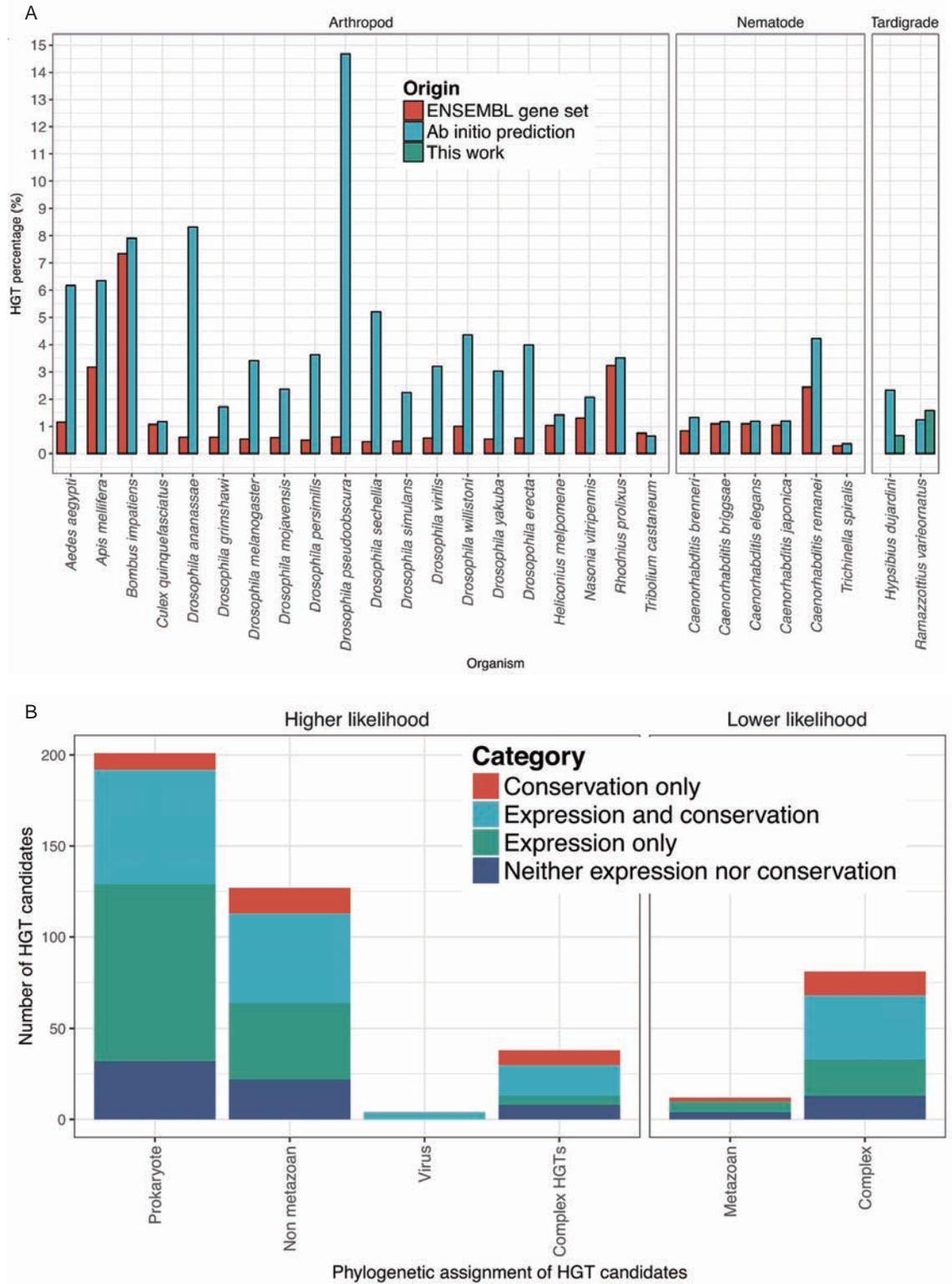


Figure 2.2: Horizontal gene transfer in *H. exemplaris*.

**(A) Horizontal gene transfer ratios in various metazoa.** For a set of assembled arthropod and nematode genomes, genes were repredicted *ab initio* with Augustus. Putative horizontal gene transfer (HGT) loci were identified using the HGT index for the longest transcript for each gene from the new and the Ensembl reference gene sets. In most species, the *ab initio* gene sets had elevated numbers of potential HGT loci compared to their Ensembl representations. **(B) Classification of HGT candidates in *H. exemplaris*.** Classification of the initial HGT candidates identified in *H. exemplaris* by their phylogenetic annotation (prokaryotic, nonmetazoan eukaryotic, viral, complex HGT, and likely non-HGT metazoan and complex), their support in RNA-Seq expression data, and the presence of a homologue in *R. varieornatus*. *H. dujardini* has been redescribed as *H. exemplaris*. Replicated from Yoshida *et al.* (2017).

### 2.3.4 The genomics of anhydrobiosis in tardigrades

We explored the predicted *H. exemplaris* proteome and the reannotated *R. varieornatus* proteome for loci implicated in anhydrobiosis. In the new *R. varieornatus* proteome, we found 16 CAHS loci and 13 SAHS loci and 1 copy each of MAHS, RvLEAM, and Dsup. In *H. exemplaris*, we identified 12 CAHS loci, 10 SAHS loci, and single members of the RvLEAM and MAHS families (Supplementary Table S13). Direct interrogation of the *H. exemplaris* genome with *R. varieornatus* loci identified an additional possible CAHS-like locus and an additional SAHS-like locus. We found no evidence for a *H. exemplaris* homologue of Dsup. Phylogenetic analyses revealed a unique duplication of CAHS3 in *R. varieornatus*. No SAHS2 orthologue was found in *H. exemplaris* (Supplementary Figure S7), and most of the *H. exemplaris* SAHS loci belonged to a species-specific expansion that was orthologous to a single *R. varieornatus* SAHS locus, RvSAHS13. SAHS1-like genes in *H. exemplaris* and SAHS1- and SAHS2-like genes in *R. varieornatus* were locally duplicated, forming SAHS clusters on single scaffolds.

*R. varieornatus* was reported to have undergone extensive gene loss in the stress-responsive transducer of mechanistic target of rapamycin (mTOR) pathway and in the peroxisome pathway, which generates H<sub>2</sub>O<sub>2</sub> during the beta-oxidation of fatty lipids. *H. exemplaris* was similarly compromised (Figure 2.3a). We identified additional gene losses in the peroxisome pathway in *H. exemplaris*, as peroxisome proteins PEX5, PEX10, and PEX12, while present in *R. varieornatus*, were not found in *H. exemplaris* (TBLASTN search against genome with an E-value threshold of 1E-3).

To identify gene functions associated with anhydrobiosis, we explored differential gene expression in fully hydrated and post desiccation samples from both species. We compared single individual RNA-Seq of *H. exemplaris* undergoing anhydrobiosis (Arakawa *et al.*, 2016) with new data for *R. varieornatus* induced to enter anhydrobiosis in 2 ways: slow desiccation (approximately 24 h) and fast desiccation (approximately 30 min). Successful anhydrobiosis was assumed when >90% of the samples prepared in the same chamber recovered after rehydration. Many more genes were differentially up-regulated by entry into anhydrobiosis in *H. exemplaris* (1,422 genes, 7.1%) than in *R. varieornatus* (fast desiccation: 64 genes, 0.5%; slow desiccation: 307 genes, 2.2%) (Supplementary Data S9). The fold change distribution of the whole transcriptome of *H. exemplaris* (mean 8.33, median 0.91 ± 69.90 SD) was significantly broader than

those of both fast (0.67,  $0.48 \pm 2.25$ ) and slow (0.77,  $0.65 \pm 0.79$ ) desiccation *R. varieornatus* (U-test,  $p$ -value < 0.001) (Figure 2.3b).

For the loci differentially expressed in anhydrobiosis (Supplementary Data S10), we investigated their membership of gene families with elevated numbers in tardigrades and functional annotations associated with anhydrobiosis. Proteins with functions related to protection from oxidants, such as SOD and peroxiredoxin, were found to have been extensively duplicated in tardigrades. In addition, the mitochondrial chaperone (BCS1), osmotic stress-related transcription factor NFAT5, and apoptosis related-gene poly (ADP-ribose) polymerase (PARP) families were expanded in tardigrades. Chaperones were extensively expanded in *H. exemplaris* (HSP70, DnaK, and DnaJ subfamily C-5, C-13, and B-12), and the DnaJ subfamily B3, B-8 was expanded in *R. varieornatus*. In *H. exemplaris*, we found 5 copies of DNA repair endonuclease XPF, which functions in the nucleotide-excision repair pathway, and, in *R. varieornatus*, 4 copies of the double-stranded break repair protein MRE11 (as reported previously in Hashimoto *et al.* (2016)) and additional copies of DNA ligase 4, from the nonhomologous end-joining pathway. In both *R. varieornatus* (Hashimoto *et al.*, 2016) and *H. exemplaris*, some of the genes with anhydrobiosis-related function appear to have been acquired through HGT. All copies of catalase were high-confidence HGTs (Supplementary Data S1), and 1 copy was differentially expressed during *H. exemplaris* anhydrobiosis (expression rises from 0 TPM to 27.5 TPM during slow dehydration in *H. exemplaris*). *R. varieornatus* had 11 trehalase loci (9 trehalases and 2 acid trehalase-like proteins). While *H. exemplaris* did not have an orthologue of trehalose-6-phosphatase synthase (TPS), a gene required for trehalose synthesis, *R. varieornatus* had a HGT-derived TPS (Supplementary Figure S6). Previous studies in *M. tardigradum* have shown that trehalose does not accumulate during anhydrobiosis, and this is supported by the low expression of the *R. varieornatus* TPS gene (10 to 20 TPM in active and tun states). We note that the *R. varieornatus* TPS had the highest similarity to TPS from bacterial species in Bacterioidetes, including Chitinophaga, which was one of the contaminating organisms in the Boothby *et al.* assembly (de Rosa *et al.*, 1999; Delmont and Eren, 2016). The *R. varieornatus* locus contains spliceosomal introns that do not compromise the TPS protein sequence and is surrounded by metazoan-affinity loci. The ascorbate synthesis pathway appears to have been acquired through HGT in *H. exemplaris*, and a horizontally acquired L-gulonolactone oxidase was identified in *R. varieornatus* (Supplementary Figure S6).

Several protection-related genes were differentially expressed in anhydrobiotic *H. exemplaris*, including CAHS (8 loci of 12), SAHS (2 of 10), RvLEAM (1 of 1), and MAHS (1 of 1). Loci involved in reactive oxygen protection (5 SOD genes, 6 glutathione-S transferase genes, a catalase gene, and a LEA gene) were up-regulated under desiccation. Interestingly, 2 trehalase loci were up-regulated, even though we were unable to identify any TPS loci in *H. exemplaris*. We also identified differentially expressed transcription factors that may regulate anhydrobiotic responses. Two calcium-signaling factors, calmodulin (CaM) and

a cyclic nucleotide-gated channel 3 (CNG-3), were both up-regulated, which may drive cAMP synthesis through adenylate cyclase. Although *R. varieornatus* is capable of rapid anhydrobiosis induction, complete desiccation is unlikely to be as rapid in natural environments, and regulation of gene expression under slow desiccation might reflect a more likely scenario. Fitting this expectation, 5 CAHS loci and a single SAHS locus were up-regulated after slow desiccation, but none were differentially expressed following rapid desiccation. Most *R. varieornatus* CAHS and SAHS orthologues had high expression in the active state, several over 1,000 TPM. In contrast, *H. exemplaris* CAHS and SAHS orthologues had low resting expression (median 0.7 TPM) and were up-regulated (median 1,916.8 TPM) on anhydrobiosis induction. Aquaporins contribute to transportation of water molecules into cells and could be involved in anhydrobiosis (Cornette and Kikawada, 2011). Aquaporin-10 was highly expressed in *R. varieornatus* and differentially expressed in anhydrobiotic *H. exemplaris*. *M. tardigradum* has at least 10 aquaporin loci (Grohme *et al.*, 2013), *H. exemplaris* has 11, and *R. varieornatus* has 10. The contributions to anhydrobiosis of additional genes identified as up-regulated (including cytochrome P450, several solute carrier families, and apolipoproteins) are unknown.

Some genes differentially expressed in both *H. exemplaris* and *R. varieornatus* slow-desiccation anhydrobiosis were homologous (Supplementary Data S11). Of the 1,422 differentially expressed genes from *H. exemplaris*, 121 genes were members of 70 protein families that also contained 115 *R. varieornatus* differentially expressed genes. These included CAHS, SAHS, glutathione-S transferase, and SOD gene families, but in each case *H. exemplaris* had more differentially expressed copies than *R. varieornatus*. Other differentially expressed gene families were annotated as metalloproteinases, calcium-binding receptors, and G-protein coupled receptors, suggesting that these functions may participate in cellular signaling during induction of anhydrobiosis. Many more (887) gene families included members that were up-regulated by anhydrobiosis in *H. exemplaris* but unaffected by desiccation in *R. varieornatus*. These gene families included 1,879 *R. varieornatus* genes; some (154) were highly expressed in the active state (TPM > 100).

In addition to gene loss, we predicted that the tardigrades might have undergone expansion in gene families active in anhydrobiotic physiology. We identified 3 gene families-each containing members with significant differential expression during anhydrobiosis, that had elevated numbers of members in the tardigrades compared to the other taxa analyzed. *H. exemplaris* and *R. varieornatus* had more members of OG000684 (33 and 8, respectively) than any other (mode of 1 and mean of 1.46 copies in the other 28 species, with a maximum of 4 in the moth *Plutella xylostella*). Proteins in OG000684 were annotated with domains associated with ciliar function. OG0002660 contained 3 proteins from *H. exemplaris* and 3 proteins from *R. varieornatus* but a mean of 1.2 from other species. OG0002660 was annotated as fumarylacetoacetase, which acts in phenylalanine metabolism. Fumarylacetoacetase has been identified as a target of the SKN-1-induced stress responses in *C. elegans* (Oliveira *et al.*, 2009). OG0002103 was also overrepresented in the tardigrades

(3 in each species), while 23 of the other species had 1 copy. Interestingly, the extremophile nematode *Plectus murrayi* had 4 copies. OG0002103 was annotated as guanosine-5'-triphosphate (GTP) cyclohydrolase, involved in formic acid metabolism, including tetrahydrobioterin synthesis. Tetrahydrobioterin is a cofactor of aromatic amino acid hydroxylases, which metabolize phenylalanine. The association of these functions with anhydrobiosis merits investigation.

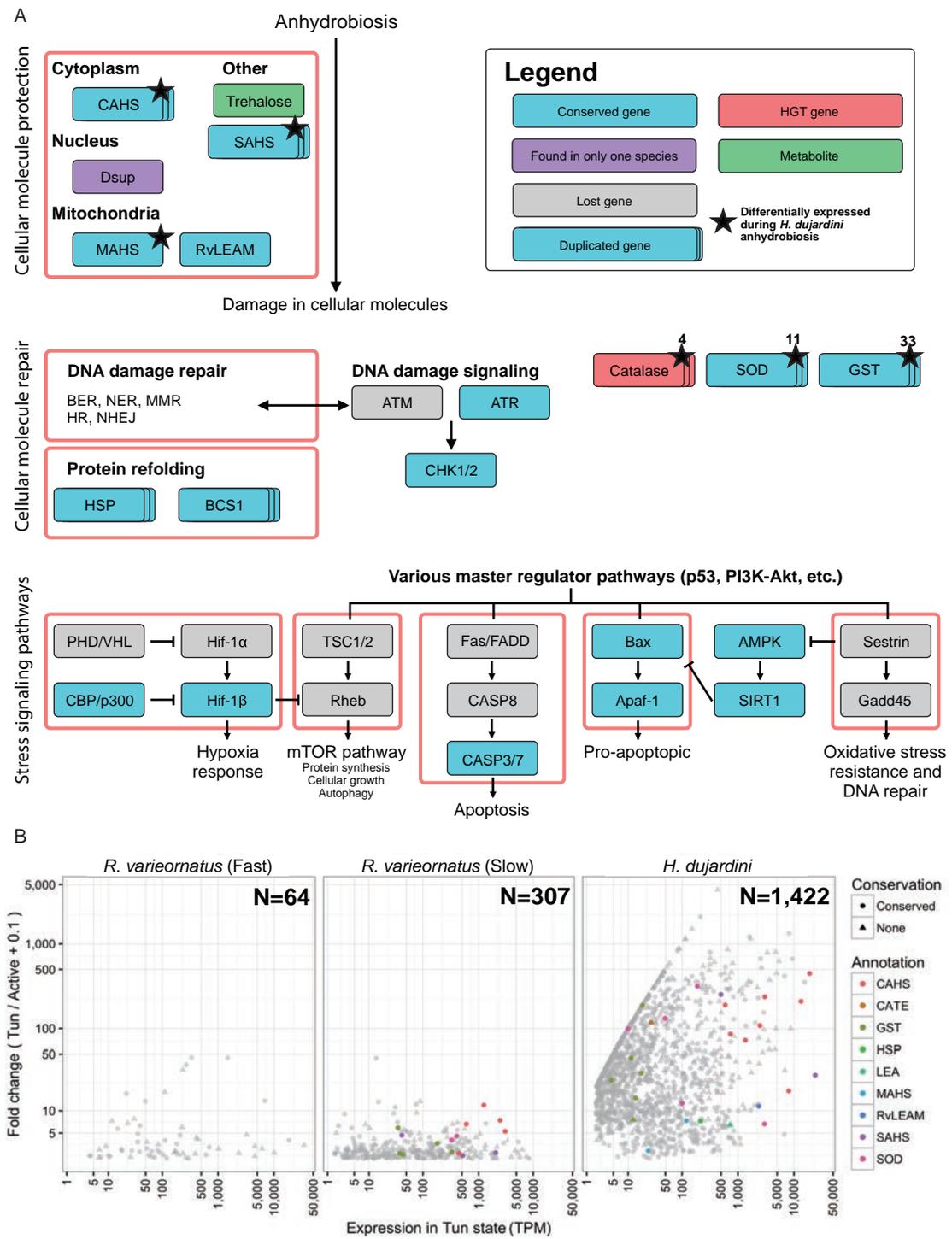


Figure 2.3: The genomics of anhydrobiosis in tardigrades.

**(A) Gene losses in Hypsibiidae.** Gene losses were detected by mapping to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways using KEGG Automatic Annotation Server (KAAS) and validated by Basic Local Alignment Search Tool (BLAST) TBLASTN search of KEGG orthologue gene amino acid sequences. Light blue and gray boxes indicate genes conserved and lost in both tardigrades, respectively. Furthermore, purple boxes represent genes retained in only 1 species, and red boxes represent genes that have been detected as horizontal gene transfer (HGT). The numbers on the top right of the boxes indicate copy numbers of multiple copy genes in *H. exemplaris*. Genes annotated as CASP3 and CDC25A have contradicting annotation with KAAS and Swiss-Prot; however, the KAAS annotation was used. **(B) Differential gene expression in tardigrades on entry to the anhydrobiotic state.** The transcript per million (TPM) expression for each sample was calculated using Kallisto, and the fold change between active and tun and the TPM expression in the tun state were plotted. Genes that likely contribute to anhydrobiosis were colored. Genes that had an orthologue in the other species are plotted as circles; other genes are plotted as triangles. *H. dujardini* has been redescribed as *H. exemplaris*. Replicated from Yoshida *et al.* (2017).

### 2.3.5 Phylogenetic relationships of Tardigrada

From the two analyses of protein families shared between *H. exemplaris*, *R. varieornatus*, taxa from other ecdysozoan phyla, and 2 lophotrochozoan outgroup taxa (one that included only taxa with whole genome data, and a second that also included taxa with transcriptome data), we selected putative orthologous protein families. These were screened to eliminate evident paralogous sequences, and alignments were concatenated into a supermatrix. The genomes-only supermatrix included 322 loci from 28 taxa spanning 67,256 aligned residues and had 12.5% missing data. The alignment was trimmed to remove low-quality regions. The genomes and transcriptomes supermatrix included 71 loci from 37 taxa spanning 68,211 aligned residues, had 27% missing data, and was not trimmed. Phylogenomic analyses were carried out in RAxML (using the General Time Reversible model with Gamma distribution of rates model, GTR+G) and PhyloBayes (using a GTR plus rate categories model, GTR-CAT+G). We also explored bipartition support from individual gene trees and RAxML and PhyloBayes analyses of 6-state Dayhoff recoded amino acid alignments using the GTR model (as GTR-CAT cannot be used on these recoded data; Supplementary Data S12).

The genomes-only phylogeny (Figure 2.4a) strongly supported Tardigrada as a sister to monophyletic Nematoda. Within Nematoda and Arthropoda, the relationships of species were congruent with previous analyses, and the earliest branching taxon in Ecdysozoa was Priapulida. RAxML bootstrap and PhyloBayes Bayes proportion support was high across the phylogeny, with only 2 internal nodes in Nematoda and Arthropoda receiving less-than-maximal support. Analysis of individual RAxML phylogenies derived from the 322 loci revealed a degradation of support deeper in the tree, with 53% of trees supporting a monophyletic Arthropoda, 56% supporting Tardigrada plus Nematoda, and 54% supporting the monophyly of Arthropoda plus Tardigrada plus Nematoda. The phylogeny derived from the genomes and transcriptomes dataset (Figure 2.4b) also recovered credibly resolved Nematoda and Arthropoda and, as expected, placed Nematomorpha as sister to Nematoda. Tardigrada was again recovered as sister to Nematoda plus Nematomorpha, with maximal support. Priapulida plus Kinorhyncha was found to arise basally in Ecdysozoa. Unexpectedly,

Onychophora, represented by 3 transcriptome datasets, was sister to an Arthropoda plus (Tardigrada, Nematomorpha, and Nematoda) clade, again with high support.

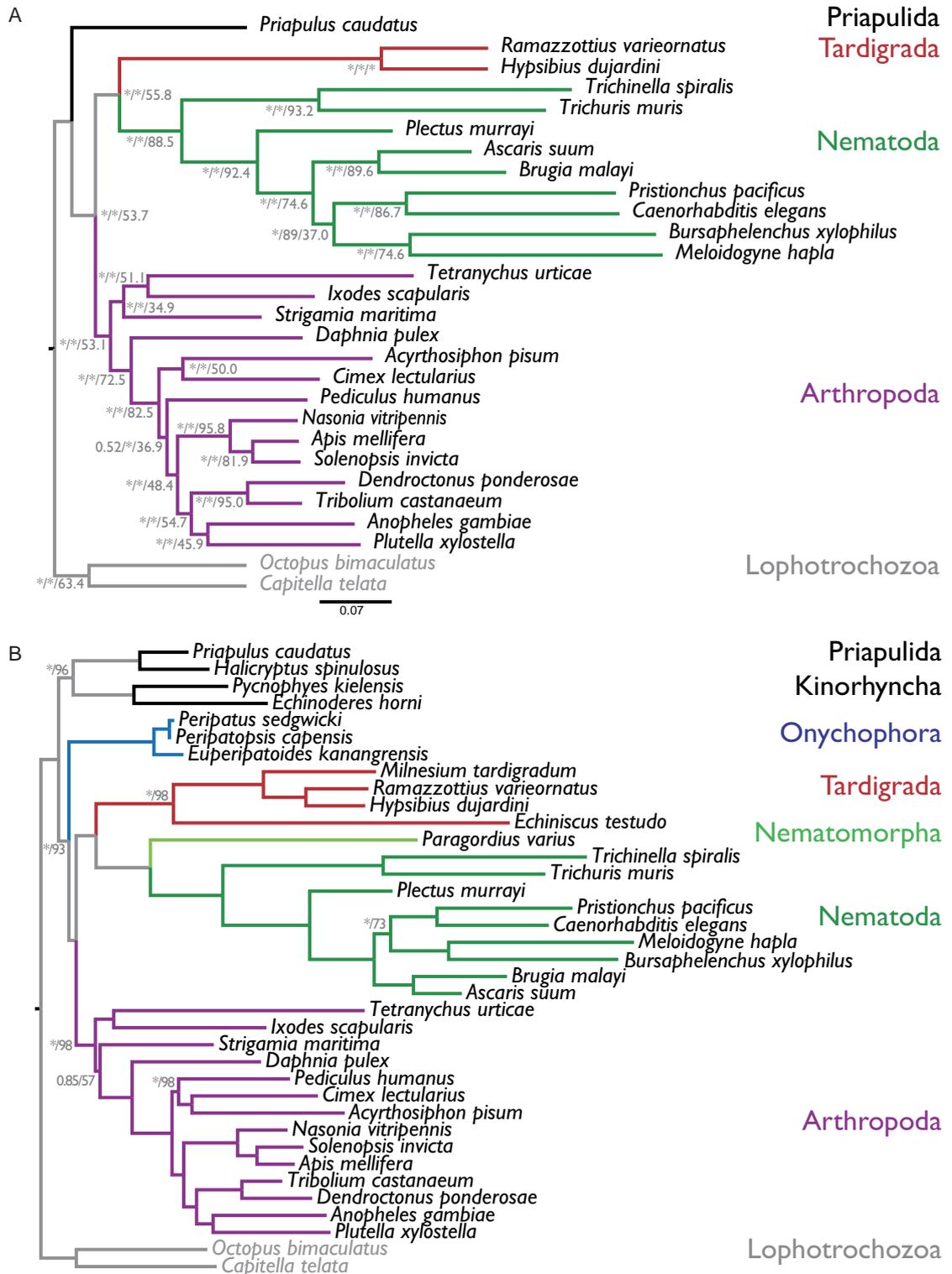


Figure 2.4: Phylogeny of Ecdysozoa.

**(A) Phylogeny of tardigrada using only genome information.** Phylogeny of 28 species from 5 phyla, based on 322 loci derived from whole genome sequences, and rooted with the lophotrochozoan outgroup. The labels on the nodes are Bayes proportions from PhyloBayes analysis / bootstrap proportions from Randomized Axelerated Maximum Likelihood (RAxML) maximum likelihood bootstraps / proportion of trees of individual loci supporting each bipartition. Note that different numbers of trees were assessed at each node, depending on the representation of the taxa at each locus. \* indicates maximal support (Bayes proportion of 1.0 or RAxML bootstrap of 1.0). **(B) Phylogeny of tardigrada using genome and transcriptome information.** Phylogeny of 36 species from 8 phyla, based on 71 loci derived using PhyloBayes from whole genome and transcriptome sequences, and rooted with the lophotrochozoan outgroup. All nodes had maximal support in Bayes proportions and RAxML bootstrap, except those labeled (Bayes proportion, \* = 1.0 / RAxML bootstrap). *H. dujardini* has been redescribed as *H. exemplaris*. Replicated from Yoshida *et al.* (2017).

### 2.3.6 Rare genomic changes and tardigrade relationships

We tested support for a Nematoda+Tardigrada clade in rare genomic changes (Rokas and Holland, 2000) in core developmental gene sets and protein family evolution. Rare genomic changes can be used as strong parsimony markers of phylogenetic relationships that are hard to resolve using model-based sequence analyses. An event shared by 2 taxa can be considered to support their relationship where the likelihood of the event is *a priori* expected to be vanishingly small.

HOX genes are involved in anterior-posterior patterning across the Metazoa, with a characteristic set of paralogous genes present in most animal genomes, organized as a tightly regulated cluster. The ancestral cluster is hypothesized to have included HOX1, HOX2, HOX3, HOX4, HOX5, and a HOX6-8 like locus and HOX9. The HOX6-8 and HOX9 types have undergone frequent, independent expansion and contraction during evolution, and HOX clustering has broken down in some species. HOX complements are generally conserved between related taxa, and gain and loss of HOX loci can be considered a rare genomic change. We surveyed HOX loci in tardigrades and relatives (Figure 2.5a). In the priapulid *Priapulius caudatus*, 9 HOX loci have been described (de Rosa *et al.*, 1999), but no HOX6-8/*AbdA*-like gene was identified. All arthropods surveyed (including representatives of the 4 classes) had a complement of HOX loci very similar to that of *D. melanogaster*, with at least 10 loci including HOX6-8 and HOX9. Some HOX loci in some species have undergone duplication, particularly HOX3/*zen*. In the mite *Tetranychus urticae* and the salmon louse *Lepeoptheirus salmonis*, we identified “missing” HOX genes in the genome. For Onychophora, the sister group to Arthropoda, HOX loci have only been identified through PCR screens (Grenier *et al.*, 1997; Janssen *et al.*, 2014), but they appear to have the same complement as Arthropoda.

In *H. exemplaris*, a reduced HOX gene complement (6 genes in 5 orthology groups) has been reported (Smith *et al.*, 2016), and we confirmed this reduction using our improved genome (Figure 2.5a). The same reduced complement was also found in the genome of *R. varieornatus* (Hashimoto *et al.*, 2016), and the greater contiguity of the *R. varieornatus* genome shows that 5 of the 6 HOX loci are on 1 large scaffold, distributed over 2.7 Mb, with 885 non-HOX genes separating them. The *H. exemplaris* loci were unlinked

in our assembly, except for the 2 HOX9/*AbdB*-like loci, and lack of gene level synteny precludes ordering of these scaffolds based on the *R. varieornatus* genome. The order of the HOX genes on the *R. varieornatus* scaffolds is not colinear with other, unfragmented clusters, as HOX6-8/*ftz* and the pair of HOX9/*AbdB* genes are inverted, and HOX4/*dfd* is present on a second scaffold (and not found between HOX3 and HOX6-8/*ftz* as would be expected).

The absences of HOX2/*pb*, HOX5/*scr*, and HOX6-8/*Ubx/AbdA* in both tardigrade species is reminiscent of the situation in Nematoda, in which these loci are also absent (Aboobaker and Blaxter, 2003a;b; 2010). HOX gene evolution in Nematoda has been dynamic. No Nematode HOX2 or HOX5 orthology group genes were identified, and only a few species had a single HOX6-8 orthologue. Duplication of the HOX9/*AbdB* locus was common, generating, for instance, the *egl-5*, *php-3*, and *nob-1* loci in *Caenorhabditis* species. The maximum number of HOX loci in a nematode was 7, deriving from 6 orthology groups. Loss of HOX3 happened twice (in *Syphacia muris* and in the common ancestor of *Tylenchomorpha* and *Rhabditomorpha*). The independent loss in *S. muris* was confirmed in 2 related pinworms, *Enterobius vermicularis* and *Syphacia oblevata*. The pattern of presence and absence of the *Antp*-like HOX6-8 locus is more complex, requiring 6 losses (in the basally arising enoplean *Enoplis brevis*, the chromadorean *Plectus sambesii*, the pinworm *S. muris*, the ancestor of *Tylenchomorpha*, the diplogasteromorph *Pristionchus pacificus*, and the ancestor of *Caenorhabditis*). We affirmed loss in the pinworms by screening the genomes of *E. vermicularis* and *S. oblevata* as above, and no HOX6-8/*Antp*-like locus was present in any of the over 20 genomes available for *Caenorhabditis*. A PCR survey for HOX loci and screening of a *de novo* assembled transcriptome from the nematomorph *Paragordius varius* identified 6 putative loci from 5 HOX groups. The presence of a putative HOX2/*pb*-like gene suggests that loss of HOX2 may be independent in Tardigrada and Nematoda.

Gene family birth can be used as another rare genomic marker. We analyzed the whole proteomes of ecdysozoan taxa for gene family births that supported either the Tardigrada+Nematoda model or the Tardigrada+Arthropoda (specifically, Panarthropoda) model. We mapped gene family presence and absence across the 2 contrasting phylogenies using KinFin (Laetsch and Blaxter, 2017) using different inflation parameters in the Markov Cluster Algorithm (MCL) step in OrthoFinder (Supplementary Data S4). Using the default inflation value of 1.5, the 2 tardigrades shared more gene families with Arthropoda than they did with Nematoda (Figure 2.5b). The numbers of gene family births synapomorphic for Arthropoda and Nematoda were identical under both phylogenies, as was expected (Table 2.3; Figure 2.5c; Supplementary Data S5). Many synapomorphic families had variable presence in the daughter taxa of the common ancestors of Arthropoda and Nematoda, likely because of stochastic gene loss or lack of prediction. However, especially in Nematoda, most synapomorphic families were present in a majority of species (Figure 2.5c).

At inflation value 1.5, we found 6 gene families present that had members in both tardigrades and all 14 arthropods under Panarthropoda, but no gene families were found in both tardigrades and all 9 ne-

matodes under the Tardigrada+Nematoda hypothesis (Supplementary Table S14). Allowing for stochastic absence, we inferred 154 families to be synapomorphic for Tardigrada+Arthropoda under the Panarthropoda hypothesis, and 99 for Tardigrada+Nematoda under the Tardigrada+Nematoda hypothesis (Figure 2.5d). More of the Tardigrada+Arthropoda synapomorphies had higher species representation than did the Tardigrada+Nematoda synapomorphies. This pattern was also observed in analyses using different inflation values and in analyses including the transcriptome from the nematomorph *P. varius*.

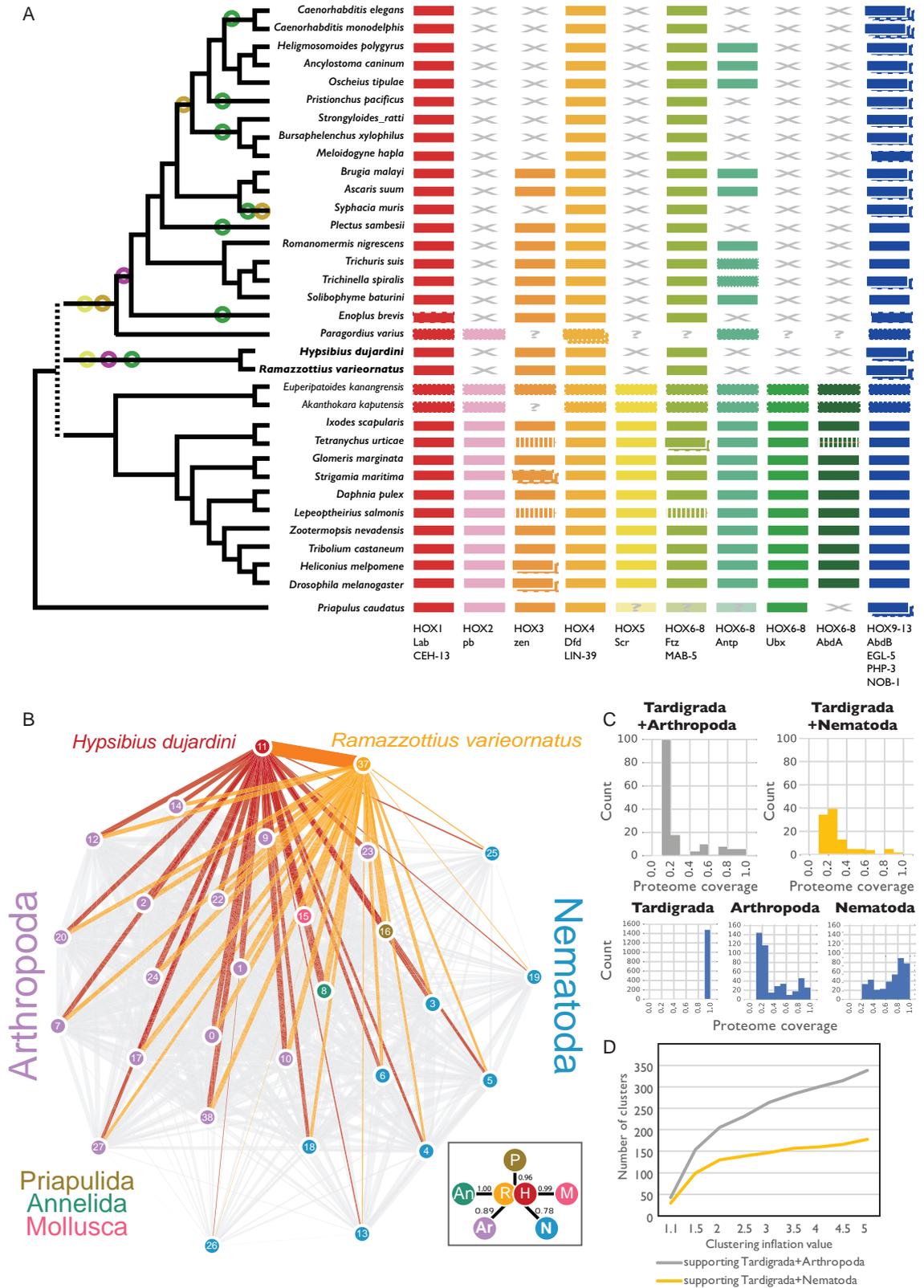
We explored the biological implications of these putative synapomorphies by examining the functional annotations of each protein family that contained members from  $\geq 70\%$  of the ingroup species (Table 2.3). Under Tardigrada+Arthropoda, 20 families had  $\geq 70\%$  of the ingroup taxa represented, and 6 were universally present. These included important components of developmental and immune pathways, neuromodulators, and others. Two families were annotated as serine endopeptidases, 1 missing in some arthropods that included *D. melanogaster Nudel* and 1 found in all species. Another synapomorphic family, found in all species, included *Spätzle* orthologues. *Spätzle* is a cysteine-knot, cytokine-like ligand involved in dorsoventral patterning and is the target of a serine protease activation cascade initiated by *Nudel* protease. The identification of more than 1 member of a single regulatory cascade as potential gene family births suggests that the pathway may have been established in a Tardigrada+Arthropoda most recent common ancestor. Other Tardigrada+Arthropoda-synapomorphic families were annotated with ommatidial apical extracellular matrix (eyes shut), adipokinetic hormone, neuromodulatory allatostatin-A, drosulfakinin, leucine-rich repeat, thioredoxin, major facilitator superfamily associated, and domain of unknown function DUF4728 domains. However, 9 of the 20 Panarthropoda synapomorphic families had no informative domain annotations. Under Tardigrada+Nematoda, only 5 putatively synapomorphic families had members from  $\geq 70\%$  of the ingroup species. Four of these had domain matches (proteolipid membrane potential modulator, zona pellucida, RUN, and amidinotransferase domains), and 1 contained no proteins with identifiable domains.

**Table 2.3: Gene family births that support different relationships of Tardigrada.**

Protein family*	Number of proteins	Proportion of proteomes represented			Domain annotations**	
		All	Nematoda (n=9)	Arthropoda (n=15)		Tardigrada (n=2)
Synapomorphies with membership $\geq 0.7$ under the Panarthropoda (Tardigrada + Arthropoda) hypothesis						
OG0000436	104	1.00	0.00	1.00	1.00	Serine proteases, trypsin domain (IPR001254)
OG0001236	54	1.00	0.00	1.00	1.00	Major facilitator superfamily associated domain (IPR024989)
OG0002592	36	1.00	0.00	1.00	1.00	Spätzle (IPR032104)
OG0006538	19	1.00	0.00	1.00	1.00	Leucine-rich repeat (IPR001611)
OG0006541	19	1.00	0.00	1.00	1.00	None
OG0006869	17	1.00	0.00	1.00	1.00	Thioredoxin domain (IPR013766)
OG0005117	27	0.88	0.00	0.93	0.50	BTB/POZ domain (IPR000210)
OG0005941	22	0.77	0.00	0.73	1.00	None
OG0006662	18	0.82	0.00	0.80	1.00	None
OG0006889	17	0.71	0.00	0.73	0.50	None
OG0006940	17	0.82	0.00	0.87	0.50	EGF-like domain (IPR000742), Laminin G domain (IPR001791)
OG0006941	17	0.71	0.00	0.67	1.00	EF-hand domain (IPR002048)
OG0006951	17	0.71	0.00	0.67	1.00	Adipokinetic hormone (IPR010475)
OG0007141	16	0.82	0.00	0.80	1.00	None
OG0007285	15	0.71	0.00	0.67	1.00	GPCR, family 2, secretin-like (IPR000832)
OG0007290	15	0.82	0.00	0.80	1.00	Allatostatin (IPR010276)
OG0007298	15	0.88	0.00	0.87	1.00	None
OG0007328	15	0.71	0.00	0.67	1.00	Sulfakinin (IPR013259), Peptidase S1A, nudel (IPR015420), Serine proteases, trypsin domain (IPR001254), Low-density lipoprotein (LDL)
OG0007463	14	0.77	0.00	0.73	1.00	receptor class A repeat (IPR002172)
OG0007689	13	0.71	0.00	0.67	1.00	Marvel domain (IPR008253)
Synapomorphies with membership $\geq 0.7$ under the Tardigrada + Nematoda hypothesis						
OG0005423	26	0.82	0.89	0.00	0.50	Amidinotransferase (PF02274)
OG0006414	20	0.82	0.78	0.00	1.00	Proteolipid membrane potential modulator (IPR000612)
OG0007199	16	0.91	1.00	0.00	0.50	Zona pellucida domain (IPR001507)
OG0007812	13	0.82	0.78	0.00	1.00	None
OG0008368	11	0.82	0.78	0.00	1.00	RUN domain (IPR004012)

\* Protein families from OrthoFinder clustering at inflation value 1.5.

\*\* Domain annotations are reported where proteins from more than one third of the proteomes in the family had that annotation. IPR = InterPro domain identifier; PF = PFam identifier.



**(A) HOX genes in tardigrades and other Ecdysozoa.** HOX gene losses in Tardigrada and Nematoda. HOX gene catalogues of tardigrades and other Ecdysozoa were collated by screening Ensembl Genomes and WormBase Parasite. HOX orthology groups are indicated by different colors. Some “missing” HOX loci were identified by Basic Local Alignment Search Tool (BLAST) search of target genomes (indicated by vertical striping of the affected HOX). “?” indicates that presence/absence could not be confirmed because the species was surveyed by PCR or transcriptomics; loci identified by PCR or transcriptomics are indicated by a dotted outline. “X” indicates that orthologous HOX loci were not present in the genome of that species. Some species have duplications of loci mapping to 1 HOX group, and these are indicated by boxes with dashed outlines. The relationships of the species are indicated by the cladogram to the left, and circles on this cladogram indicate Dollo parsimony mapping of events of HOX group loss on this cladogram. Circles are colored congruently with the HOX loci. **(B) Evolution of gene families under different hypotheses of tardigrade relationships.** Tardigrades share more gene families with Arthropoda than with Nematoda. In this network, derived from the OrthoFinder clustering at inflation value 1.5, nodes represent species (0: *Anopheles gambiae*, 1: *Apis mellifera*, 2: *Acyrtosiphon pisum*, 3: *Ascaris suum*, 4: *Brugia malayi*, 5: *Bursaphelenchus xylophilus*, 6: *Caenorhabditis elegans*, 7: *Cimex lectularius*, 8: *Capitella teleta*, 9: *Dendroctonus ponderosae*, 10: *Daphnia pulex*, 11: *Hypsibius exemplaris*, 12: *Ixodes scapularis*, 13: *Meloidogyne hapla*, 14: *Nasonia vitripennis*, 15: *Octopus bimaculoides*, 16: *Priapulus caudatus*, 17: *Pediculus humanus*, 18: *Plectus murrayi*, 19: *Pristionchus pacificus*, 20: *Plutella xylostella*, 37: *Ramazzottius varieornatus*, 22: *Solenopsis invicta*, 23: *Strigamia maritima*, 24: *Tribolium castaneum*, 25: *Trichuris muris*, 26: *Trichinella spiralis*, 27: *Tetranychus urticae*, 38: *Drosophila melanogaster*). The thickness of the edge connecting 2 nodes is weighted by the count of shared occurrences of both nodes in OrthoFinder-clusters. Links involving *H. exemplaris* (red) and *R. varieornatus* (orange) are colored. The inset box on the lower right shows the average weight of edges between each phylum and both Tardigrades, normalized by the maximum weight (specifically, count of co-occurrences of Tardigrades and the annelid *C. teleta*). **(C) Gene family birth synapomorphies at key nodes in Ecdysozoa under 2 hypotheses: Tardigrada+Nematoda versus Tardigrada+Arthropoda.** Each graph shows the number of gene families at the specified node inferred using Dollo parsimony from OrthoFinder clustering at inflation value 1.5. Gene families are grouped by the proportion of taxa above that node that contain a member. Note that to be included as a synapomorphy of a node, a gene family must contain proteins of at least 1 species of each child node of the node in question, and thus, there are no synapomorphies with <0.3 proportional proteome coverage in Nematoda and <0.2 in Arthropoda, and all synapomorphies of Tardigrada have 1.0 representation. **(D) Gene family birth synapomorphies for Tardigrada+Arthropoda (grey) and Tardigrada+Nematoda (yellow) for OrthoFinder clusterings performed at different Markov Cluster Algorithm (MCL) inflation parameters.** *H. dujardini* has been redescribed as *H. exemplaris*. Replicated from Yoshida *et al.* (2017).

## 2.4 Discussion

### 2.4.1 A robust estimate of the *H. exemplaris* genome

We have sequenced and assembled a high-quality genome for the tardigrade *H. exemplaris*, utilizing new data, including single-molecule, long-read sequencing, and heterozygosity-aware assembly methods. Comparison of genomic metrics with previous assemblies for this species showed that our assembly is more complete and more contiguous than has been achieved previously and retains minimal uncollapsed heterozygous regions. The span of this new assembly is much closer to independent estimates of the size of the *H. exemplaris* genome (75 to 100 Mb) using densitometry and staining. The *H. exemplaris* genome is thus nearly twice the size of that of the related tardigrade *R. varieornatus*. We compared the two genomes to identify differences that would account for the larger genome in *H. exemplaris*. While *H. exemplaris* had approximately 6,000

more protein coding genes than *R. varieornatus*, these accounted for only approximately 23 Mb of the additional span and are not obviously simple duplicates of genes in *R. varieornatus*. Analyses of the gene contents of the 2 species showed that while *H. exemplaris* had more species-specific genes, it also had greater numbers of loci in species-specific gene family expansions than *R. varieornatus* and had lost fewer genes whose origins could be traced to a deeper ancestor. *H. exemplaris* genes had, on average, the same structure (approximately 6 exons per gene) as did *R. varieornatus*; however, introns in *H. exemplaris* genes were on average twice the length of their orthologues in *R. varieornatus* (255 bases versus 158 bases). Finally, the *H. exemplaris* genome was more repeat rich (28.5% compared to only 21% in *R. varieornatus*). These data argue against simple whole genome duplication in *H. exemplaris*. The genome of *H. exemplaris* is larger because of expansion of noncoding DNA, including repeats and introns, and acquisition and retention of more new genes and gene duplications than *R. varieornatus*. The disparity in retention of genes with orthologues outside the Tardigrada, where *R. varieornatus* has lost more genes than has *H. exemplaris*, suggests that *R. varieornatus* may have undergone genome size reduction and that the ancestral tardigrade (or hypsibiid) genome is more likely to have been approximately 100 Mb than 54 Mb. We await additional tardigrade genomes with interest. While we identified linkage between genes in the 2 tardigrades, local synteny was relatively rare. In this, these genomes resemble those of the genus *Caenorhabditis*, in which extensive, rapid, within-chromosome rearrangement has served to break close synteny relationships while, in general, maintaining linkage (Mitreva *et al.*, 2005). The absence of chromosomal level assemblies for either tardigrade (and lack of any genetic map information) precludes definitive testing of this hypothesis.

#### **2.4.2 HGT in tardigrades: *H. exemplaris* has a normal metazoan genome**

Boothby *et al.* made the surprising assertion that 17.5% of *H. exemplaris* genes originated through HGT from a wide range of bacterial, fungal, and protozoan donors (Boothby *et al.*, 2015). Subsequently, several groups including our teams proved that this finding was the result of contamination of their tardigrade samples with cobionts and less-than-rigorous screening of HGT candidates (Arakawa, 2016; Bemm *et al.*, 2016; Delmont and Eren, 2016; Koutsovoulos *et al.*, 2016). We found that the use of uncurated gene-finding approaches yielded elevated HGT proportion estimates in many other nematode and arthropod genomes, even where contamination is unlikely to have been an issue. It is thus essential to follow up initial candidate sets of HGT loci with detailed validation. We screened our new *H. exemplaris* assembly for evidence of HGT, identifying a maximum of 2.3% of the protein coding genes as potential candidates. After careful assessment using phylogenetic analysis and expression evidence, we identified a likely high-confidence set of only 0.7% of *H. exemplaris* genes that originated through HGT. HGT was also low (1.6%) in the high-quality *R. varieornatus* genome. These proportions are congruent with similar analyses of *C. elegans* and *D. melanogaster*. Curation of the genome assemblies and gene models may decrease the proportion

further. Tardigrades do not have elevated levels of HGT.

While tardigrades do not have elevated levels of HGT in their genomes, some HGT events are of importance in anhydrobiosis. All *H. exemplaris* catalase loci were of bacterial origin, as described for *R. varieornatus* (Hashimoto *et al.*, 2016). While trehalose phosphatase synthase was absent from *H. exemplaris*, *R. varieornatus* has a TPS that likely was acquired by HGT (Supplementary Data S1). As *M. tardigradum* does not have a TPS homologue, while other ecdysozoan taxa do, this suggests that TPS may have been lost in the common ancestor of Eutardigrada and regained in *R. varieornatus* by HGT after divergence from *H. exemplaris*.

### 2.4.3 Contrasting modes of anhydrobiosis in tardigrades

Genes with likely roles in protection against extreme stress previously identified in *R. varieornatus* were largely conserved in *H. exemplaris*. Both *CAHS* and *SAHS* families had high copy numbers in both species, with independent expansions. However, we did not find a *Dsup* orthologue in *H. exemplaris*, however there has been a report of identification of a *H. exemplaris* *Dsup* ortholog (Hashimoto and Kunieda, 2017). *H. exemplaris* has similar gene losses to *R. varieornatus* in pathways that produce reactive oxygen species (ROS) and in cellular stress signaling pathways, which suggest that the gene losses occurred before the divergence of the 2 species. This loss of important signaling pathway genes may disconnect signals of stress induction from activating downstream response systems that must be suppressed if anhydrobiosis is to be achieved successfully—for example, cell cycle regulation, transcription and replication inhibition, and apoptosis. As cellular protection and repair pathways were highly conserved, damaged cell systems will still be protected and repaired. Indeed, some stress-related gene families had undergone gene family expansion in 1 or both tardigrades. *SOD* was duplicated in both species, as was a calcium-activated potassium channel, which has been implicated in cellular signaling during anhydrobiosis (Kondo *et al.*, 2015). The elevated gene family expansion in *H. exemplaris* compared to *R. varieornatus* may be related to retention and expansion of induced stress response systems.

The transcriptome response to anhydrobiosis differs between the 2 tardigrades. *H. exemplaris* has an induced transcriptomic response where *R. varieornatus* does not. We found that *H. exemplaris* had more genes differentially expressed on anhydrobiosis than *R. varieornatus*. As anticipated, more *R. varieornatus* loci were differentially expressed when desiccated at a slow pace. Genes induced by slow desiccation included *CAHS* and *SAHS* genes and antioxidant-related genes. Although most of these genes were highly expressed (>100 TPM) in the active state, the induction of these genes may enable higher recovery. *CAHS* and *SAHS* loci were also overexpressed on anhydrobiosis in *H. exemplaris*. We found a variety of calcium-related transporters and receptors were differentially expressed on anhydrobiosis. Kondo *et al.* suggested that cellular signaling using calmodulin and calcium may be required for anhydrobiosis (Kondo *et al.*,

2015), but it is still unclear how this is related to anhydrobiosis. Calcium and other metal ion concentrations could be increased during dehydration and thus could act as a desiccation signal. Trehalose is known for its role in protecting cellular systems from dehydration (da Costa Morato Nery *et al.*, 2008; Herdeiro *et al.*, 2006; Kikawada *et al.*, 2007; Yoshinaga *et al.*, 1997). It has been hypothesized that it may not be required for tardigrade anhydrobiosis, as trehalose was not detected in *M. tardigradum* (Hengherr *et al.*, 2008b). Trehalose synthesis via TPS has been lost in *H. exemplaris*, although we found an HGT-origin TPS in *R. varieornatus*. Unexpectedly, 3 *R. varieornatus* trehalase loci were differentially expressed on slow desiccation, including 2 with over 200 TPM in the anhydrobiotic state. As trehalose degradation should not be required in the absence of trehalose, there may be an alternative pathway for trehalose synthesis.

#### 2.4.4 The position of tardigrades in the metazoa

Our phylogenomic analyses found Tardigrada, represented by *H. exemplaris* and *R. varieornatus* genomes as well as transcriptomic data from *M. tardigradum* and *Echiniscus testudo*, to be sisters to Nematoda, not Arthropoda. This finding was robust to inclusion of additional phyla, such as Onychophora and Nematomorpha, and to filtering the alignment data to exclude poorly represented or rapidly evolving loci. This finding is both surprising and not new. Many previous molecular analyses have found Tardigrada to group with Nematoda, whether using single genes or ever larger gene sets derived from transcriptome and genome studies (Borner *et al.*, 2014; Campbell *et al.*, 2011; Dunn *et al.*, 2008). This phenomenon has been attributed to long branch attraction in suboptimal datasets, with elevated substitutional rates or biased compositions in Nematoda and Tardigrada mutually and robustly driving Bayesian and Maximum Likelihood algorithms to support the wrong tree. Strikingly, in our analyses including taxa for which transcriptome data are available, we found Onychophora to lie outside a ([Nematoda, Nematomorpha, and Tardigrada], Arthropoda) clade. This finding, while present in some other analyses (*e.g.*, component phylogenies summarized in (Campbell *et al.*, 2011)), conflicts with accepted systematic and many molecular analyses. We note that Onychophora was only represented by transcriptome datasets and that there is accordingly an elevated proportion of missing data in the alignment for this phylum.

Developmental and anatomical data do not, in general, support a tree linking Tardigrada with Nematoda. Tardigrades are segmented, have appendages, and have a central and peripheral nervous system anatomy that can be homologized with those of Onychophora and Arthropoda (Gross and Mayer, 2015; Martin *et al.*, 2017). In contrast, nematodes are unsegmented, have no lateral appendages, and have a simple nervous system. The myoepithelial triradiate pharynx, found in Nematoda, Nematomorpha, and Tardigrada, is 1 possible morphological link, but Nielsen has argued persuasively that the structures of this organ in nematodes and tardigrades (and other taxa) are distinct and thus nonhomologous (Nielsen, 2013).

*H. exemplaris* has a reduced complement of HOX loci, as does *R. varieornatus*. Some of the HOX

loci missing in the Tardigrada are the same as those absent from Nematoda. Whether these absences are a synapomorphy for a Nematode-Tardigrade clade or simply a product of homoplasious evolution remains unclear. It may be that miniaturization of Nematoda and Tardigrada during adaptation to life in interstitial habitats facilitated the loss of specific HOX loci involved in postcephalic patterning and that both nematodes and tardigrades can be thought to have evolved by reductive evolution from a more fully featured ancestor. It may be intrinsically easier to lose some HOX loci than others. While tardigrades retain obvious segmentation, nematodes do not, with the possible exception of repetitive cell lineages along the anterior-posterior axis during development (Sulston *et al.*, 1983). We note that until additional species were analyzed, the pattern observed in *C. elegans* was assumed to be the ground pattern for all Nematoda. More distantly related Tardigrada may have different HOX gene complements to these hypsibiids, and a pattern of staged loss similar to that in Nematoda (Aboobaker and Blaxter, 2003b; Aboobaker and Blaxter, 2003a; Aboobaker and Blaxter, 2010) may be found.

Assessment of gene family births as rare genomic changes lent support to a Tardigrada+Arthropoda clade, but the support was not striking. There were more synapomorphic gene family births when a Tardigrada+Arthropoda (Panarthropoda) clade was assumed than when a Tardigrada+Nematoda clade was assumed. However, analyses under the assumption of Tardigrada+Nematoda identified synapomorphic gene family births at 50% of the level found when Panarthropoda was assumed. We note that recognition of gene families may be compromised by the same “long branch attraction” issues that plague phylogenetic analyses and also that any taxon in which gene loss is common (such as has been proposed for Nematoda as a result of its simplified body plan) may score poorly in gene family membership metrics. The short branch lengths that separate basal nodes in the analysis of the panarthropodan-nematode part of the phylogeny of Ecdysozoa may make robust resolution very difficult. We explored the biological implications of the synapomorphies that supported Panarthropoda by examining the functional annotations of each protein family (Supplementary Table S14) and were surprised that many of these deeply conserved loci have escaped experimental, genetic, or biochemical annotation. One family included *Spätzle*, a cysteine-knot, cytokine-like family involved in dorsoventral patterning as well as immune response, and 2 others were serine endopeptidases, including *nucl*, which is part of the same pathway as *Spätzle*. This pathway may be a Panarthropod innovation. Thus, our analyses of rare genomic changes lent some support to the Panarthropoda hypothesis, as did analysis of miRNA gene birth (Campbell *et al.*, 2011), but analysis of HOX loci may conflict with this.

The question of tardigrade relationships remains open (Edgecombe, 2010). While we found support for a clade of Tardigrada, Onychophora, Arthropoda, Nematoda, and Nematomorpha, the branching order within this group remains contentious, and in particular, the positions of Tardigrada and Onychophora are poorly supported and/or variable in our and others’ analyses. Full genome sequences of representatives of Onychophora, Heterotardigrada (the sister group to the Eutardigrada including Hypsibiidae), Nematomor-

pha, and enoplian, basally arising Nematoda are required. Resolution of the conflicts between morphological and molecular data will be informative, either of the ground state of a nematode-tardigrade ancestor or of the processes that drive homoplasy in rare genomic changes and robust discovery of nonbiological trees in sequence-based phylogenomic studies.

---

---

## **Chapter 3**

---

Horizontal gene transfer in metazoa:  
examples and methods

### 3.1 Introduction

The evolutionary history of life is commonly represented as a tree-like process, with species separating and diversifying through repeated, nested bifurcations. However, it is increasingly becoming clear that the evolution of life also includes reticulation events—the exchange of genetic material across extended phylogenetic distances. In bacteria, horizontal exchange of genetic material between strains and species is common, via transducing phage, conjugation and direct transformation by exogenous DNA. This can drive rapid adaptation, for example, in resistance to anthropogenic antibiotic challenge. In eukaryotes, horizontal transfer of genetic material that does not conform to the bifurcating tree-like behaviour of the bulk of the genome takes two major forms. One is introgression or hybridisation, where one species acquires genetic material from another, usually closely related, via sexual reproduction between individuals from distinct lineages. In some cases, the event of hybridisation results in a lineage that is reproductively isolated and can thus form a new hybrid species.

The other form of reticulate evolution is horizontal gene transfer (HGT). We reserve the term HGT for transfer of genetic material into a eukaryotic nuclear genome from a distantly related donor through non-sexual mechanisms. HGT into eukaryotic genomes can be from other eukaryotes, from bacteria or from viruses. The relative importance of HGT in eukaryote evolution has been enthusiastically discussed for decades, and interest has grown since the beginning of the genome era (Salzberg *et al.*, 2001; Stanhope *et al.*, 2001). The emergence of high-throughput sequencing technology has enabled sequencing of numerous eukaryotes, and comparative genomics analyses have identified sequences that are likely to have originated by horizontal transfer in many taxa.

While HGT simply describes transfer of DNA from a donor to a recipient through non-sexual mechanisms, it is important to distinguish simple transfer of DNA and the incorporation of functional genes into the recipient genome. In eukaryotes, where gene expression is the tightly regulated outcome of complex transcriptional control, it is likely that most HGT events are neutral or mildly deleterious to the recipient (Blaxter 2007).

The mechanisms of HGT into eukaryotic genomes are unclear, though several possibilities exist. HGT may be part of the life strategy of pathogens and parasites. The tumour-inducing plasmid of *Agrobacterium rhizogenes* is transferred to the host, where it causes neoplastic diseases (crown gall and hairy roots; reviewed in (Nester *et al.*, 1984)). Viruses can transfer genes into the human genome (Crosbie *et al.*, 2013; Pett and Coleman, 2007; Sung *et al.*, 2012), and insertions of viral origin are common in sequenced eukaryotic genomes. However, these HGTs largely occur in somatic cells, causing disease in the recipient, and will not be inherited by the next generation. Most eukaryotes have effective genome surveillance systems that prevent, or at least minimise, the activities of elements in their germ lines, reducing the likelihood of HGT

acquisition from these sources.

Endogenous double-strand break (DSB) repair mechanisms can result in the insertion of DNA fragments into genomes, irrespective of their source or possible function. When these breaks are repaired in cells that go on to contribute to the next generation, the HGT event is transmitted and can become part of the core genome of the species. As an example, this neutral or non-functional HGT process has resulted in the insertion of many copies of mitochondrial genome fragments into the human and other genomes. These fragments, called nuclear-mitochondrial transfers (NUMTs), are not functional. NUMTs are not expressed (at any significant level) and do not contribute to the biology of the nuclear genome other than as bulk noncoding DNA. Once inserted, they evolve neutrally, incorporating “disabling” mutations (nonsynonymous substitutions, insertions and deletions) and have no adaptive significance. The high frequency of NUMTs suggests that these transfers arise because of the (ubiquitous) presence of the mitochondrion in germline cells. Similar abundances of chloroplast DNA transfers into photosynthetic lineage genomes and of vertically transmitted endosymbiont DNA into arthropod and nematode genomes (Dunning Hotopp *et al.*, 2007) are other examples of non-functional HGTs that arise because of the proximity of the donor genome to the germline (Blaxter, 2007).

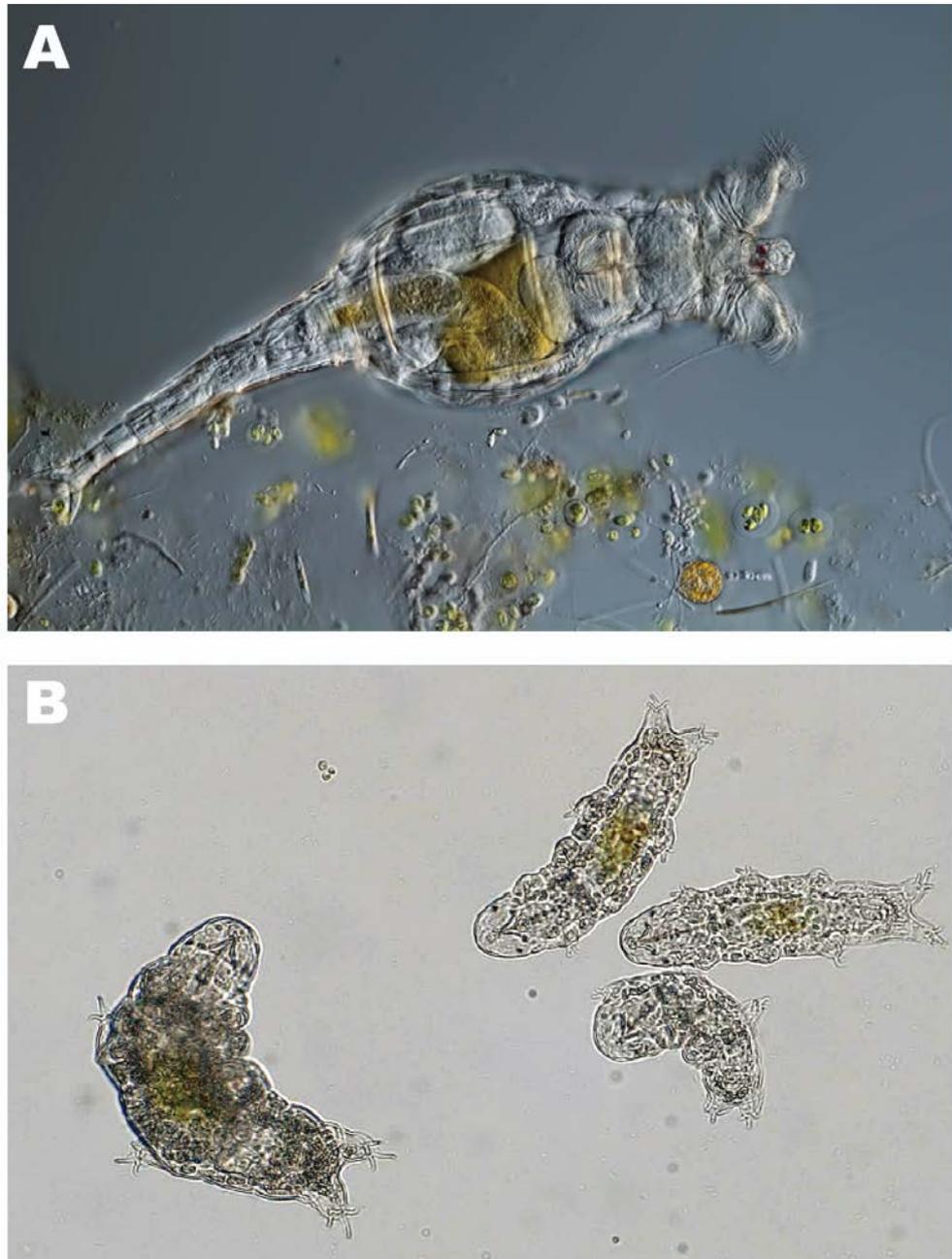
HGTs can acquire functional roles in their new host genome. Gene expression in (most) eukaryotes is complex and requires that a gene has the correct chromatin marks, has effective enhancer and promoter elements, (usually) contains spliceosomal introns and has the correct transcriptional termination and polyadenylation signals. To generate protein products at high levels, the transcript should also utilise a codon set optimised for the cellular environment. A horizontally transferred DNA segment may contain a gene that is intact based on its sequence, but because it lacks any one of these components, is not expressed and is not functional. In the case of HGT from bacteria or archaea, therefore, a gene will have to acquire enhancer and promoter elements, acquire spliceosomal introns and acquire transcription termination and poly (A) addition signals in order to be expressed. To be expressed at high levels, its codon usage will have to be ameliorated to match that of the new host genome (Blaxter 2007). HGT events into a eukaryote nucleus from other eukaryotes may result in the acquisition of loci that already carry transcriptional control signals and introns from their originating genome. Whether these function in the new host genome environment will depend on the phylogenetic relatedness of the donor and recipient taxa, but it can be imagined that functional integration might be easier. Thus, functional HGTs should display features of integration not only into the genome of the host but also into the molecular biological and physiological systems of the host.

Functional HGT loci that have their origins through HGT can have a wide range of enzymatic and biological roles. Husnik and McCutcheon (2018) offered a broad classification of the functions of horizontally transferred genes based on their contributions to the recipient cell. Maintenance HGTs complement endogenous functions, allowing endogenous gene loss, while innovative HGTs convey new phenotypes to

the host cell and enable adaptation to novel factors. For example, a major set of maintenance HGTs involved transfer to the nucleus of genes from the endosymbiotic, alphaproteobacterial ancestor of mitochondria. These loci still play roles in mitochondrial function. Similar ancient transfers occurred from ancestral plastid and chloroplast genomes in photosynthetic lineages and of nucleomorph genes after secondary symbioses (Archibald, 2015; Koonin, 2015; Ku *et al.*, 2015a;b; Pittis and Gabaldon, 2016; Spang *et al.*, 2015; Zaremba-Niedzwiedzka *et al.*, 2017). The loss of a gene in the host can be compensated by HGTs from endosymbionts and other sources, as observed in both insects and amoeba (Husnik and McCutcheon, 2016; Nikoh *et al.*, 2010; Nowack *et al.*, 2016).

Innovative HGTs have contributed to phenotypic diversity. Herbivorous animals in particular have acquired enzymes for digestion of plant cell wall and other complex carbohydrates from bacterial (and fungal) donors (Boschetti *et al.*, 2012; Ricard *et al.*, 2006; Richards *et al.*, 2011; Savory *et al.*, 2015; Schuster and Sommer, 2012; Wybouw *et al.*, 2016). To supplement restricted plant-derived diets, herbivores have acquired loci that drive synthesis of essential nutrients (Alsmark *et al.*, 2013; Husnik *et al.*, 2013). Some herbivores have acquired enzymes derived from bacterial metabolism to detoxify toxic plant compounds (Acuna *et al.*, 2012; Daimon *et al.*, 2008; Wybouw *et al.*, 2014; 2016). In return, plants have exploited bacterial genes to protect themselves from herbivores. For example, *Epichloë* fungi, which are endophytic in grass, produce a toxin from a likely HGT locus which kills larvae of herbivorous insects (Ambrose *et al.*, 2014; Daborn *et al.*, 2002). Several examples of functional HGT involve acquisition of bactericidal effectors (amidases and lysozymes) (Dunning Hotopp and Estes, 2014). HGTs also contribute to extremotolerance in various eukaryotes (Blanc *et al.*, 2012; Eyres *et al.*, 2015; Harding *et al.*, 2017; Mock *et al.*, 2017; Raymond and Kim, 2012; Schonknecht *et al.*, 2013).

Although the advent of whole genome sequence data has accelerated the detection of HGTs in eukaryotic genomes, their discovery, validation and biological significances remain controversial (Danchin, 2016). In this chapter, we describe the patterns of HGT found in genomes from two animal phyla: bdelloid rotifers, whose genomes contain many functional HGTs, and hypsibiid tardigrades, where an initial claim of high numbers of HGTs was found to be elevated due to technical errors (Figure 3.1). We review HGT detection methods and suggest solutions to the problems of robust and credible HGT detection in genome projects.



**Figure 3.1: Meiofaunal animals analysed for HGT.**

(A) The rotifer *Rotaria macrura*. Light micrograph, body length of specimen 1 mm. Image provided by Michael Plewka ([www.plingfactory.de](http://www.plingfactory.de)). (B) The tardigrade *Hypsibius exemplaris*. Light micrograph, body length of specimen 600  $\mu\text{m}$ . Image provided by Kazuharu Arakawa. Replicated from Yoshida *et al.* (2019b).

## 3.2 HGT in extremotolerant tardigrades

Tardigrades are a phylum of meiofaunal ecdysozoan animals found almost ubiquitously in marine, freshwater and limno-terrestrial sedimentary environments. Many freshwater and terrestrial tardigrades are cryptobionts, capable of entering anhydrobiosis and/or cryobiosis (Guidetti *et al.*, 2011; Keilin, 1959). Genome sequences are currently available for three species of Tardigrada, all from the class Eutardigrada: *Ramazzottius varieornatus* (Hashimoto *et al.*, 2016; Tanaka *et al.*, 2015; Yamaguchi *et al.*, 2012), *Hypsibius exemplaris* (Yoshida *et al.*, 2017) and *Milnesium tardigradum* (Bemm *et al.*, 2017). All three species are parthenogenetic, freshwater or terrestrial taxa. Analyses of their genomes identified tardigrade-specific gene families that are likely to contribute to their extremotolerant abilities (Hashimoto *et al.*, 2016; Tanaka *et al.*, 2015; Yamaguchi *et al.*, 2012).

The initial publication of a draft of the *H. exemplaris* genome (Boothby *et al.*, 2015) reported an extremely high number of loci that apparently were acquired by this species by HGT from a wide range of bacterial, protistan and other eukaryotic donors. The authors compared the best similarity scores obtained for each protein-coding locus in metazoan versus bacterial and non-metazoan databases and proposed HGT status for any where the bacterial match was much better than any match to an animal sequence. One sixth (17.5%) of the predicted protein-coding genes were reported as having sequence similarity signatures of HGT, exceeding the rates previously known in any other metazoan (including rotifers, see below). It was suggested that the tardigrades' cryptobiotic lifestyle somehow drove the high frequency of HGT, which in turn allowed the animal to acquire additional loci that played adaptive roles in cryptobiosis, a virtuous circle.

However, this claim was almost immediately refuted by multiple independent analyses that found a high level of bacterial contamination in the assembly and raw data. The genome assembly reported was over 200 Mb, while previous independent estimates of genome size using Feulgen densitometry and flow cytometry had suggested the genome was 70–100 Mb (Gabriel and Goldstein, 2007; Koutsovoulos *et al.*, 2016). It was possible to identify near-complete bacterial genomes (from Chitinophaga species, likely commensals living in or on the tardigrade cuticle, and several others) in the assembly and to exclude other assembled contigs because of biologically unlikely coverage and sequence similarity (Bemm *et al.*, 2016; Delmont and Eren, 2016; Koutsovoulos *et al.*, 2016). After applying rigorous filtering of contaminants from the published draft, many possible HGT candidates remained. Koutsovoulos *et al.* (2016) performed phylogenetic analysis of each HGT candidate, excluding any where the initial assessment based on sequence similarity alone was countered by phylogenetic signal of animal origin, or of low confidence in any placement. They identified only 0.2% of the genes as strong HGT candidates and an additional 1.5% as possible candidates. Two groups independently sequenced *H. exemplaris*, from the same source population (Arakawa, 2016; Koutsovoulos *et al.*, 2016; Yoshida *et al.*, 2017), and generated more contiguous drafts with much smaller

genome spans. Arakawa developed a novel ultralow input sequencing protocol to sequence the genome of a cleaned single individual of *H. exemplaris*, thoroughly inspected to be free of bacterial contamination, and obtained sequence data nearly free of any contamination (Arakawa *et al.*, 2016; Arakawa, 2016). By mapping these data to the published draft, nearly one third (31.7%) of the draft contigs, including the longest 11 contigs, were identified as contaminant. The majority of HGT candidates were predicted from these contaminants. In response to the above series of rebuttals, the lead authors argued that the 2.5 ~4.6% of putative HGT candidates remaining after exclusion of bacterial genomes were still high in comparison to other metazoan genomes and that their assertion that HGT and cryptobiosis were linked in the *H. exemplaris* genome could stand (Boothby and Goldstein, 2016).

We therefore re-sequenced the genome of *H. exemplaris* with single molecule real-time sequencing long reads and filtering short and long reads with the single specimen ultralow input sequencing approach (Yoshida *et al.*, 2018) and assembled these data using heterozygosity-aware protocols into a highly contiguous genome having just the expected size (104 Mb) (Yoshida *et al.*, 2017). Hashimoto and Kunieda (2017) reported the high-quality genome sequence of another tardigrade in the superfamily Hypsibioidea, *R. varieornatus*, with a span of 54 Mb and scaffold N50 length of 4.74 Mb. The genomes of *H. exemplaris* and *R. varieornatus* were annotated using transcriptomic data from developmental stages and multiple conditions including anhydrobiosis in order to have comprehensive coverage of expressed loci. Putative HGT candidates were identified from these gene sets and further validated through phylogenetic analysis. These analyses were consistent with the reinvestigation of the original draft genome data, with only 0.7% of all protein coding genes being credible candidates. The upper bound, less credible set of possible candidates in *H. exemplaris* included only 1.8%. In *R. varieornatus*, the less credible set amounted to only 1.2%, despite this species being much more desiccation tolerant.

The hypothesis that tardigrades have an extraordinary level of acquisition of functional genes through HGT and that this acquisition is related to DNA damage repair induced by anhydrobiosis thus does not hold, but is there no relationship between anhydrobiosis and HGT? Within our credible HGT candidate set, we have identified at least two instances where horizontally acquired genes are likely to contribute to the mechanisms of anhydrobiosis.

In *R. varieornatus*, Hashimoto and colleagues first reported that all four catalase genes are likely to derive from HGT (Hashimoto *et al.*, 2016). Catalases decompose hydrogen peroxide and are important in cellular defence from oxidative damage. Metazoan catalases (called clade III catalases) are typically 500 amino acids in length and have two domains: a catalase domain and a catalase-related domain. All four catalases in *R. varieornatus* contained an additional glutamine amidotransferase-like domain, making them more similar to bacterial clade II catalases. No typical clade III metazoan catalases were identified. Clade II catalases are known to tolerate higher temperatures and be more resistant to protein denaturants than

metazoan clade III catalases (Switala *et al.*, 1999). These catalases of *H. exemplaris* were also clade II (Yoshida *et al.*, 2017), and clade II catalases were found in the draft genome of *M. tardigradum* (Bemm *et al.*, 2017). These tardigrade clade II catalases may thus contribute to resistance during anhydrobiosis. *M. tardigradum* belongs to the order Apochela, while *R. varieornatus* and *H. exemplaris* are in Parachela (within the class Eutardigrada). This suggests that an original horizontal transfer event may have occurred within or before the common ancestor of Eutardigrada. Molecular dating estimated the separation of these two orders to be between 433 and 474 My, suggesting a very ancient origin of these genes (Regier *et al.*, 2005).

Another HGT candidate with functions related to anhydrobiosis are tardigrade trehalose-6-phosphate synthase (TPS) genes. The non-reducing sugar trehalose is accumulated in extremely high concentrations in a large number of anhydrobiotes, and it has been shown to be absolutely required for survival during anhydrobiosis in *Caenorhabditis elegans* (Crowe, 2014; Erkut *et al.*, 2011). However, trehalose has been found to be absent or only detected at very low levels (less than 3% of dry weight) in tardigrades (Hengherr *et al.*, 2008b). In line with these findings, the genomes of *H. exemplaris* and *M. tardigradum* lacked TPS. However, a locus encoding a protein with linked TPS and trehalose-6-phosphate phosphatase (TPP) domains was present in *R. varieornatus*. This tardigrade gene clustered within a clade of bacterial TPS loci in phylogenetic analyses. The genes in this bacterial clade predominantly derived from species in Chitinophagaceae (phylum Bacteroidetes). We note that precisely this group, the Chitinophagaceae, were major contaminants of the Boothby *et al.* (2015) draft *H. exemplaris* assembly, but this locus passes stringent tests for true HGT: it has spliceosomal introns and expression in poly(A) RNA. Many species belonging to the family Macrobiotidae display elevated amounts of trehalose during anhydrobiosis, implying that these disaccharides must be synthesised by these tardigrades (Jönsson and Persson, 2010). Further analysis covering a wider diversity of species in phylum Tardigrada will be necessary to elucidate whether such patchy conservation of TPS is the result of multiple gene loss events (*i.e.*, HGT occurred only once, more than 400 Ma) or whether multiple HGT events occurred after the separation of families within Eutardigrada (*i.e.* ca. 100 Ma).

### 3.3 Lessons from the tardigrades

The unravelling of the claim of extensive HGT in tardigrades, and its association with cryptobiosis, highlighted the importance of very careful assessment of claims of HGT. HGT is a rare event, and, as with all science, extraordinary claims require extraordinary evidence. Methods for detection of putative HGTs from a genome are inherently predictive, and predictive methods should always be assessed in terms of the possibilities of false positives. This is particularly the case with genomes newly assembled from previously

unexplored regions of the tree of life, and for species where functional genomic data such as transcriptomic datasets, are sparse.

We explored the contribution of the comprehensiveness and maturity of annotation to the level of likely false positives in putative HGT detection by comparing predicted HGT proportions between well-annotated, high-quality gene sets and *ab initio* annotations (using the 'industry standard' Augustus pipeline) of a set of non-vertebrate metazoan genomes obtained from Ensembl (Figure 3.2). Uncurated *ab initio* gene predictions resulted in extremely high predicted HGT proportions. For example, the well-studied genome of *Drosophila melanogaster* was predicted to contain ~4% putative HGTs. In every case the mature gene predictions contained many fewer potential HGT candidates. Therefore, the reality of HGT predictions must be conditioned by the understanding that such predictions contain false positives, especially when the assembly and annotations are in the early stages of curation. After thorough validation, the level of HGT was very limited in tardigrades, in keeping with proportions found in other metazoans (Figure 3.2). The over-prediction of HGTs in all uncurated *ab initio* gene sets suggests that these artefacts are produced because the gene finding systems are undertrained. Sadly, many erroneous predictions gather support from sequences already present in the public databases which are themselves the products of undertrained gene finding. This snowball of annotation error is currently unavoidable when using the bulk databases provided by the International Nucleotide Sequence Database Consortium nodes (GenBank, ENA, DDBJ) but can be minimised by only using well-validated data as evidence in gene finding (*e.g.*, the UniProt protein database or the set of reference genomes curated by Ensembl or NCBI). These findings emphasise the need for validation of HGT candidates using multiple criteria.

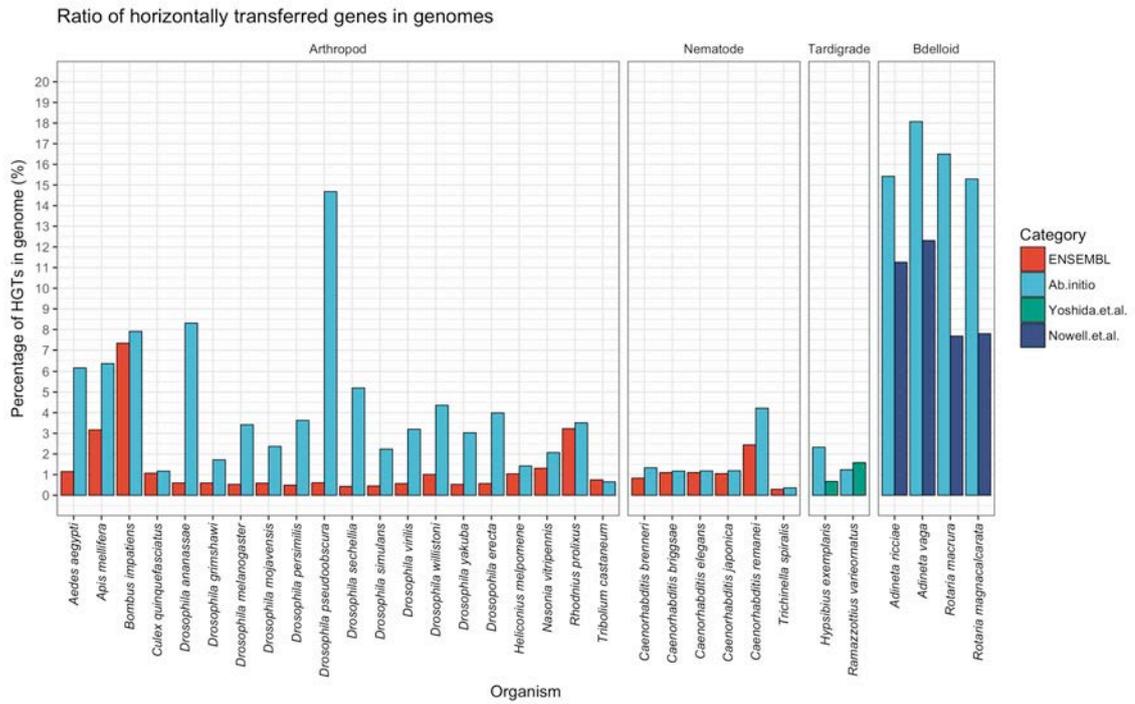


Figure 3.2: HGT proportions in metazoan genomes.

The proportions of genes identified as being likely HGT candidates were calculated using data taken from Ensembl Genomes for a range of non-vertebrate species, along with *de novo* assemblies of tardigrades and bdelloid rotifers. Additionally, *ab initio* gene predictions for the same Ensembl genomes were conducted using the Augustus toolkit, and HGT proportions were calculated. The validated HGT proportions for tardigrades and bdelloids were taken from Yoshida *et al.* (2017) and Nowell *et al.* (2018). Replicated from Yoshida *et al.* (2019b).

In particular we note that precise application of a series of technical and informatic procedures is critical to ensuring that erroneous claims of HGT do not arise and that claims must be strongly evidenced before being accepted as valid. At the outset of a genomics program, it is essential that samples of the target organism are cleaned as far as is possible of contaminants, to minimise contribution to the initial raw data from gut or endophytic microbiomes, adherent organisms, host- or food-derived material and DNA from contamination during processing. While the ‘holobiome’ of a target species is of interest (*i.e.*, it could be interesting to know the composition and likely physiology of its microbiome), a primary genome sequencing project needs to focus on specific target organism(s). The requirement for technical cleaning of samples before DNA extraction is most important, and most difficult, for specimens sourced from the wild.

Secondly, it is important to pay close attention during the assembly process to the signatures associated with contamination. Most contaminants can be robustly and rationally filtered at this stage because of their different genomic molarity compared to the target species, their unique sequence properties (bulk GC content or pattern of k-mer content) and their differential presence in independent samples.

Lastly, it is important to include several tools and approaches to detect and verify putative HGT events. These approaches should include sequence similarity and phylogenetics assessments of the putative HGT locus, accepting strong evidence only when the analyses are unequivocally in support. In addition, functional HGT should be supported by evidence of integration into the host genome, acquisition of a gene structure compatible with expression in the new location and the presence of polyadenylated transcripts derived from the HGT locus. The presence of DNA with phylogenetic evidence of origin through horizontal transfer but lacking any supporting functional evidence would suggest that the fragment is a non-functional insertion. It is important to note that the presence of a non-functional insertion cannot be argued to be a gene waiting to evolve a function or retained by the host for some future purpose. Evolution does not look forward, and a non-functional piece of DNA is not taken up and stored in the hope of future usefulness. Acquisition is random, not directed, and insertions are only selected once useful.

### **3.4 High levels of HGT in bdelloid rotifers**

While the precise levels and evolutionary impacts of HGT in eukaryotic evolution continue to be debated, it is already clear that the proportion of genes with signatures of HGT is usually low. Most maturely annotated metazoan genomes examined thus far have 1% or less of protein coding genes with likely origin through HGT. An exception to this finding are the bdelloid rotifers (Figure 3.1).

These tiny lophotrochozoans are abundant in a variety of limno-terrestrial and freshwater habitats, including permanent and temporary ponds, leaf litter and moss. In addition to analyses of their roles in ecosystem function, bdelloids are useful models for the study of two fundamental biological processes.

Firstly, they have gained fame as ancient asexuals, as neither males nor meiosis have been observed in over 500 described species, despite centuries of observation, while molecular and fossil evidence suggest that bdelloids have been evolving without sex for tens of millions of years (Welch *et al.*, 2009). The persistence and diversification of the asexual bdelloids are at odds with theory that suggests that asexual lineages should be short lived. Secondly, many bdelloid species are remarkably tolerant of desiccation and can survive extended periods of extreme water loss via anhydrobiosis (Ricci, 1998; Ricci and Covino, 2005). In this they resemble the tardigrades, but the two groups are likely to have evolved their cryptobiotic abilities independently, perhaps in association to miniaturisation.

Early explorations of bdelloid sequence data found that their genomes carried a remarkably high proportion of genes with signatures of HGT from non-metazoan origins (Gladyshev *et al.*, 2008). These putative HGT loci showed sequence similarity to a variety of bacteria, fungi and plants and were concentrated in sub-telomeric regions of the bdelloid genome. Subsequent work by Boschetti and colleagues using first expressed sequence tags (ESTs) (Boschetti *et al.*, 2011) and then transcriptome data (Boschetti *et al.*, 2012) from the bdelloid *Adineta ricciae* estimated that 8–9% of protein-coding genes were of foreign (non-metazoan) origin. These genes participated in a range of biochemical and metabolic functions. A similar proportion of putative HGT loci was reported for the first whole genome sequence for a bdelloid, from the closely related species *Adineta vaga* (Flot *et al.*, 2013). Subsequent comparative analyses of additional transcriptomes (Eyres *et al.*, 2015) and whole genomes (Nowell *et al.*, 2018) from multiple divergent bdelloid species with different ecological characteristics have also found similar levels of putative HGT into the expressed genome.

The finding that a substantial proportion of bdelloid genes are of foreign origin, and have presumably arrived in the bdelloid genome via HGT, is thus robust. The proportion of HGT identified is consistent across numerous studies from independent research groups, using different data types and multiple species, and have been estimated using methods that are fully cognizant of the potential pitfalls in HGT prediction discussed above. This remarkable pattern appears to be unique to bdelloids, as no other metazoan investigated with rigour has levels of HGT approaching the ~10% consistently reported for rotifers. Why are bdelloid rotifers so different from other Metazoa in their levels of HGT? Several hypotheses have been proposed, most of which adduce their ancient asexuality, lack of typical meiosis and cryptobiotic abilities to suggest either some adaptive role for the ability to take up and utilise foreign DNA in the absence of normal sexual genetic exchange or HGT as an inevitable by-product of processes that have evolved to maintain the integrity of the nuclear genome under pressure of damage induced by cryptobiotic stressors. The field is now exploring a suite of questions to test these hypotheses. Do bdelloids take up foreign DNA more readily than other taxa, and do these HGTs get functionalised more rapidly? Are there components of the mechanisms that bdelloids use to maintain genome integrity, such as DSB repair or gene conversion, that predispose them to HGT? The

elucidation of these problems is important not only for understanding the bdelloids, which are fascinating in themselves, but also for defining the evolutionary forces that shape the genomes of all animals.

One way to test these hypotheses is to ask how particular characteristics of bdelloid ecology or evolution may have resulted in higher levels of foreign genes, in particular the long-term lack of sex and recombination and anhydrobiosis. Previous work has placed emphasis firmly on anhydrobiosis (Boschetti *et al.*, 2012; Flot *et al.*, 2013; Gladyshev *et al.*, 2008), proposing a link between HGT and the repair of DSB damage in DNA induced by UV radiation while animals are desiccated (Gladyshev and Meselson, 2008; Hespeels *et al.*, 2014). Dehydration does cause DSB damage, and this must be repaired on rehydration for the animals to survive. The model suggests that fragments of exogenous DNA from bacteria, fungi and other organisms present in the environment or consumed as food particles might be occasionally incorporated into the host genome during the DSB repair. This model predicts a positive correlation between the frequency of desiccation and the number of foreign genes present in the genome and also an elevated level of HGT in other anhydrobiotic taxa. Some of the discussion of the extraordinary levels of apparent HGT in the initial draft genome of the tardigrade *H. exemplaris* invoked just this putative mechanism.

However, comparative analyses of desiccating and non-desiccating species of bdelloid, and other anhydrobiotic taxa, cast doubt on the role of desiccation in HGT. For example, some bdelloid species live in permanent aquatic environments and have lost the ability to survive desiccation. These taxa show levels of HGT that are comparable to anhydrobiotic species (Eyres *et al.*, 2015; Nowell *et al.*, 2018). An analysis of HGT using transcriptome data in both desiccating and non-desiccating species from the genus *Rotaria* did suggest a small increase in the cryptobiotic species (13% vs. 10%) but also found several HGT candidates that were unique to the non-cryptobiotic species (Eyres *et al.*, 2015). These loci suggest that acquisition of loci by HGT does not require cryptobiotic stress or cryptobiotic DSB repair mechanisms. Evidence from other anhydrobiotic taxa is also mixed. Few foreign genes have been detected in the tardigrades *R. varieornatus* and *H. exemplaris* (Yoshida *et al.*, 2017) or the chironomid midge, *P. vanderplanki* (Gusev *et al.*, 2014). Tardigrade DNA may be protected by tardigrade-specific nuclear proteins during anhydrobiosis and does not experience high levels of DSBs (Hashimoto *et al.*, 2016). In *P. vanderplanki*, however, there is evidence of extensive DNA damage (Gusev *et al.*, 2010) similar to that found in *A. vaga* (Hespeels *et al.*, 2014). Overall, the lack of extensive HGT in both systems does not support a hypothesised link between HGT and anhydrobiosis.

Could long-term asexuality provide a better explanation for the high levels of HGT in bdelloids? There has been speculation that horizontal transfers between rotifers and other organisms might be a side-effect of a mechanism of maintenance of genomic integrity under asexuality. Thus rotifer-to-rotifer “horizontal transfer” might operate to recombine alleles among individuals, in lieu of more conventional forms of sex. However, a study that claimed to find evidence of between-rotifer transfers (Debortoli *et al.*, 2016) was

subsequently shown to be the result of accidental cross-contamination between samples (Wilson *et al.*, 2018). There is thus no compelling evidence for frequent HGT to be a substitute for sex in bdelloids.

It is important to note that the overall rate of HGT in bdelloids is low in absolute terms, on the order of about 10 gains per million years (Eyres *et al.*, 2015). Perhaps it is not the rate of foreign import that is unusually elevated in bdelloids but the rate at which such genes are subsequently retained (Nowell *et al.*, 2018). Bdelloids might experience an equivalent level of primary transfers as other meiofauna living in similar ecological and environmental conditions, but these foreign DNA fragments are rapidly eliminated from non-bdelloid genomes because of the selective pressures imposed by both sex (meiosis) itself and by recombination more generally. In this model, HGT events are exceedingly rare occurrences in all taxa but have accumulated to relatively high levels in bdelloids over the evolutionary timescales during which recombination has been absent because of the inefficiency of elimination in the absence of sex.

Long-term asexuality could increase the rate of retention of newly inserted DNA in two ways. First, the distribution of fitness effects for HGT events in asexuals may be shifted such that foreign genes inserted into an asexual genome are on average less deleterious relative to the same events in a sexual species. This might arise because asexual transfers do not affect cellular processes and/or genomic structures that are crucial to normal meiosis, which would otherwise be highly deleterious or lethal. We note that HGT candidates are not randomly distributed in bdelloid genomes but are concentrated in sub-telomeric regions (Gladyshev *et al.*, 2008). This distribution of HGT loci resembles the partitioning of novel (largely non-HGT) genes and the majority of non-genic repeats to chromosomal arms in *Caenorhabditis* nematodes and may reflect the roles of these regions as testbeds for novel function. This insertion pattern of HGT loci resembles the patterns of insertion of mobile genetic elements (transposons, retrotransposons and related elements) in many species. Telomeres themselves play a crucial role in meiosis, particularly in the pairing of homologous chromosomes during meiotic prophase (Lee *et al.*, 2012; Siderakis and Tarsounas, 2007). It is possible that insertions into bdelloid telomeres are tolerated because purifying selection acting on telomere function in meiosis is either weakened or removed entirely. Insertions into gene-rich, non-telomeric regions are more likely to disrupt cellular functions and be deleterious regardless of the mode of reproduction.

Second, the lack of recombination in asexuals reduces the power of selection to remove transfers, even if they are initially mildly deleterious (*i.e.*, Hill–Robertson interference). The reduced effectiveness of selection in asexuals has already been shown for point mutations in a number of sexual–asexual comparisons, including in bdelloids (Barraclough *et al.*, 2007) and *Timema* stick insects (Bast *et al.*, 2018). Similarly, newly inserted foreign DNA might be able to persist longer in asexuals than in sexuals, providing sufficient time for amelioration. Even if asexuals are not subject to higher rates of initial incorporation of foreign DNA, they may realise higher rates of standing functional HGT simply because of suppressed recombination and reduced effectiveness of selection. Countering this is the fact that most asexual lineages are short lived in

evolutionary terms, and so accumulation of HGTs will not in practice be realised. That such a high proportion of genes encoded in bdelloids are of non-metazoan origin is perhaps the most convincing evidence for ancient asexuality.

It is still tempting to seek an association between cryptobiosis, ancient asexuality and HGT in the bdelloids. Could they be linked, with HGT promoted by DNA damage during cryptobiotic episodes bringing in genetic novelty otherwise unavailable to bdelloids? The non-directed nature of uptake and the long time required for amelioration of each gene argue against this hope. It may be that other mechanisms that permit longevity of the bdelloid asexual lineage have as a by-product affected rates of functional integration of foreign DNA. One candidate for such mechanisms might be high rates of gene conversion between alleles and homologues that serves to repair mutations and also results in amelioration of HGT fragments. These ideas require careful scrutiny and extensive testing and are the subject of ongoing research into bdelloids.

In prokaryotes the contributions of HGT to evolution, adaptation and differentiation in prokaryotes are richly documented and uncontroversial (Polz *et al.*, 2013; Vos *et al.*, 2015). However, in eukaryotes, and Metazoa in particular, functional HGT is much rarer, and the realised rate of HGT is orders of magnitude lower than in bacteria (reviewed in Husnik and McCutcheon (2018)). Functional HGT is an exceedingly rare event and for it to impact adaptation appears to be the exception rather than the rule. In this regard, the bdelloids may provide the single exception. Much remains to be elucidated regarding the evolutionary causes and consequences of HGT in bdelloid rotifers. Future work must consider all forms of (potential) recombination within a population genetics framework, measuring and modelling rates of DNA uptake, patterns of amelioration and, especially, the differences within and between individual lineages. It may be that bdelloids represent an edge case among metazoans and that the high level of HGT found here is a product of their unusual life history and is therefore unrepresentative of eukaryotes more generally. Thus, understanding what drives HGT in bdelloids will improve our understanding of the overall contribution of ongoing HGT to eukaryotic evolution.

### **3.5 Bioinformatic approaches to the prediction of HGT**

Cataloguing the specific loci involved in functional HGT in Metazoa, and thus measuring rates and patterns of HGT, depends sensitively on the methods used to identify HGT candidates in genomes and transcriptomes. These methods have developed rapidly in the last decade as more and more complete metazoan and eukaryote genomes have been sequenced, providing essential background in which to embed both analytic strategies and improve the credibility of predictions.

In bacteria, recently transferred genomic fragments can be distinguished from their new host genome by a range of metrics, including GC content and additional k-mer spectra, amongst which tetranucleotide

content appears most informative, CG skew in relation to the surrounding genome, and codon usage pattern. In addition, sequence similarity and phylogenetic analyses comparing the candidate HGT locus to the genomes of other, potential donor species can be used to show phylogenetic incongruence compared to the bulk genome. Less emphasis is placed in bacterial HGT analysis on more subtle but essential components of amelioration, such as acquisition of promoter elements optimally functional in the new host cell. Bacteria have very large effective population sizes, and thus amelioration at the level of GC content, k-mer spectra, CG skew and codon usage can be relatively rapid. The enduring signal of (relatively) ancient HGT is phylogenetic evidence of incongruence. An array of methods have been developed for HGT detection in bacteria (Zaneveld *et al.*, 2008).

Tools developed for bacterial metagenomics and detection of HGT in bacteria applications have been very effectively used in contamination detection in eukaryotic genome assemblies (including the draft genome of the tardigrade *H. exemplaris*) and can aid HGT detection (Clasen *et al.*, 2018; Eren *et al.*, 2015). The EuGI toolkit, developed from a bacterial HGT detection program, uses tetranucleotide spectral composition to identify extended regions that are likely to derive from HGT events (Clasen *et al.*, 2018). By identifying long (~100 kb) segments of assembled genomes that have tetranucleotide spectra that differ from that found in the local genomic region, and from the genome as a whole, EuGI was able to identify recent HGT events in protist, fungal and non-vertebrate genomes. Tetranucleotide frequency spectra are also used by the Anvi'o toolkit (Eren *et al.*, 2015) to build complex discriminators to separate segments that derive from distinct genomic sources and can identify recent HGTs. We note that tetranucleotide spectra perform poorly on short sequences and are also likely to fail to identify older, ameliorated HGTs.

In eukaryotes, functional integration of an HGT from a bacterial source is more difficult than simple acquisition of foreign DNA, as it requires acquisition of (in most cases) spliceosomal introns, mRNA initiation and polyadenylation signals and integration into the more complex transcriptional landscape of its host. Horizontal acquisition from other eukaryotes requires less amelioration. It is important to note that, for example, different eukaryotic groups can have very different spliceosomal intron systems and thus that donor introns may not be recognised by the recipient machinery. The lower effective population sizes of most eukaryotes mean the amelioration processes must be slower than is observed in bacteria. A phylogenetic incongruity signal will remain, and this is the primary source of evidence for HGT in eukaryotes. Evidence of amelioration—features of the putative HGT that closely resemble those of the host—are instead largely used to exclude contamination and to support functional integration into host systems. The process of sexual reproduction and the permeability of species barriers also mean that hybridisation between species, and thus introgression of genes from a donor to a recipient, can occur. These processes will result in sets of genes that have phylogenetic histories that differ from the species' history and could be classified as HGTs. As noted above, we reserve the term HGT for a locus that has been acquired through asexual mechanisms, but

the distinction between HGTs and introgressed genes is necessarily fuzzy: how deeply separated do two hybridising taxa have to be to generate observed loci with phylogenies indistinguishable from true HGTs? Can we tell whether a gene was acquired through ancient sexual reproduction rather than non-sexual uptake?

The identification of HGTs in bdelloid rotifers and tardigrades described above (Nowell *et al.* 2018; Yoshida *et al.* 2017; Yoshida *et al.* 2018) was based on initial candidate determination by sequence similarity followed by validation by phylogenetics, gene structure analysis and expression. The initial candidate lists were built by sequence similarity searches against taxonomically annotated sequence databases, identifying sequences that were much more similar to bacterial (or other potential donor) proteins than to any eukaryote or metazoan sequence. This HGT Index or Alien Index approach is the most commonly deployed tool for HGT identification (Table 3.1).

**Table 3.1: Methods for detection of HGT.**

Sequence similarity based				
Method	Reference	Input	Similarity search tool Database and Match threshold	Algorithm and Threshold
Alienness (Alien index)	Rancurel <i>et al.</i> (2017)	Predicted proteins	BLASTP (blastcl3) nr GenBank 1.0E-03 no filtering for low complexity regions	AI = $\ln(\text{best Metazoa E-value} + 1e^{-200}) - \ln(\text{best non-Metazoa E-value} + 1e^{-200})$ AI > 14    26
HGT index	Boschetti <i>et al.</i> (2012)	CDS (Transcript assembly)	BLASTX UniProt KB KW-0181 (complete proteomes) 1.00E-05	$hU = h_{nonmet} - h_{met}$ ; $hU >= 30$
class C	Crisp <i>et al.</i> (2015)	CDS (Transcript assembly)	BLASTX UniProt KB KW-0181 (complete proteomes) 1.00E-05	$hU = h_{nonmet} - h_{met}$ ; $hU >= 30$ && non-metazoan bitscore > 100
Integrated pipeline	Koutsovoulos <i>et al.</i> (2016) Nowell <i>et al.</i> (2018)	(either of above)	(either of above)	sum of bitscores wins (a) be found to be excluded from ingroup (b) have support for this designation from a given proportion of the rest of the hits (default is $\geq 90\%$ )
Sequence compisition based				
Method	Reference	Input	Algorithms and Threshold	
EuGI	Clasen <i>et al.</i> (2018)	Genome and database of "genomic islands"	1: OUP (oligonucleotide usage pattern of segment; tetranucleotide spectrum [k=4]) 2: D (rank order distance between two OUP spectra) 3: PS (pattern skew; D calculated for forward and reverse strands of the same segment) 4: RV (relative variance of segment normalised by GC content of segment) 5: GRV (relative variance of segment normalised by GC content of whole genome) A genome segment is classified as a HGT island when $D > 1.5$ , $PS < 55$ and $RV/GRV > 1.5$ .	

\* see <https://github.com/reubwn/hgt>

The Alien Index was introduced for the identification of HGTs in bdelloid rotifers (Gladyshev *et al.*, 2008) and uses the Expect (E) values for matches from database similarity searches as its core metric. E-values are the expectation that a match of the given quality (length, match score) would be found in a database of a particular size and base or amino acid composition. In this method, translated coding sequences (CDS) from the target genome are used as queries in a BLAST search (protein–protein, *i.e.* BLASTP) against the NCBI GenBank non-redundant protein (nr) database. High-scoring matches with an E-value cut off of less than  $1e^{-3}$  are retained and filtered to remove matches to proteins from the target species' taxonomic group (*e.g.*, Rotifera or the bdelloid searches), proteins with incomplete taxonomy and very short overlaps (less than 100 amino acids in length). For each target sequence, up to 1,000 hits are grouped based on their taxonomic origin (Bacteria, Fungi, Plantae, Metazoa and other eukaryotes including protists) and the lowest E-value recorded. If no hits are found for the particular taxonomic group, the E-value is set to 1. The lowest E-value from each taxonomic group is used to calculate the Alien Index (AI):

$$AI_{Metazoa:NonMetazoa} = \log(E - value_{Metazoa} + 1e^{-200}) - \log(E - value_{NonMetazoa} + 1e^{-200})$$

The  $AI_{Metazoa:NonMetazoa}$  is positive if the target sequence has greater similarity to non-metazoan sequences. Rather obviously, E-values are contingent on the search strategy (*e.g.*, initial k-mer or word length match value, whether simple sequence is masked and the scoring matrix used in evaluation) and on the size and composition of the database searched. The approach usually employs default program values (*e.g.*, for BLASTP a word size of 6, masking of simple sequence off, BLOSUM62 matrix for scoring). For the rotifer data (Gladyshev *et al.*, 2008), the AI varied between +460 and -460. Gladyshev and colleagues suggested that an AI > 45 was a good indicator of foreign origin, while targets with  $0 < AI < 45$  were indeterminate (Gladyshev *et al.*, 2008). The AI has been deployed widely, including in Alieness, a taxonomy aware web application for rapid detection of HGT candidates (Rancurel *et al.*, 2017). In Alieness, Rancurel and colleagues recommended an AI > 14 threshold for minimizing false positives and AI > 26 for predicting candidates that are more likely to produce phylogenies with higher support for HGT.

The HGT Index approach was first implemented for analysis of the transcriptome assembly of the bdelloid rotifer *Adineta ricciae* (Boschetti *et al.*, 2012). A gene set (in the case of *A. ricciae* a transcriptome assembly) is used as a query in a translated sequence similarity search (*i.e.*, BLASTX) against custom, taxon-specific subsets of the UniProtKB database of genic sequences (*e.g.* Metazoa, Plantae, Fungi, Eubacteria, Archaea and other Eukaryotes). Sequences from organisms belonging to the same phylum as the target organism are excluded from these taxon-specific databases. To simplify the search, only UniProtKB sequences from complete proteomes were used. High-scoring matches with an E-value cut off of less than  $1e^{-5}$  are retained. The HGT Index (hU) is calculated from the following formula:

$$hU_{Metazoa:Non-metazoa} = H_{Non-metazoa} - H_{Metazoa}$$

where  $H_{Non-metazoa}$  and  $H_{Metazoa}$  are the bitscores of the best hit against the non-metazoan and metazoan databases. While the hU approach thus avoids some of the contingent issues that arise from the simple use of E-values, it is still dependent on the specific search strategy used. Again, program defaults are standardly deployed. While target sequences that have  $hU > 0$  are more likely to have a non-metazoan origin, Boschetti and colleagues suggest using  $hU > 30$  as a filter. Using this HGT Index approach, 9.7% of the *A. ricciae* transcriptome was identified to be of foreign origin (Boschetti *et al.*, 2012). Among the 2,792 candidates, 1,884 had no significant metazoan match. Phylogenetic assessment of the remaining 908 candidates confirmed 887 as strong candidates for HGT into the *A. ricciae* genome. Using the same approach, strong HGT candidates were also identified in the rotifer *Brachionus plicatilis*, the nematode *C. elegans* and the fly *D. melanogaster* (Boschetti *et al.*, 2012).

The HGT Index approach was refined by Crisp *et al.* (2015). They proposed qualification of HGT candidates into Class C (those with  $hU > 30$  and best non-metazoan bitscore  $>100$ ), Class B (class C loci that are members of orthologue groups obtained using Markov linkage clustering where the metazoan proteins in the cluster have an average  $hU > 30$ ) and Class A (Class B proteins where their best metazoan BLASTP bitscore is  $< 100$  and their ortholog group contains no genes with a best metazoan bitscore  $>100$ ). Class A HGTs represent a minimum estimate of strongly supported HGTs for a target species. The application of these thresholds identified an average of 68 Class A genes in *Caenorhabditis* species' genomes, 4 in *Drosophila* species' genomes and an average of 32 genes in several primate species' genomes. However, as with the initial hypothesis of over 200 HGTs in the human genome (Lander *et al.*, 2001; Salzberg *et al.*, 2001), these assessments have been challenged (Salzberg, 2017) (see below).

The draft genome from the tardigrade *H. exemplaris* was submitted to public databases, and the proteins predicted from the contaminating Chitinophaga proteins thus are present in public databases annotated as metazoan in origin. Submission of data with erroneous taxonomic attribution is not confined to this assembly project. This contamination of the public databases has the dual negative effect of obscuring contamination (future projects discovering Chitinophaga sequences may be misled into believing them reliably metazoan) and confounding efforts to identify HGT. We have proposed an additional filter, which we here call Consensus Hit Support (CHS), for the AI and hU methods to improve confidence (Koutsovoulos *et al.*, 2016; Nowell *et al.*, 2018). Both AI and hU are calculated using only the 'best' hits from Metazoa and non-Metazoa (*i.e.*, those with the lowest E-value or highest bitscore) and can mislead if the top-ranked subject sequences are themselves taxonomically misclassified. The addition of a CHS filter mitigates this problem by requiring that any HGT candidate identified as non-metazoan based on similarity to the top-ranked hit should be supported by a specified proportion of the rest of the hits. Essentially, CHS asks whether there is a consensus in the taxonomic annotation derived from the top-ranked hit. The CHS filter makes HGT candidate identification less prone to errors associated with contamination issues in either the subject genome or the database itself,

as information from all hits is considered.

AI and hU approaches can only suggest candidate HGT loci. We have suggested that additional criteria should be assessed to support claims of HGT (Koutsovoulos *et al.*, 2016; Nowell *et al.*, 2018). Phylogenetic analyses are essential, as marginal differences in alignment scores in similarity searches can translate into significant AI or hU scores. Potential homologues of HGT candidates should be gathered by an iterative search process such as PHI-BLAST, where the search is widened based on the first rounds of matches to discover additional sequences that are likely to be members of the same homology group, halting when searches converge on a stable set of likely homologues. PSI-BLAST searches that fail to converge can also identify initial matches that are due to chance similarities or sequence biases and thus offer poor support for HGT status. Multiple sequence alignments should be inspected to affirm credibility of the alignment, asking if all the presumed homologues overlap with each other, or are only linked by their independent overlaps with the initial query sequence. A phylogenetic analysis of the target locus alongside all its best-matching putative homologues (from both source and recipient taxa) can reveal that there is no strong phylogenetic signal in the sequence or that the index approach has elevated a marginal difference to categorical significance.

### **3.6 All that glitters is not gold: HGTs as hypotheses**

*The argument for lateral gene transfer is essentially a statistical one ... As with all statistical arguments, great care needs to be exercised to confirm assumptions and explore alternative hypotheses. In cases where equally if not more plausible mechanisms exist, extraordinary events such as horizontal gene transfer do not provide the best explanation.*

Salzberg *et al.* (2001)

*(T)he evolutionary relationships among proteins cannot be concluded solely from the ranking of database hits in homology searches (for example BLAST reports). This is not a new conceptual point ... but one that seems to have been overlooked in this instance. Phylogenetic analysis must be a central component of any protein family or genome annotation effort.*

Stanhope *et al.* (2001)

Extraordinary claims require extraordinary evidence, and HGT should only be accepted when other, more likely explanations have been exhausted. Thus, reports that the human genome had 223 HGT candidates (Lander *et al.*, 2001), with 113 having homologues only in bacteria and vertebrates, were immediately rebutted through statistical and phylogenetic analyses (Salzberg *et al.*, 2001). Only ~40 lower-quality

candidates were left unrefuted. Detailed phylogenetic analysis of 28 of the proposed HGT candidates found that the majority were present in more anciently derived eukaryotes, rejecting HGT origin of these genes (Stanhope *et al.*, 2001). Salzberg (2017) similarly reanalysed and rejected 32 of the human genome Class A HGTs predicted by Crisp *et al.* (2015).

HGT detection through bioinformatics is predictive only, and all predictive methods result in both false negatives (missed HGTs) and false positives. The overprediction of HGT in initial *ab initio* gene predictions from new genome assemblies (see ‘Lessons from the Tardigrades’ above) is sobering. Even the current state-of-the-art gene predictors have base-level sensitivities in the range of 93–99% and specificities of 80–92% (Coghlan *et al.*, 2008) in model organism genomes, when compared to highly validated gene sets. Errors in gene prediction are directly reflected as false positives in HGT detection (Figure 3.2). While we found that curated metazoan genomes from Ensembl and our reanalysed tardigrade genomes still had around 1% HGT, this does not immediately imply that all metazoan genomes contain 1% HGT. We suggest that these numbers indicate that at the worst, our methods still contain up to 1% of false positives. These candidates must be validated.

Even where loci have a very strong AI or hU Index and phylogenetic support, we recommend using independent data for validation. Does the locus have strong evidence of incorporation into the recipient metazoan genome? Does it have spliceosomal introns? Is it physically linked in the genome assembly to other expressed loci of unproblematic metazoan origin? Is it present in all haplotypes of the recipient species or in related species? Is the locus a functional HGT or a dead-on-arrival acquisition of foreign DNA? Is the locus expressed in the recipient species at significant levels? HGT is undoubtedly real and has been an important process in organismal evolution. However, uncritical reporting of what are, in the end, low-confidence HGT candidates obscures the true rate and significance of horizontal transfer.

---

---

## Chapter 4

---

Comparison of the transcriptomes of two  
tardigrades with different hatching  
coordination

## 4.1 Introduction

Tardigrades are microscopic animals found in various meiofauna and are prominently known for their capability to tolerate various extreme environments during the ametabolic state designated as cryptobiosis (Keilin, 1959). Recent advances in the establishment of tardigrade genomic information has allowed for comprehensive molecular analysis in tardigrade species (Hashimoto *et al.*, 2016; Tanaka *et al.*, 2015; Yamaguchi *et al.*, 2012; Yoshida *et al.*, 2017).

The establishment of rearing systems for several species has enabled various quantitative zoological and molecular analyses, mainly focused on the limnic *Hypsibius exemplaris* and terrestrial *Ramazzottius varieornatus* (Gabriel *et al.*, 2007; Horikawa *et al.*, 2008). In particular, developmental research of the *H. exemplaris* embryo has provided valuable evidence regarding the evolution and phylogeny of tardigrades within Ecdysozoa (reviewed in Altiero *et al.* (2019)). A large amount of developmental research has focused on the morphology of *H. exemplaris*, due to their rapid embryogenesis and life cycle. The *H. exemplaris* embryo hatches after around 4 – 4.5 days (Gabriel *et al.*, 2007). Observations based on microscopy-based techniques have reported that the first cleavage is observed at 2 hours post laying (hpl), and gastrulation is completed by 16.5 hpl (Gabriel *et al.*, 2007). Further scanning electron microscope (SEM) observations indicate that the head and limb buds, segmental lines, claws and mouth are visible by –25 hpl, –30 hpl and 35 and 50 hpl, respectively (Gross *et al.*, 2017), suggesting that embryonic development is concluded around 3 days after oviposition. On the other hand, establishment of gene sequence databases and genetic engineering methods has enabled the molecular analysis of development related genes (Arakawa *et al.*, 2016; Hashimoto *et al.*, 2016; Yoshida *et al.*, 2017). Immunostaining-based expression analysis and *in situ* hybridization analysis has enabled the analysis of expression patterns of several genes: pair-rule gene paired (Pax3/7), the segment polarity gene engrailed (Gabriel and Goldstein, 2007), and Homeobox genes (Smith and Goldstein, 2016). Additionally, RNA interference knock-down of the expression of *mago-nashi* resulted in failure of elongation in the developing embryo, suggesting a major role of *mago-nashi* during embryogenesis (Tenlen *et al.*, 2013).

On the other hand, the hatching timing of in-lab cultures of *R. varieornatus* extended to  $5.8 \pm 1.2$  days (mean  $\pm$  SD), slightly longer than *H. exemplaris* (Horikawa *et al.*, 2008). Relatively higher diversity in hatching timing of in-lab cultures which lack non-predictable stimuli (*i.e.*, environmental factors) suggests that the diversity may be the result of adaptation to their living habitat.

To this end, we sought to identify factors that cause diversity in hatching timing. We have previously analyzed the genomes of two tardigrades from different habitats, the limnic *H. exemplaris* and terrestrial *R. varieornatus*. Genomes of both species have been sequenced, making these two species ideal for a comparative genomic/transcriptomic analysis, such as the identification of anhydrobiosis genes (Hashimoto

*et al.*, 2016; Yoshida *et al.*, 2017). We first observed hatching timing in *H. exemplaris* to calculate the diversity in this species. We then conducted a comparative transcriptomic analysis of the embryo and juvenile stages in both species to identify pathways that are induced around 1 day prior to hatching. We finally exposed embryos to the ecdysteroid 20-Hydroxyecdysone (20-E) and juvenile hormone analog Fenoxycarb to understand the effects of such substances on the embryo. We observed that *H. exemplaris* also has tightly regulated embryo hatching, and the arthropod molting pathway was induced. 20-E and Fenoxycarb exposure induced a latent state in only the *H. exemplaris* embryo. These findings support the idea that limnic species may have tighter regulation in embryo hatching and that embryo hatching may be hormone regulated.

## 4.2 Materials and methods

### 4.2.1 Tardigrade rearing and life history analysis

*R. varieornatus* was maintained using the methods previously described (Horikawa *et al.*, 2008). Briefly, individuals were placed on 2% Bacto Agar (Difco) gels prepared with Volvic mineral water and fed with *Chlorella vulgaris* (Chlorella Industry). Individuals were moved to a new plate every week. The cultures were examined daily to obtain eggs and were observed every day to obtain juvenile samples for days 3–7.

*H. exemplaris* was also maintained using the same method with minor modifications (1.2% Bacto Agar gel plates). For the life history analysis, 21 hatchlings were isolated from the culture plates, and were imaged every day. Body length was measured using ImageJ, and the days required for vitellogenesis, laying eggs days, and the number of eggs in each clutch were noted. 73 eggs were used for hatching day observation. Standard deviation was calculated for the hatching timing.

### 4.2.2 Transcriptome sequencing and data preparation

Isolated *R. varieornatus* specimens were collected and were subjected to transcriptome sequencing using the methods previously described with three replicates (Arakawa *et al.*, 2016). In brief, specimens were thoroughly washed with Milli-Q water on a sterile nylon mesh (Millipore), and were placed in a low-binding PCR tube. Pipette tips were used to homogenize the animals and lysed with the TRIzol reagent (Life Technologies). Total RNA was extracted using a Direct-zol RNA kit (Zymo Research), and the quality was validated with High Sensitivity RNA ScreenTape on TapeStation 2200 (Agilent Technologies). The mRNA was amplified with the SMARTer Ultra Low Input RNA Kit for Sequencing v.3, and Illumina libraries were prepared with the KAPA HyperPlus Kit (KAPA Biosystems). The library was quantified using the Qubit Fluorometer (Life Technologies), and size distribution was validated with TapeStation D1000 ScreenTape

(Agilent Technologies). The library was size selected for 200 bp by manually cutting out an agarose gel and was purified with a NucleoSpin Gel and PCR Clean-up kit (Clontech). The samples were sequenced using the NextSeq 500 High Output Mode 75 cycles kit (Illumina) as single end reads. Adaptor sequences were removed, and the reads were de-multiplexed using the bcl2fastq v.2 software (Illumina). All mRNA-Seq data have been uploaded to GEO under the accession ID GSE130111.

### 4.2.3 Informatics analysis

We previously performed time-series transcriptome analysis using the developmental stages of both *H. exemplaris* and *R. varieornatus* (Egg 1–5 d), in addition to the juvenile stages of *H. exemplaris* (Juvenile 1–7 d) and *R. varieornatus* (Juvenile 1 d) (Yoshida *et al.*, 2017). These RNA-Seq data contained 4–17M reads per sample. mRNA-Seq and small RNA-Seq data were downloaded from SRP098585 and SRR5179573, respectively. Sequence data were validated for sequencing quality using the FastQC software (Andrews, 2015).

Amino acid and coding sequences and annotations for *R. varieornatus* and *H. exemplaris* were downloaded from [www.ensembl.tardigrades.org](http://www.ensembl.tardigrades.org) (Yoshida *et al.*, 2017). Gene expression values were quantified using Kallisto v.0.42.4 with the options: `-bias -b 100 -t 31 -single -l 350 -s 50` (Bray *et al.*, 2016). The expression profiles were clustered by self-organizing map (SOM,2x3) using transcripts with an average over 1 TPM in all samples. To identify differentially expressed genes (DEGs), the RNA-Seq data were mapped to coding sequences with BWA MEM v.0.7.12-r1039 (Li *et al.*, 2009), after summarizing mapped read counts with in-house Perl scripts, statistical tests were conducted with DESeq2 v.1.8.2. (Love *et al.*, 2014). Genes with BH method adjusted  $p$ -values  $\leq 0.05$  were defined as DEGs.

Query genes were obtained from UniProt KB (UniProt Consortium, 2019), and were subjected to a BLASTP v2.2.22 (Altschul *et al.*, 1997) search against the amino acid sequences to determine tardigrade orthologs. For hormone receptor related gene identification, we screened the gene annotations for receptor genes that had gene ontology terms that contained “hormone”. Additionally, we obtained single embryo RNA-Seq data of *H. exemplaris* (Levin *et al.*, 2016), and assembled a 3' biased stranded transcriptome with Trinity v.2.4.0 (default parameters) (Grabherr *et al.*, 2011). This transcriptome assembly was subjected to a BLASTn search against a previous assembly of the whole transcriptome of *H. exemplaris*, which then were further subjected to BLASTn searches against the coding sequences to identify the originating genes (Yoshida *et al.*, 2017). To quantify gene expression, the RNA-Seq data were mapped against the assembled 3' biased stranded transcriptome with BWA MEM, and after summarizing mapped read counts with SAMtools idxstat v.1.4 (Li and Durbin, 2009), TPM values were calculated in R.

To identify PIWI-interacting RNA (piRNA) clusters, we used our previously sequenced small RNA-Seq

data (Yoshida *et al.*, 2017). Sequenced reads were mapped to known non-coding RNA sequences of *H. exemplaris* (tRNA, rRNA, miRNA from Yoshida *et al.* (2017)), and reads that were not mapped were used as follows. After removal of redundant and low-complexity reads using scripts from proTRAC (Rosenkranz and Zischler, 2012), we mapped the remaining reads to piRBase (Wang *et al.*, 2019) to identify piRNA sequences from *C. elegans* and *D. melanogaster*. The reads that did not map to piRBase were further mapped to the *H. exemplaris* genome, and proTRAC v.2.4.2 and PILFER (downloaded at 2018.11.05) were used for piRNA cluster identification (Ray and Pandey, 2018; Rosenkranz and Zischler, 2012). File conversion was conducted using BEDtools v.2.25.0, SAMtools and G-language Genome Analysis Environment v 1.9.1 (Arakawa *et al.*, 2003; Arakawa and Tomita, 2006; Arakawa *et al.*, 2010; Li *et al.*, 2009; Quinlan and Hall, 2010), and read mapping was conducted using Bowtie2 v.2.2.9 (Langmead and Salzberg, 2012).

#### 4.2.4 Chemical exposure

Fenoxycarb (Wako, Japan) and 20-Hydroxyecdysone (20-E, Tokyo Chemical Industry Co., Ltd., Japan) were used as juvenile hormone analog and ecdysone, respectively. Chemicals were eluted in 15  $\mu$ L 100% Ethanol (Nacalai tesque, Japan), and were further diluted with 45 mL of Volvic water to obtain the exposure concentrations (20-E : 8.32  $\mu$ M, 100  $\mu$ M, Fenoxycarb : 3.3  $\mu$ M, 100  $\mu$ M) following previous research on *Daphnia* (Baldwin *et al.*, 2001; Tatarazako *et al.*, 2003). For controls, embryos were exposed to 15  $\mu$ L 100% ethanol in 45 mL Volvic solution (0.03% Ethanol) at day 0. Embryos (0–3 days after oviposition) with exuvium (*H. exemplaris*) or freely laid eggs (*R. varieornatus*) were exposed to the chemical solution in 12 well dishes, and the number of days required for the eggs to hatch was observed for six days after oviposition. *R. varieornatus* embryos were exposed to only low concentrations. We also exposed embryos to 100  $\mu$ M of ecdysone receptor antagonist Cucurbitacin B (Tokyo Chemical Industry Co., Ltd., Japan) to assess whether inhibiting the molting pathway affected embryo hatching (Vellichirammal *et al.*, 2017; Zou *et al.*, 2018). We assayed six and three clutches for low and high concentration exposures, respectively. Three clutches were assayed for the 100  $\mu$ M Cucurbitacin B treatment. For *R. varieornatus*, we assayed three clutches for both chemicals. Treated samples (including control samples) were incubated at 18 degrees Celcius. Furthermore, to validate if unhatched embryos were capable of hatching, embryos were washed with Volvic water three times to remove chemicals and then cultured until hatching could be confirmed. To assess whether the unhatched embryo had completed development, the cuticle shells of embryos that did not hatch at Egg 7 d were peeled off with a 26G needle and were observed with light microscopy for claw formation.

### 4.2.5 Statistics

All statistics analysis were conducted in the R software. Standard deviation for each data are shown throughout this thesis. Heatmaps and line plots were generated using R and ggplot2 library.

## 4.3 Results

Previous studies have shown that embryonic development occurs in 4–4.5 days in *H. exemplaris*, and generation time takes 13–14 days. We collected embryos immediately after oviposition, and observed the development time, as well as generation time (Figure 4.1a). Embryonic development took  $4.03 \pm 0.50$  days from oviposition to hatch. The 1st and 2nd oviposition occurred around 7–11 days and 16–22 days after hatching, with  $2.68 \pm 0.75$  and  $8.34 \pm 4.12$  eggs per individual, respectively (Figure 4.1b). The body length reached the average length of an adult (240  $\mu\text{m}$ ) at around 15h after oviposition (Supplementary Figure S8). In comparison, *R. varieornatus* requires  $5.72 \pm 1.13$  days to hatch and the 1st and 2nd oviposition occurs at 8–12 days and 13–18 days after hatching. Additionally, we observed that the average number of eggs per individual was  $7.37 \pm 2.83$ , similar to 7.8 in previous studies (Horikawa *et al.*, 2008). We therefore assumed that hatching related genes may be upregulated around Egg 3 d and 5 d in *H. exemplaris* and *R. varieornatus*, respectively.

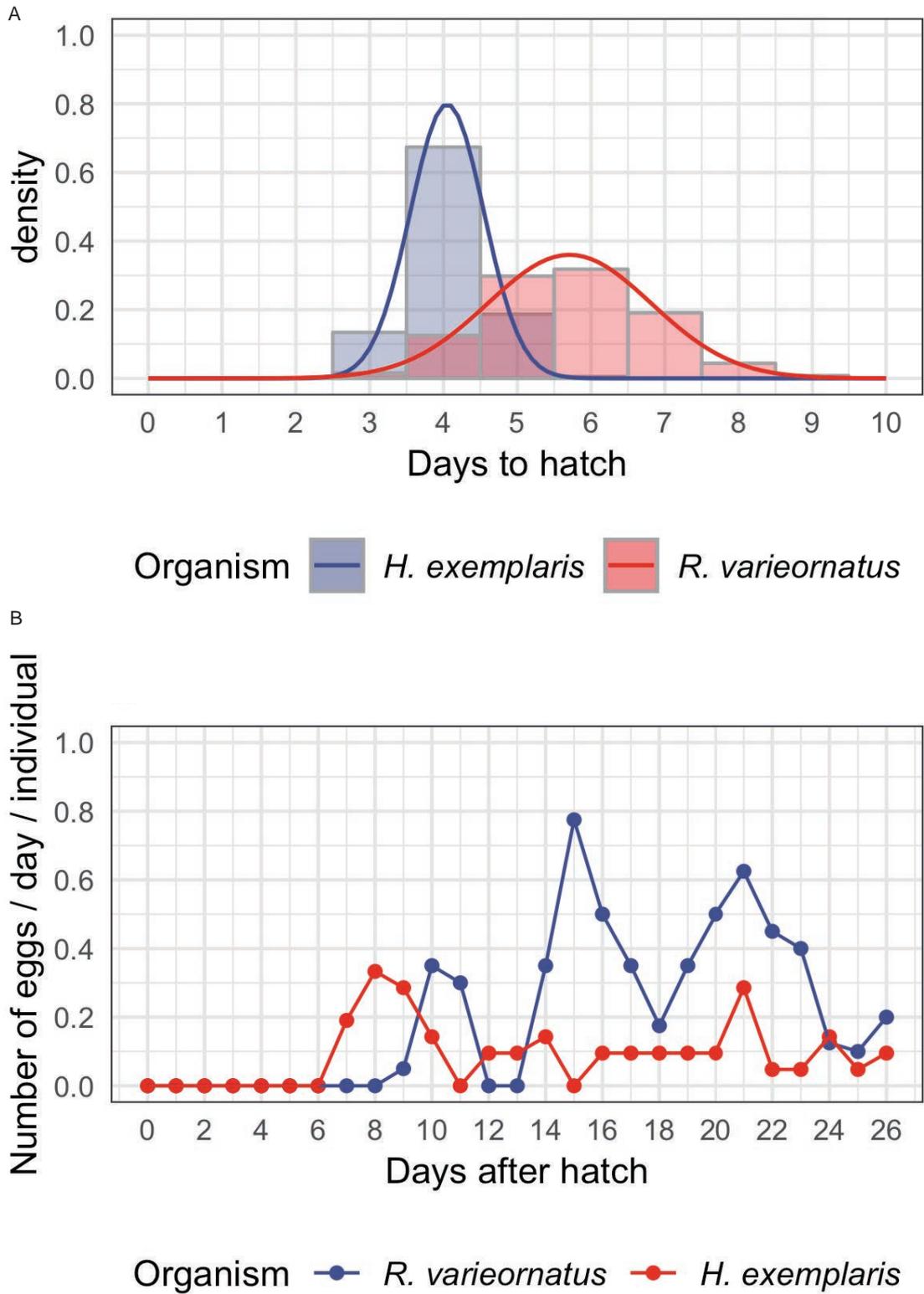
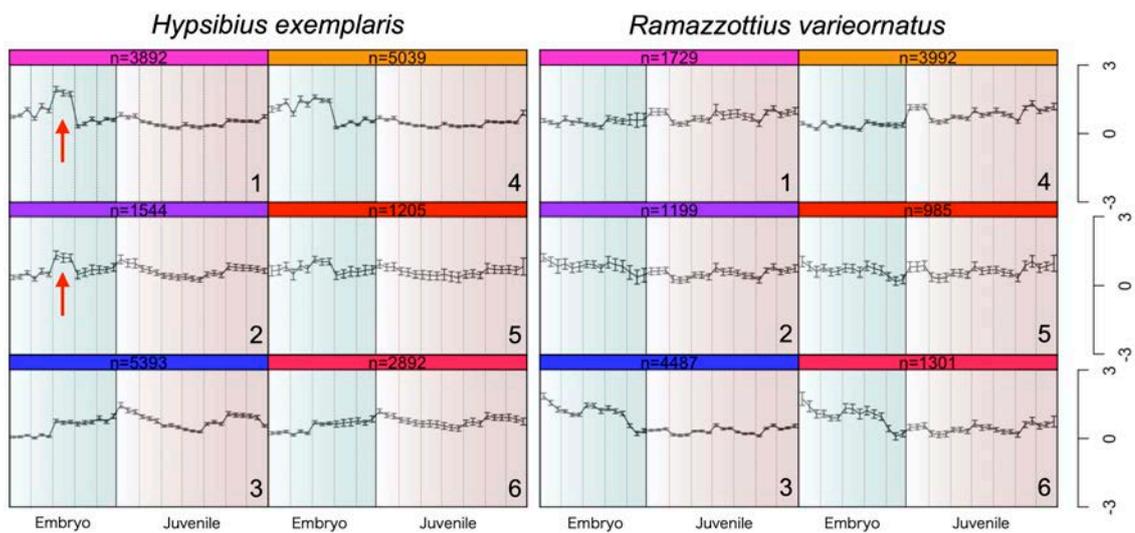


Figure 4.1: Comparison of hatching timing between *H. exemplaris* and *R. varieornatus*.

**(A) Probability density plot of the days required for egg hatching in both species.** 73 and 259 embryos were observed for *H. exemplaris* and *R. varieornatus*, respectively. A normal distribution curve is overlaid for each species. **(B) Number of eggs that were oviposited in each day per individual in *H. exemplaris* and *R. varieornatus*.** 21 and 40 individuals were observed for each species, respectively. Data from Horikawa *et al.* (2008) was used for *R. varieornatus*. Replicated from Yoshida *et al.* (2019a).

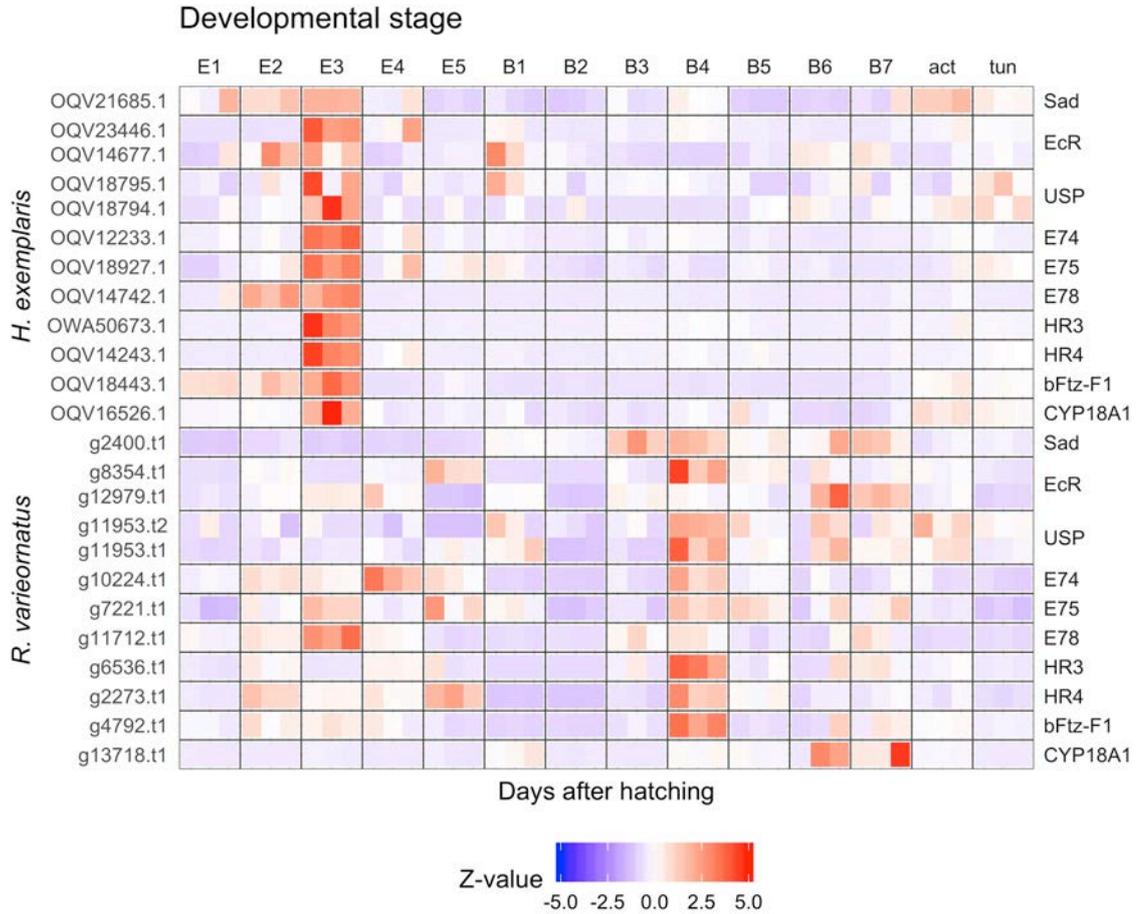
We have previously sequenced the transcriptome of the embryo and juvenile stages in *H. exemplaris*, and, in part, *R. varieornatus*. Hence, we additionally sequenced the juvenile stages of *R. varieornatus*, and conducted a comprehensive analysis of developmental and juvenile stages of both species (Supplementary Table S15). High correlation was observed between replicates (N=3). In order to validate our sequencing data, we first compared the expression profiles of HOX genes observed previously by single embryo RNA-Seq methods (Levin *et al.*, 2016). In our data, HOX genes showed an Egg 2-3 d specific expression in *H. exemplaris*, while a wide expression was observed at the Egg 2-4 d in *R. varieornatus* (Supplementary Figure S9). In particular, *caudal* was expressed in Egg 2 d, while other HOX genes were expressed at Egg 2-3 d. In comparison, *caudal* was mainly expressed around 500 minutes (8.3h), while other HOX genes (*i.e.*, *lab*, *ftz*, *disformed*, *zen*) were expressed around 1,600 minutes (26.7h) in the CEL-Seq data (Supplementary Figure S10). To obtain a general view of the transcriptome profile of both species, we conducted SOM clustering of gene expression profiles. We found two out of six clusters showed an Egg 3 d-specific expression in *H. exemplaris* (Figure 4.2, groups 1 and 2, Supplementary Data S13), suggesting dynamic shift in the embryonic transcriptome prior to hatching in only *H. exemplaris*. No such shift was observed in *R. varieornatus*.

In order to identify potential mediators of hatching, we screened the 5,436 genes induced at the Egg 3 d in *H. exemplaris* (Group 1 and 2, Supplementary Data S14). It is commonly known that many factors are regulated during embryonic development; however, most of these mechanisms have yet to be validated in tardigrades. We focused on signaling through neuropeptides and hormones, of which we identified 232 receptor genes in *H. exemplaris* with gene ontology annotations containing “hormone”. 82 of these orthologs were contained in groups 1 and 2. We obtained 20 orthologs that were significantly expressed in only Egg 3 d. Although a majority of these genes were atrial natriuretic peptide receptors, we identified multiple orthologs of the arthropod molting pathway (*i.e.*, ecdysone receptor (*EcR*), *RXR*, *E75*, *HR3*, *HR4*). Therefore, we collected other genes in the molting pathway identified in previous studies and observed their expression patterns during development. The expression patterns of these genes were highly upregulated in Egg 3 d in *H. exemplaris* and constitutively moderately between Egg 2-5 d in *R. varieornatus* (Figure 4.3).



**Figure 4.2: SOM clustering of gene expression profiles during development.**

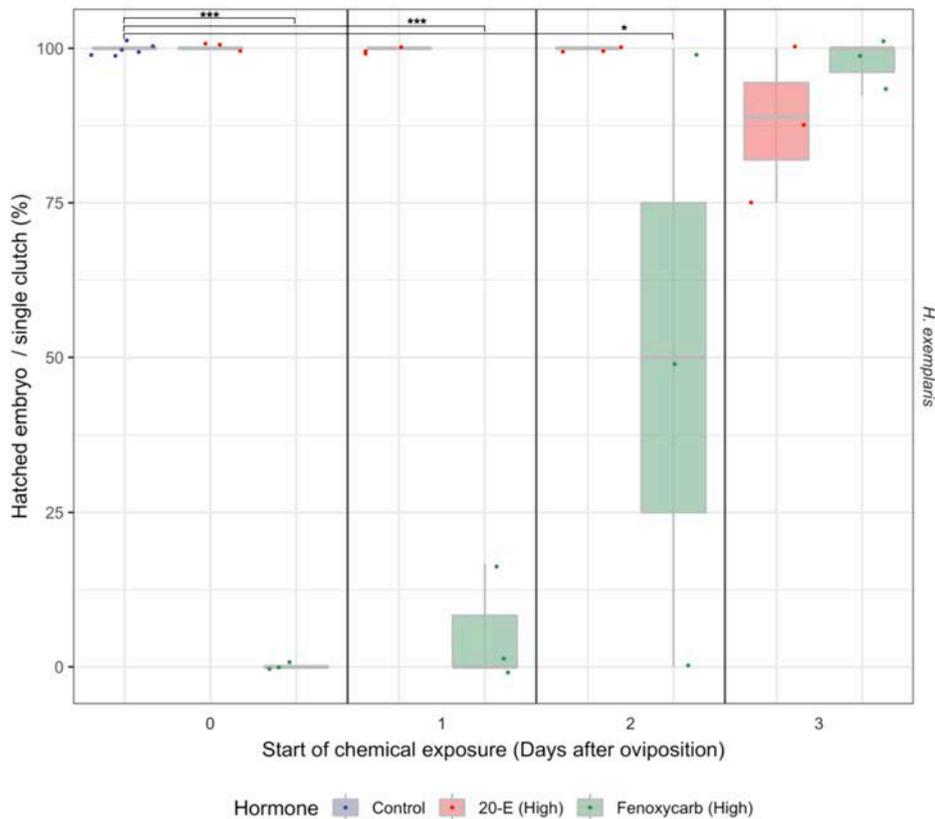
SOM clustering of TPM values were performed in R. Genes with TPM values over 1 were used. The areas colored in blue and red represent embryo and juvenile stages, respectively. The Y axis indicates Z-score normalized TPM values. The arrows indicate an increase in Egg 3 d of *H. exemplaris*. Replicated from Yoshida *et al.* (2019a).



**Figure 4.3: Gene expression patterns for molting pathways during development.**

Genes identified in (Schumann *et al.*, 2018) were identified by BLASTP searches against amino acid sequences of *H. exemplaris* and *R. variegornatus*, and Z-score normalized TPM values were visualized. E: Egg, B: Juvenile, act/tun: Adult stages. The identified orthologs are as follows; sad (OQV21685.1, g2400.t1), EcR (OQV14677.1, OQV23446.1, g12979.t1, g8354.t1), USP/RXR (OQV18794.1, OQV18795.1, g11953.t1, g11953.t2), E74 (OQV12233.1, g10224.t1), E75 (OQV18927.1, g7221.t1), E78 (OQV14742.1, g11712.t1), HR3 (OWA50673.1, g6536.t1), HR4 (OQV14243.1, g2273.t1), bFtz-F1 (OQV18443.1, g4792.t1), CYP18A1 (OQV16526.1, g13718.t1). An additional copy of *EcR* was found during this analysis (*H. exemplaris* OQV23446.1, *R. variegornatus* g8354.t1). Replicated from Yoshida *et al.* (2019a).

We then validated whether disrupting the molting pathway inhibits hatching in *H. exemplaris*, we exposed *H. exemplaris* embryos to chemicals (20-E: 8.42  $\mu$ M and 100  $\mu$ M, Fenoxycarb: 3.3  $\mu$ M and 100  $\mu$ M) and observed whether they affected embryo development. While we observed that approximately 90% of the embryos exposed to low concentrations hatched within 5 days of oviposition (Supplementary Figure S11), there were several cases where all of the embryos in the same clutch did not hatch (Supplementary Figure S12). We observed claw formation at this stage, which is consistent with previous body developmental observations. Washing these embryos to remove chemical exposure induced hatching in 57.1% (4/7, five E0 embryos and two E3 embryos, of which three E0 embryos died) and 71.4% (5/7, one F0 and F1 embryos, three F2 embryo and two F3 embryos, of which one F0 and F2 embryos died) of the embryos exposed to 20-E and Fenoxycarb, respectively. Exposure of *R. varieornatus* embryo to low concentrations did not show any effect in hatching (Supplementary Figure S12). On the contrary, exposure to high concentrations of Fenoxycarb significantly suppressed hatching in embryos exposed at Egg 0-2 d (Figure 4.4, Supplementary Figure S13, ANOVA and Tukey HSD, FDR < 0.05). For the embryos that did not hatch, we also observed claw formation and movement inside the eggshells regardless of exposure time, suggesting that the embryos had finished development and did not die (Supplementary Data S15). Exposure to the EcR antagonist cucurbitacin B had no effect on embryo hatching (Data not shown).



**Figure 4.4: High concentration Fenoxycarb exposure inhibited hatching in *H. exemplaris*.**

Percentages of successful hatching embryos exposed in 100  $\mu$ M Fenoxycarb or 100  $\mu$ M 20-E until hatch. Embryo exposed to 0.03% ethanol in Volvic mineral water at Egg 0 d were used as controls. Hatching ratio for each clutch are overlaid on the boxplot. The X axis indicates the day to start exposing the chemicals after oviposition. High concentration treatment of Fenoxycarb from 0-2 days after oviposition significantly impaired hatching, whereas treatment at 3 days after oviposition and treatment with 20-E did not. ANOVA and Tukey HSD, \* = FDR < 0.05, \*\* = FDR < 0.01, \*\*\* = FDR < 0.001. Replicated from Yoshida *et al.* (2019a).

## 4.4 Discussion

Previous studies on the life history of *H. exemplaris* have observed that the embryo hatches around 4–4.5 days after oviposition (Gabriel *et al.*, 2007). We have observed that the *H. exemplaris* embryo hatches in a strictly controlled manner within a short time range ( $4.03 \pm 0.50$  days) compared to the wider range observed in *R. varieornatus* ( $5.72 \pm 1.13$  day, see Horikawa *et al.* (2008)). The small variance in *H. exemplaris* is similar to the limnic species *Isohypsibius myrops* (Ito *et al.*, 2016). Therefore, we hypothesized that transcriptome regulation affects hatching timing, and factors that are related to hatching regulation may be induced around a day prior to hatching; day 3 in *H. exemplaris* and day 5 in *R. varieornatus*.

To understand the molecular pathways regulated prior to hatching, we analyzed our previously sequenced transcriptome data of the developmental stages of two tardigrade species, *H. exemplaris* and *R. varieornatus*, with newly generated data for the juvenile stages of *R. varieornatus* (Arakawa *et al.*, 2016; Yoshida *et al.*, 2017). First, we validated our sequencing data by comparison of the expression profiles of developmental genes (*i.e.*, Homeobox genes) assayed in previous single embryo sequencing data and *in-situ* hybridization experimental data (Levin *et al.*, 2016; Smith *et al.*, 2016). In our data, caudal and engrailed expressed earlier than HOX orthologs, at around 24–48 h post laying, and other orthologs were expressed around 48–72 h, approximately 36 hours prior to hatching. These data are similar to those of previous reports of HOX expression profiles, supporting the validity of our data. In comparison, *R. varieornatus* HOX orthologs were expressed around Egg 3d–4 d, 24 hours later than *H. exemplaris*. This period is also approximately 36 hours prior to hatching (Horikawa *et al.*, 2008). Clustering based analysis of the embryo–juvenile time course indicated an abrupt shift in transcriptome profile at Egg 3 d in *H. exemplaris*. The Egg 3 d stage is immediately before the hatch in *H. exemplaris* ( $4.03 \pm 0.53$  days), which suggests that *H. exemplaris* may have some sort of trigger that drastically changes the transcriptome profile for embryo hatching. The lack of such a shift in *R. varieornatus* remains a surprise, as we have sequencing data until approximately 20 hours prior to hatching. This may be due to the low quality of the sequencing data, observed by the low mapping ratio, or pathways induced in *H. exemplaris* may be expressed less abruptly, but rather gradually during the last few days of embryogenesis.

Analysis of genes induced at Egg 3 d of *H. exemplaris* revealed the arthropod molting pathway was upregulated at Egg 3 d. We first observed that the expression of small RNA-related genes (Argonaute, OQV17169.1/ OQV21162.1, g3665.t1) was induced at the developmental stages, which supports the implication of piRNA existence suggested in previous studies. We used our previously sequenced small RNA-Seq data and observed that 1.5M reads mapped to the *H. exemplaris* genome, but only 27K reads to known piRNA sequences from *C. elegans* and *D. melanogaster* in piRbase. Using these piRNA candidates, PILFER and proTRAC, predicted 894 and 187 piRNA clusters, respectively. We designated the 201 regions

that were identified in both methods as putative piRNA clusters in *H. exemplaris*, which overlapped with 490 genes (504 transcripts), of which 385 transcripts were hypothetical genes (Supplementary Data S16). Using our previously sequenced *H. exemplaris* small RNA-Seq data (Yoshida *et al.*, 2017), we predicted piRNA clusters in the *H. exemplaris* genome, consistent with previous studies (Sarkies *et al.*, 2015). Most of these piRNA clusters showed no homology with *D. melanogaster* nor *C. elegans* piRNAs, suggesting most of these piRNAs may have been acquired after divergence from the common ancestors of Arthropoda and Nematoda. We also observed that several copies of CAHS and SAHS genes were expressed exclusively in the developmental stages (Supplementary Figure S14A,B, Supplementary Table S16). Additionally, we have found that the anti-oxidative stress-related thioredoxin was also induced in Egg 3 d in both species (Supplementary Figure S15). In comparison, glutathione S-transferase and superoxide dismutase were most expressed during the early juvenile stages, and catalase at late juvenile to adult stages. Tardigrades are famous for their tolerance capabilities, such as desiccation tolerance (Keilin, 1959). Although we found several CAHS and SAHS orthologs to have relatively high expression during the embryonic stages in both species (Yamaguchi *et al.*, 2012), we found one SAHS ortholog that showed particularly high expression in both embryonic stages, suggesting that the regulation of an embryo-specific ortholog may be conserved between Hypsibiidae and Ramazzottiidae. Along with the expression profiles of antioxidative genes, these data suggest the fundamental anhydrobiosis mechanism during the embryonic stages in Hypsibiidae and Ramazzottiidae. Additional analysis of the developmental stages in *Milnesium inceptum*, a terrestrial species capable of anhydrobiosis in the embryonic stages (Schill and Fritz, 2008), may further support our findings. Finally, molting is an essential event for the development of animals in the Ecdysozoa, however, the fact that the arthropod molting pathway is induced in embryos of *Daphnia magna*, *Bombyx mori* and *D. melanogaster* (Bownes and Rembold, 1987; Kato *et al.*, 2007; Sekimoto *et al.*, 2006; Sonobe and Yamada, 2004), suggests a mechanism conserved within Arthropoda that may be related to embryo hatching. In ecdysozoans that have been investigated, the molting pathway is initiated by the signaling hormone Ecdysone, synthesized by the Halloween genes in hexapods. It is released as the active state 20-hydroxyecdysone (20-E, (Karlson, 1996)). The released hormone binds to the EcR and ultraspiracle (USP) heterodimer, which regulates the expression of the early genes (HR3, HR4, E75, Ftz-F1). These respond and transmits the signal to the late response genes, leading to new cuticle biosynthesis. Although most of the downstream genes of EcR/USP were conserved, only shadow (*sad*) gene in the 20-E biosynthesis pathway were found in both *H. exemplaris* and *R. varieornatus* (Schumann *et al.*, 2018). These genes were induced at Egg 3 d in *H. exemplaris* and relatively constitutively expressed (most expressed at Egg 5 d) in *R. varieornatus*, approximately 1 day prior to egg hatching. We did not observe a lag between early and late genes since we sampled the embryos every 24 hours. Interestingly, these genes were not expressed prior to the initial molting around Juvenile 7 d, whereas they were in *R. varieornatus* (Juvenile 4 d). Lack of expression during molting in *H. exemplaris* may imply

that this pathway may not participate in the *H. exemplaris* molting, but in other pathways. On the other hand, the effects of metamorphosis negative regulator juvenile hormones (JH) are well studied (Kayukawa *et al.*, 2017). The downstream gene of JH signaling Krüppel homolog-1 (Kr-h1) negatively regulates E75 (Li *et al.*, 2019). Additionally, expression of Kr-h1 inhibits ecdysone biosynthesis in prothoracic glands, where 20-E precursor ecdysone are produced. These suggests that juvenile hormones inhibit the arthropod molting pathway in both ecdysone production and signal transition. Fenoxycarb, a versatile substance that functions non-dependently against juvenile hormones that are species specific, has been used intensively as a JH analog for pathway inhibition (Gordon *et al.*, 1996; Metwally *et al.*, 2017). Although the conservation of signaling pathway of JH have not been identified for tardigrades, Fenoxycarb functions as a JH analog for a variety of metazoan species, which implies that this may also have an effect on tardigrades. Similarly, we hypothesized that 20-E may function in tardigrades considering that the EcR/USP genes were conserved. To validate if the molting pathway regulates hatching behavior, we evaluated the effect of exposure to the ecdysteroid 20-E and juvenile hormone analog Fenoxycarb. Exposure to high concentrations of Fenoxycarb or 20-E induced un-hatched embryos, suggesting that hormone exposure may have affected embryonic development or hatching. We confirmed that the majority of the un-hatched embryos hatched after hormone removal developed claws, suggesting that exposure to 20-E and Fenoxycarb regulates only hatching and not embryonic development. On the contrary, exposure of both chemicals to *R. varieornatus* did not affect hatching timing. This was not the result of the lack of expression of related genes; the expression of EcR/USP downstream genes were constitutively expressed at moderate levels in the *R. varieornatus* embryo (higher levels prior to the initial molting). These suggests that similar hatching regulation mechanisms may not be conserved in *R. varieornatus*. Additionally, previous studies have suggested that tardigrade embryos may enter a “rest state” induced by environmental perturbations (Altiero *et al.*, 2006; 2010). Although there have been no reports of a rest state induced in *H. exemplaris*, hatching of hatch-inhibited embryo after the removal of hormones is remarkably similar. These observations suggest that the embryo hatching process can be regulated by chemicals, both positively and negatively. The differences between the time of effects between chemicals may be due to the differences of incorporation rates or the expression of receptors. Observation of inhibition of embryo hatch in high concentration *H. exemplaris* exposure suggests that a portion of the chemicals are penetrating the eggshell, of which such concentrations do not affect development but only the hatch regulation mechanism.

Together, these findings suggest that the variation in embryo hatching may suggest an adaptation against their living habitat; xerophilic *R. varieornatus* inhabits a desiccation-prone environment, and hygrophilic *H. exemplaris* is less likely to face environmental desiccation on a regular basis. Therefore, *R. varieornatus* may possess variation in hatching timing in order to avoid loss of entire broods that faces environmental stress close to hatching, therefore hatching regulation by the molting pathway may not be required. On the

contrary, *H. exemplaris* may allow for controlled hatching regulation through internal hormones to enhance proliferation by a more rapid development. However, we have yet to validate if the phenotypic changes to the embryo originates from the chemical exposure; the lack of phenotypic change in low concentration exposure in both species may be result of failure of chemicals to penetrate the eggshell or 20-E/Fenoxycarb may not be the proper chemicals to use as juvenile hormone analogs. Furthermore, our findings are based on only two species from two families, Hypsibiidae and Ramazzotiidae; molecular data on other species or families will be required to validate our hypothesis. We await further studies to confirm if these findings are general to the Tardigrada phylum.

In conclusion, we conducted a comparative transcriptomic study using two tardigrade species to identify pathways induced for hatching regulation, and then validated whether these critical points could be interfered with by chemical exposure. We observed that the 3rd day of the embryo in *H. exemplaris* has a unique transcriptome profile compared to other stages, approximately 1 day prior to hatching, presumably for the strict control of its timing. Mirroring this observation, exposure of eggs to chemicals before the 3rd day influenced hatching only in *H. exemplaris*, which suggests that related pathways may be a critical component of *H. exemplaris* embryo hatching.

---

---

## Chapter 5

---

A novel Mn-dependent peroxidase  
contributes to tardigrade anhydrobiosis

## 5.1 Introduction

Tardigrades are microscopic animals that are renowned for their ability to enter an ametabolic state known as cryptobiosis (Keilin, 1959) or, more particularly, anhydrobiosis (life without water), which is cryptobiosis upon almost complete desiccation. Tardigrades can withstand extreme conditions in this dormant state, including extreme temperature, pressure, high doses of ionizing radiation, and exposure to the vacuum of space (Beltran-Pardo *et al.* 2015; Heidemann *et al.* 2016; Hengherr *et al.* 2009a; Hengherr *et al.* 2009b; Horikawa *et al.* 2013; Jönsson *et al.* 2008; Ono *et al.* 2016), yet they quickly resume life upon rehydration. Anhydrobiosis has been acquired in multiple lineages encompassing all kingdoms of life, but tardigrades are unique in multi-cellular animals that they can enter anhydrobiosis within minutes (Wright, 2001), and that the mechanism does not rely on trehalose and LEA proteins (Banton and Tunnacliffe, 2012; Sakurai *et al.*, 2008). The molecular machinery of tardigrade anhydrobiosis is beginning to be uncovered due to the availability of genomic resources (Hashimoto *et al.*, 2016; Koutsovoulos *et al.*, 2016; Yoshida *et al.*, 2017), leading to the identification of several tardigrade-specific genes, such as CAHS, SAHS, MAHS, LEAM, and Dsup, that have been suggested to play critical roles in cellular protection upon anhydrobiosis (Hashimoto *et al.*, 2016; Tanaka *et al.*, 2015; Yamaguchi *et al.*, 2012). Notably, Dsup is a nucleus-localizing DNA-binding protein that is reported to protect DNA molecules from hydroxyl radicals, where the induction of this single protein in mammalian cells and plants can increase its radiation tolerance (Chavez *et al.*, 2019; Hashimoto *et al.*, 2016; Kirke *et al.*, 2020; Minguez-Toral *et al.*, 2020). However, these proteins are not conserved across the phylum Tardigrada (as they are not conserved in the class Heterotardigrada) (Kamilari *et al.*, 2019), and the necessary and sufficient set of genes and pathways enabling anhydrobiosis remains elusive.

Of the many adverse extremities tardigrades can tolerate in anhydrobiosis, radiation is unique in that tardigrades can better tolerate it in the active hydrated state than in the inactive desiccated state (Horikawa *et al.*, 2008; Jönsson *et al.*, 2016), suggesting the existence of efficient repair pathways in addition to the protective mechanisms identified thus far. Tolerance to radiation in tardigrades is a cross-tolerance of anhydrobiosis (Jönsson, 2019), and the overlapping pathway is presumably the defense against reactive oxygen species (ROS) that mediates protein oxidation and DNA damage (Cadet and Wagner, 2013; Rieger and Chu, 2004). To this end, we employed ultraviolet C (UVC), a low-level energy stressor that causes oxidative stress, to screen for tardigrade-unique components for ROS defense in *Ramazzottius varieornatus*. Tardigrades are capable of tolerating approximately 1,000-fold higher dosages of ultraviolet B (UVB) and UVC than human cell lines (Altiero *et al.*, 2011; Horikawa *et al.*, 2013).

## 5.2 Materials and methods

### 5.2.1 Tardigrade culturing conditions and UVC exposure

In this study, we used the *Ramazzottius varieornatus* strain YOKOZUNA-1. The animals were reared in an environment established by our previous study (Horikawa *et al.*, 2008) and were maintained in 90 mm petri dish plates. Approximately 600 animals were placed on 2.0% Bacto Agar layered dishes and fed with *Chlorella vulgaris* mixed in Volvic water. The petri dish plates were lidded and placed in a dark incubator, set at 22°C. The animals were transferred to new plates every 7 days. Irradiation of the tardigrades was conducted using the UVC lamp of a drying rack (AS ONE) following the protocol in our previous study (Horikawa *et al.*, 2013). Tardigrades were set on a Volvic gel layered in a plastic plate, and excess water was removed. This procedure does not induce anhydrobiosis and is required to minimize UVC absorbance by the layer of water. The plates were then set for exposure to UVC by a UVC lamp. The average dosage was 0.54 mW/cm<sup>2</sup> (0.0054 kJ/m<sup>2</sup>), quantified by UV intensity meter (Fuso, YK-37UVSD). Water was added to the plates immediately after irradiation and set to incubate in a dark chamber set at 22°C until sampling.

### 5.2.2 Quantification of movement in irradiated samples

Ten individuals were exposed to 2.5 kJ/m<sup>2</sup> of UVC radiation and were placed into a single well of the 6x12 plate supplied with 2 µL of 2% agar gel. Five replicates were prepared for UVC exposed and control individuals. A 30 second movie was taken at 0, 1, 2, 3, 6, 9, 12, 18, 24, 36, 48, 72 hours after exposure for each individual with VHX-5000 (Keyence). An individual was classified as moving if any movement was observed within this thirty-second time frame. The quantified movement was statistically compared with a non-irradiated control group by ANOVA and Tukey HSD in R. Conditions with FDR < 0.05 were classified as significant changes.

### 5.2.3 RNA extraction and library preparation

We conducted RNA-Seq on two time-courses with three biological replicates, the first 0-12 hours post exposure (1 h, 2 h, 3 h, 4 h, 5 h, 6 h, 9 h, 12 h), designated “Short”, and the second 0-72 hours post exposure (0h, 12h, 24h, 36h, 48h, 72h), designated “Long”. Non-exposed individuals were sampled at 0 hours without UVC exposure. With each sampling, tardigrades were immersed in 100 µL of TRIzol reagent in a PCR tube (30 animals per sample) and preserved at -80°C. Biomasher II (Nippi. Inc.) was used for the homogenization of tardigrades and 200 µL of TRIzol reagent (Thermo Fisher Scientific) was added to the preservation PCR tube for total RNA extraction with Direct-zol (Zymo research). Total RNA was submitted to mRNA Isolation and RNA fragmentation with NEBNext Oligo d(T)25 beads (NEW ENGLAND BioLabs) and

cDNA synthesis, adapter ligation and PCR enrichment with NEBNext Ultra RNA Library Prep Kit (NEW ENGLAND BioLabs). RNA and cDNA concentrations were determined using Qubit RNA/DNA (Thermo Fisher Scientific), and electrophoresis of the synthesized library was conducted with TapeStation D1000 tape (Agilent Technologies). After pooling all of the samples (70 ng of cDNA per sample), the library was size-selected for fragments between 300-1,000 bp with an E-Gel EX 1% (Thermo Fisher Scientific) and NucleoSpin Gel and PCR Clean-up (Takara). The sequencing library was sequenced with the Illumina Next-Seq 500 instrument (Illumina) with 75 bp single end settings.

#### 5.2.4 Informatics analysis

Prior to informatics analysis, preparation of sequenced reads was conducted. After de-multiplexing sequenced raw files with bcl2fastq (Illumina), quality was checked with FastQC (Andrews, 2015). We then used Kallisto (Bray *et al.*, 2016) v0.44.0 (-b 1000, -bias) to quantify gene expression, using the *R. varieornatus* genome (Hashimoto *et al.*, 2016; Yoshida *et al.*, 2017). To determine differentially expressed genes (DEGs), we mapped the RNA-Seq reads to the coding sequences with BWA MEM v0.7.12-r1039 and conducted statistical test with DESeq2 v1.22.2 (Li and Durbin, 2009; Love *et al.*, 2014). Genes with FDR below 0.05 and fold change over 2 were determined as DEGs. The expression profiles were validated using Spearman correlation in between samples and clustering. DEGs of slow-dried *R. varieornatus* anhydrobiotic samples were obtained from our previous study (Yoshida *et al.*, 2017), and differentially expressed genes in both conditions were obtained.

For the identification of g12777 and g241 orthologs, we first prepared published transcriptomes and our in-house tardigrade genome database (Arakawa, Personal communication). We obtained transcriptome assemblies of *Richtersius coronifer* and *Echiniscoides sigismundi* from previous studies (Kamilari *et al.*, 2019). Additionally, we re-assembled each transcriptome using Bridger v2014-12-01 (Chang *et al.* (2015), *R. coronifer* : SRR7340056, *E. sigismundi* : SRR7309271, default parameters). Each tardigrade genome/transcriptome was submitted to BUSCO genome completeness validation against the eukaryote lineage (Simao *et al.*, 2015). The gene model obtained during this assessment was used with autoAugPred.pl script in the Augustus v3.3.3 package to predict genes in each genome (Keller *et al.*, 2011). The amino acid sequences of each genome were submitted to a TBLASTX or BLASTP search v2.2.2 (Altschul *et al.*, 1997) of g12777.t1 and g241.t1 coding sequences and amino acid sequences to obtain orthologs in each species. Additionally, the g241 amino acid sequence was submitted to a Diamond BLASTP v0.9.24.125 search against NCBI NR and NCBI Bacteria RefSeq to obtain bacterial orthologs (Buchfink *et al.*, 2015; O'Leary *et al.*, 2016). Furthermore, we submitted predicted proteome sequences for each tardigrade species and additional metazoan species used in our previous study (Yoshida *et al.*, 2017) to OrthoFinder v2.3.7 (mcl-I 2.0) to validate our BLASTP search results (Emms and Kelly, 2015). For g12777 orthologs, amino

acid sequences were subjected to MAFFT v7.271 (auto) multiple sequence alignment and phylogenetic tree construction by FastTree v2.1.8 SSE3 (1000, -gamma) (Kato and Standley, 2013; Price *et al.*, 2010), which was visualized in iTOL (Letunic and Bork, 2019). Orthologs were also submitted to MEME analysis v5.1.0 for a novel motif search (Bailey *et al.*, 2009). G-language Genome Analysis Environment v1.9.1 was used for data manipulation (Arakawa *et al.*, 2003; Arakawa and Tomita, 2006).

### 5.2.5 Transmission electron microscope observation

Anhydrobiosis was induced by placing *R. varieornatus* in 37% RH and 0% RH for 24 hours each following our previous study (Yoshida *et al.*, 2017). Sample preparation and transmission electron microscope (TEM) imaging was conducted at Tokai Electronic Microscope Analysis. Samples were sandwiched between copper disks and rapidly frozen in liquid propane at  $-175^{\circ}\text{C}$ . Frozen samples were freeze-substituted with 2% glutaraldehyde, 1% tannic acid in ethanol and 2% distilled water at  $-80^{\circ}\text{C}$  for 2 days,  $-20^{\circ}\text{C}$  for 3 h, and  $4^{\circ}\text{C}$  for 4 h. For dehydration, samples were incubated in ethanol for 30 minutes three times, and overnight at room temperature. These samples were infiltrated with propylene oxide (PO) twice for 30 minutes each and were put into a 70:30 mixture of PO and resin (Quetol-812; Nisshin EM Co.) for 1 h. PO was volatilized by keeping the tube cap open overnight. These samples were further transferred to fresh 100% resin and polymerized at  $60^{\circ}\text{C}$  for 48 hours before being ultra-thin sectioned at 70 nm with a diamond knife using a ultramicrotome (ULTRACUT, UCT, Leica), after which sections were placed on copper grids. Sections were stained with 2% uranyl acetate at room temperature for 15 minutes and rinsed with distilled water. The sections were further secondary-stained with Lead stain solution (Sigma-Aldrich) at room temperature for 3 minutes. The grids were observed under a transmission electron microscope (JEM-400Plus1 JEOL Ltd.) at an acceleration voltage of 100 kV. Digital images (3,296 x 2,472 pixels) were taken with a CCD camera (EM-14830RUBY2; JEOL, Ltd.). The images were analyzed for mitochondrial size using ImageJ and area and circularity values were tested with ANOVA and Tukey HSD in R.

### 5.2.6 Sample preparation of recombinant proteins

For biochemical and biophysical characterizations of the g12777 protein with deletions of the N-terminal signal sequence and putative intrinsically disordered region, the gene encoding residues Gly63–Leu231 were cloned into the *Nde*I and *Eco*RI sites of the pET-28b vector (Novagen - Merck Millipore). *Escherichia coli* BL21-CodonPlus(DE3) harboring the plasmids were cultured in LB medium containing 15  $\mu\text{g/ml}$  kanamycin and subsequently harvested after induction with 0.5 mM isopropyl  $\beta$ -D-thiogalactoside for 4 h at  $37^{\circ}\text{C}$ . The harvested cells were resuspended with buffer A [20 mM Tris-HCl (pH 8.0), 150 mM NaCl, and 1 mM EDTA] and lysed by sonication. The insoluble inclusion bodies were extensively washed with buffer A in

the presence and absence of 2% Triton X-100, and then solubilized with 6 M guanidinium chloride, 50 mM Tris-HCl (pH 8.0), and 0.5 mM dithiothreitol. The solubilized proteins were refolded by dilution (0.05 mg/mL) in 50 mM Tris-HCl (pH 8.0), 400 mM L-arginine, 5 mM reduced glutathione, 0.5 mM oxidized glutathione, and 2 mM CaCl<sub>2</sub> at 4°C for 12 h. The refolded protein was concentrated and then dialyzed against 20 mM Tris-HCl (pH 8.0), 150 mM NaCl, and 2 mM CaCl<sub>2</sub>. Subsequently, the N-terminal His<sub>6</sub>-tag peptide was removed by thrombin digestion. The nontagged proteins were incubated at 4°C for 30 min in the presence of 10 mM EDTA and 0.1 mM AEBSF [4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride], and further purified with a HiLoad Superdex 75 pg (GE Healthcare) equilibrated with buffer A.

### 5.2.7 Crystallization, X-ray data collection and structure determination

The catalytic domain of g12777 protein (10 mg/mL) was dissolved in 20 mM Tris-HCl (pH 8.0), 150 mM NaCl and 2 mM CaCl<sub>2</sub>. The crystals of Zn<sup>2+</sup>-bound g12777 protein complex were obtained in a precipitant containing 10% PEG3350, 100 mM imidazole-HCl buffer (pH 7.5), 300 mM zinc acetate, and 50 mM sodium fluoride upon incubation at 20°C for 3 days. In contrast, the crystals of Mn<sup>2+</sup>-bound complex were obtained by a soaking with a buffer containing 11% PEG3350, 100 mM imidazole-HCl buffer (pH 7.5), 50 mM manganese chloride, and 50 mM sodium fluoride for 5 h using the Zn<sup>2+</sup>-bound crystals. The crystals were cryoprotected with the crystallization or soaking buffer supplemented with 20% glycerol. The crystals of Zn<sup>2+</sup>- and Mn<sup>2+</sup>-bound g12777 protein complexes belonged to space groups P1 and P21 with two g12777 protein molecules per asymmetric unit and diffracted up to resolutions of 1.60Å and 2.30Å, respectively. Diffraction data were scaled and integrated using XDS GPLv2 (Kabsch, 2010) and AIMLESS v1.12.1 (Evans, 2011).

The crystal structure of the catalytic domain of the g12777 protein crystallized in the presence of excess amount of Zn<sup>2+</sup> ion (300 mM) was solved by the single-wavelength anomalous dispersion (SAD) method using 1.1200 wavelength (Aichi-SR BL2S1) with the program Autosol in the Phenix suite v1.11.1-2575 (Adams *et al.*, 2010). As for the Mn<sup>2+</sup>-bound g12777 protein complex, the crystal structure was solved by the molecular replacement method using the Zn<sup>2+</sup>-bound structure as a search model. The diffraction data set was collected at Osaka University using BL44XU beamline at SPring-8 (Japan). Automated model building and manual model fitting to the electron density maps were performed using ARP/wARP v8.0 (Langer *et al.*, 2008) and COOT v0.9 (Emsley *et al.*, 2010), respectively. REFMAC5 v5.8.0258 (Murshudov *et al.*, 1997) was used to refine the crystal structure, and the stereochemical quality of the final models were validated using MolProbity v4.2 (Chen *et al.*, 2010), showing that no amino acids were located in the disallowed regions of the Ramachandran plot. The final model of the Zn<sup>2+</sup>-bound g12777 protein complexes had Rwork of 19.2 and Rfree of 24.5%, whereas that of Mn<sup>2+</sup>-bound form had Rwork of 26.2 and Rfree of 32.5% (Supplementary Data S17). The molecular graphics were prepared using PyMOL

v2.4.0a0 (<https://pymol.org/2/>). These structures were compared with known protein structures using the DALI server (Holm, 2020).

### 5.2.8 Measurement of enzymatic activity of the g12777 protein

The catalase/oxidase activities of the g12777 protein were measured using Cayman Catalase Assay Kit (Cayman Chemical) following the manufacturer's protocol. For the enzymatic assay, 20 ng of the wild-type g12777 protein catalytic domain was incubated in the presence and absence of 2 mM ZnCl<sub>2</sub> and MnCl<sub>2</sub> for 30 min at 25°C. Bovine liver catalase (Sigma Aldrich) was used as a positive control. Catalase/oxidase activity were tested with ANOVA and Tukey HSD in R.

### 5.2.9 Isothermal titration calorimetry

Purified catalytic domain of the g12777 protein dissolved in 20 mM Bis-Tris-HCl (pH 6.5) containing 150 mM NaCl was used for isothermal titration calorimetry (ITC). In this experiment, a syringe containing 1 mM ZnCl<sub>2</sub>, MnCl<sub>2</sub> or CaCl<sub>2</sub> was titrated into a sample cell containing 0.1 mM catalytic domain of the g12777 protein using an iTC200 calorimeter (GE Healthcare).

### 5.2.10 Expression of GFP-tagged g12777 in HEK293

For expression of GFP-fused g12777 proteins, we inserted the coding sequence between the *SalI* and *BamHI* restriction sites of the pAcGFP1-N1 plasmid (Takara). The total RNA was extracted from *R. varieornatus* with Direct-zol Ultra RNA (Zymo Research) and was reverse transcribed with PrimeScript Reverse Transcriptase (Takara). g12777 mRNA was selected by PCR using Tks Gflex DNA Polymerase (Takara) with the following primers:

g12777-F: 5'-ATTGTCGACATGGCATTATCTTTGTGGATGACTG-3',

g12777-R: 5'-TAAGGATCCTTTAGGGAAGAAGGTGCCCGACAG-3'.

For construction of Δ62aa construct, the forward primer was changed to:

5'-ATTGTCGACGGCCGCTTTGCCGATTTCTTCAGAA-3'.

The corresponding g12777-D2A (D92A, D98A, D161A, D163A mutations) coding sequences were adapted for human codon usage and synthesized at Eurofin Genomics. g12777-D2A was inserted into the pAcGFP1-N1 plasmid (Takara) between *SalI* and *BamHI*. Constructs were transfected into HEK293 cells with the X-tremeGENE 9 reagent (Sigma-Aldrich) and submitted to selection with G418 (Sigma-Aldrich) at 400 μg/mL for more than two weeks, and further passaged at 100 μg/mL. HEK293 cell lines were cultured with Minimum Essential Media (MEM) medium (Sigma-Aldrich) supplied with FBS (Funakoshi), NEAA (Thermo Fisher Scientific) and Antibiotic-Antimycotic Mixed Stock Solution (Nacalai Tesque) and were

passed every 3-4 days with TryPLE (Thermo Fisher Scientific). For microscopy observations, cells were fixed with 4% Paraformaldehyde (Wako, in phosphate-buffered saline (PBS)), and co-stained with 300 nM 4',6-diamidino-2-phenylindole (DAPI, Thermo Fisher Scientific) and 20 µL of CellLight Golgi RFP, BacMan 2.0 (Thermo Fisher Scientific) to visualize the nucleus and Golgi apparatus, respectively. Fluorescent signals were observed under the SZ-5000 (Keyence) at 40x or 100x magnification.

Additionally, the localization of six high expressed g12777 orthologs were validated using similar methods. cDNA was obtained from *R. varieornatus* with RNeasy Mini Kit (Qiagen) and PrimeScript II 1st strand cDNA Synthesis Kit (Takara). The coding sequences of the corresponding genes were amplified by PrimeSTAR Max DNA Polymerase (Takara) using the following primers: g243-F: 5'-ATTCTGCAGTCGACGGTATGCAGGATGTTTCCGA-3',  
g243-R: 5'-catgaccgggtggatcCTATTTGGTGAACCATCCG-3',  
g244-F: 5'-ATTCTGCAGTCGACGATGGCCAAGGCGGCAATC-3',  
g244-R: 5'-catgaccgggtggatcTCATCTAGAGAAAAGTCCGC-3',  
g245-F: 5'-ATTCTGCAGTCGACGGTATGAACTTTCTCTGCTGG-3',  
g245-R: 5'-catgaccgggtggatcTCAGGAGAACATGCCCGA-3',  
g246-F: 5'-ATTCTGCAGTCGACGATGATGCAGCTGACAATCTT-3',  
g246-R: 5'-catgaccgggtggatcTTATGAGAACATGCCGTCG-3',  
g779-F: 5'-ATTCTGCAGTCGACGGTATGGATCTGGACAGGG-3',  
g779-R: 5'-catgaccgggtggatcTCATTTTCCAATAAAGGGAATG-3',  
g2856-F: 5'-ATTCTGCAGTCGACGGTATGTGGGAATACTGTG-3',  
g2856-R: 5'-catgaccgggtggatcTCATGCCATAGCGTGCG-3',  
g12777-F: 5'-ATTCTGCAGTCGACGGTATGGCATTATCTTTGTGG-3',  
g12777-R: 5'-catgaccgggtggatcTTATAGGGAAGAAGGTGCC-3',

These coding sequences were inserted into the pAcGFP1-N1 plasmid by In-Fusion HD Cloning Kit. Constructs were transfected into HEK293 cells with the X-tremeGENE 9 reagent (Sigma Aldrich) in opti-MEM solution for 4 hours. After incubation for 24-48 hours in MEM medium, cells were co-stained with Hoechst33342 (Thermo Fisher Scientific) and CellLight Golgi-RFP (Thermo Fisher Scientific) and observed under the Leica SP8 microscope with x100 magnification.

### 5.2.11 Validation of cellular tolerance against oxidative stress

To evaluate cellular tolerance of g12777 transfected cells, we conducted MTT analysis of H<sub>2</sub>O<sub>2</sub> exposed cell lines expressing only GFP, g12777, g12777-Δ62aa, and g12777-D2A (3 technical replicates). In brief, approximately 10,000 cells were plated into 96 well plates and were incubated at 37°C at 5% CO<sub>2</sub> for 24 hours. Cells were exposed to 0-1.0 mM H<sub>2</sub>O<sub>2</sub> (Wako) or 0-64.0 µM Bleomycin (Funakoshi) for 30 minutes

(3-5 replicates). The culture medium was removed to stop exposure, and 100  $\mu\text{L}$  MEM medium was added to the wells and were further incubated for 24 hours. The culture medium was replaced with 100  $\mu\text{L}$  fresh culture medium supplied with 10  $\mu\text{L}$  of 5  $\mu\text{g}/\mu\text{L}$  Thiazolyl Blue Tetrazolium Bromide (Sigma-Aldrich) and incubated for 2 hours for formazan formation. The culture medium was removed, and 100  $\mu\text{L}$  DMSO (Wako) was added to melt the formazan. The 96 well plate was measured with 570 nm and 670 nm using SpectraMax PLus 384 (Molecular Devices). Three technical replicates were measured. Absorbance values were calculated by  $\text{Ab}570 - \text{Ab}670 - \text{Ab-BLANK}$ . Absorbance values were statistically tested with ANOVA and Tukey HSD in the R program.

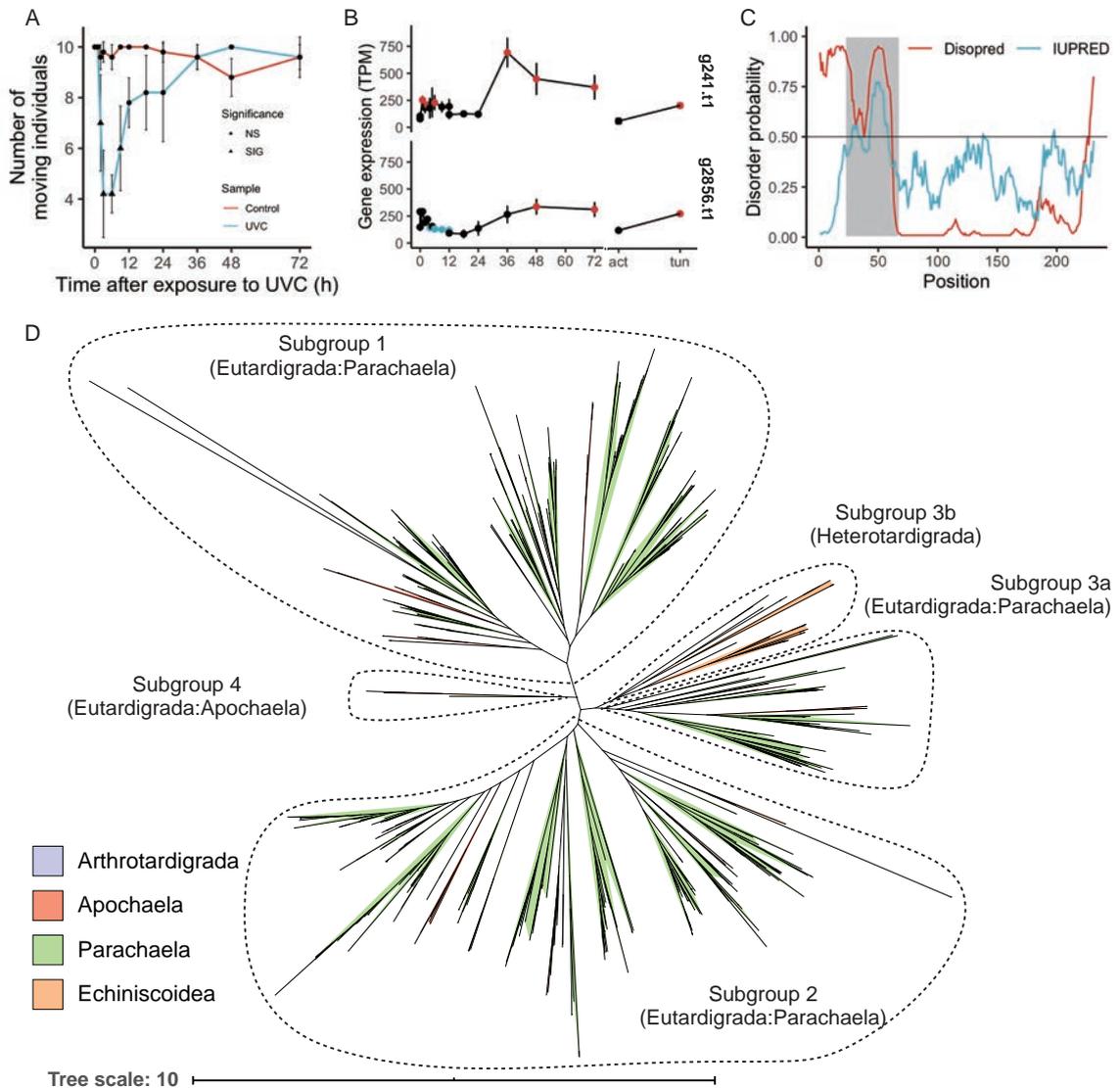
For MACS flow cytometry, cells were cultured and exposed to  $\text{H}_2\text{O}_2$  similar to the MTT assay protocol. After 24-hour incubation after exposure, the MEM medium (100  $\mu\text{L}$ ) was moved to a clean well and each well were washed with 100  $\mu\text{L}$  PBS (-). All of the PBS were moved to the corresponding well. 50  $\mu\text{L}$  TryPLE was added for trypsinization and incubated at 37°C in the  $\text{CO}_2$  incubator for 4 minutes. The MEM culture and PBS mixture were moved to their original wells and mixed thoroughly to release the cells. The 96 well plate was centrifuged at 2,000 rpm (750g) for 6 minutes at 4°C, and the supernatant was removed. Each well was washed with 1x Annexin binding buffer (Thermo Fisher Scientific), and was centrifuged at 2,000 rpm (750g) for 6 minutes at 4°C. The buffer was removed, and each well was supplied with 100  $\mu\text{L}$  of Annexin binding buffer supplied with 0.1  $\mu\text{L}$  SYTOX Blue Nucleic Acid Stain (Thermo Fisher Scientific) and 2  $\mu\text{L}$  Annexin V Alexa Fluor 657 (Thermo Fisher Scientific). The cells were resuspended with pipetting and was set to stain for 10 minutes on ice. The MACSQuant Analyzer 10 instrument (Miltenyi Biotec) was set for measurement of DAPI, Alexa 657 and GFP. Three technical replicates were measured. Approximately 100-1,500 cells were measured for each cell line. The ratio of healthy cells (Annexin-V (-) SYTOX blue (-) cells) were compared with ANOVA and Tukey HSD in the R program. Adjusted  $p$ -values (FDR) below 0.05 were designated as significant differences.

## 5.3 Results and discussions

### 5.3.1 Transcriptome sequencing of UVC exposed *R. varieornatus*

Previous studies have observed that *R. varieornatus* exposed to 2.5  $\text{kJ}/\text{m}^2$  UVC showed a prolonged decrease in movement for approximately 1 day, which presumably is the critical period for the ROS response. We first validated this observation on a finer time scale, where individuals who were exposed to 2.5  $\text{kJ}/\text{m}^2$  showed significantly lower movements from 2-9 hours after exposure (Figure 5.1a). We then conducted transcriptome profiling from 0-12 hours and 0-72 hours to screen for genes induced in this period (Supplementary Data S18). Initial clustering of expression values using Spearman correlation indicated that the transcriptome profiles

drastically shifted between 3-4 and 24-36 hours post exposure (Supplementary Figure S16), suggesting consistency with the motility of the animals. We found a total of 3,324 differentially expressed genes following exposure to UVC (DESeq2, FDR<0.05), of which 1,314 and 2,110 genes to be up-regulated and down-regulated, accordingly (Supplementary Figure S16). Genes with high fold change (>4) were comprised of various genes, including chaperones (Mitochondrial chaperone BCS1), DNA damage repair pathways (XRCC4, PARP), metalloproteases (NAS-13), anti-oxidative pathways (GST), and previously identified tardigrade specific protection-related genes (CAHS, SAHS). Interestingly, DEGs that had high expression values included several of those with high fold change, additionally anti-oxidative stress genes thioredoxin and peroxiredoxin (Supplementary Data S19, S20). These genes were induced during tardigrade anhydrobiosis (Yoshida *et al.*, 2017), suggesting similar pathways are being regulated between desiccation and UVC exposure. Additionally, we found that the zinc metalloprotease NAS, apolipoproteins, autophagy-related sequestosome, and the mitochondrial chaperone BCS1 were highly expressed (TPM>1,000) after exposure (Supplementary Figure S17, Supplementary Data S21).



**Figure 5.1: Identification of a novel stress responsive gene family conserved within Tardigrada phylum.**

(A) *R. varieornatus* specimens were exposed to 2.5 kJ/m<sup>2</sup> of UVC and movements were observed at each time point. Dot shapes indicate significant differences (Tukey HSD test,  $p$ -value < 0.05). (B) Expression values of two tardigrade-specific genes with no known annotations. (C) Probability plot of disordered regions in g12777 predicted by DISOPRED or IUPRED. Disordered regions predicted are indicated with a gray highlight. (D) Phylogenetic tree of g12777 orthologs in Tardigrada. Each ortholog was colored by the corresponding lineage. Four subfamilies were found and named according to the major lineage. Recreated from Yoshida *et al.* (2020).

### 5.3.2 Identification of the g12777 gene family as a novel stress responsive gene family

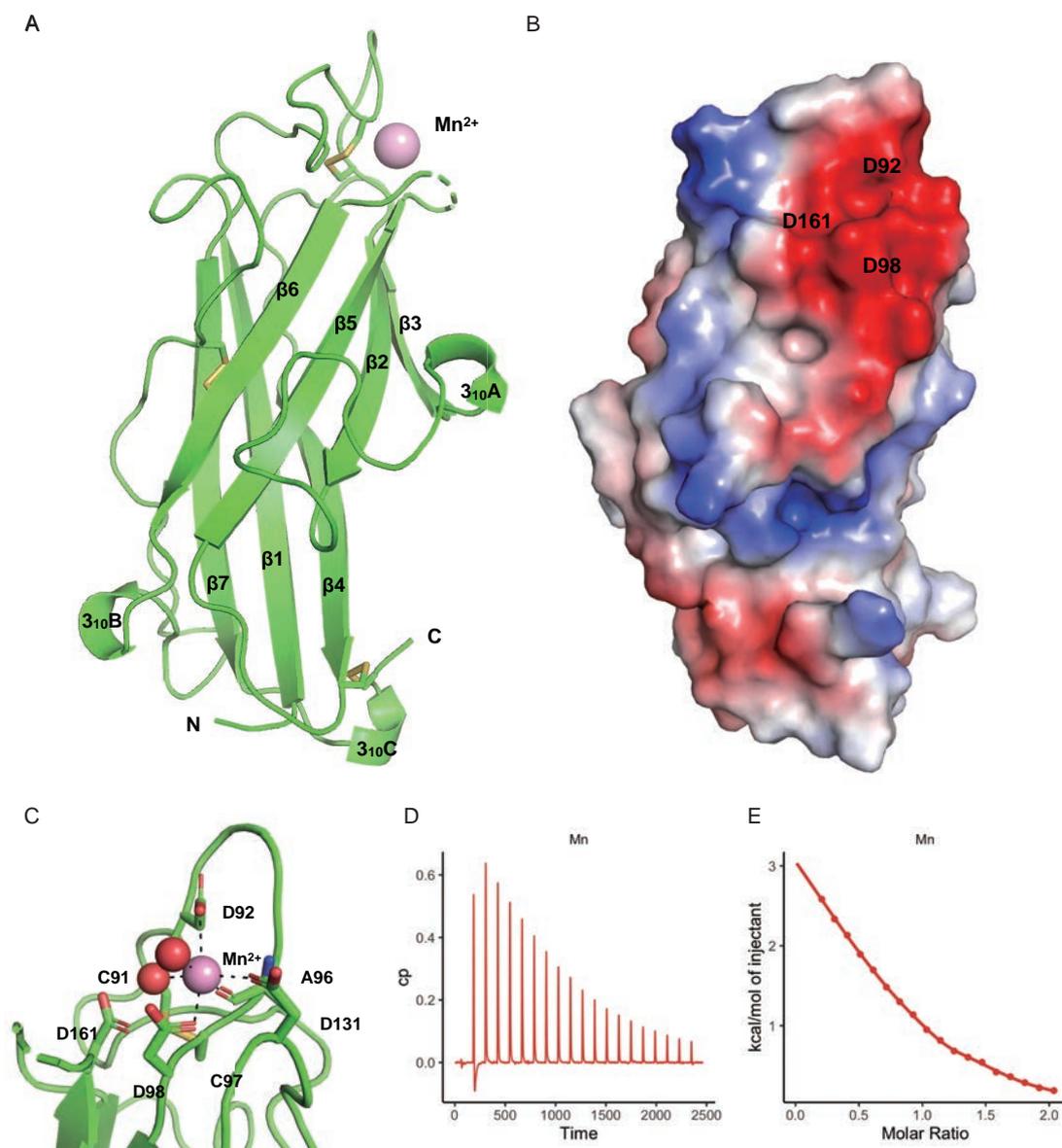
To screen for genes responsible for the cross-tolerance of anhydrobiosis and UVC exposure, we analyzed the intersection of DEGs in the above result and our previous differential transcriptome analysis during slow desiccation. We found 141 genes that were upregulated in both conditions (Supplementary Data S22) that were significantly enriched in Gene Ontology terms related to antioxidative stress (e.g., glutathione transferase activity, superoxide dismutase activity, *etc.*, Supplementary Data S23), and these genes also included the previously identified tardigrade-specific heat soluble proteins CAHS and SAHS. Seventy-five of these genes were hypothetical genes, and only two (g2856/RvY\_14843, g241/RvY\_00334) contained no known functional domains (InterProScan, CDD, Pfam-A, Superfamily, SMART, Figure 5.1b, Supplementary Data S24, S25, S26, S27, S28, S29, S30). The g241 gene showed similarity with bacterial genes, and phylogenetic analysis of this gene suggested a possible horizontal gene transfer event at the early stages of the Tardigrada lineage (Supplementary Figure S18, Supplementary Data S31, S32, S33, S34). The remaining g2856 was strongly multiplied within the *R. varieornatus* genome (total of 35 copies), and several of the orthologs were duplicated in tandem (Supplementary Data S35) and were highly conserved throughout the phylum Tardigrada, including in nonanhydrobiotic species of Heterotardigrada. Phylogenetic analysis of g2856 orthologs indicated that there are 4 subgroups (Figure 5.1c), where a single subfamily was comprised of Heterotardigrada species. Within the 35 copies of the g2856 gene family, the g12777 gene was found to have 2.5-fold induction of expression during slow desiccation of *R. varieornatus*. Additionally, informatics-based analysis predicted a signal peptide and a disordered region in the N-terminus of this protein (Figure 5.1d), similar to various tardigrade-specific anhydrobiosis-related genes. These characteristics suggested that this gene may play an important role in tardigrade anhydrobiosis; therefore, we submitted this g12777 gene for further functional analysis.

### 5.3.3 g12777 has a highly conserved Mn<sup>2+</sup> binding site

We crystallized the putative globular domain that lacked the N-terminus 62 amino acids of g12777 protein and solved the crystal structures as two forms containing Mn<sup>2+</sup> or Zn<sup>2+</sup> ions at 2.30Å and 1.60Å resolutions, respectively (Figure 5.2a, Supplementary Figure S19A,B,C,D, Supplementary Data S17, Supplementary Data S36). These crystallographic data revealed that g12777 protein possesses a  $\beta$ -sandwich domain sharing a common Mn<sup>2+</sup> and Zn<sup>2+</sup> binding site. In this site, a Mn<sup>2+</sup> ion was coordinated by three aspartic acids with high electrostatic surface potential (Figure 5.2b and 5.2c). In the Mn<sup>2+</sup>-binding site located at the negatively charged patch comprising  $\beta$ 1 -  $\beta$ 2 and  $\beta$ 3 -  $\beta$ 4 loops, the side-chain oxygen atoms of Asp92, Asp98, and Asp131 and the main-chain carbonyl atom of Ala96 and two water molecules coordinate with Mn<sup>2+</sup> ion at distances of 2.2–2.9Å (Figure 5.2c). As for the Zn<sup>2+</sup>-binding site located at the almost same region as that

of  $\text{Mn}^{2+}$ , the side-chain oxygen atoms of Asp92, Asp98, Asp161, and Asp163 coordinate with  $\text{Zn}^{2+}$  ion at distances of 1.9–2.0 Å (Supplementary Figure S19D). Considerable conformational difference was observed between  $\text{Mn}^{2+}$ - and  $\text{Zn}^{2+}$ -bound forms in terms of the metal-binding site, in which Asp92 and Asp98 were commonly involved in  $\text{Mn}^{2+}$  and  $\text{Zn}^{2+}$  binding (Figure 5.2c and Supplementary Figure S19D). These aspartic acid residues were highly conserved within tardigrade orthologs (Supplementary Figure S19E,F), and the highly conserved CD-CD motif containing two Asp residues (D92 and D98, Supplementary Figure S19E,F,G) was used in the metal ion binding in both structures, suggesting the importance of these residues. The cysteine residues in this motif formed a disulfide bond in the crystal structure, suggesting that this disulfide bond may also contribute to this protein's function. We also observed another region that was highly conserved (W194 and R216, Supplementary Figure S19E,F), suggesting that another functional site may be present.

ITC experiment demonstrated that the estimated dissociation constants of catalytic domain of g12777 and  $\text{Zn}^{2+}$ ,  $\text{Mn}^{2+}$ , and  $\text{Ca}^{2+}$  were  $1.92 \times 10^{-6} \pm 6.00 \times 10^{-8}$  M,  $2.42 \times 10^{-5} \pm 1.75 \times 10^{-6}$  M, or  $1.39 \times 10^{-4} \pm 3.02 \times 10^{-5}$  M, respectively, indicating that the binding affinity of  $\text{Zn}^{2+}$  was the highest among these divalent cations (Figure 5.2d, Supplementary Figure S19G, Supplementary Data S37). Importantly, the binding stoichiometry ( $n$ ) of  $\text{Zn}^{2+}$ ,  $\text{Mn}^{2+}$ , and  $\text{Ca}^{2+}$  were  $1.00 \pm 0.00$ ,  $0.77 \pm 0.02$ , and  $0.80 \pm 0.17$ , indicating that catalytic domain of g12777 binds to the one metal ions in solution (Figure 5.2d, Supplementary Figure S19G, Supplementary Data S37).



**Figure 5.2: g12777 protein has a SOD-like -sandwich fold and binds to  $\text{Mn}^{2+}$  ion.**

(A) **The crystal structure of the catalytic domain of the g12777 protein.** Bound  $\text{Mn}^{2+}$  ion and residues involved in disulfide bond formation are indicated with a pink sphere and stick, respectively. Positions of the N and C termini are indicated as letters. (B) **Electrostatic surface potential of g12777 protein.** The surface model of g12777 is colored according to the electrostatic surface potential (blue, positive; red, negative; scale from  $-50$  to  $+50$  kT/e). (C) **Close-up view of the  $\text{Mn}^{2+}$ -binding site.** The binding site is comprised from three aspartic acid residues (D92, D98, and D131) and a disulfide bond with close proximity (C91 and C97). (D,E)  **$\text{Mn}^{2+}$  binding affinity measured by isothermal titration calorimetry.** (D) Raw data from ITC measurement, in seconds. (E) integrated heat values corrected for the heat of dilution and fit to a one-site binding model (solid line). Recreated from Yoshida *et al.* (2020).

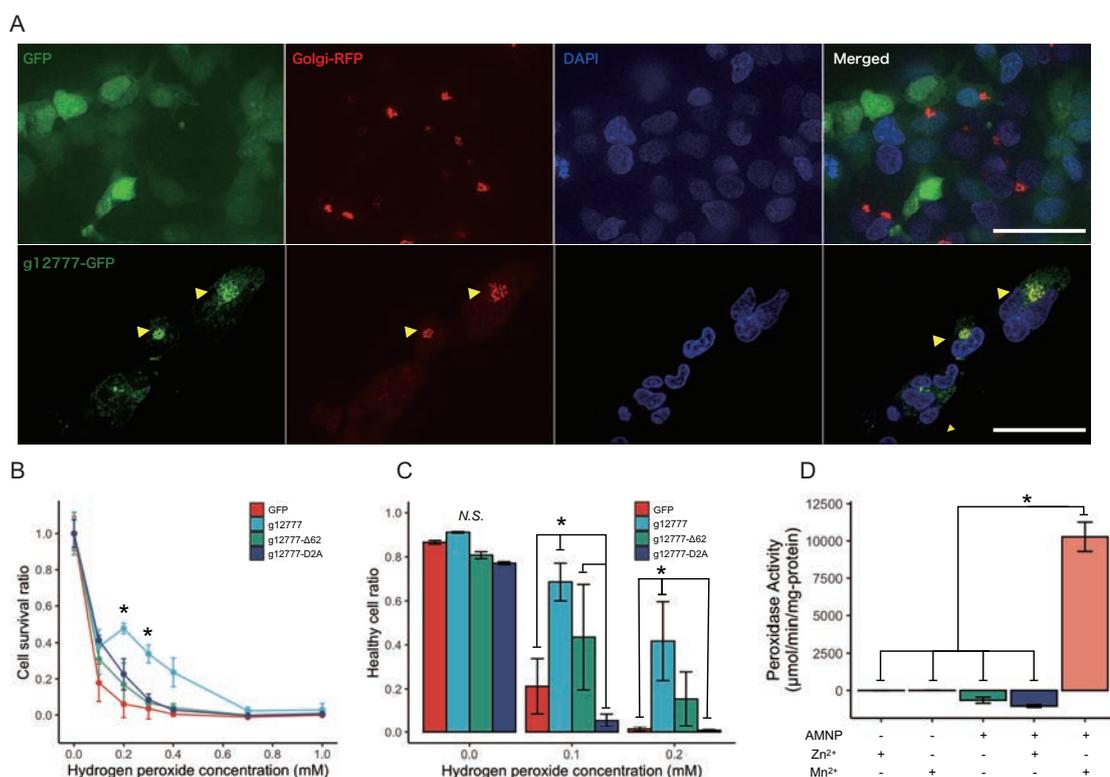
### 5.3.4 g12777 is a Mn<sup>2+</sup> dependent peroxidase

To validate where this protein functions in cultured cells, we constructed a GFP-tagged recombinant g12777 protein in the pAcGFP1-N1 plasmid and expressed these proteins in the HEK293 cell line. Costaining with DAPI and CellLight GolgiRFP suggested that this protein localizes in the Golgi apparatus (Figure 5.3a). Six of the most highly expressed orthologs of the g12777 family of proteins (highly expressed in the UVC time course and differentially expressed during slow-dry anhydrobiosis) also showed Golgi localization (Supplementary Figure S20). To validate this protein's antioxidative stress functions, we subjected cells expressing GFP-tagged g12777 proteins to H<sub>2</sub>O<sub>2</sub> exposure. The transfected cells showed increased survival at 0.2-0.3 mM H<sub>2</sub>O<sub>2</sub> as measured by MTT assays (Figure 5.3b). Similar results were obtained (0.1-0.2 mM H<sub>2</sub>O<sub>2</sub>) by flow cytometry measurements (Annexin V-negative + SYTOX Blue-negative cells, Figure 5.3c). Substitution of the Asp residues comprising Zn-1 to alanine (g12777-D2A) showed a decrease in cell survival (Figure 5.3b, Figure 5.3c), indicating that this Mn-1 binding site is required for antioxidative stress function. We then assayed whether recombinant proteins also have an antioxidative stress function. Recombinant  $\beta$ -sandwich domain of g12777 protein showed significant peroxidase activity when Mn<sup>2+</sup> ions but not Zn<sup>2+</sup> ions were present (Figure 5.3d). The peroxidase activity of g12777 in Mn<sup>2+</sup> conditions were 1/20 of that of bovine liver catalase (Supplementary Figure S21). Removal of the signal peptide and disorder region did not significantly reduce tolerance against oxidative stress (Figure 5.3b, Figure 5.3c), suggesting that the disordered region does not directly contribute to the anti-oxidative stress function.

These findings suggest that this protein is a manganese-dependent peroxidase, independent of other components of the cell. Hence, we named this gene family Anhydrobiosis-related Mn-dependent Peroxidase (AMNP). We hypothesize that the presence of Mn-1 and the disulfide bond at the active site is suggestive of the utilization of both mechanisms for peroxidase activity, which could categorize AMNP as a new class of peroxidase. Constructs lacking the disordered region showed anti-oxidative stress functions. Intrinsically disordered proteins (IDP) are implied to contribute to cellular tolerance in tardigrades (Tanaka *et al.*, 2015; Yamaguchi *et al.*, 2012) and *Drosophila* (Tsuboyama *et al.*, 2020). IDPs have been proposed to participate in protein stability (van der Lee *et al.*, 2014), therefore we hypothesize this disordered region may contribute to protein stability of AMNP proteins. We have also observed that the AMNP gene utilizes Mn<sup>2+</sup> ions for peroxidase function, while it has a higher affinity for Zn<sup>2+</sup> ions. We speculate that this affinity is related to the localization of this protein in the Golgi apparatus, since the Golgi apparatus has a strictly controlled high manganese concentration (Alejandro *et al.*, 2017; Carmona *et al.*, 2010; Hirata, 2002; Reddi *et al.*, 2009), suggesting the antioxidative stress functionality in the Golgi seems to be crucial for anhydrobiotic survival. The importance of the Golgi apparatus in stress response and regulation has been studied (Alborzinia *et al.*, 2018; Jiang *et al.*, 2011; Yoshida, 2009), but its contribution to anhydrobiosis has not been discussed in detail to date. We believe that the results of this work can be a starting point for a more comprehensive study

of the intracellular mechanisms of anhydrobiosis.

In conclusion, our results provide a global image of the transcriptomic response against UVC in an extremophile tardigrade and have showed that the oxidative stress response partly comprised of the novel  $Mn^{2+}$ -dependent peroxidase family is one of the central mechanisms in tardigrade anhydrobiosis. Moreover, AMNP is the first tardigrade-specific anhydrobiosis-related gene that is conserved throughout the phylum, in contrast to previously identified tardigrade-specific elements that are conserved in only Eutardigrada at most. This finding underlies a fundamental basis of the cellular protection against oxidative stress during anhydrobiosis and the contribution of the Golgi apparatus.



**Figure 5.3: g12777 is a Golgi apparatus localizing protein that enhances anti-oxidative stress.**

(A) **Localization of g12777 protein in HEK293 cells.** The g12777 protein was fused to GFP and transfected to HEK293 cells. Cells were co-transfected with CellLight Golgi-RFP and DAPI. Yellow arrows indicate the co-localization of g12777 and Golgi-RFP. Scale bar 25μm. (B) **Tolerance of g12777-expressing HEK293 cells to H<sub>2</sub>O<sub>2</sub> measured by MTT assay.** Cells transfected with g12777-GFP were exposed to hydrogen peroxide for 30 minutes and subjected to MTT assay after 24 hours incubation. Only the full-length g12777-GFP show an increase in survival at around 0.2-0.3 mM. ANOVA + Tukey HSD (\* FDR < 0.05). (C) **Tolerance of g12777-expressing HEK293 cells to H<sub>2</sub>O<sub>2</sub> measured by flow cytometry.** Cells transfected with g12777-GFP were exposed to hydrogen peroxide for 30 minutes and subjected to MACS flow cytometry to detect SYTOX blue and AnnexinV-Alexa 657 fluorescence after 24 hours incubation. Only full-length g12777-GFP shows an increase in survival. ANOVA + Tukey HSD (\* FDR < 0.05). (D) **Peroxidase activity of the recombinant protein lacking the N terminal 62 aa region.** Peroxidase function was present only when manganese ions were present. Error bars indicate standard errors. Recreated from Yoshida *et al.* (2020).

---

---

## Chapter 6

---

Final remarks

## 6.1 Establishing a high-quality genome of microscopic invertebrates

Anhydrobiosis is a fascinating physiological phenomenon, particularly in tardigrades, as they can tolerate complete desiccation at any point in their life cycle and are capable of tolerating various extreme environments. However, the lack of genetic information and experimental methods, as well as the difficulty in rearing these species, have made research on tardigrades problematic. Therefore, others and I first designed a method of sequencing the genome or transcriptome from a single individual (Arakawa *et al.*, 2016; Yoshida *et al.*, 2018), eliminating bacterial contamination. Combining this method with state-of-the-art long-read sequencing methods resulted in a high-quality genome of the hydrobiotic *H. exemplaris* (Yoshida *et al.*, 2017), free of the contaminants and heterozygotic regions that hampered previous attempts (Arakawa *et al.*, 2016; Boothby *et al.*, 2015; Koutsovoulos *et al.*, 2016). Using this genome, I compared the genome content and gene regulation of two tardigrade species, first conducting a genome-based, multi-replicate transcriptome analysis. I obtained approximately 1,400 candidate transcript sequences of which over half were novel genes, which represent targets for functional analysis of anhydrobiosis. These candidates were screened using transcriptome sequencing data from various developmental stages and the response to UVC exposure (Yoshida *et al.* 2019a; Yoshida *et al.* 2020), resulting in the identification of embryonic stage-specific anhydrobiosis-related genes and a novel peroxidase protein family. This illustrates the effectiveness of *de novo* genome screening at improving our understanding of anhydrobiosis.

Additionally, this high-quality genome enabled extensive analysis of horizontally transferred genes (Yoshida *et al.* 2017; Yoshida *et al.* 2019b). Many studies have observed apparently high levels of HGT in newly sequenced genomes, including the human genome (Lander *et al.*, 2001), though close examination of these candidates revealed multiple false positives, possibly due to mistakes during screening (Salzberg *et al.*, 2001; Stanhope *et al.*, 2001). I provided evidence that, in invertebrates, this was an artifact of gene prediction, as detection from *ab initio* predictions inferred high HGT levels (and misprediction of non-gene regions). However, it is important to note that not all HGT candidates were false positives: screening the contaminant-free *H. exemplaris* genome suggested the whole ascorbate synthesis pathway was obtained through HGT, as well as the previously identified catalase of bacterial origin. Additionally, *R. varieornatus* has obtained the TPS gene through HGT. These studies show the importance of genome/transcriptome sequencing methodologies, as they influence how one understands the resulting genome.

The data obtained in my study has led to multiple studies utilizing the genome of *H. exemplaris*: resolving phylogeny within Metazoa (Game and Smith, 2020; Giribet and Edgecombe, 2017; Laumer *et al.*, 2019; Stec *et al.*, 2020), gene conservation analysis (Carrero *et al.*, 2019; Koziol, 2018; Mapalo *et al.*, 2020; Shabardina *et al.*, 2020; Treffkorn *et al.*, 2018; Vizueta *et al.*, 2020; Zieger *et al.*, 2021), comparative genomics (Bemm *et al.*, 2017; Guijarro-Clarke *et al.*, 2020; Kamilari *et al.*, 2019; Nowell *et al.*, 2018; 2020;

Schiffer *et al.*, 2019; Yamada *et al.*, 2018; Yoshida *et al.*, 2019a), an understanding of body plan development (Smith *et al.*, 2018) and anhydrobiosis (Kondo *et al.* 2019; Kondo *et al.* 2020; Minguez-Toral *et al.* 2020; Murai *et al.* 2020; Yoshida *et al.* 2020). The *H. exemplaris* genome would continue to be a platform for tardigrade genomics.

## 6.2 A model of the anhydrobiosis machinery derived from the genome analysis

The establishment of a high-quality genome along with a comparative genomics approach in tardigrades enabled various gene sets to be identified, in particular, genes governing anhydrobiosis. In this study, I have focused on three vital points in tardigrade anhydrobiosis: conservation of anhydrobiosis-related genes, the regulation of these genes during entry into anhydrobiosis, and genes involved in anhydrobiosis during embryogenesis. I also identified genes participating in cross-tolerance.

I first conducted a comparative genome and transcriptome analysis to identify loci that contribute to tardigrade anhydrobiosis. *H. exemplaris* and *R. varieornatus* differ in their degree of desiccation tolerance (Horikawa *et al.*, 2008; Kondo *et al.*, 2015), and this is hypothesized to be related to the expression of anhydrobiosis-related genes (Hashimoto *et al.*, 2016; Kondo *et al.*, 2015). Comparisons of the genetic content of these two species and the desiccation transcriptome response have revealed a detailed repertoire of anhydrobiosis machinery, including previously unknown factors that may be the subject of further analyses (Yoshida *et al.*, 2017).

The models of anhydrobiosis proposed in the pre-genome era have much in common with the model I propose, *e.g.*, the expression of certain anhydrobiosis-related genes (anti-oxidative stress proteins, chaperones, *etc.*). Building upon this framework, the new omics data have expanded our knowledge of anhydrobiosis. For example, I have shown adaptations involving genome size reduction in the more desiccation-tolerant species, *R. varieornatus*, compared to *H. exemplaris*; a similar genome compaction was also noted in anhydrobiotic bdelloid rotifers compared to obligately aquatic species (Nowell *et al.*, 2018). Expansion of gene families encoding multiple anhydrobiosis-related protein families was also observed (Yoshida *et al.*, 2017). Previous studies have suggested that the duplication of LEA protein genes in bdelloid rotifers and chironomids has enabled sub-functionalization (Gusev *et al.*, 2014; Pouchkina-Stantcheva *et al.*, 2007). It is possible that the anhydrobiosis-related protein families identified may show different functions during the life cycle or anhydrobiosis trajectory.

I then focused on how these candidate genes are expressed in other conditions, *e.g.*, the embryonic stages. Tardigrade embryos are known to be capable of anhydrobiosis (Horikawa *et al.*, 2008), and thus

anhydrobiosis-related genes should be expressed. This is indeed the case; for example, our transcriptome analysis showed expression of an embryonic stage-specific SAHS ortholog (Yoshida *et al.*, 2019a). Further bioinformatics analysis (co-expression analysis, mutation analysis) may show how these orthologs differ from other paralogs. Additionally, I have found that thioredoxins are the major antioxidative stress response protein in both species. Thioredoxins have been suggested to be important in anhydrobiosis in tardigrades and *P. vanderplanki* (Förster *et al.*, 2009; Gusev *et al.*, 2014; Schokraie *et al.*, 2010). We have yet to understand the reasons for the differential usage of multiple antioxidants, but the expression profiles of these antioxidative stress response genes and IUPs are consistent with the ability of embryonic stages to undergo anhydrobiosis.

Furthermore, although somewhat limited, tardigrades show extremotolerance in the active state as well, *e.g.*, to ultraviolet and ionizing radiation. Studies have suggested this is a result of cross-tolerance (Gade *et al.*, 2020; Jönsson, 2003). If this is the case, genes regulated during anhydrobiosis should also be regulated during the response to these stresses. Anhydrobiosis is a complicated phenomenon, with multiple factors involved (*i.e.*, protection of the lipid bilayer, proteins, DNA and RNA, oxidative stress, organelle integrity), but the response to ultraviolet and ionizing radiation is well-studied in model organisms, and thus analyzing these responses may provide a catalogue of genes that contribute to anhydrobiosis. By using UVC as the stress inducer, I screened the genome for genes that are regulated in both anhydrobiosis and the UVC response (Yoshida *et al.*, 2020). Among the several hundred genes in common, I identified a novel protein family that localizes to the Golgi apparatus and shows antioxidative stress functions. Previously identified tardigrade-specific anhydrobiosis-related proteins are only conserved within Eutardigrada or are specific to Parachaela. On the other hand, the identified protein family, anhydrobiosis-related Mn<sup>2+</sup>-dependent peroxidase, is widely conserved, even in the non-anhydrobiotic tardigrade, *Batillipes* sp. It suggests this gene family may have evolved before the diversion of the two orders and has largely expanded in the Parachaela lineage. The duplication in Parachaela may have enabled greater tolerance of cellular hydrogen peroxide, as this protein is a Mn<sup>2+</sup>-dependent peroxidase. The requirement of such proteins in the Golgi remains to be clarified, but recent studies have suggested the importance of the Golgi stress response (Alborzinia *et al.*, 2018; Jiang *et al.*, 2011; Yoshida, 2009).

These quantitative analyses highlight the possibility of damage during anhydrobiosis to two cellular compartments, the mitochondria and the Golgi apparatus. The mitochondria are the major source of reactive oxygen species (ROS), as they are a byproduct of ATP synthesis. The loss of cellular water molecules due to impending desiccation would increase ROS concentration within the mitochondria, which would therefore be susceptible to oxidative stress damage. Others and I (Richaud *et al.*, 2020; Yoshida *et al.*, 2020) have observed abnormal mitochondrial morphology in tardigrade specimens recovering from anhydrobiosis, suggesting such stress may cause mitochondrial damage. Metabolomic analysis of *P. vanderplanki* larvae

recovering from anhydrobiosis suggested mitochondria may be functioning at low levels during the first few hours after water becomes available again (Ryabova *et al.*, 2020). The increase in mitochondrial metabolism after full recovery will cause an increase in ROS, possibly causing mitochondrial damage. One factor that might protect mitochondria from such stress may be the MAHS protein, as it localizes to the mitochondria and confers osmotic stress tolerance on heterologous cells (Tanaka *et al.*, 2015). Osmotic stress mimics the initial stages of anhydrobiosis, when cells lose internal water molecules. However, the expression of MAHS genes alone does not confer full protection, suggesting other factors may be present. A more detailed analysis of proteins that localize to mitochondria is necessary to investigate these factors.

Using all the above information, I build upon the anhydrobiosis models of the pre-genome era, adding further mechanisms to extend what we know about this phenomenon. While previous studies have highlighted various antioxidants (Mali *et al.*, 2010; Rebecchi, 2013; Rizzo *et al.*, 2010; Savic *et al.*, 2015), I have provided additional evidence that support the importance of anti-oxidative stress: duplication of anti-oxidative stress proteins, the regulation of life-stage specific expression of antioxidants, and the acquirement of a novel multi-copy peroxidase family. Eukaryotic cells face oxidative stress constantly; ROS are a byproduct of ATP synthesis in the mitochondria, and are created during lipid metabolism by beta oxidation and by exposure to environmental stress (oxygen, UV). Although cells have evolved various countermeasures, they are still susceptible to ROS damage, which results in aging (Liguori *et al.*, 2018; Vanfleteren, 1993) and the development of various abnormalities (Pizzino *et al.*, 2017). Simply removing all oxidative stress from cells is not a solution, however. Recent studies have emphasized that oxidants are utilized as signaling molecules (Finkel, 2003), and therefore the balance between excess and insufficient oxidants needs to be strictly regulated. Indeed, there is potential for the extensive anti-oxidative mechanisms in tardigrades to distort this balance. Do tardigrades need to compensate for this distortion? And, if so, how? These questions remain to be answered.

### 6.3 Future prospects

Although I have employed genomics and transcriptomics to identify factors that may contribute to anhydrobiosis, we have yet to synthesize these findings into a more general understanding of the phenomenon.

The question “*are the mechanisms currently identified sufficient to explain anhydrobiosis?*” still remains. The abnormal mitochondrial morphology (Richaud *et al.*, 2020; Yoshida *et al.*, 2020) and DNA damage (Gusev *et al.*, 2010; Hespels *et al.*, 2014; Neumann *et al.*, 2009; Rebecchi *et al.*, 2009) observed during recovery from anhydrobiosis suggests not. Most studies have focused on entry into anhydrobiosis (Förster *et al.*, 2012; Mali *et al.*, 2010; Wang *et al.*, 2014). For example, exposure of *H. exemplaris* specimens to the AMPK activator D942 enables some survival of rapid desiccation at 10% RH, but not

to the extent that results from preconditioning and slow desiccation. This suggests that regulation of anhydrobiosis-related genes through the AMPK signaling cascade is not sufficient; still other factors are required for successful anhydrobiosis. One important requirement is the rapid repair of cellular damage. In my studies, I employed UVC to mimic recovery from stress exposure and identified many inducible proteins involved in cellular repair, including those related to mitochondrial stress responses (*i.e.*, sequestosome, mitochondrial chaperones) and a novel  $Mn^{2+}$ -dependent peroxidase. There are probably many novel and known proteins that have so far been overlooked, but which may contribute to cellular repair during recovery from anhydrobiosis. Thus, further screening experiments are required to clarify the repair machinery in tardigrades.

Additionally, further analysis of the relationships between anhydrobiosis-related genes is required, *i.e.*, signaling pathways, gene regulation networks, *etc.* By identifying the regulatory factors that control anhydrobiosis-related genes, we should be able to understand how the cells in anhydrobiotes recognize desiccation stress and how they differ from those of non-anhydrobiotic organisms. Studies employing phosphoproteomics and inhibitors of signaling factors have identified PP2A and AMPK as the major signaling factors in *H. exemplaris* during anhydrobiosis (Kondo *et al.* 2015; Kondo *et al.* 2019; Kondo *et al.* 2020). Activating AMPK in *H. exemplaris* significantly increases desiccation tolerance, showing the importance of this signaling pathway (Kondo *et al.*, 2020). However, we have yet to determine how desiccation is sensed and through what factors AMPK regulates anhydrobiosis-related genes. Bioinformatics analysis on the sleeping chironomid *P. vanderplanki* has identified NF-Y as the master regulator of anhydrobiosis-related genes (Yamada *et al.*, 2020). NFY regulates several trans-factors, including the HSF, a transcription factor previously implicated in *P. vanderplanki*, and in the regulation of anhydrobiosis in *Artemia franciscana* (Mazin *et al.*, 2018; Tan and MacRae, 2018). The methods used in these analyses include genomic and transcriptomic resources; therefore, application to tardigrades, *H. exemplaris* in particular, should be possible. Additionally, a recent preprint has reported utilization of the CRISPR/Cas9 system in Pv11, a cell line derived from *P. vanderplanki*, to investigate calcium signaling in the initial stages of anhydrobiosis preconditioning (Miyata *et al.*, 2020). This study has linked calmodulin and nuclear factor of activated T-cells (NFAT) to trehalose exposure, indicating that these proteins form part of a regulatory system during preconditioning. NFAT5 and calmodulin are regulated during anhydrobiosis in *H. exemplaris*, suggesting a similar signaling pathway in this tardigrade species. The development of genetic engineering tools and the establishment of a culturable cell line are outstanding problems in tardigrades, meaning that validation of cellular processes identified through omics analysis is not yet possible.

The genomic analysis conducted in my studies is based on only two eutardigrade species. Recent studies on heterotardigrade omics have identified similar genomic adaptations (Bemm *et al.*, 2017; Kamilari *et al.*, 2019; Murai *et al.*, 2020). A multi-omics analysis using genomics, transcriptomics, and proteomics

identified a CAHS structural analog in the *Echiniscidae* lineage (Murai *et al.*, 2020), which supports the notion of convergent evolution of IUPs between Heterotardigrada and Eutardigrada lineages. This also implies that there are still important genes we have yet to identify. Further analysis of other tardigrade species and anhydrobiotic invertebrates, as well as experimental validation of key components, is required to determine how these genes contribute to the anhydrobiosis machinery.

The findings I report in this dissertation suggest the importance of the protection and repair of cellular compartments. Previous studies have focused more on the molecular level of protection, identifying individual elements that protect or repair particular (categories of) molecules. However, organisms, in this case, tardigrades, are an ensemble of cellular compartments and cell types, which are further arranged into differentiated tissues, forming organs, comprising a single individual. Further analysis at the organelle, tissue and organ levels would broaden our knowledge of anhydrobiosis.

## 6.4 Conclusions

*“How is anhydrobiosis possible?”* To provide a basis for solving this ultimate question, I have (1) constructed a foundation for comparative genomics in Tardigrada; (2) applied this foundation to two tardigrade physiological states, anhydrobiosis and embryogenesis; and (3) identified and validated the function of a novel peroxidase. Accumulation of such genomic data provides non-tardigradologists with a foundation on which to initiate tardigrade research. I have provided a comprehensive list of anhydrobiosis genes, nearly half of which are novel genes, as a basis for future anhydrobiosis research. I have also suggested the importance of analysis at the organelle level, specifically here the Golgi and mitochondria. However, just identifying the pieces of the anhydrobiosis machinery is not *“understanding”* anhydrobiosis itself. A systematic biological approach to the integration of data at various levels may shine a light into the abyss, but numerous gaps between these highlights will inevitably remain in our knowledge of anhydrobiosis. Nevertheless, these highlights are essential; comprehensive identification of individual elements, as achieved in my studies, will be the foundation for further research of anhydrobiosis. Hence, in this dissertation, I provide the knowledge obtained through genomics and transcriptomics of tardigrades as the basis of discussion of the anhydrobiosis mechanism, leading to an understanding of anhydrobiosis, the robustness of cells, and ultimately the essence of life itself.

---

## References

- Aboobaker, A. and Blaxter, M. (2003a) Hox gene evolution in nematodes: novelty conserved. *Curr Opin Genet*, **13**(6):593–598.
- Aboobaker, A. A. and Blaxter, M. L. (2003b) Hox gene loss during dynamic evolution of the nematode cluster. *Curr Biol*, **13**(1):37–40.
- Aboobaker, A. and Blaxter, M. (2010) *The nematode story: Hox gene loss and rapid evolution*, pages 101–110. Springer, New York.
- Acuna, R., Padilla, B. E., Florez-Ramos, C. P., Rubio, J. D., Herrera, J. C., Benavides, P., Lee, S. J., Yeats, T. H., Egan, A. N., Doyle, J. J., and Rose, J. K. (2012) Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. *Proc Natl Acad Sci U S A*, **109**(11):4197–4202.
- Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*, **66**(Pt 2):213–221.
- Aken, B. L., Achuthan, P., Akanni, W., Amode, M. R., Bernsdorff, F., Bhai, J., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Gil, L., Giron, C. G., Gordon, L., Hourlier, T., Hunt, S. E., Janacek, S. H., Juettemann, T., Keenan, S., Laird, M. R., Lavidas, I., Maurel, T., McLaren, W., Moore, B., Murphy, D. N., Nag, R., Newman, V., Nuhn, M., Ong, C. K., Parker, A., Patricio, M., Riat, H. S., Sheppard, D., Sparrow, H., Taylor, K., Thormann, A., Vullo, A., Walts, B., Wilder, S. P., Zadissa, A., Kostadima, M., Martin, F. J., Muffato, M., Perry, E., Ruffier, M., Staines, D. M., Trevanion, S. J., Cunningham, F., Yates, A., Zerbino, D. R., and Flicek, P. (2017) Ensembl 2017. *Nucleic Acids Res*, **45**(D1):D635–D642.
- Alborzinia, H., Ignashkova, T. I., Dejure, F. R., Gendarme, M., Theobald, J., Wolf, S., Lindemann, R. K., and Reiling, J. H. (2018) Golgi stress mediates redox imbalance and ferroptosis in human cells. *Commun Biol*, **1**:210.
- Alejandro, S., Cailliatte, R., Alcon, C., Dirick, L., Domergue, F., Correia, D., Castaings, L., Briat, J. F., Mari, S., and Curie, C. (2017) Intracellular distribution of manganese by the trans-golgi network transporter NRAMP2 is critical for photosynthesis and cellular redox homeostasis. *Plant Cell*, **29**(12):3068–3084.
- Alsmark, C., Foster, P. G., Sicheritz-Ponten, T., Nakjang, S., Martin Embley, T., and Hirt, R. P. (2013) Patterns of prokaryotic lateral gene transfers affecting parasitic microbial eukaryotes. *Genome Biol*, **14**(2):R19.
- Altiero, T. and Rebecchi, L. (2001) Rearing tardigrades: results and problems. *Zool Anz*, **240**(3-4):217–221.

- Altiero, T., Rebecchi, L., and Bertolani, R. (2006) Phenotypic variations in the life history of two clones of *Macrobiotus richtersi* (Eutardigrada, Macrobiotidae). *Hydrobiologia*, **558**(1):33–40.
- Altiero, T., Bertolani, R., and Rebecchi, L. (2010) Hatching phenology and resting eggs in tardigrades. *J Zool*, **280**(3):290–296.
- Altiero, T., Guidetti, R., Caselli, V., Cesari, M., and Rebecchi, L. (2011) Ultraviolet radiation tolerance in hydrated and desiccated eutardigrades. *J Zool Syst Evol Res*, **49**:104–110.
- Altiero, T., Suzuki, A., and Rebecchi, L. (2019) Reproduction, development and life cycles. *Pest Manag Sci*, **2**(6):211–247.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**(17):3389–3402.
- Ambrose, K. V., Koppenhofer, A. M., and Belanger, F. C. (2014) Horizontal gene transfer of a bacterial insect toxin gene into the *Epichloe* fungal symbionts of grasses. *Sci Rep*, **4**:5562.
- Andrews, A. (2015) FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Anno, G. H., Young, R. W., Bloom, R. M., and Mercier, J. R. (2003) Dose response relationships for acute ionizing-radiation lethality. *Health Phys*, **84**(5):565–575.
- Arakawa, K., Mori, K., Ikeda, K., Matsuzaki, T., Kobayashi, Y., and Tomita, M. (2003) G-language Genome Analysis Environment: a workbench for nucleotide sequence data mining. *Bioinformatics*, **19**(2):305–306.
- Arakawa, K. and Tomita, M. (2006) G-language system as a platform for large-scale analysis of high-throughput omics data. *J Pestic Sci*, **31**(3):282–288.
- Arakawa, K., Kido, N., Oshita, K., and Tomita, M. (2010) G-language genome analysis environment with REST and SOAP web service interfaces. *Nucleic Acids Res*, **38**(Web Server issue):W700–W705.
- Arakawa, K. (2013) Comparative metabolomics of anhydrobiosis in tardigrade *Ramazzottius varieornatus*. *J Jpn So Extremophiles*, **11**(2):30–37.
- Arakawa, K., Yoshida, Y., and Tomita, M. (2016) Genome sequencing of a single tardigrade *Hypsibius dujardini* individual. *Sci Data*, **3**:160063.
- Arakawa, K. (2016) No evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. *Proc Natl Acad Sci U S A*, **113**(22):E3057.
- Arakawa, K. and Numata, K. (2021) Reconsidering the “glass transition” hypothesis of the contribution of intrinsically unstructured CAHS proteins to the desiccation tolerance of tardigrades. *Mol Cell*, **In press**.
- Archibald, J. M. (2015) Endosymbiosis and eukaryotic cell evolution. *Curr Biol*, **25**(19):R911–R921.
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., and Noble, W. S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*, **37**(Web Server issue):W202–W208.
- Bakri, A., Heather, N., Hendrichs, J., and Ferris, I. (2005) Fifty years of radiation biology in entomology: lessons learned from IDIDAS. *Ann Entomol Soc Am*, **98**(1):1–12.
- Baldwin, W. S., Bailey, R., Long, K. E., and Klaine, S. (2001) Incomplete ecdysis is an indicator of ecdysteroid exposure in *Daphnia magna*. *Environ Toxicol Chem*, **20**(7):1564–1569.

- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*, **19**(5):455–477.
- Banton, M. C. and Tunnacliffe, A. (2012) MAPK phosphorylation is implicated in the adaptation to desiccation stress in nematodes. *J Exp Biol*, **215**(Pt 24):4288–4298.
- Barracough, T. G., Fontaneto, D., Ricci, C., and Herniou, E. A. (2007) Evidence for inefficient selection against deleterious mutations in cytochrome oxidase I of asexual bdelloid rotifers. *Mol Biol Evol*, **24**(9):1952–1962.
- Barrett, J. (1982) Metabolic responses to anabiosis in the fourth stage juveniles of *Ditylenchus dipsaci* (Nematoda). *Proc R Soc Lond B*, **216**:159–177.
- Bast, J., Parker, D. J., Dumas, Z., Jalvingh, K. M., Tran Van, P., Jaron, K. S., Figuet, E., Brandt, A., Galtier, N., and Schwander, T. (2018) Consequences of asexuality in natural populations: insights from stick insects. *Mol Biol Evol*, **35**(7):1668–1677.
- Baumann, H. (1922) Die anabiose der tardigraden. *Zoologische jahrbücher: abteilung für systematik, geographie und biologie der tiere*, **45**:501–556.
- Becquerel, P. (1950) La suspension de la vie au dessous de 1/20 K absolu par demagnetization adiabatique de l'alun de fer dans le vide les plus élève. *C R Hebd Seance Acad Sci, Paris*, **231**:261–264.
- Beisser, D., Grohme, M. A., Kopka, J., Frohme, M., Schill, R. O., Hengherr, S., Dandekar, T., Klau, G. W., Dittrich, M., and Muller, T. (2012) Integrated pathway modules using time-course metabolic profiles and EST data from *Milnesium tardigradum*. *BMC Syst Biol*, **6**(72):1–13.
- Belott, C., Janis, B., and Menze, M. A. (2020) Liquid-liquid phase separation promotes animal desiccation tolerance. *Proc Natl Acad Sci U S A*, **117**(44):27676–27684.
- Beltran-Pardo, E., Jönsson, K. I., Wojcik, A., Haghdoost, S., Harms-Ringdahl, M., Bermudez-Cruz, R. M., and Bernal Villegas, J. E. (2013) Effects of ionizing radiation on embryos of the tardigrade *Milnesium cf. tardigradum* at different stages of development. *PLoS One*, **8**(9):e72098.
- Beltran-Pardo, E., Jönsson, K. I., Harms-Ringdahl, M., Haghdoost, S., and Wojcik, A. (2015) Tolerance to gamma radiation in the tardigrade *Hypsibius dujardini* from embryo to adult correlate inversely with cellular proliferation. *PLoS One*, **10**(7):e0133658.
- Bemm, F., Weiss, C. L., Schultz, J., and Förster, F. (2016) Genome of a tardigrade: horizontal gene transfer or bacterial contamination? *Proc Natl Acad Sci U S A*, **113**(22):E3054–E3056.
- Bemm, F. M., Burleigh, L., Förster, F., Schmucki, R., Ebeling, M., Janzen, C., Dandekar, T., Schill, R., Certa, U., and Schultz, J. (2017) Draft genome of the Eutardigrade *Milnesium tardigradum* sheds light on ecdysozoan evolution. *bioRxiv*, **122309**.
- Bertolani, R. (1981) A new genus and five new species of Italian fresh-water tardigrades. *Boll Mus Civ Stor Nat Verona*, **8**:249–254.
- Bertolani, R. and Kinchin, I. M. (1993) A new species of *Ramazzottius* (Tardigrada, Hypsibiidae) in a rain gutter sediment from England. *Zool J Linnean Soc*, **109**(3):327–333.
- Bertolani, R., Guidetti, R., Marchioro, T., Altiero, T., Rebecchi, L., and Cesari, M. (2014) Phylogeny of Eutardigrada: new molecular data and their morphological support lead to the identification of new evolutionary lineages. *Mol Phylogenet Evol*, **76**:110–126.

- Blanc, G., Agarkova, I., Grimwood, J., Kuo, A., Brueggeman, A., Dunigan, D. D., Gurnon, J., Ladunga, I., Lindquist, E., Lucas, S., Pangilinan, J., Proschold, T., Salamov, A., Schmutz, J., Weeks, D., Yamada, T., Lomsadze, A., Borodovsky, M., Claverie, J. M., Grigoriev, I. V., and Van Etten, J. L. (2012) The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome Biol*, **13**(5):R39.
- Blaxter, M. (2007) Symbiont genes in host genomes: fragments with a future? *Cell Host Microbe*, **2**(4):211–213.
- Boetzer, M. and Pirovano, W. (2014) SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics*, **15**:211.
- Bonifacio, A., Guidetti, R., Altiero, T., Sergo, V., and Rebecchi, L. (2012) Nature, source and function of pigments in tardigrades: *in vivo* raman imaging of carotenoids in *Echiniscus blumi*. *PLoS One*, **7**(11):e50162.
- Boothby, T. C., Tenlen, J. R., Smith, F. W., Wang, J. R., Patanella, K. A., Nishimura, E. O., Tintori, S. C., Li, Q., Jones, C. D., Yandell, M., Messina, D. N., Glasscock, J., and Goldstein, B. (2015) Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. *Proc Natl Acad Sci U S A*, **112**(52):15976–15981.
- Boothby, T. C. and Goldstein, B. (2016) Reply to Bemm *et al.* and Arakawa: identifying foreign genes in independent *Hypsibius dujardini* genome assemblies. *Proc Natl Acad Sci U S A*, **113**(22):E3058–E3061.
- Boothby, T. C. and Pielak, G. J. (2017) Intrinsically disordered proteins and desiccation tolerance: elucidating functional and mechanistic underpinnings of anhydrobiosis. *Bioessays*, **39**(11):1700119.
- Boothby, T. C., Tapia, H., Brozena, A. H., Piszkiwicz, S., Smith, A. E., Giovannini, I., Rebecchi, L., Pielak, G. J., Koshland, D., and Goldstein, B. (2017) Tardigrades use intrinsically disordered proteins to survive desiccation. *Mol Cell*, **65**(6):975–984.e5.
- Borner, J., Rehm, P., Schill, R. O., Ebersberger, I., and Burmester, T. (2014) A transcriptome approach to ecdysozoan phylogeny. *Mol Phylogenet Evol*, **80**:79–87.
- Borodovsky, M. and Lomsadze, A. (2011) Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. *Curr Protoc Bioinformatics*, **4**:1–10.
- Boschetti, C., Pouchkina-Stantcheva, N., Hoffmann, P., and Tunnacliffe, A. (2011) Foreign genes and novel hydrophilic protein genes participate in the desiccation response of the bdelloid rotifer *Adineta ricciae*. *J Exp Biol*, **214**(Pt 1):59–68.
- Boschetti, C., Carr, A., Crisp, A., Eyres, I., Wang-Koh, Y., Lubzens, E., Barraclough, T. G., Micklem, G., and Tunnacliffe, A. (2012) Biochemical diversification through foreign gene expression in bdelloid rotifers. *PLoS Genet*, **8**(11):e1003035.
- Bownes, M. and Rembold, H. (1987) The titre of juvenile hormone during the pupal and adult stages of the life cycle of *Drosophila melanogaster*. *Eur J Biochem*, **164**(3):709–712.
- Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*, **34**(5):525–527.
- Brian, H. and Papanicolaou, A. (2017) Transdecoder (find coding regions within transcripts). <http://transdecoder.github.io>.
- Broca, P. (1860) Rapport sur la question soumise a la Société de biologie au sujet de la réviviscence des animaux desséchés. *C R et Mèm Soc Biol*, **3 Sér**, **2**:1–139.
- Buchfink, B., Xie, C., and Huson, D. H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods*, **12**(1):59–60.

- Cadet, J. and Wagner, J. R. (2013) DNA base damage by reactive oxygen species, oxidizing agents, and UV radiation. *Cold Spring Harb Perspect Biol*, **5**(2):a012559.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**:421.
- Campbell, L. I., Rota-Stabelli, O., Edgecombe, G. D., Marchioro, T., Longhorn, S. J., Telford, M. J., Philippe, H., Rebecchi, L., Peterson, K. J., and Pisani, D. (2011) MicroRNAs and phylogenomics resolve the relationships of Tardigrada and suggest that velvet worms are the sister group of Arthropoda. *Proc Natl Acad Sci U S A*, **108**(38):15920–15924.
- Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**(15):1972–1973.
- Carmona, A., Deves, G., Roudeau, S., Cloetens, P., Bohic, S., and Ortega, R. (2010) Manganese accumulates within golgi apparatus in dopaminergic cells as revealed by synchrotron X-ray fluorescence nanoimaging. *ACS Chem Neurosci*, **1**(3):194–203.
- Carrero, D., Perez-Silva, J. G., Quesada, V., and Lopez-Otin, C. (2019) Differential mechanisms of tolerance to extreme environmental conditions in tardigrades. *Sci Rep*, **9**(1):14938.
- Challis, R. J., Kumar, S., Stevens, L., and Blaxter, M. (2017) GenomeHubs: simple containerized setup of a custom Ensembl database and web server for any species. *Database*, **2017**. bax039.
- Chang, Z., Li, G., Liu, J., Zhang, Y., Ashby, C., Liu, D., Cramer, C. L., and Huang, X. (2015) Bridger: a new framework for *de novo* transcriptome assembly using RNA-seq data. *Genome Biol*, **16**:30.
- Chavez, C., Cruz-Becerra, G., Fei, J., Kassavetis, G. A., and Kadonaga, J. T. (2019) The tardigrade damage suppressor protein binds to nucleosomes and protects DNA from hydroxyl radicals. *eLife*, **8**:e47682.
- Chen, V. B., Arendall, W. B., r., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, **66**(Pt 1):12–21.
- Chin, C. S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., Cramer, G. R., Delledonne, M., Luo, C., Ecker, J. R., Cantu, D., Rank, D. R., and Schatz, M. C. (2016) Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods*, **13**(12):1050–1054.
- Chinnasri, B., Moy, J. H., Sipes, B. S., and Schmitt, D. P. (1997) Effect of gamma-irradiation and heat on root-knot nematode, *Meloidogyne javanica*. *J Nematol*, **29**(1):30–34.
- Clasen, F. J., Pierneef, R. E., Slippers, B., and Reva, O. (2018) EuGI: a novel resource for studying genomic islands to facilitate horizontal gene transfer detection in eukaryotes. *BMC Genomics*, **19**(1):323.
- Clegg, J. S. (1967) Metabolic studies of cryptobiosis in encysted embryos of *Artemia salina*. *Comp Biochem Physiol*, **20**(3):801–809.
- Clegg, J. S. (2001) Cryptobiosis—a peculiar state of biological organization. *Comp Biochem Physiol B Biochem Mol Biol*, **128**(4):613–624.
- Clegg, J. S. (2005) Desiccation tolerance in encysted embryos of the animal extremophile, *Artemia*. *Integr Comp Biol*, **45**(5):715–724.
- Coghlan, A., Fiedler, T. J., McKay, S. J., Flicek, P., Harris, T. W., Blasiar, D., n, G. C., and Stein, L. D. (2008) nGASP—the nematode genome annotation assessment project. *BMC Bioinformatics*, **9**:549.
- Cornette, R. and Kikawada, T. (2011) The induction of anhydrobiosis in the sleeping chironomid: current status of our knowledge. *IUBMB Life*, **63**(6):419–429.

- Cornette, R., Yamamoto, N. A. O., Yamamoto, M., Kobayashi, T., Petrova, N. A., Gusev, O., Shimura, S., Kikawada, T., Pemba, D., and Okuda, T. (2017) A new anhydrobiotic midge from Malawi, *Polypedilum pembai* sp.n. (Diptera: Chironomidae), closely related to the desiccation tolerant midge, *Polypedilum vanderplanki* Hinton. *Syst Entomol*, **42**(4):814–825.
- Corti, B. (1774) *Osservazioni microscopiche sulla Tremella e sulla circolazione del fluido in una pianta acquajuola (Chara)*. appresso Giuseppe Rocchi, Lucca.
- Crisp, A., Boschetti, C., Perry, M., Tunnacliffe, A., and Micklem, G. (2015) Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome Biol*, **16**:50.
- Crosbie, E. J., Einstein, M. H., Franceschi, S., and Kitchener, H. C. (2013) Human papillomavirus and cervical cancer. *Lancet*, **382**(9895):889–899.
- Crowe, J. H., Crowe, L. M., Carpenter, J. F., and Aurell Wistrom, C. (1987) Stabilization of dry phospholipid bilayers and proteins by sugars. *Biochem J*, **242**(1):1–10.
- Crowe, J. H. (2014) Anhydrobiosis: an unsolved problem. *Plant Cell Environ*, **37**(7):1491–1493.
- Czerneková, M., Jönsson, K. I., Chajec, L., Student, S., and Poprawa, I. (2017) The structure of the desiccated *Richtersius coronifer* (Richters, 1903). *Protoplasma*, **254**(3):1367–1377.
- Czernik, M., Fidanza, A., Luongo, F. P., Valbonetti, L., Scapolo, P. A., Patrizio, P., and Loi, P. (2020) Late embryogenesis abundant (LEA) proteins confer water stress tolerance to mammalian somatic cells. *Cryobiology*, **92**(1):189–196.
- da Costa Morato Nery, D., da Silva, C. G., Mariani, D., Fernandes, P. N., Pereira, M. D., Panek, A. D., and Eleutherio, E. C. (2008) The role of trehalose and its transporter in protection against reactive oxygen species. *Biochim Biophys Acta*, **1780**(12):1408–1411.
- Daborn, P. J., Waterfield, N., Silva, C. P., Au, C. P., Sharma, S., and Ffrench-Constant, R. H. (2002) A single *Photorehabdus* gene, makes caterpillars floppy (mcf), allows *Escherichia coli* to persist within and kill insects. *Proc Natl Acad Sci U S A*, **99**(16):10742–10747.
- Daimon, T., Taguchi, T., Meng, Y., Katsuma, S., Mita, K., and Shimada, T. (2008) Beta-fructofuranosidase genes of the silkworm, *Bombyx mori*: insights into enzymatic adaptation of *B. mori* to toxic alkaloids in mulberry latex. *J Biol Chem*, **283**(22):15271–15279.
- Daly, M. J., Gaidamakova, E. K., Matrosova, V. Y., Vasilenko, A., Zhai, M., Leapman, R. D., Lai, B., Ravel, B., Li, S. M., Kemner, K. M., and Fredrickson, J. K. (2007) Protein oxidation implicated as the primary determinant of bacterial radioresistance. *PLoS Biol*, **5**(4):e92.
- Daly, M. J. (2012) Death by protein damage in irradiated cells. *DNA Repair (Amst)*, **11**(1):12–21.
- Danchin, E. G. (2016) Lateral gene transfer in eukaryotes: tip of the iceberg or of the ice cube? *BMC Biol*, **14**(1):101.
- Darling, A. C., Mau, B., Blattner, F. R., and Perna, N. T. (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*, **14**(7):1394–1403.
- de Rosa, R., Grenier, J. K., Andreeva, T., Cook, C. E., Adoutte, A., Akam, M., Carroll, S. B., and Balavoine, G. (1999) Hox genes in brachiopods and priapulids and protostome evolution. *Nature*, **399**(6738):772–776.
- Debortoli, N., Li, X., Eyres, I., Fontaneto, D., Hespels, B., Tang, C. Q., Flot, J. F., and Van Doninck, K. (2016) Genetic exchange among bdelloid rotifers is more likely due to horizontal gene transfer than to meiotic sex. *Curr Biol*, **26**(6):723–732.
- Degma, P., Bertolani, R., and Guidetti, R. (2017) Actual checklist of Tardigrada species (2009–2017, 32nd Edition: 01-06-2017). DOI: 10.25431/11380\_1178608.

- Degma, P., Bertolani, R., and Guidetti, R. (2020) Actual checklist of Tardigrada species (2009-2020, 38th Edition: 18-08-2020). DOI: 10.25431/11380\_1178608.
- Delmont, T. O. and Eren, A. M. (2016) Identifying contamination with advanced visualization and analysis practices: metagenomic approaches for eukaryotic genome assemblies. *PeerJ*, **4**:e1839.
- Doyère, L. M. F. (1840) Mémoire sur les tardigrades. *Annales des Sciences Naturelles, Série 2, Zoologie*, **14**:269–361.
- Doyère, L. M. F. (1842) Mémoire sur les tardigrades. Sur le facilité que possèdent les tardigrades, les rotifères, les anguillules des toits et quelques autres animalcules, de revenir à la vie après être complètement desséchés. *Ann Sci Nat Zool*, **18**:5–35.
- Dunn, C. W., Hejnol, A., Matus, D. Q., Pang, K., Browne, W. E., Smith, S. A., Seaver, E., Rouse, G. W., Obst, M., Edgecombe, G. D., Sorensen, M. V., Haddock, S. H., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R. M., Wheeler, W. C., Martindale, M. Q., and Giribet, G. (2008) Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*, **452**(7188):745–749.
- Dunning Hotopp, J. C., Clark, M. E., Oliveira, D. C., Foster, J. M., Fischer, P., Munoz Torres, M. C., Giebel, J. D., Kumar, N., Ishmael, N., Wang, S., Ingram, J., Nene, R. V., Shepard, J., Tomkins, J., Richards, S., Spiro, D. J., Ghedin, E., Slatko, B. E., Tettelin, H., and Werren, J. H. (2007) Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science*, **317**(5845):1753–1756.
- Dunning Hotopp, J. C. and Estes, A. M. (2014) Biology wars: the eukaryotes strike back. *Cell Host Microbe*, **16**(6):701–703.
- Dure, L., Greenway, S. C., and Galau, G. A. (1981) Developmental biochemistry of cottonseed embryogenesis and germination: changing messenger ribonucleic acid populations as shown by *in vitro* and *in vivo* protein synthesis. *Biochemistry*, **20**(14):4162–4168.
- Edgar, R. C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**(19):2460–2461.
- Edgecombe, G. D. (2010) Arthropod phylogeny: an overview from the perspectives of morphology, molecular data and the fossil record. *Arthropod Struct Dev*, **39**(2-3):74–87.
- Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc*, **2**(4):953–971.
- Emms, D. M. and Kelly, S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol*, **16**:157.
- Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, **66**(Pt 4):486–501.
- English, A. C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D. M., Reid, J. G., Worley, K. C., and Gibbs, R. A. (2012) Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One*, **7**(11):e47768.
- Eren, A. M., Esen, O. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., and Delmont, T. O. (2015) Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ*, **3**:e1319.
- Erkut, C., Penkov, S., Khesbak, H., Vorkel, D., Verbavatz, J. M., Fahmy, K., and Kurzchalia, T. V. (2011) Trehalose renders the dauer larva of *Caenorhabditis elegans* resistant to extreme desiccation. *Curr Biol*, **21**(15):1331–1336.
- Erkut, C., Vasilij, A., Boland, S., Habermann, B., Shevchenko, A., and Kurzchalia, T. V. (2013) Molecular strategies of the *Caenorhabditis elegans* dauer larva to survive extreme desiccation. *PLoS One*, **8**(12):e82473.

- Evans, P. R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr*, **67**(Pt 4):282–292.
- Eyres, I., Boschetti, C., Crisp, A., Smith, T. P., Fontaneto, D., Tunnacliffe, A., and Barraclough, T. G. (2015) Horizontal gene transfer in bdelloid rotifers is ancient, ongoing and more frequent in species from desiccating habitats. *BMC Biol*, **13**:90.
- Finkel, T. (2003) Oxidant signals and oxidative stress. *Curr Opin Cell Biol*, **15**(2):247–254.
- Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., and Bateman, A. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*, **44**(D1):D279–D285.
- Flot, J. F., Hespeels, B., Li, X., Noel, B., Arkhipova, I., Danchin, E. G., Hejnl, A., Henrissat, B., Koszul, R., Aury, J. M., Barbe, V., Barthelemy, R. M., Bast, J., Bazykin, G. A., Chabrol, O., Couloux, A., Da Rocha, M., Da Silva, C., Gladyshev, E., Gouret, P., Hallatschek, O., Hecox-Lea, B., Labadie, K., Lejeune, B., Piskurek, O., Poulain, J., Rodriguez, F., Ryan, J. F., Vakhrusheva, O. A., Wajnberg, E., Wirth, B., Yushenova, I., Kellis, M., Kondrashov, A. S., Mark Welch, D. B., Pontarotti, P., Weissenbach, J., Wincker, P., Jaillon, O., and Van Doninck, K. (2013) Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. *Nature*, **500**(7463):453–457.
- Förster, F., Liang, C., Shkumatov, A., Beisser, D., Engelmann, J. C., Schnölzer, M., Frohme, M., Müller, T., Schill, R. O., and Dandekar, T. (2009) Tardigrade workbench: comparing stress-related proteins, sequence-similar and functional protein clusters as well as RNA elements in tardigrades. *BMC Genomics*, **10**(469):1–10.
- Förster, F., Beisser, D., Grohme, M. A., Liang, C., Mali, B., Siegl, A. M., Engelmann, J. C., Shkumatov, A. V., Schokraie, E., Müller, T., Schnölzer, M., Schill, R. O., Frohme, M., and Dandekar, T. (2012) Transcriptome analysis in tardigrade species reveals specific molecular pathways for stress adaptations. *Bioinform Biol Insights*, **6**:69–96.
- Franks, F. (2000) *Water 2nd Edition, A Matrix of Life*. RSC Paperbacks. The Royal Society of Chemistry, 2 edition.
- Friedlander, M. R., Mackowiak, S. D., Li, N., Chen, W., and Rajewsky, N. (2012) miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res*, **40**(1):37–52.
- Fu, Z., Agudelo, P., and Wells, C. E. (2020) Detoxification-related gene expression accompanies anhydrobiosis in the foliar nematode (*Aphelenchoides fragariae*). *J Nematol*, **52**:1–12.
- Gabriel, W. N. and Goldstein, B. (2007) Segmental expression of Pax3/7 and engrailed homologs in tardigrade development. *Dev Genes Evol*, **217**(6):421–433.
- Gabriel, W. N., McNuff, R., Patel, S. K., Gregory, T. R., Jeck, W. R., Jones, C. D., and Goldstein, B. (2007) The tardigrade *Hypsibius dujardini*, a new model for studying the evolution of development. *Dev Biol*, **312**(2):545–559.
- Gade, V. R., Traikov, S., Oertel, J., Fahmy, K., and Kurzhalia, T. V. (2020) *C. elegans* possess a general program to enter cryptobiosis that allows dauer larvae to survive different kinds of abiotic stress. *Sci Rep*, **10**(1):13466.
- Game, M. and Smith, F. W. (2020) Loss of intermediate regions of perpendicular body axes contributed to miniaturization of tardigrades. *Proc Biol Sci*, **287**(1931):20201135.
- Giard, A. (1894) L'anhydrobiose ou ralentissement des phénomènes vitaux. *C R Soc Biol, Paris*, **46**:497–500.
- Giribet, G. and Edgecombe, G. D. (2017) Current Understanding of Ecdysozoa and its Internal Phylogenetic Relationships. *Integr Comp Biol*, **57**(3):455–466.

- Gąsiorek, P., Stec, D., Morek, W., and Michalczyk, Ł. (2018) An integrative redescription of *Hypsibius dujardini* (Doyère, 1840), the nominal taxon for Hypsibioidea (Tardigrada: Eutardigrada). *Zootaxa*, **4415**(1):45.
- Gladyshev, E. and Meselson, M. (2008) Extreme resistance of bdelloid rotifers to ionizing radiation. *Proc Natl Acad Sci U S A*, **105**(13):5139–5144.
- Gladyshev, E. A., Meselson, M., and Arkhipova, I. R. (2008) Massive horizontal gene transfer in bdelloid rotifers. *Science*, **320**(5880):1210–1213.
- Goeze, J. A. E. (1773) *Beobachtung: Über den kleinen Wasserbären*. JJ Gebauers Wittwe und Joh Jacob Gebauer, Halle.
- Gordon, R., Chippett, J., and Tilley, J. (1996) Effects of two carbamates on infective juveniles of *Stemernema carpopapsae* All strain and *Steinernema feltiae* Umeå strain. *J Nematol*, **28**(3):310–317.
- Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res*, **38**(Web Server issue):W695–W699.
- Gouy, M., Guindon, S., and Gascuel, O. (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*, **27**(2):221–214.
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*, **29**(7):644–652.
- Grenier, J. K., Garber, T. L., Warren, R., Whittington, P. M., and Carroll, S. (1997) Evolution of the entire arthropod Hox gene set predated the origin and radiation of the onychophoran/arthropod clade. *Curr Biol*, **7**(8):547–553.
- Grohme, M. A., Mali, B., Welnicz, W., Michel, S., Schill, R. O., and Frohme, M. (2013) The aquaporin channel repertoire of the tardigrade *Milnesium tardigradum*. *Bioinform Biol Insights*, **7**:153–165.
- Gross, V. and Mayer, G. (2015) Neural development in the tardigrade *Hypsibius dujardini* based on anti-acetylated alpha-tubulin immunolabeling. *EvoDevo*, **6**:1–12.
- Gross, V., Minich, I., and Mayer, G. (2017) External morphogenesis of the tardigrade *Hypsibius dujardini* as revealed by scanning electron microscopy. *J Morphol*, **278**(4):563–573.
- Grothman, G. T., Johansson, C., Chilton, G., Kagoshima, H., Tsujimoto, M., and Suzuki, A. C. (2017) Gilbert Rahm and the status of Mesotardigrada Rahm, 1937. *Zoolog Sci*, **34**(1):5–10.
- Guidetti, R., Boschini, D., Rebecchi, L., and Bertolani, R. (2006) Encystment processes and the “Matrioshka-like Stage” in a moss-dwelling and in a limnic species of eutardigrades (Tardigrada). *Hydrobiologia*, **558**(1):9–21.
- Guidetti, R., Altiero, T., and Rebecchi, L. (2011) On dormancy strategies in tardigrades. *J Insect Physiol*, **57**(5):567–576.
- Guidetti, R., Cesari, M., Bertolani, R., Altiero, T., and Rebecchi, L. (2019) High diversity in species, reproductive modes and distribution within the *Paramacrobiotus richtersi* complex (Eutardigrada, Macrobiotidae). *Zoological Lett*, **5**:1.
- Guijarro-Clarke, C., Holland, P. W. H., and Paps, J. (2020) Widespread patterns of gene loss in the evolution of the animal kingdom. *Nat Ecol Evol*, **4**(4):519–523.

- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*, **59**(3):307–321.
- Gusev, O., Nakahara, Y., Vanyagina, V., Malutina, L., Cornette, R., Sakashita, T., Hamada, N., Kikawada, T., Kobayashi, Y., and Okuda, T. (2010) Anhydrobiosis-associated nuclear DNA damage and repair in the sleeping chironomid: linkage with radioresistance. *PLoS One*, **5**(11):e14008.
- Gusev, O., Suetsugu, Y., Cornette, R., Kawashima, T., Logacheva, M. D., Kondrashov, A. S., Penin, A. A., Hatanaka, R., Kikuta, S., Shimura, S., Kanamori, H., Katayose, Y., Matsumoto, T., Shagimardanova, E., Alexeev, D., Govorun, V., Wisecaver, J., Mikheyev, A., Koyanagi, R., Fujie, M., Nishiyama, T., Shigenobu, S., Shibata, T. F., Golygina, V., Hasebe, M., Okuda, T., Satoh, N., and Kikawada, T. (2014) Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge. *Nat Commun*, **5**(4784):1–9.
- Halberg, K. A., Persson, D., Ramløv, H., Westh, P., Kristensen, R. M., and Møbjerg, N. (2009) Cyclomor- phosis in Tardigrada: adaptation to environmental constraints. *J Exp Biol*, **212**(17):2803–2811.
- Harding, T., Roger, A. J., and Simpson, A. G. B. (2017) Adaptations to high salt in a halophilic protist: differential expression and gene acquisitions through duplications and gene transfers. *Front Microbiol*, **8**:944.
- Hashimoto, T., Horikawa, D. D., Saito, Y., Kuwahara, H., Kozuka-Hata, H., Shin, I. T., Minakuchi, Y., Ohishi, K., Motoyama, A., Aizu, T., Enomoto, A., Kondo, K., Tanaka, S., Hara, Y., Koshikawa, S., Sagara, H., Miura, T., Yokobori, S. I., Miyagawa, K., Suzuki, Y., Kubo, T., Oyama, M., Kohara, Y., Fujiyama, A., Arakawa, K., Katayama, T., Toyoda, A., and Kunieda, T. (2016) Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. *Nat Commun*, **7**:12808.
- Hashimoto, T. and Kunieda, T. (2017) DNA protection protein, a novel mechanism of radiation tolerance: lessons from tardigrades. *Life (Basel)*, **7**(26):1–11.
- Hatanaka, R., Gusev, O., Cornette, R., Shimura, S., Kikuta, S., Okada, J., Okuda, T., and Kikawada, T. (2015) Diversity of the expression profiles of late embryogenesis abundant (LEA) protein encoding genes in the anhydrobiotic midge *Polypedilum vanderplanki*. *Planta*, **242**(2):451–459.
- Heidemann, N. W. T., Smith, D. K., Hygum, T. L., Stapane, L., Clausen, L. K. B., Jørgensen, A., Hélix-Nielsen, C., and Møbjerg, N. (2016) Osmotic stress tolerance in semi-terrestrial tardigrades. *Zool J Linn Soc*, **178**(4):912–918.
- Hengherr, S., Brummer, F., and Schill, R. O. (2008a) Anhydrobiosis in tardigrades and its effects on longevity traits. *J Zoology*, **275**(3):216–220.
- Hengherr, S., Heyer, A. G., Kohler, H. R., and Schill, R. O. (2008b) Trehalose and anhydrobiosis in tardigrades—evidence for divergence in responses to dehydration. *FEBS J*, **275**(2):281–288.
- Hengherr, S., Worland, M. R., Reuner, A., Brummer, F., and Schill, R. O. (2009a) Freeze tolerance, supercooling points and ice formation: comparative studies on the subzero temperature survival of limno-terrestrial tardigrades. *J Exp Biol*, **212**(Pt 6):802–807.
- Hengherr, S., Worland, M. R., Reuner, A., Brummer, F., and Schill, R. O. (2009b) High-temperature tolerance in anhydrobiotic tardigrades is limited by glass transition. *Physiol Biochem Zool*, **82**(6):749–755.
- Hengherr, S., Reuner, A., Brummer, F., and Schill, R. O. (2010) Ice crystallization and freeze tolerance in embryonic stages of the tardigrade *Milnesium tardigradum*. *Comp Biochem Physiol A Mol Integr Physiol*, **156**(1):151–155.
- Herdeiro, R. S., Pereira, M. D., Panek, A. D., and Eleutherio, E. C. (2006) Trehalose protects *Saccharomyces cerevisiae* from lipid peroxidation during oxidative stress. *Biochim Biophys Acta*, **1760**(3):340–346.

- Hespeels, B., Knapen, M., Hanot-Mambres, D., Heuskin, A. C., Pineux, F., Lucas, S., Koszul, R., and Van Doninck, K. (2014) Gateway to genetic exchange? DNA double-strand breaks in the bdelloid rotifer *Adineta vaga* submitted to desiccation. *J Evol Biol*, **27**(7):1334–1345.
- Hinton, H. E. (1960) Cryptobiosis in the Larva of *Polypedilum vanderplanki* Hint. (Chironomidae). *J Insect Physiol*, **5**(3-4):286–300.
- Hirata, Y. (2002) Manganese-induced apoptosis in PC12 cells. *Neurotoxicol Teratol*, **24**(5):639–653.
- Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., and Stanke, M. (2016) BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*, **32**(5):767–769.
- Holm, L. (2020) DALI and the persistence of protein shape. *Protein Sci*, **29**(1):128–140.
- Holt, C. and Yandell, M. (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, **12**:491.
- Horikawa, D. D. and Higashi, S. (2004) Desiccation tolerance of the tardigrade *Milnesium tardigradum* collected in Sapporo, Japan, and Bogor, Indonesia. *Zoolog Sci*, **21**(8):813–816.
- Horikawa, D. D., Sakashita, T., Katagiri, C., Watanabe, M., Kikawada, T., Nakahara, Y., Hamada, N., Wada, S., Funayama, T., Higashi, S., Kobayashi, Y., Okuda, T., and Kuwabara, M. (2006) Radiation tolerance in the tardigrade *Milnesium tardigradum*. *Int J Radiat Biol*, **82**(12):843–848.
- Horikawa, D. D., Kunieda, T., Abe, W., Watanabe, M., Nakahara, Y., Yukuhiro, F., Sakashita, T., Hamada, N., Wada, S., Funayama, T., Katagiri, C., Kobayashi, Y., Higashi, S., and Okuda, T. (2008) Establishment of a rearing system of the extremotolerant tardigrade *Ramazzottius varieornatus*: a new model animal for astrobiology. *Astrobiology*, **8**(3):549–556.
- Horikawa, D. D., Iwata, K., Kawai, K., Koseki, S., Okuda, T., and Yamamoto, K. (2009) High hydrostatic pressure tolerance of four different anhydrobiotic animal species. *Zoolog Sci*, **26**(3):238–242.
- Horikawa, D. D., Yamaguchi, A., Sakashita, T., Tanaka, D., Hamada, N., Yukuhiro, F., Kuwahara, H., Kunieda, T., Watanabe, M., Nakahara, Y., Wada, S., Funayama, T., Katagiri, C., Higashi, S., Yokobori, S., Kuwabara, M., Rothschild, L. J., Okuda, T., Hashimoto, H., and Kobayashi, Y. (2012) Tolerance of anhydrobiotic eggs of the tardigrade *Ramazzottius varieornatus* to extreme environments. *Astrobiology*, **12**(4):283–289.
- Horikawa, D. D., Cumbers, J., Sakakibara, I., Rogoff, D., Leuko, S., Harnoto, R., Arakawa, K., Katayama, T., Kunieda, T., Toyoda, A., Fujiyama, A., and Rothschild, L. J. (2013) Analysis of DNA repair and protection in the tardigrade *Ramazzottius varieornatus* and *Hypsibius dujardini* after exposure to UVC radiation. *PLoS One*, **8**(6):e64793.
- Howe, K. L., Bolt, B. J., Cain, S., Chan, J., Chen, W. J., Davis, P., Done, J., Down, T., Gao, S., Grove, C., Harris, T. W., Kishore, R., Lee, R., Lomax, J., Li, Y., Muller, H. M., Nakamura, C., Nuin, P., Paulini, M., Raciti, D., Schindelman, G., Stanley, E., Tuli, M. A., Van Auken, K., Wang, D., Wang, X., Williams, G., Wright, A., Yook, K., Berriman, M., Kersey, P., Schedl, T., Stein, L., and Sternberg, P. W. (2016a) WormBase 2016: expanding to enable helminth genomic research. *Nucleic Acids Res*, **44**(D1):D774–D780.
- Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P., and Berriman, M. (2016b) WormBase ParaSite - a comprehensive resource for helminth genomics. *Mol Biochem Parasitol*, **215**:2–10.
- Huang, X. (1999) CAP3: a DNA sequence assembly program. *Genome Res*, **9**(9):868–877.
- Hubley, R., Finn, R. D., Clements, J., Eddy, S. R., Jones, T. A., Bao, W., Smit, A. F., and Wheeler, T. J. (2016) The Dfam database of repetitive DNA families. *Nucleic Acids Res*, **44**(D1):D81–D89.

- Husnik, F., Nikoh, N., Koga, R., Ross, L., Duncan, R. P., Fujie, M., Tanaka, M., Satoh, N., Bachtrog, D., Wilson, A. C., von Dohlen, C. D., Fukatsu, T., and McCutcheon, J. P. (2013) Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell*, **153**(7):1567–1578.
- Husnik, F. and McCutcheon, J. P. (2016) Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. *Proc Natl Acad Sci U S A*, **113**(37):E5416–E5424.
- Husnik, F. and McCutcheon, J. P. (2018) Functional horizontal gene transfer from bacteria to eukaryotes. *Nat Rev Microbiol*, **16**(2):67–79.
- Hygum, T. L., Clausen, L. K. B., Halberg, K. A., Jørgensen, A., and Møbjerg, N. (2016) Tun formation is not a prerequisite for desiccation tolerance in the marine tidal tardigrade *Echiniscoides sigismundi*. *Zool J Linn Soc*, **178**(4):907–911.
- Hygum, T. L., Fobian, D., Kamilari, M., Jørgensen, A., Schiøtt, M., Grosell, M., and Møbjerg, N. (2017) Comparative investigation of copper tolerance and identification of putative tolerance related genes in tardigrades. *Front Physiol*, **8**:95.
- Ito, M., Saigo, T., Abe, W., Kubo, T., and Kunieda, T. (2016) Establishment of an isogenic strain of the desiccation-sensitive tardigrade *Isohypsibius myrops* (Parachela, Eutardigrada) and its life history traits. *Zool J Linn Soc*, **178**(4):863–870.
- Janis, B., Belott, C., and Menze, M. A. (2018) Role of intrinsic disorder in animal desiccation tolerance. *Proteomics*, **18**:e1800067.
- Janssen, R., Eriksson, B. J., Tait, N. N., and Budd, G. E. (2014) Onychophoran Hox genes and the evolution of arthropod Hox gene expression. *Front Zool*, **11**(1):22.
- Jiang, Z., Hu, Z., Zeng, L., Lu, W., Zhang, H., Li, T., and Xiao, H. (2011) The role of the Golgi apparatus in oxidative stress: is this organelle less significant than mitochondria? *Free Radic Biol Med*, **50**(8):907–917.
- Jiang, H., Lei, R., Ding, S. W., and Zhu, S. (2014) Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*, **15**:182.
- Jönsson, K. I. (2003) Causes and consequences of excess resistance in cryptobiotic metazoans. *Physiol Biochem Zool*, **76**(4):429–435.
- Jönsson, K. I., Harms-Ringdahl, M., and Torudd, J. (2005) Radiation tolerance in the eutardigrade *Richtersius coronifer*. *Int J Radiat Biol*, **81**(9):649–656.
- Jönsson, K. I. and Schill, R. O. (2007) Induction of Hsp70 by desiccation, ionising radiation and heat-shock in the eutardigrade *Richtersius coronifer*. *Comp Biochem Physiol B Biochem Mol Biol*, **146**(4):456–460.
- Jönsson, K. I. (2007) Tardigrades as a potential model organism in space research. *Astrobiology*, **7**(5):757–766.
- Jönsson, K. I., Rabbow, E., Schill, R. O., Harms-Ringdahl, M., and Rettberg, P. (2008) Tardigrades survive exposure to space in low Earth orbit. *Curr Biol*, **18**(17):R729–R731.
- Jönsson, K. I. and Persson, O. (2010) Trehalose in three species of desiccation tolerant tardigrades. *Open Biol*, **3**(1):1–5.
- Jönsson, I., Beltran-Pardo, E., Haghdoost, S., Wojcik, A., Bermúdez-Cruz, R. M., Bernal Villegas, J. E., and Harms-Ringdahl, M. (2013) Tolerance to gamma-irradiation in eggs of the tardigrade *Richtersius coronifer* depends on stage of development. *J Limnol*, **72**(1s):73–79.
- Jönsson, K. I., Hygum, T. L., Andersen, K. N., Clausen, L. K., and Møbjerg, N. (2016) Tolerance to gamma radiation in the marine heterotardigrade, *Echiniscoides sigismundi*. *PLoS One*, **11**(12):e0168884.

- Jönsson, K. I. and Wojcik, A. (2017) Tolerance to X-rays and heavy ions (Fe, He) in the tardigrade *Richtersius coronifer* and the bdelloid rotifer *Mniobia russeola*. *Astrobiology*, **17**(2):163–167.
- Jönsson, K. I. (2019) Radiation tolerance in tardigrades: current knowledge and potential applications in medicine. *Cancers (Basel)*, **11**(9):1333.
- Kabsch, W. (2010) Xds. *Acta Crystallogr D Biol Crystallogr*, **66**(Pt 2):125–132.
- Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., Yabana, M., Harada, M., Nagayasu, E., Maruyama, H., Kohara, Y., Fujiyama, A., Hayashi, T., and Itoh, T. (2014) Efficient *de novo* assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res*, **24**(8):1384–1395.
- Kamilari, M., Jørgensen, A., Schiøtt, M., and Møbjerg, N. (2019) Comparative transcriptomics suggest unique molecular adaptations within tardigrade lineages. *BMC Genomics*, **20**(1):607.
- Karolson, P. (1996) On the hormonal control of insect metamorphosis. A historical review. *Int J Dev Biol*, **40**(1):93–96.
- Kato, Y., Kobayashi, K., Oda, S., Tatarazako, N., Watanabe, H., and Iguchi, T. (2007) Cloning and characterization of the ecdysone receptor and ultraspiracle protein from the water flea *Daphnia magna*. *J Endocrinol*, **193**(1):183–194.
- Katoh, K. and Standley, D. M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*, **30**(4):772–780.
- Kayukawa, T., Jouraku, A., Ito, Y., and Shinoda, T. (2017) Molecular mechanism underlying juvenile hormone-mediated repression of precocious larval–adult metamorphosis. *Proc Natl Acad Sci U S A*, **114**(5):1057–1062.
- Keilin, D. (1959) The Leeuwenhoek lecture - the problem of anabiosis or latent life - history and current concept. *Proc R Soc B*, **150**(939):149–191.
- Keller, O., Kollmar, M., Stanke, M., and Waack, S. (2011) A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics*, **27**(6):757–763.
- Kikawada, T., Saito, A., Kanamori, Y., Nakahara, Y., Iwata, K., Tanaka, D., Watanabe, M., and Okuda, T. (2007) Trehalose transporter 1, a facilitated and high-capacity trehalose transporter, allows exogenous trehalose uptake into cells. *Proc Natl Acad Sci U S A*, **104**(28):11585–11590.
- Kikawada, T., Saito, A., Kanamori, Y., Fujita, M., Snigorska, K., Watanabe, M., and Okuda, T. (2008) Dehydration-inducible changes in expression of two aquaporins in the sleeping chironomid, *Polypedilum vanderplanki*. *Biochim Biophys Acta*, **1778**(2):514–520.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*, **14**(4):R36.
- Kirke, J., Jin, X. L., and Zhang, X. H. (2020) Expression of a tardigrade Dsup gene enhances genome protection in plants. *Mol Biotechnol*, **62**(11-12):563–571.
- Kondo, K., Kubo, T., and Kunieda, T. (2015) Suggested involvement of PP1/PP2A activity and *de novo* gene expression in anhydrobiotic survival in a tardigrade, *Hypsibius dujardini*, by chemical genetic approach. *PLoS One*, **10**(12):e0144803.
- Kondo, K., Mori, M., Tomita, M., and Arakawa, K. (2019) AMPK activity is required for the induction of anhydrobiosis in a tardigrade *Hypsibius exemplaris*, and its potential up-regulator is PP2A. *Genes Cells*, **24**(12):768–780.

- Kondo, K., Mori, M., Tomita, M., and Arakawa, K. (2020) Pre-treatment with D942, a furancarboxylic acid derivative, increases desiccation tolerance in an anhydrobiotic tardigrade *Hypsibius exemplaris*. *FEBS Open Bio*, **10**(9):1774–1781.
- Koonin, E. V. (2015) Archaeal ancestors of eukaryotes: not so elusive any more. *BMC Biol*, **13**:84.
- Koshland, D. E., J. (2002) Special essay. The seven pillars of life. *Science*, **295**(5563):2215–2216.
- Koutsovoulos, G., Kumar, S., Laetsch, D. R., Stevens, L., Daub, J., Conlon, C., Maroon, H., Thomas, F., Aboobaker, A. A., and Blaxter, M. (2016) No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*. *Proc Natl Acad Sci U S A*, **113**(18):5053–5058.
- Kozioł, U. (2018) Precursors of neuropeptides and peptide hormones in the genomes of tardigrades. *Gen Comp Endocrinol*, **267**:116–127.
- Kozomara, A. and Griffiths-Jones, S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*, **42**(Database issue):D68–73.
- Ku, C., Nelson-Sathi, S., Roettger, M., Garg, S., Hazkani-Covo, E., and Martin, W. F. (2015a) Endosymbiotic gene transfer from prokaryotic pangenomes: inherited chimerism in eukaryotes. *Proc Natl Acad Sci U S A*, **112**(33):10139–10146.
- Ku, C., Nelson-Sathi, S., Roettger, M., Sousa, F. L., Lockhart, P. J., Bryant, D., Hazkani-Covo, E., McInerney, J. O., Landan, G., and Martin, W. F. (2015b) Endosymbiotic origin and differential loss of eukaryotic genes. *Nature*, **524**(7566):427–432.
- Kuck, P. and Longo, G. C. (2014) FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front Zool*, **11**(1):81.
- Kumar, S., Jones, M., Koutsovoulos, G., Clarke, M., and Blaxter, M. (2013) Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Front Genet*, **4**:237.
- Kuzmic, M., Richaud, M., Cuq, P., Frelon, S., and Galas, S. (2018) Carbonylation accumulation of the *Hypsibius exemplaris* anhydrobiote reveals age-associated marks. *PLoS One*, **13**(12):e0208617.
- Laetsch, D. R. and Blaxter, M. L. (2017) KinFin: software for taxon-aware analysis of clustered protein sequences. *G3 (Bethesda)*, **7**(10):3349–3357.
- Lagesen, K., Hallin, P., Rodland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. W. (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res*, **35**(9):3100–3108.
- Lammer, H., Bredehöft, J. H., Coustenis, A., Khodachenko, M. L., Kaltenecker, L., Grasset, O., Prieur, D., Raulin, F., Ehrenfreund, P., Yamauchi, M., Wahlund, J. E., Griebmeier, J. M., Stangl, G., Cockell, C. S., Kulikov, Y. N., Grenfell, J. L., and Rauer, H. (2009) What makes a planet habitable? *Astron Astrophys Rev*, **17**(2):181–249.
- Lance, D. (1896) *Contribution à l'étude anatomique et biologique des tardigrades (genre Macrobiotus Schultze)*. Thesis, Impr. de E. Mauchaussat.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Showkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A.,

- Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al. (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**(6822):860–921.
- Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc*, **3**(7):1171–1179.
- Langmead, B. and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*, **9**(4):357–359.
- Lapinski, J. and Tunnacliffe, A. (2003) Anhydrobiosis without trehalose in bdelloid rotifers. *FEBS Letters*, **553**(3):387–390.
- Lartillot, N. and Philippe, H. (2004) A bayesian mixture model for across-site heterogeneities in the amino acid replacement process. *Mol Biol Evol*, **21**(6):1095–1109.
- Laumer, C. E., Fernandez, R., Lemer, S., Combosch, D., Kocot, K. M., Riesgo, A., Andrade, S. C. S., Sterrer, W., Sorensen, M. V., and Giribet, G. (2019) Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proc Biol Sci*, **286**(1906):20190831.
- Lee, C.-Y., Conrad, M. N., and Dresser, M. E. (2012) Meiotic chromosome pairing is promoted by telomere-led chromosome movements independent of bouquet formation. *PLoS Genetics*, **8**(5):e1002730.
- Letunic, I. and Bork, P. (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*, **47**(W1):W256–W259.
- Levin, M., Anavy, L., Cole, A. G., Winter, E., Mostov, N., Khair, S., Senderovich, N., Kovalev, E., Silver, D. H., Feder, M., Fernandez-Valverde, S. L., Nakanishi, N., Simmons, D., Simakov, O., Larsson, T., Liu, S. Y., Jerafi-Vider, A., Yaniv, K., Ryan, J. F., Martindale, M. Q., Rink, J. C., Arendt, D., Degnan, S. M., Degnan, B. M., Hashimshony, T., and Yanai, I. (2016) The mid-developmental transition and the evolution of animal body plans. *Nature*, **531**(7596):637–641.
- Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, **25**(14):1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**(16):2078–2079.
- Li, K., Jia, Q. Q., and Li, S. (2019) Juvenile hormone signaling - a mini review. *Insect Sci*, **26**(4):600–606.
- Liguori, I., Russo, G., Curcio, F., Bulli, G., Aran, L., Della-Morte, D., Gargiulo, G., Testa, G., Cacciatore, F., Bonaduce, D., and Abete, P. (2018) Oxidative stress, aging, and diseases. *Clin Interv Aging*, **13**:757–772.
- Love, M. I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, **15**(12):550.
- Lowe, T. M. and Eddy, S. R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*, **25**(5):955–964.
- Madin, K. A. C. and Crowe, J. H. (1975) Anhydrobiosis in nematodes: carbohydrate and lipid metabolism during dehydration. *J Exp Zool*, **193**(3):335–342.
- Mali, B., Grohme, M. A., Förster, F., Dandekar, T., Schnölzer, M., Reuter, D., Welnicz, W., Schill, R. O., and Frohme, M. (2010) Transcriptome survey of the anhydrobiotic tardigrade *Milnesium tardigradum* in comparison with *Hypsibius dujardini* and *Richtersius coronifer*. *BMC Genomics*, **11**(168):1–11.

- Mapalo, M. A., Arakawa, K., Baker, C. M., Persson, D. K., Mirano-Bascos, D., and Giribet, G. (2020) The unique antimicrobial recognition and signaling pathways in tardigrades with a comparison across ecdysozoa. *G3 (Bethesda)*, **10**(3):1137–1148.
- Marcus, E. (1929) *Tardigrada*, volume 5 (4, 3) of *Klassen und ordnungen des thier-reichs: wissenschaftlich dargestellt in wort und nild*. Akademische Verlagsgesellschaft, Leipzig.
- Marcus, E. d. B.-R. (1944) Notes on fresh-water Oligochaeta from Brazil. *Comun Zool Mus Montevideo*, **1**(20):1–8.
- Martin, C., Gross, V., Pfluger, H. J., Stevenson, P. A., and Mayer, G. (2017) Assessing segmental versus non-segmental features in the ventral nervous system of onychophorans (velvet worms). *BMC Evol Biol*, **17**(1):3.
- Marunde, M. R., Samarajeewa, D. A., Anderson, J., Li, S., Hand, S. C., and Menze, M. A. (2013) Improved tolerance to salt and water stress in *Drosophila melanogaster* cells conferred by late embryogenesis abundant protein. *J Insect Physiol*, **59**(4):377–386.
- May, M. (1953) L'évolution des Tardigrades de la vie aquatique à la vie terrestre. *Bulletin Français de Pisciculture*, **168**:93–100.
- May, R. M., Maria, M., and Gumard, J. (1964) Action différentielle des rayons x et ultraviolets sur le tardigrade *Macrobiotus areolatus*, a l'état actif et desséché. *Bull Biol Fr Belg*, **98**:349–367.
- Mazin, P. V., Shagimardanova, E., Kozlova, O., Cherkasov, A., Sutormin, R., Stepanova, V. V., Stupnikov, A., Logacheva, M., Penin, A., Sogame, Y., Cornette, R., Tokumoto, S., Miyata, Y., Kikawada, T., Gelfand, M. S., and Gusev, O. (2018) Cooption of heat shock regulatory system for anhydrobiosis in the sleeping chironomid *Polypedium vanderplanki*. *Proc Natl Acad Sci U S A*, **115**(10):E2477–E2486.
- Metwally, H. M. S., Abdel-Hakim, E., and Abdou, W. L. (2017) Effects of the juvenoid (Fenoxycarb) on entomopathogenic nematode *Steinernema carpocapsae* S2 strain. *Biosci Res*, **14**(4):966–975.
- Minguez-Toral, M., Cuevas-Zuñiga, B., Garrido-Arandia, M., and Pacios, L. F. (2020) A computational structural study on the DNA-protecting role of the tardigrade-unique Dsup protein. *Sci Rep*, **10**(1):13424.
- Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., and Punta, M. (2013) Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res*, **41**(12):e121.
- Mitreva, M., Blaxter, M. L., Bird, D. M., and McCarter, J. P. (2005) Comparative genomics of nematodes. *Trends Genet*, **21**(10):573–581.
- Miyata, Y., Fuse, H., Tokumoto, S., Hiki, Y., Deviatiiarov, R., Yoshida, Y., Yamada, T. G., Cornette, R., Gusev, O., Shagimardanova, E., Funahashi, A., and Kikawada, T. (2020) Cas9-mediated genome editing reveals a significant contribution of calcium signaling pathways to anhydrobiosis in Pv11. *bioRxiv*, **2020.10.15.340281**.
- Mock, T., Otiillar, R. P., Strauss, J., McMullan, M., Paajanen, P., Schmutz, J., Salamov, A., Sanges, R., Toseland, A., Ward, B. J., Allen, A. E., Dupont, C. L., Frickenhaus, S., Maumus, F., Veluchamy, A., Wu, T., Barry, K. W., Falciatore, A., Ferrante, M. I., Fortunato, A. E., Glockner, G., Gruber, A., Hipkin, R., Janech, M. G., Kroth, P. G., Leese, F., Lindquist, E. A., Lyon, B. R., Martin, J., Mayer, C., Parker, M., Quesneville, H., Raymond, J. A., Uhlig, C., Valas, R. E., Valentin, K. U., Worden, A. Z., Armbrust, E. V., Clark, M. D., Bowler, C., Green, B. R., Moulton, V., van Oosterhout, C., and Grigoriev, I. V. (2017) Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature*, **541**(7638):536–540.
- Moore, D. S., Hansen, R., and Hand, S. C. (2016) Liposomes with diverse compositions are protected during desiccation by LEA proteins from *Artemia franciscana* and trehalose. *Biochim Biophys Acta*, **1858**(1):104–115.
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*, **35**(Web Server issue):W182–W185.

- Murai, Y., Yagi-Utsumi, M., Fujiwara, M., Tomita, M., Kato, K., and Arakawa, K. (2020) Multiomics study of a heterotardigrade, *Echiniscus testudo*, suggests convergent evolution of anhydrobiosis-related proteins in Tardigrada. *bioRxiv*, **2020.10.27.358333**.
- Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*, **53**(Pt 3):240–255.
- Needham, J. (1743) Concerning chalky tubulous concretions called malm: with some microscopical observations on the farina of red lily, and of worms discovered in smutty corn. *Phil Trans*, **42**:634–641.
- Nelson, D. R. (2002) Current status of the tardigrada: evolution and ecology. *Integr Comp Biol*, **42**(3):652–659.
- Nester, E. W., Gordon, M. P., Amasino, R. M., and Yanofsky, M. F. (1984) Crown gall - a molecular and physiological analysis. *Annu Rev Plant Physiol*, **35**:387–413.
- Neumann, S., Reuner, A., Brummer, F., and Schill, R. O. (2009) DNA damage in storage cells of anhydrobiotic tardigrades. *Comp Biochem Physiol A Mol Integr Physiol*, **153**(4):425–429.
- Nguyen Ba, A. N., Pogoutse, A., Provar, N., and Moses, A. M. (2009) NLStradamus: a simple hidden markov model for nuclear localization signal prediction. *BMC Bioinformatics*, **10**(202):1–11.
- Nielsen, C. (2013) The triradiate sucking pharynx in animal phylogeny. *Invertebr Biol*, **132**(1):1–13.
- Nikoh, N., McCutcheon, J. P., Kudo, T., Miyagishima, S. Y., Moran, N. A., and Nakabachi, A. (2010) Bacterial genes in the aphid genome: absence of functional gene transfer from *Buchnera* to its host. *PLoS Genet*, **6**(2):e1000827.
- Nilsson, E. J., Jönsson, K. I., and Pallon, J. (2010) Tolerance to proton irradiation in the eutardigrade *Richtersius coronifer*—a nuclear microprobe study. *Int J Radiat Biol*, **86**(5):420–427.
- Nowack, E. C., Price, D. C., Bhattacharya, D., Singer, A., Melkonian, M., and Grossman, A. R. (2016) Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of *Paulinella chromatophora*. *Proc Natl Acad Sci U S A*, **113**(43):12214–12219.
- Nowell, R. W., Almeida, P., Wilson, C. G., Smith, T. P., Fontaneto, D., Crisp, A., Micklem, G., Tunnacliffe, A., Boschetti, C., and Barraclough, T. G. (2018) Comparative genomics of bdelloid rotifers: insights from desiccating and nondesiccating species. *PLoS Biol*, **16**(4):e2004830.
- Nowell, R. W., Wilson, C. G., Almeida, P., Schiffer, P. H., Fontaneto, D., Becks, L., Rodriguez, F., Arkhipova, I. R., and Barraclough, T. G. (2020) Evolutionary dynamics of transposable elements in bdelloid rotifers. *bioRxiv*, **2020.09.15.297952**.
- Okonechnikov, K., Conesa, A., and Garcia-Alcalde, F. (2016) Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, **32**(2):292–294.
- Okuda, S., Yamada, T., Hamajima, M., Itoh, M., Katayama, T., Bork, P., Goto, S., and Kanehisa, M. (2008) KEGG Atlas mapping for global analysis of metabolic pathways. *Nucleic Acids Res*, **36**(Web Server issue):W423–W426.
- O’Leary, N. A., Wright, M. W., Brister, J. R., Ciufu, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C. M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V. S., Kodali, V. K., Li, W., Maglott, D., Masterson, P., McGarvey, K. M., Murphy, M. R., O’Neill, K., Pujar, S., Rangwala, S. H., Rausch, D., Riddick, L. D., Schoch, C., Shkeda, A., Storz, S. S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R. E., Vatsan, A. R., Wallin, C., Webb, D., Wu, W., Landrum, M. J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T. D., and Pruitt, K. D. (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*, **44**(D1):D733–D745.

- Oliveira, R. P., Porter Abate, J., Dilks, K., Landis, J., Ashraf, J., Murphy, C. T., and Blackwell, T. K. (2009) Condition-adapted stress and longevity gene regulation by *Caenorhabditis elegans* SKN-1/Nrf. *Aging Cell*, **8**(5):524–541.
- Ono, F., Mori, Y., Takarabe, K., Fujii, A., Saigusa, M., Matsushima, Y., Yamazaki, D., Ito, E., Galas, S., Saini, N. L., and Ahuja, R. (2016) Effect of ultra-high pressure on small animals, tardigrades and *Artemia*. *Cogent Physics*, **3**(1):1167575.
- Parra, G., Bradnam, K., and Korf, I. (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, **23**(9):1061–1067.
- Pearson, W. R. and Lipman, D. J. (1988) Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A*, **85**(8):2444–2448.
- Persson, D., Halberg, K. A., Jørgensen, A., Ricci, C., Møbjerg, N., and Kristensen, R. M. (2011) Extreme stress tolerance in tardigrades: surviving space conditions in low earth orbit. *J Zool Syst Evol Res*, **49**:90–97.
- Pett, M. and Coleman, N. (2007) Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis? *J Pathol*, **212**(4):356–367.
- Pigon, A. and Weglarska, B. (1955) Rate of metabolism in tardigrades during active life and anabiosis. *Nature*, **176**(4472):121–122.
- Piskiewicz, S., Gunn, K. H., Warmuth, O., Propst, A., Mehta, A., Nguyen, K. H., Kuhlman, E., Guseman, A. J., Stadmiller, S. S., Boothby, T. C., Neher, S. B., and Pielak, G. J. (2019) Protecting activity of desiccated enzymes. *Protein Sci*, **28**(5):941–951.
- Pittis, A. A. and Gabaldon, T. (2016) Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature*, **531**(7592):101–104.
- Pizzino, G., Irrera, N., Cucinotta, M., Pallio, G., Mannino, F., Arcoraci, V., Squadrito, F., Altavilla, D., and Bitto, A. (2017) Oxidative stress: harms and benefits for human health. *Oxid Med Cell Longev*, **2017**:8416763.
- Polz, M. F., Alm, E. J., and Hanage, W. P. (2013) Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet*, **29**(3):170–175.
- Popendorf, K., Tsuyoshi, H., Osana, Y., and Sakakibara, Y. (2010) Murasaki: a fast, parallelizable algorithm to find anchors from multiple genomes. *PLoS One*, **5**(9):e12651.
- Pouchkina-Stantcheva, N. N., McGee, B. M., Boschetti, C., Tolleter, D., Chakrabortee, S., Popova, A. V., Meersman, F., Macherel, D., Hinch, D. K., and Tunnacliffe, A. (2007) Functional divergence of former alleles in an ancient asexual invertebrate. *Science*, **318**(5848):268–271.
- Preyer, W. (1891) Über die anabiose. *Biologisches Centralblatt*, **11**(1):1–5.
- Price, A. L., Jones, N. C., and Pevzner, P. A. (2005) *De novo* identification of repeat families in large genomes. *Bioinformatics*, **21 Suppl 1**:i351–i358.
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010) FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One*, **5**(3):e9490.
- Quinlan, A. R. and Hall, I. M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**(6):841–842.
- Rahm, G. (1921) Biologische und physiologische beiträge zur kenntnis de moosfauna. *Z allgem Physiol*, **29**:1–36.
- Rahm, G. (1926) Die trockenstarre (anabiose) der moostierwelt. *Biologisches Zentralblatt*, **46**(8):452–477.

- Rahm, G. (1937) Eine neue tardigraden-ordnung aus den heissen quellen von Unzen, insel Kyushu, Japan. *Zool Anz*, **120**(3-4):65–71.
- Ramazzotti, G. and Maucci, W. (1983) *II Phylum Tardigrada*, volume 41. Istituto Italiano di Idrobiologia, Istituto Italiano di Idrobiologia, Verbania Pallanza.
- Rambaut, A. (2016) FigTree. <http://tree.bio.ed.ac.uk/software/figtree>.
- Ramløy, H. and Westh, P. (2001) Cryptobiosis in the eutardigrade *Adorybiotus (Richtersius) coronifer*: tolerance to alcohols, temperature and *de novo* protein synthesis. *Zool Anz*, **240**(3-4):517–523.
- Rancurel, C., Legrand, L., and Danchin, E. G. J. (2017) Alieness: rapid detection of candidate horizontal gene transfers across the tree of life. *Genes (Basel)*, **8**(10):1–18.
- Ray, R. and Pandey, P. (2018) piRNA analysis framework from small RNA-Seq data by a novel cluster prediction tool - PILFER. *Genomics*, **110**(6):355–365.
- Raymond, J. A. and Kim, H. J. (2012) Possible role of horizontal gene transfer in the colonization of sea ice by algae. *PLoS One*, **7**(5):e35968.
- Rebecchi, L., Cesari, M., Altiero, T., Frigieri, A., and Guidetti, R. (2009) Survival and DNA degradation in anhydrobiotic tardigrades. *J Exp Biol*, **212**(Pt 24):4033–4039.
- Rebecchi, L. (2013) Dry up and survive: the role of antioxidant defences in anhydrobiotic organisms. *J Limnol*, **72**:62–72.
- Reddi, A. R., Jensen, L. T., and Culotta, V. C. (2009) Manganese homeostasis in *Saccharomyces cerevisiae*. *Chem Rev*, **109**(10):4722–4732.
- Regier, J. C., Shultz, J. W., Kambic, R. E., and Nelson, D. R. (2005) Robust support for tardigrade clades and their ages from three protein-coding nuclear genes. *Invertebr Biol*, **123**(2):93–100.
- Renaud-Mornant, J. (1982) *Species diversity in marine Tardigrada*, pages 149–178. East Tennessee State University Press, Johnson City, Tennessee.
- Reuner, A., Hengherr, S., Mali, B., Förster, F., Arndt, D., Reinhardt, R., Dandekar, T., Frohme, M., Brummer, F., and Schill, R. O. (2010) Stress response in tardigrades: differential gene expression of molecular chaperones. *Cell Stress Chaperones*, **15**(4):423–430.
- Ricard, G., McEwan, N. R., Dutilh, B. E., Jouany, J. P., Macheboeuf, D., Mitsumori, M., McIntosh, F. M., Michalowski, T., Nagamine, T., Nelson, N., Newbold, C. J., Nsabimana, E., Takenaka, A., Thomas, N. A., Ushida, K., Hackstein, J. H., and Huynen, M. A. (2006) Horizontal gene transfer from bacteria to rumen ciliates indicates adaptation to their anaerobic, carbohydrates-rich environment. *BMC Genomics*, **7**:22.
- Ricci, C. and M., P. (1997) Desiccation of *Panagrolaimus rigidus* (Nematoda): survival, reproduction and the influence on the internal clock. *Hydrobiologia*, **347**:1–13.
- Ricci, C. (1998) Anhydrobiotic capabilities of bdelloid rotifers. *Hydrobiologia*, **387**:321–326.
- Ricci, C. and Covino, C. (2005) Anhydrobiosis of *Adineta ricciae*: costs and benefits. *Hydrobiologia*, **546**(1):307–314.
- Richards, T. A., Leonard, G., Soanes, D. M., and Talbot, N. J. (2011) Gene transfer into the fungi. *Fungal Biol Rev*, **25**(2):98–110.
- Richaud, M., Le Goff, E., Cazevielle, C., Ono, F., Mori, Y., Saini, N. L., Cuq, P., Baghdiguian, S., Godefroy, N., and Galas, S. (2020) Ultrastructural analysis of the dehydrated tardigrade *Hypsibius exemplaris* unveils an anhydrobiotic-specific architecture. *Sci Rep*, **10**(4324):1–11.
- Rieger, K. E. and Chu, G. (2004) Portrait of transcriptional responses to ultraviolet and ionizing radiation in human cells. *Nucleic Acids Res*, **32**(16):4786–4803.

- Rizzo, A. M., Negroni, M., Altiero, T., Montorfano, G., Corsetto, P., Berselli, P., Berra, B., Guidetti, R., and Rebecchi, L. (2010) Antioxidant defences in hydrated and desiccated states of the tardigrade *Paramacrobiotus richtersi*. *Comp Biochem Physiol B Biochem Mol Biol*, **156**(2):115–121.
- Rokas, A. and Holland, P. W. H. (2000) Rare genomic changes as a tool for phylogenetics. *Trends Ecol Evol*, **15**(11):454–459.
- Rosenkranz, D. and Zischler, H. (2012) proTRAC—a software for probabilistic piRNA cluster detection, visualization and analysis. *BMC Bioinformatics*, **13**(5):1–10.
- Ryabova, A., Cornette, R., Cherkasov, A., Watanabe, M., Okuda, T., Shagimardanova, E., Kikawada, T., and Gusev, O. (2020) Combined metabolome and transcriptome analysis reveals key components of complete desiccation tolerance in an anhydrobiotic insect. *Proc Natl Acad Sci U S A*, **117**(32):19209–19220.
- Sakurai, M., Furuki, T., Akao, K., Tanaka, D., Nakahara, Y., Kikawada, T., Watanabe, M., and Okuda, T. (2008) Vitrification is essential for anhydrobiosis in an African chironomid, *Polypedilum vanderplanki*. *Proc Natl Acad Sci U S A*, **105**(13):5093–5098.
- Salzberg, S. L., White, O., Peterson, J., and Eisen, J. A. (2001) Microbial genes in the human genome: lateral transfer or gene loss? *Science*, **292**(5523):1903–1906.
- Salzberg, S. L. (2017) Horizontal gene transfer is not a hallmark of the human genome. *Genome Biol*, **18**(1):85.
- Sarkies, P., Selkirk, M. E., Jones, J. T., Blok, V., Boothby, T., Goldstein, B., Hanelt, B., Ardila-Garcia, A., Fast, N. M., Schiffer, P. M., Kraus, C., Taylor, M. J., Koutsovoulos, G., Blaxter, M. L., and Miska, E. A. (2015) Ancient and novel small RNA pathways compensate for the loss of piRNAs in multiple independent nematode lineages. *PLoS Biol*, **13**(2):e1002061.
- Savic, A. G., Guidetti, R., Turi, A., Pavicevic, A., Giovannini, I., Rebecchi, L., and Mojovic, M. (2015) Superoxide anion radical production in the tardigrade *Paramacrobiotus richtersi*, the first electron paramagnetic resonance spin-trapping study. *Physiol Biochem Zool*, **88**(4):451–454.
- Savory, F., Leonard, G., and Richards, T. A. (2015) The role of horizontal gene transfer in the evolution of the oomycetes. *PLoS Pathog*, **11**(5):e1004805.
- Schiffer, P. H., Danchin, E. G. J., Burnell, A. M., Creevey, C. J., Wong, S., Dix, I., O’Mahony, G., Culleton, B. A., Rancurel, C., Stier, G., Martinez-Salazar, E. A., Marconi, A., Trivedi, U., Kroiher, M., Thorne, M. A. S., Schierenberg, E., Wiehe, T., and Blaxter, M. (2019) Signatures of the evolution of parthenogenesis and cryptobiosis in the genomes of panagrolaimid nematodes. *iScience*, **21**:587–602.
- Schill, R. O., Steinbruck, G. H. B., and Kohler, H.-R. (2004) Stree gene (Hsp70) sequences and quantitative expression in *Milnesium tardigradum* (Tardigrada) during active and cryptobiotic stages. *J Exp Biol*, **207**:1607–1613.
- Schill, R. O., Huhn, F., and Köhler, H.-R. (2007) The first record of tardigrades (Tardigrada) from the Sinai Peninsula, Egypt. *Zool Middle East*, **42**(1):83–88.
- Schill, R. O. and Fritz, G. B. (2008) Desiccation tolerance in embryonic stages of the tardigrade. *J Zool*, **276**(1):103–107.
- Schill, R. O. (2013) Life-history traits in the tardigrade species *Paramacrobiotus kenianus* and *Paramacrobiotus palaui*. *J Limnol*, **72**(1s):e20.
- Schmitt, H. (1973) Mitochondrial biogenesis during differentiation of *Artemia salina* cysts. *J Cell Biol*, **58**(3):643–649.
- Schokraie, E., Hotz-Wagenblatt, A., Warnken, U., Mali, B., Frohme, M., Förster, F., Dandekar, T., Hengherr, S., Schill, R. O., and Schnölzer, M. (2010) Proteomic analysis of tardigrades: towards a better understanding of molecular mechanisms by anhydrobiotic organisms. *PLoS One*, **5**(3):e9502.

- Schokraie, E., Warnken, U., Hotz-Wagenblatt, A., Grohme, M. A., Hengherr, S., Förster, F., Schill, R. O., Frohme, M., Dandekar, T., and Schnölzer, M. (2012) Comparative proteome analysis of *Milnesium tardigradum* in early embryonic state versus adults in active and anhydrobiotic state. *PLoS One*, **7**(9):e45682.
- Schonknecht, G., Chen, W. H., Ternes, C. M., Barbier, G. G., Shrestha, R. P., Stanke, M., Brautigam, A., Baker, B. J., Banfield, J. F., Garavito, R. M., Carr, K., Wilkerson, C., Rensing, S. A., Gagneul, D., Dickenson, N. E., Oesterhelt, C., Lercher, M. J., and Weber, A. P. (2013) Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science*, **339**(6124):1207–1210.
- Schumann, I., Kenny, N., Hui, J., Hering, L., and Mayer, G. (2018) Halloween genes in panarthropods and the evolution of the early moulting pathway in Ecdysozoa. *R Soc Open Sci*, **5**(9):180888.
- Schuster, L. N. and Sommer, R. J. (2012) Expressional and functional variation of horizontally acquired cellulases in the nematode *Pristionchus pacificus*. *Gene*, **506**(2):274–282.
- Seki, K. and Toyoshima, M. (1998) Preserving tardigrades under pressure. *Nature*, **395**(6705):853–854.
- Sekimoto, T., Iwami, M., and Sakurai, S. (2006) Coordinate responses of transcription factors to ecdysone during programmed cell death in the anterior silk gland of the silkworm, *Bombyx mori*. *Insect Mol Biol*, **15**(3):281–292.
- Shabardina, V., Kashima, Y., Suzuki, Y., and Makalowski, W. (2020) Emergence and evolution of ERM proteins and merlin in metazoans. *Genome Biol Evol*, **12**(1):3710–3724.
- Shannon, A. J., Browne, J. A., Boyd, J., Fitzpatrick, D. A., and Burnell, A. M. (2005) The anhydrobiotic potential and molecular phylogenetics of species and strains of *Panagrolaimus* (Nematoda, Panagrolaimidae). *J Exp Biol*, **208**(Pt 12):2433–2445.
- Sheng, H. P. and Huggins, R. A. (1979) A review of body composition studies with emphasis on total body water and fat. *Am J Clin Nutr*, **32**(3):630–647.
- Siderakis, M. and Tarsounas, M. (2007) Telomere regulation and function during meiosis. *Chromosome Res*, **15**(5):667–679.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, **7**(539):1–6.
- Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**(19):3210–3212.
- Smit, A. F. A., Hubley, R., and Green, P. (2015) RepeatMasker Open-4.0. <http://www.repeatmasker.org>.
- Smith, F. W., Boothby, T. C., Giovannini, I., Rebecchi, L., Jockusch, E. L., and Goldstein, B. (2016) The compact body plan of tardigrades evolved by the loss of a large body region. *Curr Biol*, **26**(2):224–229.
- Smith, F. W. and Goldstein, B. (2016) Segmentation in Tardigrada and diversification of segmental patterns in Panarthropoda. *Arthropod Struct Dev*, **46**(3):328–340.
- Smith, F. W., Cumming, M., and Goldstein, B. (2018) Analyses of nervous system patterning genes in the tardigrade *Hypsibius exemplaris* illuminate the evolution of panarthropod brains. *EvoDevo*, **9**:19.
- Sobczyk, M., Michno, K., Kosztyla, P., Stec, D., and Michalczyk, Ł. (2015) Tolerance to ammonia of *Thulinus ruffoi* (Bertolani, 1981), a tardigrade isolated from a sewage treatment plant. *Bull Environ Contam Toxicol*, **95**(6):721–727.
- Sonobe, H. and Yamada, R. (2004) Ecdysteroids during early embryonic development in silkworm *Bombyx mori*: metabolism and functions. *Zoolog Sci*, **21**(5):503–516.

- Spallanzani, L. (1776) *Opuscoli di fisica animale, e vegetabile*. Società Tipografica, Modena.
- Spallanzani, L. (1777) *Opuscules de physique VI: animale et vegetale*, volume 2. Chez Pierre J. Duplain, libraire, Modena.
- Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., van Eijk, R., Schleper, C., Guy, L., and Ettema, T. J. G. (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, **521**(7551):173–179.
- Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**(9):1312–1313.
- Stanhope, M. J., Lupas, A., Italia, M. J., Koretke, K. K., Volker, C., and Brown, J. R. (2001) Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. *Nature*, **411**(6840):940–944.
- Stec, D., Arakawa, K., and Michalczyk, Ł. (2018) An integrative description of *Macrobiotus shonaicus* sp. nov. (Tardigrada: Macrobiotidae) from Japan with notes on its phylogenetic position within the Hufelandi group. *PLoS One*, **13**(2):e0192210.
- Stec, D., Krzywanski, L., Arakawa, K., and Michalczyk, L. (2020) A new redescription of *Richtersius coronifer*, supported by transcriptome, provides resources for describing concealed species diversity within the monotypic genus *Richtersius* (Eutardigrada). *Zoological Lett*, **6**:2.
- Su, R. Z., Kost, M., and Kiefer, J. (1990) The effect of krypton ions on *Artemia* cysts. *Radiat Environ Biophys*, **29**(2):103–107.
- Sugiura, K., Minato, H., Suzuki, A. C., Arakawa, K., Kunieda, T., and Matsumoto, M. (2019) Comparison of sexual reproductive behaviors in two species of macrobiotidae (Tardigrada: Eutardigrada). *Zoolog Sci*, **36**(2):120–127.
- Sugiura, K., Minato, H., Matsumoto, M., and Suzuki, A. C. (2020) *Milnesium* (Tardigrada: Apochela) in Japan: the first confirmed record of *Milnesium tardigradum* s.s. and description of *Milnesium pacificum* sp. nov. *Zoolog Sci*, **37**(5):476–495.
- Sulston, J. E., Schierenberg, E., White, J. G., and Thomson, J. N. (1983) The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev Biol*, **100**(1):64–119.
- Sung, W. K., Zheng, H., Li, S., Chen, R., Liu, X., Li, Y., Lee, N. P., Lee, W. H., Ariyaratne, P. N., Tennakoon, C., Mulawadi, F. H., Wong, K. F., Liu, A. M., Poon, R. T., Fan, S. T., Chan, K. L., Gong, Z., Hu, Y., Lin, Z., Wang, G., Zhang, Q., Barber, T. D., Chou, W. C., Aggarwal, A., Hao, K., Zhou, W., Zhang, C., Hardwick, J., Buser, C., Xu, J., Kan, Z., Dai, H., Mao, M., Reinhard, C., Wang, J., and Luk, J. M. (2012) Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet*, **44**(7):765–769.
- Suzuki, A. C. (2003) Life history of *Milnesium tardigradum* Doyère (Tardigrada) under a rearing environment. *Zoolog Sci*, **20**(1):49–57.
- Switala, J., O’Neil, J. O., and Loewen, P. C. (1999) Catalase HPII from *Escherichia coli* exhibits enhanced resistance to denaturation. *Biochemistry*, **38**(13):3895–3901.
- Tan, J. and MacRae, T. H. (2018) Stress tolerance in diapausing embryos of *Artemia franciscana* is dependent on heat shock factor 1 (Hsf1). *PLoS One*, **13**(7):e0200153.
- Tanaka, S., Tanaka, J., Miwa, Y., Horikawa, D. D., Katayama, T., Arakawa, K., Toyoda, A., Kubo, T., and Kunieda, T. (2015) Novel mitochondria-targeted heat-soluble proteins identified in the anhydrobiotic tardigrade improve osmotic tolerance of human cells. *PLoS One*, **10**(2):e0118272.
- Tatarazako, N., Oda, S., Watanabe, H., Morita, M., and Iguchi, T. (2003) Juvenile hormone agonists affect the occurrence of male *Daphnia*. *Chemosphere*, **53**(8):827–833.

- Tenlen, J. R., McCaskill, S., and Goldstein, B. (2013) RNA interference can be used to disrupt gene function in tardigrades. *Dev Genes Evol*, **223**(3):171–181.
- Treffkorn, S., Kahnke, L., Hering, L., and Mayer, G. (2018) Expression of NK cluster genes in the onychophoran *Euperipatoides rowelli*: implications for the evolution of NK family genes in nephrozoans. *EvoDevo*, **9**:1–17.
- Tsuboyama, K., Osaki, T., Matsuura-Suzuki, E., Kozuka-Hata, H., Okada, Y., Oyama, M., Ikeuchi, Y., Iwasaki, S., and Tomari, Y. (2020) A widespread family of heat-resistant obscure (Hero) proteins protect against protein instability and aggregation. *PLoS Biol*, **18**(3):e3000632.
- Tsujimoto, M., Suzuki, A. C., and Imura, S. (2015) Life history of the Antarctic tardigrade, *Acutuncus antarcticus*, under a constant laboratory environment. *Polar Biol*, **38**(10):1575–1581.
- UniProt Consortium (2015) UniProt: a hub for protein information. *Nucleic Acids Res*, **43**(Database issue):D204–D212.
- UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*, **47**(D1):D506–D515.
- van der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D. T., Kim, P. M., Kriwacki, R. W., Oldfield, C. J., Pappu, R. V., Tompa, P., Uversky, V. N., Wright, P. E., and Babu, M. M. (2014) Classification of intrinsically disordered regions and proteins. *Chem Rev*, **114**(13):6589–6631.
- van Leeuwenhoek, A. (1702) *On certain animalcules found in the sediment in gutters of the roofs of houses*, Letter 144, pages 207–213. The select works of Antony van Leeuwenhoek. G. Sidney, London.
- Vanfleteren, J. R. (1993) Oxidative stress and ageing in *Caenorhabditis elegans*. *Biochem J*, **292**(2):605–608.
- Vellichirammal, N. N., Gupta, P., Hall, T. A., and Brisson, J. A. (2017) Ecdysone signaling underlies the pea aphid transgenerational wing polyphenism. *Proc Natl Acad Sci U S A*, **114**(6):1419–1423.
- Vizueta, J., Escuer, P., Frias-Lopez, C., Guirao-Rico, S., Hering, L., Mayer, G., Rozas, J., and Sanchez-Gracia, A. (2020) Evolutionary history of major chemosensory gene families across panarthropoda. *Mol Biol Evol*, **37**(12):3601–3615.
- Vos, M., Hesselman, M. C., te Beek, T. A., van Passel, M. W. J., and Eyre-Walker, A. (2015) Rates of lateral gene transfer in prokaryotes: high but why? *Trends in Microbiol*, **23**(10):598–605.
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., and Earl, A. M. (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, **9**(11):e112963.
- Walz, B. (1979) The morphology of cells and cell organelles in the anhydrobiotic tardigrade, *Macrobotus hufelandi*. *Protoplasma*, **99**:19–30.
- Wang, C., Grohme, M. A., Mali, B., Schill, R. O., and Frohme, M. (2014) Towards decrypting cryptobiosis—analyzing anhydrobiosis in the tardigrade *Milnesium tardigradum* using transcriptome sequencing. *PLoS One*, **9**(3):e92663.
- Wang, J., Zhang, P., Lu, Y., Li, Y., Zheng, Y., Kan, Y., Chen, R., and He, S. (2019) piRBase: a comprehensive database of piRNA sequences. *Nucleic Acids Res*, **47**(D1):D175–D180.
- Watanabe, M., Kikawada, T., Minagawa, N., Yukuhiro, F., and Okuda, T. (2002) Mechanism allowing an insect to survive complete dehydration and extreme temperatures. *J Exp Biol*, **205**:2799–2802.
- Watanabe, M., Kikawada, T., Fujita, A., Forczek, E., Adati, T., and Okuda, T. (2004) Physiological traits of invertebrates entering cryptobiosis in a post-embryonic stage. *Eur J Entomol*, **101**(3):439–444.

- Watanabe, M., Sakashita, T., Fujita, A., Kikawada, T., Horikawa, D. D., Nakahara, Y., Wada, S., Funayama, T., Hamada, N., Kobayashi, Y., and Okuda, T. (2006) Biological effects of anhydrobiosis in an African chironomid, *Polypedilum vanderplanki* on radiation tolerance. *Int J Radiat Biol*, **82**(8):587–592.
- Watanabe, M., Nakahara, Y., Sakashita, T., Kikawada, T., Fujita, A., Hamada, N., Horikawa, D. D., Wada, S., Kobayashi, Y., and Okuda, T. (2007) Physiological changes leading to anhydrobiosis improve radiation tolerance in *Polypedilum vanderplanki* larvae. *J Insect Physiol*, **53**(6):573–579.
- Welch, D. B. M., Ricci, C., and Meselson, M. (2009) *Bdelloid rotifers: progress in understanding the success of an evolutionary scandal*, 13, pages 259–279. Lost Sex. Springer, New York.
- Westh, P. and Ramløv, H. (1991) Trehalose accumulation in the tardigrade *Adorybiotus coronifer* during anhydrobiosis. *J Exp Zool*, **258**:303–311.
- Wharton, D. A. and Barrett, J. (1985) Ultrastructural-changes during recovery from anabiosis in the plant parasitic nematode, *Ditylenchus*. *Tissue & Cell*, **17**(1):79–96.
- Wilson, C. G., Nowell, R. W., and Barraclough, T. G. (2018) Cross-contamination explains “inter and intraspecific horizontal genetic transfers” between asexual bdelloid rotifers. *Cur Biol*, **28**(15):2436–2444.e14.
- Wright, J. C. (1989a) Desiccation tolerance and water-retentive mechanisms in tardigrades. *J Exp Biol*, **142**:267–292.
- Wright, J. C. (1989b) The tardigrade cuticle II. Evidence for a dehydration-dependent permeability barrier in the intracuticle. *Tissue & Cell*, **21**(2):263–279.
- Wright, J. C. (2001) Cryptobiosis 300 years on from van Leuwenhoek: what have we learned about tardigrades? *Zool Anz*, **240**(3-4):563–582.
- Wybouw, N., Dermauw, W., Tirry, L., Stevens, C., Grbić, M., Feyereisen, R., and van Leeuwen, T. (2014) A gene horizontally transferred from bacteria protects arthropods from host plant cyanide poisoning. *eLife*, **3**:e02365.
- Wybouw, N., Pauchet, Y., Heckel, D. G., and Van Leeuwen, T. (2016) Horizontal gene transfer contributes to the evolution of arthropod herbivory. *Genome Biol Evol*, **8**(6):1785–1801.
- Yamada, T. G., Suetsugu, Y., Deviatiiarov, R., Gusev, O., Cornette, R., Nesmelov, A., Hiroi, N., Kikawada, T., and Funahashi, A. (2018) Transcriptome analysis of the anhydrobiotic cell line Pv11 infers the mechanism of desiccation tolerance and recovery. *Sci Rep*, **8**(1):17941.
- Yamada, T. G., Hiki, Y., Hiroi, N. F., Shagimardanova, E., Gusev, O., Cornette, R., Kikawada, T., and Funahashi, A. (2020) Identification of a master transcription factor and a regulatory mechanism for desiccation tolerance in the anhydrobiotic cell line Pv11. *PLoS One*, **15**(3):e0230218.
- Yamaguchi, A., Tanaka, S., Yamaguchi, S., Kuwahara, H., Takamura, C., Imajoh-Ohmi, S., Horikawa, D. D., Toyoda, A., Katayama, T., Arakawa, K., Fujiyama, A., Kubo, T., and Kunieda, T. (2012) Two novel heat-soluble protein families abundantly expressed in an anhydrobiotic tardigrade. *PLoS One*, **7**(8):e44209.
- Yates, A., Beal, K., Keenan, S., McLaren, W., Pignatelli, M., Ritchie, G. R., Ruffier, M., Taylor, K., Vullo, A., and Flicek, P. (2015) The Ensembl REST API: Ensembl data for any language. *Bioinformatics*, **31**(1):143–145.
- Yoshida, H. (2009) ER stress response, peroxisome proliferation, mitochondrial unfolded protein response and Golgi stress response. *IUBMB Life*, **61**(9):871–879.
- Yoshida, Y., Koutsovoulos, G., Laetsch, D. R., Stevens, L., Kumar, S., Horikawa, D. D., Ishino, K., Komine, S., Kunieda, T., Tomita, M., Blaxter, M., and Arakawa, K. (2017) Comparative genomics of the tardigrades *Hypsibius dujardini* and *Ramazzottius varieornatus*. *PLoS Biol*, **15**(7):e2002266.

- Yoshida, Y., Konno, S., Nishino, R., Murai, Y., Tomita, M., and Arakawa, K. (2018) Ultralow input genome sequencing library preparation from a single tardigrade specimen. *J Vis Exp*, **137**:e57615.
- Yoshida, Y., Sugiura, K., Tomita, M., Matsumoto, M., and Arakawa, K. (2019a) Comparison of the transcriptomes of two tardigrades with different hatching coordination. *BMC Dev Biol*, **19**(1):24.
- Yoshida, Y., Nowell, R. W., Arakawa, K., and Blaxter, M. (2019b) *Horizontal gene transfer in Metazoa: examples and methods*, chapter 7, pages 203–226. Springer.
- Yoshida, Y., Satoh, T., Ota, C., Tanaka, S., Horikawa, D. D., Tomita, M., Kato, K., and Arakawa, K. (2020) A novel Mn-dependent peroxidase contributes to tardigrade anhydrobiosis. *bioRxiv*, **2020.11.06.370643**.
- Yoshinaga, K., Yoshioka, H., Kurosaki, H., Hirasawa, M., Uritani, M., and Hasegawa, K. (1997) Protection by trehalose of DNA from radiation damage. *Biosci Biotechnol Biochem*, **61**(1):160–161.
- Zaneveld, J. R., Nemergut, D. R., and Knight, R. (2008) Are all horizontal gene transfers created equal? Prospects for mechanism-based studies of HGT patterns. *Microbiol*, **154**(1):1–15.
- Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Backstrom, D., Juzokaite, L., Vancaester, E., Seitz, K. W., Anantharaman, K., Starnawski, P., Kjeldsen, K. U., Stott, M. B., Nunoura, T., Banfield, J. F., Schramm, A., Baker, B. J., Spang, A., and Ettema, T. J. (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature*, **541**(7637):353–358.
- Zieger, E., Robert, N. S. M., Calcino, A., and Wanninger, A. (2021) Ancestral role of ecdysis-related neuropeptides in animal life cycle transitions. *Curr Biol*, **31**(1):207–213.e4.
- Zou, C., Liu, G., Liu, S., Liu, S., Song, Q., Wang, J., Feng, Q., Su, Y., and Li, S. (2018) Cucurbitacin B acts a potential insect growth regulator by antagonizing 20-hydroxyecdysone activity. *Pest Manag Sci*, **74**(6):1394–1403.

---

# Achievements

## Research Publications

- Yoshida Y., Satoh T., Ota, C., Tanaka S., Horikawa D. D., Tomita M., Kato K., and Arakawa K. (2020) A novel Mn-dependent peroxidase contributes to tardigrade anhydrobiosis. *bioRxiv*, **2020.11.06.370643**. <https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1>.
- Miyata Y., Fuse H., Tokumoto S., Hiki Y., Deviatiiarov R., Yoshida Y., Yamada T. G., Cornette R., Gusev O., Shagimardanova E., Funahashi A., Kikawada T. (2020) Cas9-mediated genome editing reveals a significant contribution of calcium signaling pathways to anhydrobiosis in Pv11. *bioRxiv*, **2020.10.15.340281**. <https://www.biorxiv.org/content/10.1101/2020.10.15.340281v1>.
- Yoshida Y., Sugiura K., Tomita M., Matsumoto M., and Arakawa K. (2019) Comparison of the transcriptomes of two tardigrades with different hatching coordination. *BMC Dev Biol*, **19**(1):24. <https://doi.org/10.1186/s12861-019-0205-9>.
- Yoshida Y., Nowell R. W., Arakawa K., and Blaxter M. (2019) Horizontal gene transfer in Metazoa: examples and methods. In *Horizontal Gene Transfer*. pages 203–226. Villa T., Viñas M., Springer, Cham. [https://doi.org/10.1007/978-3-030-21862-1\\_7](https://doi.org/10.1007/978-3-030-21862-1_7).
- Kono N., Nakamura H., Ohtoshi R., Moran D. A. P., Shinohara A., Yoshida Y., Fujiwara M., Mori M., Tomita M., and Arakawa K. (2019) Orb-weaving spider *Araneus ventricosus* genome elucidates the spidroin gene catalogue. *Sci Rep*, **9**(1):8380. <https://doi.org/10.1038/s41598-019-44775-2>.
- Yoshida Y., Konno S., Nishino R., Murai Y., Tomita M., and Arakawa K. (2018) Ultralow input genome sequencing library preparation from a single tardigrade specimen. *J Vis Exp*, **137**:e57615. <https://doi.org/10.3791/57615>.
- Yoshida Y., Koutsovoulos G., Laetsch D. R., Stevens L., Kumar S., Horikawa D. D., Ishino K., Komine S., Kunieda T., Tomita M., Blaxter M., and Arakawa K. (2017) Comparative genomics of

the tardigrades *Hypsibius dujardini* and *Ramazzottius varieornatus*. *PLoS Biol*, **15**(7):e2002266. <https://doi.org/10.1371/journal.pbio.2002266>.

- Arakawa K., Yoshida Y., and Tomita M. (2016) Genome sequencing of a single tardigrade *Hypsibius dujardini* individual. *Sci Data*, **3**:160063. <https://doi.org/10.1038/sdata.2016.63>.

## Talks

### International

- Yoshida Y., Koutsovoulos G., Laetsch D. R., Stevens L., Kumar S., Horikawa D. D., Ishino K., Komine S., Kunieda T., Tomita M., Blaxter M., Arakawa K. “Comparative genomics of the tardigrades *Hypsibius dujardini* and *Ramazzottius varieornatus*”. *14th International Symposium on Tardigrada*, 2018. Copenhagen, Denmark.
- Blaxter M., Laetsch D., Stevens L., Koutsovoulos G., Kumar S., Arakawa K., Yoshida Y., Lunt D., Bachmann L., Michalczyk Ł., Fujimoto S. “Phylogenomics, gene families and the phylogenetic position of Tardigrada within Ecdysozoa” *14th International Symposium on Tardigrada*, 2018. Copenhagen, Denmark.
- Yoshida Y., Koutsovoulos G., Laetsch D. R., Stevens L., Kumar S., Horikawa D. D., Ishino K., Komine S., Kunieda T., Tomita M., Blaxter M., Arakawa K. “Comparative genomics of the tardigrades *Hypsibius dujardini* and *Ramazzottius varieornatus*” *ISMB2017*, 2017. Prague, Czech Republic.
- Tsujimoto M., Yoshida Y., Imura S., Arakawa K. “Comparative transcriptome analysis of the Antarctic tardigrade, *Acutuncus antarcticus*, in response to changes in temperature” *SCAR Biology Symposium 2017*, 2017. Leuven, Belgium.
- Watabe T., Yoshida Y., Suzuki H., Arakawa K., Tomita M., Fukuda S. “Draft genome sequences of *Bifidobacterium longum* strains from two healthy Japanese adults” *The 2nd Annual Conference on Metagenomics and MetaSUB*, 2016. Shanghai, China.

### Domestic

- 吉田祐貴, 佐藤匡史, 太田知世, 堀川大樹, 富田勝, 加藤晃一, 荒川和晴. “オミクス解析による緩歩動物特異的な新規抗酸化遺伝子の探索” 第5回クマムシ学研究会 2020, 2020. オンライン.
- 吉田祐貴, Olga K., 伊藤昌可, 田上道平, 山田貴大, Cherkasov A., Deviatiiarov R., 徳本翔子, 宮田佑吾, Cornette R., 村田光義, 末木広美, 石山美樹, 野間将平, 舟橋啓, 富田勝, Gusev O., 黄川田隆洋. “Cap Trap RNA-Seq を用いた乾眠性ユスリカのトランスクリプトーム解析” 生命情報科学若手の会 第12回年会, 2020. オンライン.

- 吉田祐貴, 堀川大樹, 富田勝, 荒川和晴. “短波長紫外線に曝露したヨコヅナクマムシのトランスクリプトーム解析” 生命情報科学若手の会 第 11 回年会, 2019. 山梨.
- 吉田祐貴, 堀川大樹, 富田勝, 荒川和晴. “短波長紫外線に曝露したヨコヅナクマムシのトランスクリプトーム解析” 第 4 回クマムシ学研究会 2019, 2019. 大阪.
- 吉田祐貴, 杉浦健太, 富田勝, 松本緑, 荒川和晴. “発生 3 日目<sup>3</sup>がクマムシ *Hypsibus exemplaris* の孵化に重要である” 日本動物学会, 2019. 大阪.
- 吉田祐貴, Koutsovoulos G., Laetsch D. R., Stevens L., Kumar S., 堀川大樹, 石野響子, 小峰菜, 國枝武和, 富田勝, Blaxter M., 荒川和晴. “ヨコヅナクマムシとドウジャルダンヤクマムシの比較ゲノム解析” IIBMP2018, 2018. 山形 (ハイライトトラック).
- 吉田祐貴, 堀川大樹, 富田勝, 荒川和晴. “短波長紫外線へ曝露されたヨコヅナクマムシのトランスクリプトーム解析” 第 3 回クマムシ学研究会 2018, 2018. 神奈川.
- 西野稜介, 吉田祐貴, 堀川大樹, 富田勝, 荒川和晴. “経時的遺伝子発現解析によるクマムシたる形成の意義の究明” 第 3 回クマムシ学研究会 2018, 2018. 神奈川.
- 吉田祐貴, 堀川大樹, 富田勝, 荒川和晴. “短波長紫外線へ曝露された横綱クマムシのトランスクリプトーム解析” 生命科学系フロンティアミーティング 2018, 2018. 静岡.
- 岩井碩慶, 河野暢明, 吉田祐貴, 富田勝, 堀川大樹, 荒川和晴. “一時的社会寄生種であるトゲアリ (*Polyrhachis lamellidens*) が行う馬乗り行動についての役割の解明” 生命科学系フロンティアミーティング 2018, 2018. 静岡.
- 飯井虹之介, 吉田祐貴, 河野暢明, Paulino-Lima I. G., 富田勝, Rothschild L. J., 荒川和晴. “比較ゲノム解析による Micrococcaceae 科における紫外線耐性関連遺伝子群の探索” 生命科学系フロンティアミーティング 2018, 2018. 静岡.
- 西野稜介, 吉田祐貴, 堀川大樹, 富田勝, 荒川和晴. “クマムシ樽形成の意義の解明に向けた経時的トランスクリプトーム解析”. 生命科学系フロンティアミーティング 2018, 2018. 静岡.
- 荒川和晴, 河野暢明, 藤原正幸, 中村浩之, 大利麟太郎, 高井幸, 斧澤佑紀, 吉田祐貴, 阿部望美, 石井菜穂子, 富田勝. “クモ類網羅的シーケンシングによる超高機能発現メカニズムの解明” 日本蜘蛛学会第 49 回大会, 2017. 沖縄.
- 河野暢明, 藤原正幸, 中村浩之, 大利麟太郎, 高井幸, 斧澤佑紀, 吉田祐貴, 阿部望美, 石井菜穂子, 富田勝, 荒川和晴. “マルチオミクス解析から見た蜘蛛系の系統的多様性” 日本蜘蛛学会第 49 回大会, 2017. 沖縄.
- 吉田祐貴, Koutsovoulos G., Laetsch D. R., Stevens L., Kumar S., 堀川大樹, 石野響子, 小峰菜, 國枝武和, 富田勝, Blaxter M., 荒川和晴. “ドウジャルダンヤクマムシとヨコヅナクマムシの比較ゲノム解析” 第 88 回日本動物学会, 2017. 富山 (招待講演).
- 杉浦健太, 吉田祐貴, 國枝武和, 鈴木忠, 荒川和晴, 松本緑. “緩歩動物クマムシの有性生殖—雌雄の違いと求愛・交尾行動の観察—” 第 88 回日本動物学会, 2017. 富山.

- 吉田 祐貴, Koutsovoulos G., Laetsch D. R., Stevens L., Kumar S., 堀川大樹, 石野響子, 小峰栞, 國枝武和, 富田勝, Blaxter M., 荒川 和晴. “乾眠能力の異なる二種のクマムシの比較ゲノム解析” 第2回クマムシ学研究会 2017, 2017. 東京.
- 堀川 大樹, 西野 綾介, 吉田 祐貴, 富田 勝, 荒川 和晴. “熱ショックタンパク質はクマムシの乾燥耐性を向上させるか?” 第2回クマムシ学研究会 2017, 2017. 東京.
- 吉田祐貴, Koutsovoulos G., Laetsch D. R., Stevens L., Kumar S., 堀川大樹, 石野響子, 小峰栞, 國枝武和, 富田勝, Blaxter M., 荒川和晴. “ドゥジャルダンヤマクマムシとヨコヅナクマムシの比較ゲノム解析” SIG-MBI, 2017. 石川.
- 吉田祐貴, 堀川大樹, 坂下哲哉, 國枝武和, 桑原宏和, 豊田敦, 片山俊明, 小林泰彦, 富田勝, 荒川和晴. “ヨコヅナクマムシの乾眠関連遺伝子の網羅的同定へむけて” 第1回クマムシ学研究会 2016, 2016. 神奈川.
- 吉田祐貴, 堀川大樹, 坂下哲哉, 小林泰彦, 富田勝, 荒川和晴. “高線量 $\gamma$ 線に対するヨコヅナクマムシの応答解明に向けた経時的微量トランスクリプトーム解析” 生命情報若手の会 第7回研究会, 2015. 山形.

## Posters

### International

- Yoshida Y., Sugiura K., Tomita M., Matsumoto M., Arakawa K. “Transcriptomic analysis of the developmental stages of tardigrades reveals piRNA and hormone regulation” *ISMB2019*, 2019. Basel, Switzerland.
- Kono N., Nakamura H., Ohtoshi R., Moran D. A. P., Shinohara A., Yoshida Y., Mori M., Tomita M., Numata K., Arakawa K. “Spider and bagworm genomes reveal the diversity of silk protein motifs and its mechanical properties” *ISMB2019*, 2019. Basel, Switzerland.
- Yoshida Y., Koutsovoulos G., Laetsch D. R., Stevens L., Kumar S., Horikawa D. D., Ishino K., Komine S., Kunieda T., Tomita M., Blaxter M., Arakawa K. “Comparative genomics of two tardigrades with different cryptocytotic capacities” *ICSB2019*, 2019. Okinawa, Japan.
- Iwai H., Kono N., Yoshida Y., Tomita M., Horikawa D. D., Arakawa K. “Cuticular hydrocarbon camouflage: a social parasite strategy in *Polyrhachis lamellidens* (Hymenoptera: Formicidae)” *9th Congress of International Society of Hymenopterists*, 2018. Matsuyama, Japan.
- Sugiura K., Yoshida Y., Onoda K., Suzuki A. C., Arakawa K., Matsumoto M. “Sex determination in Tardigrada for evolutionary insight” *The 22nd International Congress of Zoology*, 2016. Okinawa, Japan.
- Yoshida Y., Horikawa D. D., Sakashita T., Yokota Y., Kuwahara H., Toyoda A., Katayama T., Kunieda

T., Kobayashi Y., Tomita M., Arakawa K. “Transcriptome analysis of extremophile tardigrade *Ramazzottius varieornatus* for the comprehensive identification of genes related to gamma ray irradiation response” *3R Symposium*, 2016. Shimane, Japan.

- Yoshida Y., Horikawa D. D., Kunieda T., Kuwabara H., Toyota A., Katayama T., Tomita M., Arakawa K. “Transcriptome study of *Ramazzottius varieornatus* for the comprehensive identification of genes related to DNA damage response” *13th International Symposium on Tardigrada*, 2015. Modena, Italy.

## Domestic

- 吉田祐貴, 佐藤匡史, 太田知世, 堀川大樹, 富田勝, 加藤晃一, 荒川和晴. “オミクス解析による緩歩動物特異的な新規抗酸化タンパクファミリーの探索 2. 第 43 回日本分子生物学会年会, 2020. オンライン.
- 吉田祐貴, 堀川大樹, 富田勝, 荒川和晴. “短波長紫外線に曝露したヨコヅナクマムシのトランスクリプトーム解析” 第 42 回日本分子生物学会年会, 2019. 福岡.
- 河野暢明, 中村浩之, 大利麟太郎, Moran D. A. P., 篠原麻夏, 吉田祐貴, 森大, 富田勝, 沼田圭司, 荒川和晴. “クモとミノガのマルチオミクス解析から見た糸遺伝子の多様性” 第 42 回日本分子生物学会年会, 2019. 福岡.
- 吉田祐貴, Koutsovoulos G., Laetsch D. R., Stevens L., Kumar S., 堀川大樹, 石野響子, 小峰栞, 國枝武和, 富田勝, Blaxter M., 荒川和晴. “ヨコヅナクマムシとドゥジャルダンヤマクマムシの比較ゲノム解析” 2019 年度 慶大先端生命研-薬学研究科 合同リトリート「統合システム適塾」, 2019. 山形. 最優秀ポスター賞
- 吉田祐貴, 堀川大樹, 富田勝, 荒川和晴 “短波長紫外線へ曝露されたヨコヅナクマムシのトランスクリプトーム解析” 第 41 回日本分子生物学会年会, 2018. 神奈川.
- 飯井虹之介, 吉田祐貴, 河野暢明, Paulino-Lima I. G., 富田勝, Rothschild L. J., 荒川和晴. “比較ゲノム解析による Micrococcaceae 科における紫外線耐性関連遺伝子群の探索” IIBMP2018, 2018. 山形.
- 吉田祐貴, Koutsovoulos G., Laetsch D. R., Stevens L., Kumar S., 堀川大樹, 石野響子, 小峰栞, 國枝武和, 富田勝, Blaxter M., 荒川和晴. “乾眠能力の異なる二種のクマムシの比較ゲノム解析” ConBio2017, 2017. 兵庫.
- 渡部翔, 吉田祐貴, 鈴木治夫, 富田勝, 荒川和晴, 福田真嗣. “異なる個人由来の *Bifidobacterium longum* 菌株の比較ゲノム解析” ConBio2017, 2017. 兵庫.
- 土澤優里, 吉田祐貴, 石井侑樹, 富田勝, 石川孝博, 荒川和晴. “*Euglena gracilis* の葉緑体欠損変異株の遺伝子発現変動解析” ConBio2017, 2017. 兵庫.
- 岩井碩慶, 河野暢明, 吉田祐貴, 富田勝, 堀川大樹, 荒川和晴. “マルチオミクス解析による一時

的社会寄生種であるトゲアリ (*Polyrhachis lamellidens*) が行う馬乗り行動についての役割の解明” ConBio2017, 2017. 兵庫.

- 堀川大樹, 西野稜介, 吉田祐貴, 富田勝, 荒川和晴. “熱ショックタンパク質はクマムシの乾燥耐性を向上させるか?” ConBio2017, 2017. 兵庫.
- 西野稜介, 吉田祐貴, 富田勝, 堀川大樹, 荒川和晴. “トランスクリプトーム解析によるクマムシ樽形成の意義の解析” ConBio2017, 2017. 兵庫.
- 吉田祐貴, Koutsovoulos G., Laetsch D. R., Stevens L., Kumar S., 堀川大樹, 石野響子, 小峰栞, 國枝武和, 富田勝, Blaxter M., 荒川和晴. “ヨコヅナクマムシとドウジャルダンヤマクマムシの比較ゲノム解析” 生命情報科学若手の会 第9回年会, 2017. 愛知.
- 土澤優里, 吉田祐貴, 石井侑樹, 富田勝, 石川孝博, 荒川和晴. “*Euglena gracilis* の葉緑体欠損変異株の遺伝子発現変動解析” 生命情報科学若手の会 第9回年会, 2017. 愛知.
- 吉田祐貴, Koutsovoulos G., Laetsch D. R., Stevens L., Kumar S., 堀川大樹, 石野響子, 小峰栞, 國枝武和, 富田勝, Blaxter M., 荒川和晴. “ヨコヅナクマムシとドウジャルダンヤマクマムシの比較ゲノム解析” 慶應ライフサイエンスシンポジウム, 2017. 神奈川. 優秀ポスター賞.
- 吉田祐貴, Koutsovoulos G., Laetsch D. R., Stevens L., Kumar S., 堀川大樹, 石野響子, 小峰栞, 國枝武和, 富田勝, Blaxter M., 荒川和晴. “乾眠能力の異なる二種のクマムシの比較ゲノム解析” NGS 現場の会 第五回研究会, 2017. 福島.
- 土澤優里, 吉田祐貴, 石井侑樹, 富田勝, 石川孝博, 荒川和晴. “*Euglena gracilis* の独立・従属栄養条件における遺伝子発現変動解析” NGS 現場の会 第五回研究会, 2017. 福島.
- 飯井虹之介, Ivan P. L., 吉田祐貴, 河野暢明, 富田勝, 荒川和晴. “比較ゲノム解析による *Micrococcaceae* 科における紫外線耐性関連遺伝子群の探索” ゲノム微生物学会, 2017. 神奈川.
- 吉田祐貴, 富田勝, 荒川和晴. “Genome sequencing and assembly of the extremophile tardigrade *Hypsibius dujardini*” MBSJ2016, 2016. 神奈川.
- 吉田祐貴, 富田勝, 荒川和晴. “極限環境生物クマムシ *Hypsibius dujardini* のゲノム解析とアセンブリー” 生命情報科学若手の会 第8研究会, 2016. 北海道.
- 吉田祐貴, 堀川大樹, 坂下哲哉, 國枝武和, 桑原宏和, 豊田敦, 片山俊明, 小林泰彦, 富田勝, 荒川和晴. “高線量ガンマ線に対するヨコヅナクマムシの応答解明に向けた経時的微量トランスクリプトーム解析” BMB2015, 2015. 兵庫.
- 堀川大樹, 吉田祐貴, 國枝武和, 桑原宏和, 豊田敦, 片山俊明, 富田勝, 荒川和晴. “経時的微量トランスクリプトーム解析を用いたヨコヅナクマムシの熱耐性機構の解明” BMB2015, 2015. 兵庫.
- 吉田祐貴, 堀川大樹, 坂下哲哉, 小林泰彦, 富田勝, 荒川和晴. “ヨコヅナクマムシの染色体損傷応答解明に向けた経時的遺伝子発現解析” NGS 現場の会 第四回研究会, 2013. 茨城.
- 吉田祐貴, 荒川和晴, 富田勝. “乾眠能力の弱いドウジャルダンヤマクマムシのパラログ解析” MBSJ2014, 2014. 兵庫.

- 吉田祐貴, 石黒宗, 荒川和晴, 國枝武和, 桑原宏和, 堀川大樹, 豊田敦, 片山俊明, 藤山秋佐夫, 冨田勝. “RNA-Seq データを用いたヨコヅナクマムシにおける細胞防御・修復関連遺伝子の同定” NGS 現場の会 第三回研究会, 2013. 兵庫.
- 吉田祐貴, 石黒宗, 荒川和晴, 冨田勝. “ヨコヅナクマムシとその近縁種における DNA 修復機構の遺伝子レパートリー解析” 生命情報科学若手の会 第 5 回研究会, 2013. 千葉.

## Awards

- Sep, 2019 2019 Keio University IAB-Grad. Sch. Pharm. Sci. Join Retreat in TTCK, Best Poster Presentation  
**Keio University**, 10,000 JPY.
- Aug, 2019 Research Grant  
**Keio SFC Academic Society**, 105,000 JPY.
- Apr, 2018 – Mar, 2021 JSPS Research Fellowship DC1  
**Japan Society for the Promotion of Science**, 2,500,000 JPY.
- Dec, 2018 Taikichiro Mori Memorial Research Grants 2018  
**Keio Research Institute at SFC**, 66,211 JPY.
- Mar, 2018 SFC Aiso Award  
**Keio University**, 40,000 JPY.
- Mar, 2018 Representative of M.S. Graduates  
**Keio University, Graduate School of Media and Governance.**
- Mar, 2018 Master thesis award  
**Keio University**, Systems Biology Program.
- Sep, 2017 Keio Life Science Symposium, Presentation Award  
**Keio University**, 10,000 JPY.
- Apr, 2017 – Mar, 2018 Taikichiro Mori Memorial Research Grants  
**Keio University**, 299,000 JPY.
- Apr, 2016 – Mar, 2017 Koizumi Memorial Graduate School Special Scholarship  
**Keio University**, 360,000 JPY.
- Mar, 2016 Graduation Thesis Award  
**Keio University**, Advanced Biosciences Research group.
- Apr, 2015 – Mar, 2016 Yamagishi Student Project Support Program  
**Keio University**, 146,000 JPY.
- Apr, 2013 – Mar, 2021 Research Fellowship

**Keio University, 8,880,000 JPY.**

---

# Supplementary Files

## Supplementary Tables

### List of Supplementary Tables

S1	RNA-Seq data of <i>H. exemplaris</i> used in this analysis. . . . .	156
S2	RNA-Seq data of <i>R. varieornatus</i> used in this analysis. . . . .	157
S3	HGT content calculation of Ecdysozoa. . . . .	158
S4	Proteomes used for protein family clustering. . . . .	159
S5	Software used in this study. . . . .	160
S6	DNA-Seq data used in this analysis. . . . .	162
S7	Mapping statistics of DNA sequencing data. . . . .	163
S8	Repeat content of the genomes of <i>H. exemplaris</i> and <i>R. varieornatus</i> . . . . .	164
S9	Low-complexity and possible telomeric repeats in the <i>H. exemplaris</i> genome. . . . .	165
S10	Mapping proportions of RNA-Seq data (Anhydrobiosis samples). . . . .	166
S11	Mapping proportions of RNA-Seq data (Developmental samples). . . . .	167
S12	Mapping proportion of the <i>H. exemplaris</i> Trinity assembled transcriptome. . . . .	168
S13	Tardigrade-specific, protection-related proteins. . . . .	169
S14	Synapomorphies identified under different systematic hypotheses. . . . .	170
S15	Summary of RNA-Seq data. . . . .	171
S16	SOM groups of tardigrade specific anhydrobiosis genes. . . . .	172

## Chapter 2

Supplementary Table S1: RNA-Seq data of *H. exemplaris* used in this analysis.

Individuals	Platform	sample	replicate	Sample ID	Reads	Accession
<i>H. exemplaris</i>						
10,000	HiSeq2000	Active	1	Active_1	25,172,359	SRR5218239
10,000	HiSeq2000	Active	2	Active_2	26,497,216	SRR5218240
10,000	HiSeq2000	Active	3	Active_3	28,141,582	SRR5218241
10,000	HiSeq2000	Tun	1	Tun_1	25,782,478	SRR5218242
10,000	HiSeq2000	Tun	2	Tun_2	27,832,551	SRR5218243
10,000	HiSeq2000	Tun	3	Tun_3	27,001,002	SRR5218244
30	NextSeq 500	Active	1	act-1	11,399,144	SRR5218245
30	NextSeq 500	Active	2	act-2	10,744,670	SRR5218246
30	NextSeq 500	Active	3	act-3	10,939,323	SRR5218247
30	NextSeq 500	Tun	1	tun-1	10,325,677	SRR5218248
30	NextSeq 500	Tun	2	tun-2	10,689,489	SRR5218249
30	NextSeq 500	Tun	3	tun-3	10,455,913	SRR52182450
30	NextSeq 500	Egg 1st day after laying	1	H-E1-1	8,822,054	SRR5218203
30	NextSeq 500	Egg 1st day after laying	2	H-E1-2	10,286,604	SRR5218204
30	NextSeq 500	Egg 1st day after laying	3	H-E1-3	8,319,242	SRR5218205
30	NextSeq 500	Egg 2nd day after laying	1	H-E2-1	11,794,526	SRR5218206
30	NextSeq 500	Egg 2nd day after laying	2	H-E2-2	11,086,054	SRR5218207
30	NextSeq 500	Egg 2nd day after laying	3	H-E2-3	10,151,210	SRR5218208
30	NextSeq 500	Egg 3rd day after laying	1	H-E3-1	10,057,550	SRR5218209
30	NextSeq 500	Egg 3rd day after laying	2	H-E3-2	9,253,951	SRR5218210
30	NextSeq 500	Egg 3rd day after laying	3	H-E3-3	11,871,780	SRR5218211
30	NextSeq 500	Egg 4th day after laying	1	H-E4-1	11,622,113	SRR5218212
30	NextSeq 500	Egg 4th day after laying	2	H-E4-2	12,386,383	SRR5218213
30	NextSeq 500	Egg 4th day after laying	3	H-E4-3	9,654,776	SRR5218214
30	NextSeq 500	Egg 5th day after laying	1	H-E5-1	11,921,100	SRR5218215
30	NextSeq 500	Egg 5th day after laying	2	H-E5-2	11,569,382	SRR5218216
30	NextSeq 500	Egg 5th day after laying	3	H-E5-3	10,503,387	SRR5218217
30	NextSeq 500	Juvenile 1st day	1	H-B1-1	12,440,551	SRR5218218
30	NextSeq 500	Juvenile 1st day	2	H-B1-2	12,306,138	SRR5218219
30	NextSeq 500	Juvenile 1st day	3	H-B1-3	12,734,126	SRR5218220
30	NextSeq 500	Juvenile 2nd day	1	H-B2-1	13,107,156	SRR5218221
30	NextSeq 500	Juvenile 2nd day	2	H-B2-2	14,437,609	SRR5218222
30	NextSeq 500	Juvenile 2nd day	3	H-B2-3	13,870,809	SRR5218223
30	NextSeq 500	Juvenile 3rd day	1	H-B3-1	8,360,076	SRR5218224
30	NextSeq 500	Juvenile 3rd day	2	H-B3-2	6,542,790	SRR5218225
30	NextSeq 500	Juvenile 3rd day	3	H-B3-3	9,775,113	SRR5218226
30	NextSeq 500	Juvenile 4th day	1	H-B4-1	9,824,335	SRR5218227
30	NextSeq 500	Juvenile 4th day	2	H-B4-2	16,666,875	SRR5218228
30	NextSeq 500	Juvenile 4th day	3	H-B4-3	15,995,271	SRR5218229
30	NextSeq 500	Juvenile 5th day	1	H-B5-1	6,928,823	SRR5218230
30	NextSeq 500	Juvenile 5th day	2	H-B5-2	8,857,975	SRR5218231
30	NextSeq 500	Juvenile 5th day	3	H-B5-3	12,901,947	SRR5218232
30	NextSeq 500	Juvenile 6th day	1	H-B6-1	9,843,726	SRR5218233
30	NextSeq 500	Juvenile 6th day	2	H-B6-2	12,913,346	SRR5218234
30	NextSeq 500	Juvenile 6th day	3	H-B6-3	11,745,564	SRR5218235
30	NextSeq 500	Juvenile 7th day	1	H-B7-1	12,384,307	SRR5218236
30	NextSeq 500	Juvenile 7th day	2	H-B7-2	9,182,107	SRR5218237
30	NextSeq 500	Juvenile 7th day	3	H-B7-3	13,626,269	SRR5218238
5,000	HiSeq2000	miRNA-Seq	1	HD_miRNA	32,254,413	SRR5179573

**Supplementary Table S2: RNA-Seq data of *R. varieornatus* used in this analysis.**

Individuals	Platform	sample	replicate	Sample ID	Reads	Accession
1-2.5	NextSeq500	Active-Fast	1	Y-active_slow_1	12,146,289	SRR5218269
1-2.5	NextSeq500	Active-Fast	2	Y-active_slow_2	11,076,841	SRR5218270
1-2.5	NextSeq500	Active-Fast	3	Y-active_slow_3	11,211,443	SRR5218271
1-2.5	NextSeq500	Tun-Fast	1	Y-tun_slow_1	11,781,529	SRR5218272
1-2.5	NextSeq500	Tun-Fast	2	Y-tun_slow_2	11,966,104	SRR5218273
1-2.5	NextSeq500	Tun-Fast	3	Y-tun_slow_3	12,361,848	SRR5218274
30	NextSeq500	Active-Slow	1	Y-active_fast_1	31,330,380	SRR5218275
30	NextSeq500	Active-Slow	2	Y-active_fast_2	35,320,831	SRR5218276
30	NextSeq500	Active-Slow	3	Y-active_fast_3	36,895,441	SRR5218277
30	NextSeq500	Tun-Slow	1	Y-tun_fast_1	35,469,871	SRR5218278
30	NextSeq500	Tun-Slow	2	Y-tun_fast_2	38,879,671	SRR5218279
30	NextSeq500	Tun-Slow	3	Y-tun_fast_3	31,835,650	SRR52182780
30	NextSeq500	Egg 1st day after laying	1	Y-E1-1	11,688,367	SRR5218251
30	NextSeq500	Egg 1st day after laying	2	Y-E1-2	13,064,048	SRR5218252
30	NextSeq500	Egg 1st day after laying	3	Y-E1-3	13,389,666	SRR5218253
30	NextSeq500	Egg 2nd day after laying	1	Y-E2-1	12,702,879	SRR5218254
30	NextSeq500	Egg 2nd day after laying	2	Y-E2-2	14,385,811	SRR5218255
30	NextSeq500	Egg 2nd day after laying	3	Y-E2-3	13,101,271	SRR5218256
30	NextSeq500	Egg 3rd day after laying	1	Y-E3-1	14,348,899	SRR5218257
30	NextSeq500	Egg 3rd day after laying	2	Y-E3-2	13,640,410	SRR5218258
30	NextSeq500	Egg 3rd day after laying	3	Y-E3-3	8,817,117	SRR5218259
30	NextSeq500	Egg 4th day after laying	1	Y-E4-1	12,606,663	SRR5218260
30	NextSeq500	Egg 4th day after laying	2	Y-E4-2	15,271,225	SRR5218261
30	NextSeq500	Egg 4th day after laying	3	Y-E4-3	12,517,722	SRR5218262
30	NextSeq500	Egg 5th day after laying	1	Y-E5-1	12,599,958	SRR5218263
30	NextSeq500	Egg 5th day after laying	2	Y-E5-2	14,476,417	SRR5218264
30	NextSeq500	Egg 5th day after laying	3	Y-E5-3	17,324,895	SRR5218265
30	NextSeq500	Juvenile 1st day	1	Y-B1-1	4,811,886	SRR5218266
30	NextSeq500	Juvenile 1st day	2	Y-B1-2	6,210,798	SRR5218267
30	NextSeq500	Juvenile 1st day	3	Y-B1-3	5,637,785	SRR5218268

Supplementary Table S3: HGT content calculation of Ecdysozoa.

Augustus3 Model	Organism	Ensembl (DL:2016.09.20)			ab initio Prediction (Augustus 3.2.2)		
model		# Gene	HGT	%	# Gene	HGT	%
aedes	<i>Aedes aegypti</i>	15,796	182/11,906	1.53	103,215	6,364/70,452	9.03
honeybee1	<i>Apis mellifera</i>	15,314	485/9,773	4.96	14,115	897/10,073	8.9
bombus_impatiens1	<i>Bombus impatiens</i>	15,896	1,166/11,329	10.29	18,793	1,487/10,910	13.63
culex	<i>Culex quinquefasciatus</i>	18,968	203/14,015	1.45	25,343	297/18,296	2.17
fly	<i>Drosophila ananassae</i>	15,069	91/10,336	0.88	21,842	1,816/15,694	11.58
fly	<i>Drosophila erecta</i>	15,044	85/10,121	0.84	15,447	616/11,491	5.36
fly	<i>Drosophila grimshawi</i>	14,982	90/10,448	0.86	15,293	263/11,015	2.39
fly	<i>Drosophila melanogaster</i>	13,918	74/19,191	0.73	15,535	532/11,401	4.67
fly	<i>Drosophila mojavensis</i>	14,594	86/9,918	0.87	15,677	372/11,338	3.28
fly	<i>Drosophila persimilis</i>	16,874	82/10,800	0.76	21,673	786/14,385	5.46
fly	<i>Drosophila pseudoobscura</i>	15,864	96/10,649	0.9	16,705	2,452/11,387	2.21
fly	<i>Drosophila sechellia</i>	16,465	71/10,787	0.66	22,119	1,151/16,788	6.86
fly	<i>Drosophila simulans</i>	15,413	70/9,820	0.71	16,148	362/11,393	3.18
fly	<i>Drosophila virilis</i>	14,491	82/10,081	0.81	15,991	512/11,862	4.32
fly	<i>Drosophila willistoni</i>	15,512	156/10,638	1.47	16,942	738/12,259	6.02
fly	<i>Drosophila yakuba</i>	16,077	85/10,463	0.81	17,774	539/12,469	4.32
heliconius_melpomene1	<i>Heliconius melpomene</i>	12,669	132/9,421	1.4	20,333	289/14,640	1.97
nasonia	<i>Nasonia vitripennis</i>	17,083	223/12,130	1.92	26,010	538/15,457	3.48
rhodnius	<i>Rhodnius prolixus</i>	15,438	498/10,733	4.63	52,161	1,833/39,130	4.68
tribolium2012	<i>Tribolium castaneum</i>	16,524	124/1,115	1.12	16,160	105/10,442	1
caenorhabditis	<i>Caenorhabditis brenneri</i>	30,660	253/14,748	1.7	38,953	518/17,298	2.99
caenorhabditis	<i>Caenorhabditis briggsae</i>	21,814	239/10,936	2.19	20,745	242/11,335	2.13
caenorhabditis	<i>Caenorhabditis elegans</i>	20,362	223/10,574	2.11	18,177	215/10,278	2.09
caenorhabditis	<i>Caenorhabditis japonica</i>	29,931	315/15,260	2.06	29,556	352/15,842	2.22
caenorhabditis	<i>Caenorhabditis remanei</i>	31,437	766/14,483	5.29	30,506	1,288/14,291	9.01
trichinella	<i>Trichinella spiralis</i>	16,380	47/8,616	0.55	11,310	41/8,079	0.51
BRAKER	<i>Hypsibius exemplaris</i>	NA	NA	NA	19,913	463/12,616	3.67
BRAKER	<i>Ramazzottius varieornatus*</i>	19,521	242/10,760	2.25	13,917	220/9,894	2.22

\* Genomes of these species were not registered in Ensembl, we used the released CDS sequences from each genome project

Supplementary Table S4: Proteomes used for protein family clustering.

Index*	Species	GeneBuild ID / Accession	Source
0	<i>Anopheles gambiae</i>	2014-08-VectorBase	EnsemblMetazoa
1	<i>Apis mellifera</i>	2014-05-BeeBase	EnsemblMetazoa
2	<i>Acyrtosiphon pisum</i>	2013-07-AphidBase	EnsemblMetazoa
3	<i>Ascaris suum</i>	PRJNA80881	WormbaseParasite5
4	<i>Brugia malayi</i>	PRJNA10729	WormbaseParasite5
5	<i>Bursaphelenchus xylophilus</i>	PRJEA64437	WormbaseParasite5
6	<i>Caenorhabditis elegans</i>	PRJNA13758	WormbaseParasite5
7	<i>Cimex lectularius</i>	v0.5.3	I5K
8	<i>Capitella teleta</i>	2012-12-JGI	EnsemblMetazoa
38	<i>Drosophila melanogaster</i>	r6_09	FLYBASE
9	<i>Dendroctonus ponderosae</i>	2013-04-EnsemblMetazoa	EnsemblMetazoa
10	<i>Daphnia pulex</i>	2011-02-EnsemblMetazoa	EnsemblMetazoa
11	<i>Hypsibius exemplaris</i>	nHd.3.0	this study
12	<i>Ixodes scapularis</i>	2014-08-VectorBase	EnsemblMetazoa
13	<i>Meloidogyne hapla</i>	PRJNA29083	
14	<i>Nasonia vitripennis</i>	2010-12-NasoniaBase	EnsemblMetazoa
15	<i>Octopus bimaculoides</i>	2016-03-OIST	EnsemblMetazoa
16	<i>Priapulus caudatus</i>	GCF_000485595	NCBI
17	<i>Pediculus humanus</i>	2014-04-VectorBase	EnsemblMetazoa
18	<i>Plectus murrayi</i>	nPm.2.0	ngenomes.org
19	<i>Pristionchus pacificus</i>	PRJNA12644	WormbaseParasite5
20	<i>Plutella xylostella</i>	DBM_FJ_v1_1	LEPBASE
37	<i>Ramazzottius varieornatus</i>	nRv.1.1	this study
22	<i>Solenopsis invicta</i>	2013-10-AntGenomesPortal	EnsemblMetazoa
23	<i>Strigamia maritima</i>	2013-02-EG	EnsemblMetazoa
24	<i>Tribolium castaneum</i>	2012/9/24	EnsemblMetazoa
25	<i>Trichuris muris</i>	PRJEB126	WormbaseParasite5
26	<i>Trichinella spiralis</i>	PRJNA12603	WormbaseParasite5
27	<i>Tetranychus urticae</i>	2012-11-ORCAE	EnsemblMetazoa

\* The index assigned for OrthoFinder clustering

Supplementary Table S5: Software used in this study.

Tool name	Ref	Version	Relevant parameters
<b>Raw data processing and filtering</b>			
FastQC	Andrews (2015)	v0.11.3	
Skewer	Jiang <i>et al.</i> (2014)	0.2.2	-n -q 30 -l 51 -m pe
BWA	Li <i>et al.</i> (2009)	0.7.12-r1039	
blobtools	Kumar <i>et al.</i> (2013)	v0.9.19	
SAMtools	Li and Durbin (2009)	Version: 1.2 (using htlib 1.2.1)	view: 30 -bS sort: 30 index: None
ncbi-blast+	Camacho <i>et al.</i> (2009)	ncbi-blast-2.4.0+	
<b>Genome assembly</b>			
Usearch	Edgar (2010)	v. 8.0.1517	
SPAdes	Bankevich <i>et al.</i> (2012)	v. 3.8.1	k 21,33,55,77,99,127, only-assembler, careful None
Bowtie 2	Langmead and Salzberg (2012)	v. 2.2.4	
Platanus	Kajitani <i>et al.</i> (2014)	v. 1.2.3-u 0.2	
Falcon	Chin <i>et al.</i> (2016)	v.0.2.2	<b>Error correction</b> daligner cutoff of 4,000 bp -k16 -e0.70 -s1000 -t16 -l1000 -h64 -w7 <b>Second daligner</b> -k20 -e.96 -s1000 -t32 -l1500 -h256 <b>Final assembly</b> min coverage of 2 max coverage of 80, max diff coverage of 40 -m 50
SSPACE-LongReads	Boetzer and Pirovano (2014)	v. 1.1	-m 50
PBJelly	English <i>et al.</i> (2012)	v. 13.10	-m 50
Pilon	Walker <i>et al.</i> (2014)	v. 1.17	
Qualimap	Okonechnikov <i>et al.</i> (2016)	v. 2.2	<b>bamqc</b> -bam input.sorted.bam -outformat pdf java-mem-size=16G None Lineage : eukaryota_odb9, genome mode sp
CEGMA	Parra <i>et al.</i> (2007)	v. 2.5	
BUSCO	Simao <i>et al.</i> (2015)	v. 2.0.1	
<b>Phylogenetic analyses</b>			
RAxML	Stamatakis (2014)	8.2.8	-b 12345 -# 100 -T 62 -p 12345 -m PROTGAMMAGTR
ClustalW2	Goujon <i>et al.</i> (2010)	v2.1	
ClustalO	Goujon <i>et al.</i> (2010)	v1.1.0	
PhyloBayes	Lartillot and Philippe (2004)	pb_mpi1.7a	
MAFFT	Katoh and Standley (2013)	v7.271 (2016/1/6)	
FastTree	Price <i>et al.</i> (2010)	2.1.8 SSE3	-boot 1000
PhyML	Guindon <i>et al.</i> (2010)	v20120412	
trimAl	Capella-Gutierrez <i>et al.</i> (2009)	v1.4.rev15 build [2013-12-17]	
fasconcat-G	Kuck and Longo (2014)	FASconCAT_v1.0.pl	
FigTree	Rambaut (2016)	v. 1.4.2	
Seaview	Gouy <i>et al.</i> (2010)	4.5.2	
MEME	Bailey <i>et al.</i> (2009)	v4.10.1	
<b>Annotation and databasing</b>			
Ensembl	Aken <i>et al.</i> (2017)	Version 85	
EasyMirror	Challis <i>et al.</i> (2017)	Version 0.9	
and EasyImport		Version 0.9	
Hmmsearch	Mistry <i>et al.</i> (2013)	HMMER 3.1b2 (February 2015)	cpu 32 domE 1e-15
Braker	Hoff <i>et al.</i> (2016)	v1.9	
Augustus	Keller <i>et al.</i> (2011)	v3.2.2	
GeneMark-ES	Borodovsky and Lomsadze (2011)	v.4.21	
RepeatScout	Price <i>et al.</i> (2005)	Version 1.0.5	
RepeatMasker	Smit <i>et al.</i> (2015)	version open-4.0.5	
tRNAscan-SE	Lowe and Eddy (1997)	tRNAscan-SE 1.3.1 (January 2012)	

Continued on next page

Continued on previous page

Tool name	Ref.	Version	Relevant parameters
RNAmmer	Lagesen <i>et al.</i> (2007)	v1.2	-multi -S euk -m lsu,ssu,tsu
Legacy blast	Altschul <i>et al.</i> (1997)	v2.2.22	-e 1e-15
Diamond	Buchfink <i>et al.</i> (2015)	v1.8.2	-e 10 sensitive
KAAS	Moriya <i>et al.</i> (2007)	Automatic Annotation	Representative set
TargetP	Emanuelsson <i>et al.</i> (2007)	v1.1b, March 2006	
NLStradamus	Nguyen Ba <i>et al.</i> (2009)	v1.8 Server Ver. 2.1	for GENES Via web interface
InterProScan	Goujon <i>et al.</i> (2010)	5.19-58.0	goterms -appl TIGRFAM-15.0, ProDom-2006.1 SMART-7.1, SignalP-EUK-4.1, PrositePatterns-20.119, PRINTS-42.0, SuperFamily-1.75 Pfam-29.0, -f TSV
miRDeep	Friedlander <i>et al.</i> (2012)	v2.0.0.8	
SSEARCH	Pearson and Lipman (1988)	v36.3.8e Sep, 2016(preload9)	
Genome comparison			
murasaki	Popendorf <i>et al.</i> (2010)	version 1.68.6 (LARGESEQ)	-p[28:36] -M 100
Mauve	Darling <i>et al.</i> (2004)	mauve_snapshot 2015-02-13	Progressive alignment with GUI
Databases used in annotation**			
Swiss-Prot	UniProt Consortium (2015)	2016/5/23	
TrEMBL	UniProt Consortium (2015)	2016/1/17	
Pfam-A	Finn <i>et al.</i> (2016)	2016/7/22	
Dfam	Hubley <i>et al.</i> (2016)	2016/9/26	
miRBase	Kozomara and Griffiths-Jones (2014)	2016/12/20	
Transcriptome analyses			
Trinity	Grabherr <i>et al.</i> (2011)	2.2.0	Default params
TransDecoder	Brian and Papanicolaou (2017)	3.0.0	<b>TransDecoder.LongOrfs</b> default <b>TransDecoder.Predict</b> retain_blastp_hits single_best_orf default
Tophat2	Kim <i>et al.</i> (2013)	v2.1.1	default
DESeq2	Love <i>et al.</i> (2014)	v1.8.2	
Bowtie2	Langmead and Salzberg (2012)	v2.2.8	
Kallisto	Bray <i>et al.</i> (2016)	v0.42.4	<b>index</b> None <b>quant (single)</b> bias -b 100 single -l 200 -s 50 <b>quant (paired)</b> bias b 100
Gene family analyses			
Gephi		v0.9.1	"Scaling" = 1000.0, "Stronger Gravity" = True, "Gravity" = 1.0 "Dissuade hubs" = False, "LinLog mode" = True "Prevent overlap" = False "Edge Weight Influence" = 1.0
KinFin	Laetsch and Blaxter (2017)	v0.8.2	
OrthoFinder	Emms and Kelly (2015)	v1.1.2	<b>Inflation values</b> 1.1, 1.5, 2.0, 2.5, 3.5, 4.0, 4.5, 5.0
Others			
G-language Genome Analysis Environment	Arakawa <i>et al.</i> (2003) Arakawa and Tomita (2006)	v.1.9.1	

End of Table

\* Where no entry is made, the program was used with default settings.

\*\* For databases, the date of download is given as version.

**Supplementary Table S6: DNA-Seq data used in this analysis.**

Origin	Previous work	This work	Boothby, <i>et al.</i> PNAS, 2015		
Accession ID	DRR055040	SRR5179577	SRR2986339	SRR2986435	SRR2986451
Platform	MiSeq	PacBio	HiSeq 2000	HiSeq 2000	HiSeq 2000
Number of Reads	25,050,780	779,905	43,741,412	31,789,554	22,671,626
Number of raw bases (Gbase)	14.7	5.9	8.8	6.4	4.6
Read Length (Paired)	300 b	NA	100 b	100 b	100 b
Maximum length	NA	49,455	NA	NA	NA
N50 length	NA	10,657	NA	NA	NA
Average length	NA	7,536	NA	NA	NA
Insert length	477.1	NA	347.6	496.9	749.3

We generated new sequencing data using PacBio SMRT technology. In addition, we have used sequence data from Boothby *et al.* (Boothby *et al.*, 2015) for assembly, and single individual sequencing data from our previous report (Arakawa *et al.*, 2016).

**Supplementary Table S7: Mapping statistics of DNA sequencing data.**

Origin*	Accession ID	Mapped Reads	Coverage (1 <sup>st</sup> quantile, Median, 3 <sup>rd</sup> quantile)	# Insertions/Deletions
UNC	SRR2986339	47.6M (54.24%)	36/48/56	6.85M/7.70M
	SRR2986435	53.8M (83.95%)	40/53/62	1.84M/2.02M
	SRR2986451	39.2M (85.68%)	26/37/46	2.82M/2.90M
Edinburgh	ERR1147177	116.0M (77.15%)	61/98/134	3.2M/3.32M
	ERR1147178	99.2M (80.03%)	22/50/70	2.30M/2.37M
Keio	DRR055040	50.0M (96.91%)	66/109/153	8.59M/8.48M

\* UNC = University of North Carolina (Koutsovoulos *et al.*, 2016); Edinburgh = University of Edinburgh (Boothby *et al.*, 2015), Keio = Keio University (this work and Arakawa *et al.* (2016))

**Supplementary Table S8: Repeat content of the genomes of *H. exemplaris* and *R. varieornatus*.**

Category	Term	<i>H. exemplaris</i>	<i>R. varieornatus</i>
Simple	#elements	65,638	3,301
	#length (bp)	5,391,682	137,297
	%genome	5.18	0.25
Unclassified	#elements	158,522	65,730
	#length (bp)	24,232,698	11,020,138
	%genome	23.271	9.74
Total	#elements	224,160	69,031
	#length (bp)	29,624,380	11,157,435
	%genome	28.45	19.99

**Supplementary Table S9: Low-complexity and possible telomeric repeats in the *H. exemplaris* genome.**

Scaffold	Start	End	Repeat	Length from End	Length
scaffold0088*	327,857	336,800	TTGATGGGTT	49	8,943
scaffold0114*	15	7,307	ATCAAAAACCC	15	7,292
scaffold0012*	1	5,955	CATCAAAAACC	1	5,954
scaffold0363*	14	4,481	ATCAAAAACCC	14	4,467
scaffold0321*	157	4,157	AAAACCCATC	157	4,000
scaffold0005*	52	3,367	CAAAAACCCAT	52	3,315
scaffold0128*	239,844	242,599	GGTTTTGATG	823	2,755
scaffold0001*	7,702	9,727	AAACCCATCA	7,702	2,025
scaffold0192*	164,482	165,621	GGGTTTTGAT	49	1,139
scaffold0343	22,767	23,510	TTTTGATGGG	22,767	743
scaffold0023	54,017	54,373	ATCAAAAACCC	54,017	356
scaffold0287	57,943	58,286	ATGGGTTTTG	36,758	343
scaffold0212	65,340	65,622	TTTTGATGGG	65,340	282
scaffold0093	201,990	202,227	TTGATGGGTT	107,275	237
scaffold0072	51,074	51,288	TGGGTTTTGA	51,074	214
scaffold0070	189,916	190,113	CATCAAAAACC	189,916	197
scaffold0031	383,706	383,897	GATGGGTTTT	227,004	191
scaffold0090	105,640	105,790	TCAAAAACCCA	105,640	150
scaffold0092	52,492	52,623	ATGGGTTTTG	52,492	131
scaffold0005	3,799	3,922	ATCAAAAACCC	3,799	123
scaffold0036	185,539	185,660	ATCAAAAACCC	185,539	121
scaffold0036	17,528	17,649	ATCAAAAACCC	17,528	121
scaffold0070	137,262	137,380	AAACCCATCA	137,262	118
scaffold0117	133,706	133,813	ATCAAAAACCC	120,682	107
scaffold0171	121,075	121,176	AAACCCATCA	66,249	101
scaffold0268	65,742	65,843	ATCAAAAACCC	36,714	101

\* Regions close to scaffold ends; these may represent telomeric ends.

**Supplementary Table S10: Mapping proportions of RNA-Seq data (Anhydrobiosis samples).**

<i>Hypsibius exemplaris</i>			
10k Individuals	transcriptome	<i>Hypsibius exemplaris</i> genome	
act-1	35,536,088 (70.08%)	51,790,544 (92.95%)	
act-2	38,335,541 (71.78%)	55,208,279 (93.93%)	
act-3	40,570,560 (71.53%)	58,667,531 (94.03%)	
tun-1	37,104,384 (71.41%)	54,280,566 (94.77%)	
tun-2	39,611,771 (70.58%)	58,602,174 (94.87%)	
tun-3	38,118,839 (69.87%)	57,146,961 (95.28%)	
30 Individuals	<i>H. exemplaris</i> transcriptome		
act-1	4,276,518 (66.19%)		
act-2	5,604,724 (52.04%)		
act-3	2,628,272 (24.98%)		
tun-1	6,410,871 (61.75%)		
tun-2	4,524,015 (61.87%)		
tun-3	5,562,719 (61.55%)		
<i>Ramazzottius varieornatus</i>			
Slow-dry	transcriptome	<i>R. varieornatus</i> genome	Hashimoto <i>et al.</i> (2016) gene model
act-1	6,414,974 (52.68%)	11,459,570 (90.77%)	6,592,920 (54.14%)
act-2	6,052,043 (54.49%)	10,525,516 (91.22%)	6,199,903 (55.82%)
act-3	5,587,457 (49.73%)	10,581,005 (90.96%)	5,735,575 (51.04%)
tun-1	5,327,210 (45.14%)	10,816,699 (88.77%)	5,484,698 (46.47%)
tun-2	5,593,452 (46.66%)	11,003,773 (88.82%)	5,754,535 (48.00%)
tun-3	6,108,721 (49.31%)	11,614,986 (90.71%)	6,287,399 (50.75%)
Fast-dry	<i>R. varieornatus</i> transcriptome		
act-1	15,088,390 (48.06%)		
act-2	17,576,037 (49.66%)		
act-3	17,942,032 (48.53%)		
tun-1	11,430,186 (32.18%)		
tun-2	13,773,138 (35.37%)		
tun-3	12,529,323 (39.30%)		

**Supplementary Table S11: Mapping proportions of RNA-Seq data (Developmental samples).**

Development	<i>Hypsibius exemplaris</i>		<i>Ramazottius varieornatus</i>	
	Transcriptome	Genome	Transcriptome	Genome
E1-1	4,378,145 (49.51%)	7,962,696 (87.17%)	5,191,424 (44.32%)	9,436,055 (77.92%)
E1-2	5,253,306 (50.92%)	9,311,381 (87.33%)	5,299,355 (40.49%)	9,693,795 (71.78%)
E1-3	3,958,477 (47.46%)	7,341,290 (84.97%)	5,051,568 (37.66%)	9,497,572 (68.66%)
E2-1	4,818,034 (40.76%)	10,213,359 (83.77%)	2,591,777 (20.38%)	5,835,138 (45.13%)
E2-2	4,777,963 (42.98%)	9,834,385 (85.84%)	3,098,601 (21.52%)	6,685,797 (45.59%)
E2-3	4,525,795 (44.45%)	8,777,764 (83.40%)	2,514,170 (19.17%)	5,823,252 (43.70%)
E3-1	4,529,697 (44.92%)	8,738,778 (83.96%)	3,504,071 (24.40%)	7,761,912 (53.03%)
E3-2	3,661,789 (39.49%)	7,335,588 (76.93%)	2,935,065 (21.50%)	6,961,489 (50.14%)
E3-3	4,941,305 (41.53%)	10,139,902 (82.78%)	1,551,655 (17.59%)	3,826,220 (42.74%)
E4-1	4,446,480 (38.18%)	8,251,277 (68.94%)	2,286,988 (18.13%)	5,335,818 (41.70%)
E4-2	5,057,162 (40.74%)	9,514,858 (74.52%)	1,805,923 (11.82%)	4,599,738 (29.82%)
E4-3	3,622,851 (37.44%)	7,249,760 (73.06%)	1,451,975 (11.59%)	3,634,522 (28.75%)
E5-1	5,008,318 (41.91%)	9,568,902 (77.70%)	929,244 ( 7.37%)	2,786,443 (21.98%)
E5-2	5,415,199 (46.66%)	10,173,646 (84.90%)	299,688 ( 2.07%)	1,680,860 (11.58%)
E5-3	4,816,060 (45.72%)	9,050,720 (83.26%)	630,014 ( 3.64%)	2,605,069 (14.98%)
B1-1	6,372,485 (51.09%)	11,862,133 (91.80%)	2,043,701 (42.39%)	3,563,016 (71.91%)
B1-2	3,753,075 (30.46%)	7,981,613 (63.46%)	2,871,856 (46.14%)	4,832,785 (75.38%)
B1-3	4,574,819 (35.86%)	9,315,606 (71.27%)	2,684,092 (47.50%)	4,539,890 (77.94%)
B2-1	2,727,779 (20.79%)	6,397,464 (48.10%)	NA	NA
B2-2	2,293,308 (15.87%)	5,574,352 (38.18%)	NA	NA
B2-3	2,244,975 (16.17%)	5,219,405 (37.19%)	NA	NA
B3-1	3,009,036 (35.94%)	6,134,894 (71.71%)	NA	NA
B3-2	1,800,045 (27.48%)	3,885,947 (58.34%)	NA	NA
B3-3	3,070,949 (31.38%)	6,447,666 (64.60%)	NA	NA
B4-1	3,213,530 (32.66%)	6,355,544 (63.35%)	NA	NA
B4-2	3,718,153 (22.29%)	8,188,239 (48.43%)	NA	NA
B4-3	3,808,703 (23.79%)	8,102,210 (49.89%)	NA	NA
B5-1	2,366,172 (34.10%)	4,775,355 (67.40%)	NA	NA
B5-2	2,507,768 (28.28%)	5,510,508 (61.17%)	NA	NA
B5-3	3,797,968 (29.40%)	7,968,128 (60.63%)	NA	NA
B6-1	4,176,593 (42.35%)	8,477,432 (83.80%)	NA	NA
B6-2	5,616,490 (43.40%)	11,116,147 (83.69%)	NA	NA
B6-3	4,280,233 (36.37%)	8,557,430 (71.17%)	NA	NA
B7-1	5,380,476 (43.35%)	10,719,826 (84.09%)	NA	NA
B7-2	4,177,942 (45.40%)	8,118,862 (85.79%)	NA	NA
B7-3	5,987,512 (43.85%)	11,144,850 (79.27%)	NA	NA

**Supplementary Table S12: Mapping proportion of the *H. exemplaris* Trinity assembled transcriptome.**

<i>H. exemplaris</i>			
10k Individuals	Total number of transcripts assembled	Number (proportion) mapped to the <i>H. exemplaris</i> transcriptome	Number (proportion) mapped to the <i>H. exemplaris</i> genome
act-1	66,886	43,608 (65.20%)	59,283 (88.63%)
act-2	68,941	45,168 (65.52%)	60,998 (88.48%)
act-3	73,982	52,217 (70.58%)	68,767 (99.74%)
tun-1	63,670	43,667 (68.58%)	59,503 (93.46%)
tun-2	77,919	54,768 (61.59%)	74,304 (95.37%)
tun-3	149,853	72,078 (48.10%)	117,883 (78.67%)

**Supplementary Table S13: Tardigrade-specific, protection-related proteins.**

Gene ID	Scaffold	Category	%Identity	Length	E-value	Bitscore
bHd17608.1	scaffold0022	CAHS1	54.55	165	5.00E-45	158
bHd17663.1	scaffold0023	CAHS1	49.46	184	2.00E-45	161
bHd04182.1	scaffold0087	CAHS1	61.99	221	7.00E-65	209
bHd04184.1	scaffold0087	CAHS1	78.18	55	4.00E-19	85.5
bHd06166.1	scaffold0123	CAHS1	45.13	195	2.00E-51	174
bHd16038.1	scaffold0013	CAHS2	56.55	168	8.00E-57	187
bHd17504.1	scaffold0022	CAHS2	53.33	105	3.00E-28	117
bHd17505.1	scaffold0022	CAHS2	47.4	192	3.00E-47	162
bHd17506.1	scaffold0022	CAHS2	67.71	192	1.00E-82	253
bHd18862.1	scaffold0032	CAHS2	51.63	184	9.00E-41	148
bHd01486.1	scaffold0050	CAHS2	55.09	167	8.00E-54	179
bHd02925.1	scaffold0069	CAHS3	57.33	75	2.00E-20	90.5
bHd19902.1	scaffold0018	MAHS	42.51	167	5.00E-24	107
bHd16514.1	scaffold0016	RvLEAM	41.45	234	3.00E-40	147
bHd00493.1	scaffold0002	SAHS1	36.3	135	2.00E-24	100
bHd07979.1	scaffold0005	SAHS1	34.5	171	2.00E-25	103
bHd10755.1	scaffold0239	SAHS1	45.88	170	2.00E-44	152
bHd10756.1	scaffold0239	SAHS1	38.01	171	2.00E-32	121
bHd10757.1	scaffold0239	SAHS1	38.01	171	2.00E-32	121
bHd10758.1	scaffold0239	SAHS1	54.23	142	3.00E-47	159
bHd10759.1	scaffold0239	SAHS1	54.68	139	2.00E-49	164
bHd10762.1	scaffold0239	SAHS1	47.37	171	5.00E-47	159
bHd10763.1	scaffold0239	SAHS1	46.47	170	3.00E-47	159
bHd10764.1	scaffold0239	SAHS1	69.63	135	2.00E-65	205

**Supplementary Table S14: Synapomorphies identified under different systematic hypotheses.**

cluster_id	proteins	hypothesis*	taxon coverage	Proportion of proteomes present		
				Nematode (n=9)	Arthropod (n=15)	Tardigrade (n=2)
OG0000436	104	Panarthropoda	1	0	1	1
OG0001236	54	Panarthropoda	1	0	1	1
OG0002592	36	Panarthropoda	1	0	1	1
OG0006538	19	Panarthropoda	1	0	1	1
OG0006541	19	Panarthropoda	1	0	1	1
OG0006869	17	Panarthropoda	1	0	1	1
OG0005117	27	Panarthropoda	0.88	0	0.93	0.5
OG0005941	22	Panarthropoda	0.77	0	0.73	1
OG0006662	18	Panarthropoda	0.82	0	0.8	1
OG0006889	17	Panarthropoda	0.71	0	0.73	0.5
OG0006940	17	Panarthropoda	0.82	0	0.87	0.5
OG0006941	17	Panarthropoda	0.71	0	0.67	1
OG0006951	17	Panarthropoda	0.71	0	0.67	1
OG0007141	16	Panarthropoda	0.82	0	0.8	1
OG0007285	15	Panarthropoda	0.71	0	0.67	1
OG0007290	15	Panarthropoda	0.82	0	0.8	1
OG0007298	15	Panarthropoda	0.88	0	0.87	1
OG0007328	15	Panarthropoda	0.71	0	0.67	1
OG0007463	14	Panarthropoda	0.77	0	0.73	1
OG0007689	13	Panarthropoda	0.71	0	0.67	1
OG0005423	26	Nematoda+Tardigrada	0.82	0.89	0	0.5
OG0006414	20	Nematoda+Tardigrada	0.82	0.78	0	1
OG0007199	16	Nematoda+Tardigrada	0.91	1	0	0.5
OG0007812	13	Nematoda+Tardigrada	0.82	0.78	0	1
OG0008368	11	Nematoda+Tardigrada	0.82	0.78	0	1

\* Panarthropoda = Tardigrada+Arthropoda

## Chapter 4

Supplementary Table S15: Summary of RNA-Seq data.

		<i>H. exemplaris</i>		<i>R. varieornatus</i>		
Embryo	1	E1-1	8,822,054	47.41%	11,688,367	44.32%
	1	E1-2	10,286,604	49.17%	13,064,048	40.49%
	1	E1-3	8,319,242	45.60%	13,389,666	37.65%
	2	E2-1	11,794,526	37.91%	12,702,879	20.38%
	2	E2-2	11,086,054	39.59%	14,385,811	21.52%
	2	E2-3	10,151,210	42.08%	13,101,271	19.17%
	3	E3-1	10,057,550	41.73%	14,348,899	24.40%
	3	E3-2	9,253,951	36.34%	13,640,410	21.50%
	3	E3-3	11,871,780	37.90%	8,817,117	17.59%
	4	E4-1	11,622,113	36.30%	12,606,663	18.13%
	4	E4-2	12,386,383	38.42%	15,271,225	11.82%
	4	E4-3	9,654,776	34.39%	12,517,722	11.59%
	5	E5-1	11,921,100	39.54%	12,599,958	7.37%
	5	E5-2	11,569,382	43.81%	14,476,417	2.07%
	5	E5-3	10,503,387	42.94%	17,324,895	3.64%
Juvenile	1	B1-1	12,440,551	47.15%	4,811,886	42.39%
	1	B1-2	12,306,138	27.03%	6,210,798	46.14%
	1	B1-3	12,734,126	32.31%	5,637,785	47.50%
	2	B2-1	13,107,156	17.69%	13,180,143	34.47%
	2	B2-2	14,437,609	13.51%	12,954,906	48.30%
	2	B2-3	13,870,809	13.87%	14,582,323	45.10%
	3	B3-1	8,360,076	32.57%	14,650,261	45.44%
	3	B3-2	6,542,790	24.53%	13,978,468	42.19%
	3	B3-3	9,775,113	28.38%	16,490,072	45.82%
	4	B4-1	9,824,335	29.83%	13,590,710	44.14%
	4	B4-2	16,666,875	19.37%	13,075,961	40.12%
	4	B4-3	15,995,271	21.19%	14,390,604	43.57%
	5	B5-1	6,928,823	30.79%	13,725,567	42.15%
	5	B5-2	8,857,975	24.78%	13,794,248	39.91%
	5	B5-3	12,901,947	26.09%	13,559,155	42.12%
	6	B6-1	9,843,726	37.75%	16,517,715	50.24%
	6	B6-2	12,913,346	38.81%	14,203,625	36.03%
	6	B6-3	11,745,564	32.43%	13,279,602	36.40%
	7	B7-1	12,384,307	38.91%	13,787,335	43.84%
	7	B7-2	9,182,107	41.05%	13,605,441	44.13%
	7	B7-3	13,626,269	41.16%	14,745,727	13.75%
Adult	-	Active-1	11,399,144	47.57%	12,146,289	52.68%
	-	Active-2	10,744,670	45.53%	11,076,841	54.49%
	-	Active-3	10,939,323	43.59%	11,211,443	49.73%
	-	Tun-1	10,325,677	36.06%	11,781,529	45.14%
	-	Tun-2	10,689,489	38.11%	11,966,104	46.66%
	-	Tun-3	10,455,913	39.83%	12,361,848	49.31%

Supplementary Table S16: SOM groups of tardigrade specific anhydrobiosis genes.

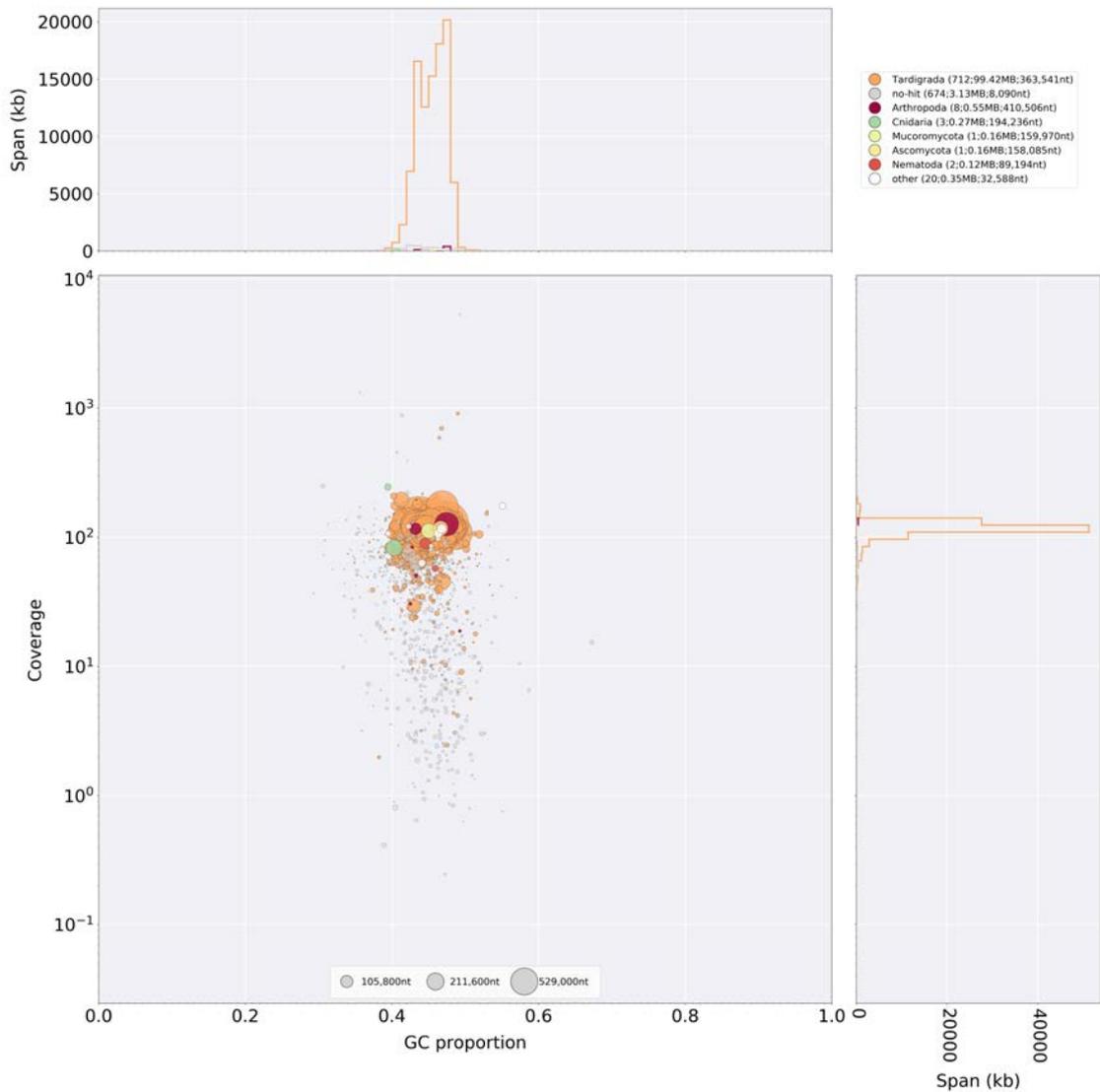
Organism	Category	Gene	X	Y	Group
<i>H. exemplaris</i>	CAHS	OQV13470.1	0	2	1
<i>H. exemplaris</i>	CAHS	OQV15590.1	0	0	3
<i>H. exemplaris</i>	CAHS	OQV15592.1	0	0	3
<i>H. exemplaris</i>	CAHS	OQV16865.1	1	2	4
<i>H. exemplaris</i>	CAHS	OQV16866.1	1	2	4
<i>H. exemplaris</i>	CAHS	OQV18383.1	0	2	3
<i>H. exemplaris</i>	CAHS	OQV20263.1	1	2	4
<i>H. exemplaris</i>	CAHS	OQV21398.1	0	2	1
<i>H. exemplaris</i>	CAHS	OQV21545.1	0	0	3
<i>H. exemplaris</i>	CAHS	OQV21546.1	1	1	5
<i>H. exemplaris</i>	CAHS	OQV21547.1	1	2	3
<i>H. exemplaris</i>	CAHS	OQV21654.1	0	0	3
<i>H. exemplaris</i>	CAHS	OQV21655.1	0	0	3
<i>H. exemplaris</i>	CAHS	OQV21656.1	0	0	3
<i>H. exemplaris</i>	CAHS	OQV23218.1	1	2	4
<i>H. exemplaris</i>	SAHS	OQV24729.1	1	0	6
<i>H. exemplaris</i>	SAHS	OQV25556.1	0	0	3
<i>H. exemplaris</i>	SAHS	OWA51789.1	1	2	4
<i>H. exemplaris</i>	SAHS	OWA51790.1	0	0	3
<i>H. exemplaris</i>	SAHS	OWA51791.1	0	0	3
<i>H. exemplaris</i>	SAHS	OWA51792.1	1	0	6
<i>H. exemplaris</i>	SAHS	OWA51793.1	1	0	6
<i>H. exemplaris</i>	SAHS	OWA51795.1	0	0	3
<i>H. exemplaris</i>	SAHS	OWA51796.1	0	0	3
<i>H. exemplaris</i>	SAHS	OWA51797.1	1	0	6
<i>H. exemplaris</i>	SAHS	OWA51798.1	0	2	1
<i>H. exemplaris</i>	MAHS	OQV22286.1	1	2	4
<i>H. exemplaris</i>	MAHS	OQV22287.1	0	0	3
<i>H. exemplaris</i>	MAHS	OQV22288.1	0	0	3
<i>H. exemplaris</i>	LEAM	OQV22604.1	1	2	4
<i>R. variegatus</i>	CAHS	g612.t1	0	2	1
<i>R. variegatus</i>	CAHS	g673.t1	1	2	4
<i>R. variegatus</i>	CAHS	g674.t1	1	2	4
<i>R. variegatus</i>	CAHS	g675.t1	0	1	2
<i>R. variegatus</i>	CAHS	g791.t1	0	0	3
<i>R. variegatus</i>	CAHS	g2231.t1	0	0	3
<i>R. variegatus</i>	CAHS	g2595.t1	0	1	2
<i>R. variegatus</i>	CAHS	g3884.t1	1	2	4
<i>R. variegatus</i>	CAHS	g4307.t1	1	2	4
<i>R. variegatus</i>	CAHS	g4308.t1	1	2	4
<i>R. variegatus</i>	CAHS	g5971.t1	1	2	4
<i>R. variegatus</i>	CAHS	g7798.t1	0	1	2
<i>R. variegatus</i>	CAHS	g12391.t1	0	0	3
<i>R. variegatus</i>	CAHS	g12902.t1	0	0	3
<i>R. variegatus</i>	CAHS	g13080.t1	0	0	3
<i>R. variegatus</i>	SAHS	g1122.t1	0	2	1
<i>R. variegatus</i>	SAHS	g1624.t1	1	2	4
<i>R. variegatus</i>	SAHS	g1668.t1	0	0	3
<i>R. variegatus</i>	SAHS	g1669.t1	1	2	4
<i>R. variegatus</i>	SAHS	g1671.t1	0	2	1
<i>R. variegatus</i>	SAHS	g1672.t1	0	0	3
<i>R. variegatus</i>	SAHS	g1675.t1	0	1	2
<i>R. variegatus</i>	SAHS	g1676.t1	1	2	4
<i>R. variegatus</i>	SAHS	g1677.t1	0	2	1
<i>R. variegatus</i>	SAHS	g1680.t1	0	1	2
<i>R. variegatus</i>	SAHS	g1681.t1	1	2	4
<i>R. variegatus</i>	SAHS	g1807.t1	0	0	3
<i>R. variegatus</i>	SAHS	g5852.t1	1	2	4
<i>R. variegatus</i>	MAHS	g6834.t1	1	2	4
<i>R. variegatus</i>	LEAM	g2978.t1	0	0	3
<i>R. variegatus</i>	Dsup	g4591.t1	0	0	3

## Supplementary Figures

### List of Supplementary Figures

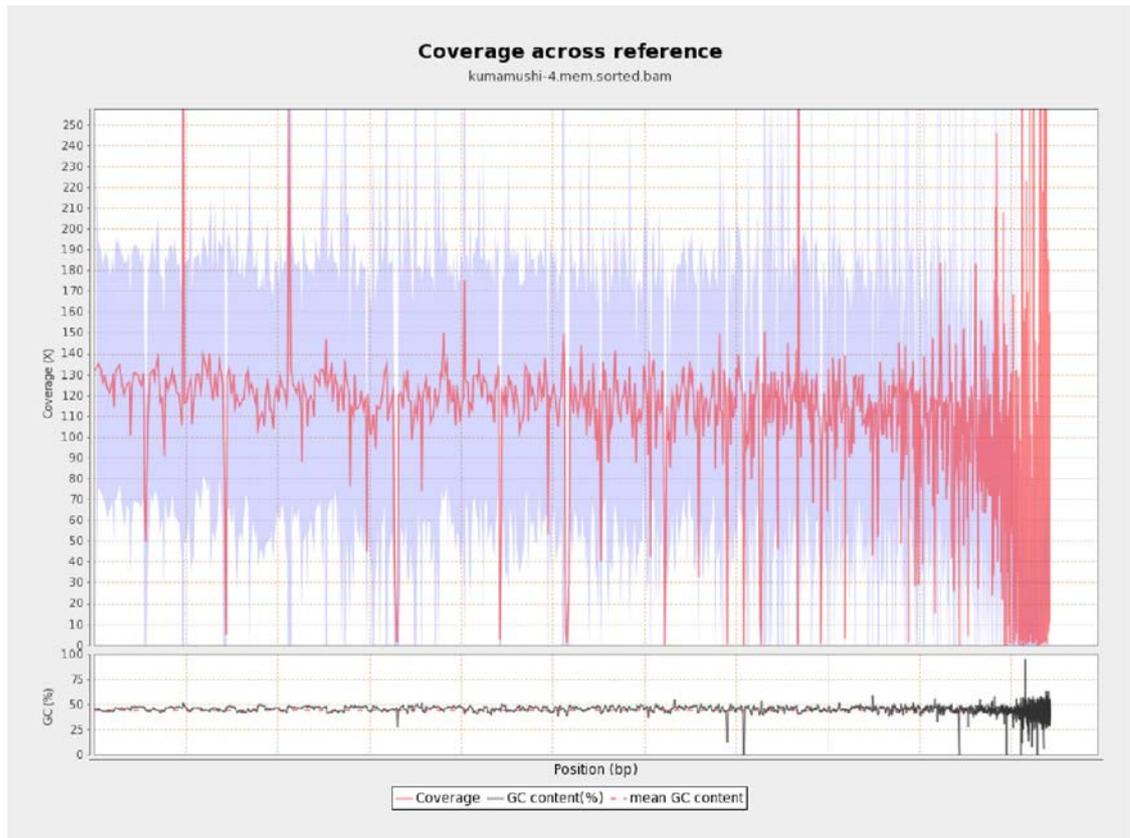
S1	Blobplot analysis of the <i>H. exemplaris</i> genome. . . . .	174
S2	DNA sequencing coverage of the <i>H. exemplaris</i> genome. . . . .	175
S3	tRNA genes in <i>H. exemplaris</i> and <i>R. varieornatus</i> . . . . .	176
S4	Comparisons of genic features between <i>H. exemplaris</i> and <i>R. varieornatus</i> . . . . .	177
S5	Clustered HGT loci in <i>H. exemplaris</i> and <i>R. varieornatus</i> . . . . .	178
S6	Biochemical pathways acquired or supplemented by HGT in <i>H. exemplaris</i> and <i>R. varieornatus</i> . . . . .	179
S7	Phylogeny of protection-related proteins. . . . .	180
S8	Body length in developing individuals of <i>H. exemplaris</i> . . . . .	181
S9	HOX genes were upregulated around Egg 2-3 d in both species. . . . .	182
S10	Expression of HOX genes in the CEL-Seq data set. . . . .	183
S11	Duration of <i>H. exemplaris</i> embryo exposed to low concentration chemicals. . . . .	184
S12	Exposure to low concentration chemicals in both tardigrades. . . . .	185
S13	Duration of <i>H. exemplaris</i> embryogenesis exposed to high-concentration chemicals. . . . .	186
S14	Expression profiles of CAHS and SAHS orthologs. . . . .	187
S15	Profiles of anti-oxidative stress related genes. . . . .	188
S16	Expression profiles of UVC exposed <i>R. varieornatus</i> . . . . .	189
S17	Abnormal mitochondrial structures in anhydrobiosis recovering <i>R. varieornatus</i> . . . . .	190
S18	Phylogenetic analysis g241 orthologs within Tardigrada. . . . .	191
S19	Metal-binding property and structural detail of g12777 globular domain. . . . .	192
S20	Localization of other g12777 orthologs. . . . .	193
S21	Comparison of peroxidase function with bovine liver catalase. . . . .	194

## Chapter 2



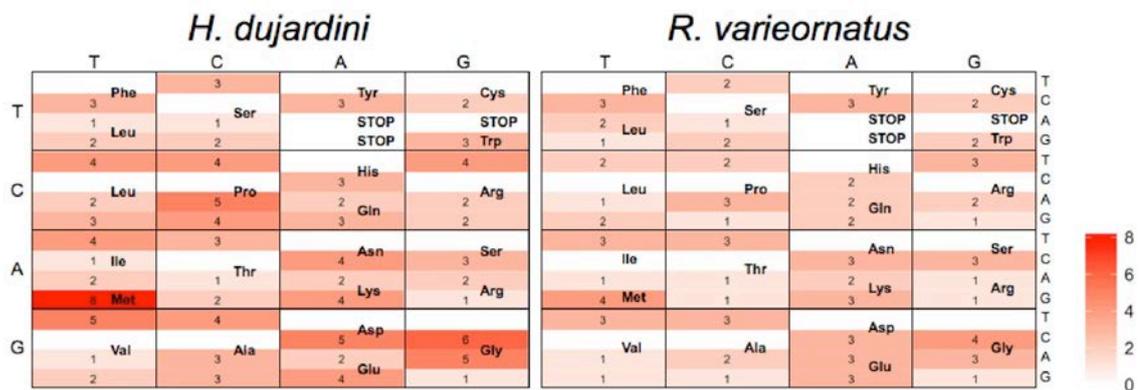
**Supplementary Figure S1: Blobplot analysis of the *H. exemplaris* genome.**

The assembled genome was subjected to Blobplot analysis for possible contamination identification. Scaffolds were submitted to DIAMOND BLASTX analysis against UniProt Reference Proteomes (2017\_11 Version) for taxonomy identification, and this information and genome coverage were analyzed by Blootools.



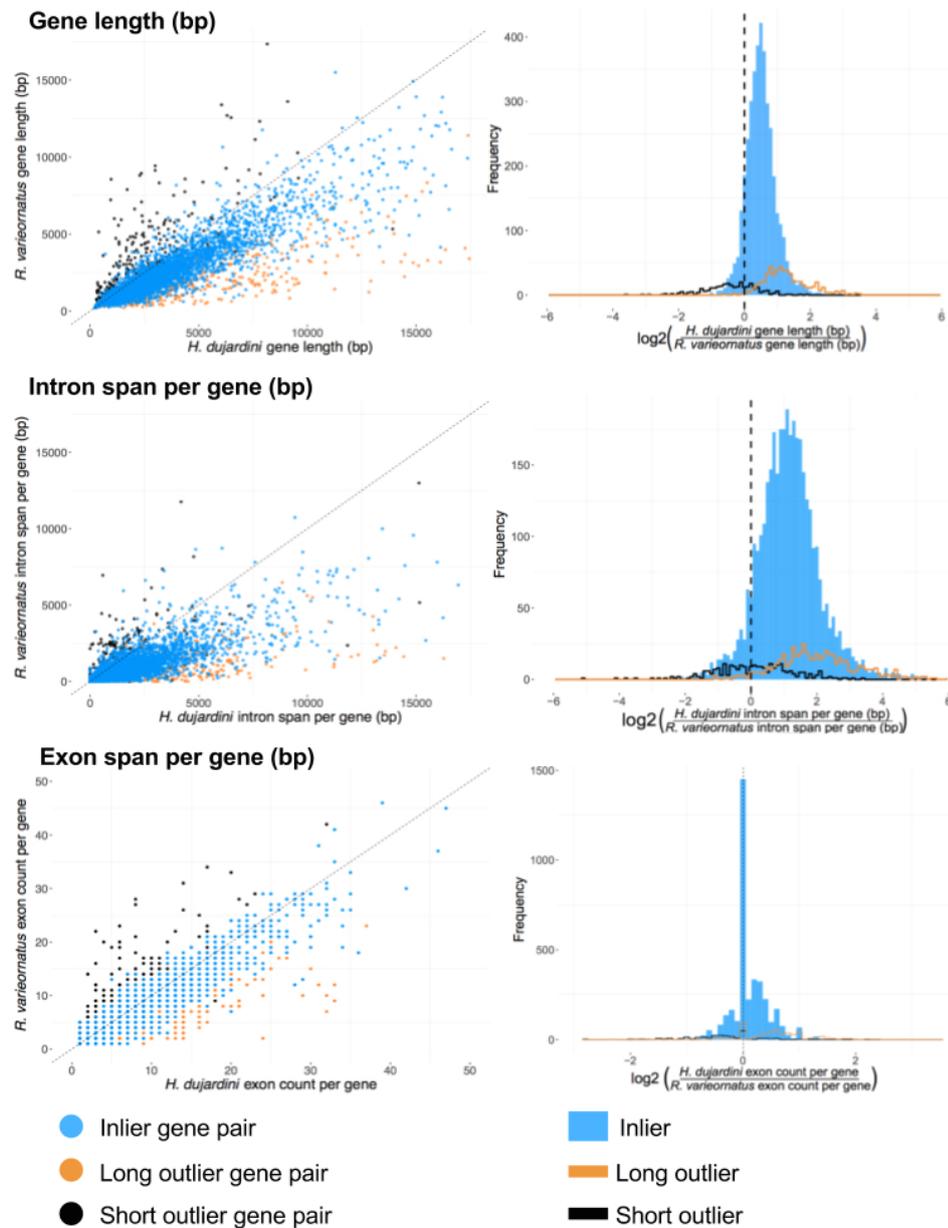
**Supplementary Figure S2: DNA sequencing coverage of the *H. exemplaris* genome.**

Single individual DNA sequencing data (DRR055040) were mapped to the *H. exemplaris* genome with BWA MEM, and the genomic coverage was calculated with Qualimap bamqc. Scaffolds were concatenated, sorted by length (longest first).



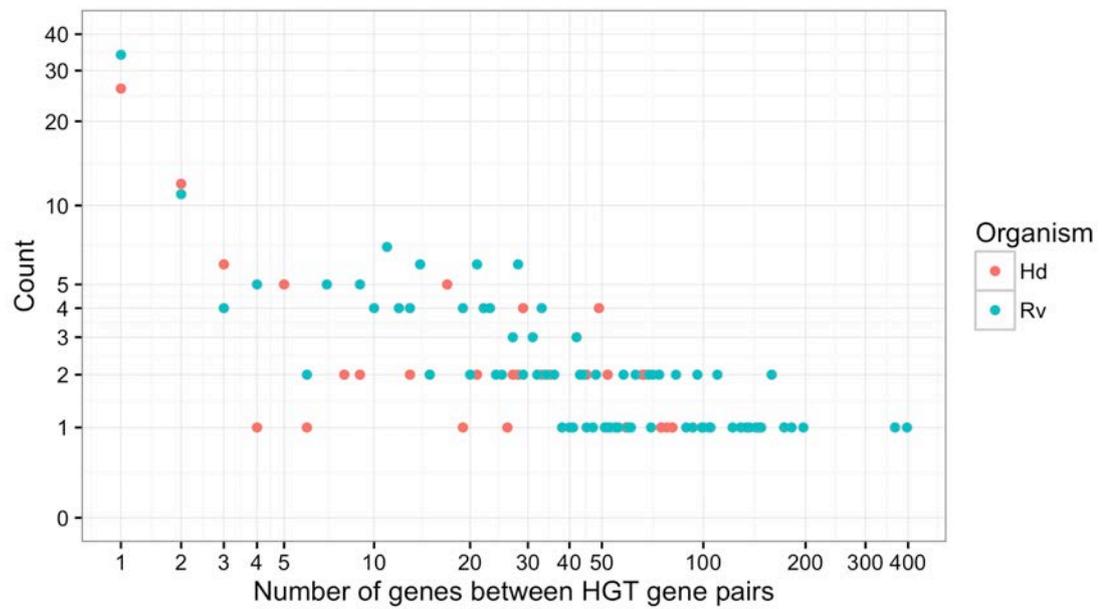
**Supplementary Figure S3: tRNA genes in *H. exemplaris* and *R. varieornatus*.**

tRNA loci were predicted with tRNAScan SE. Each codon is colored by the number of tRNA loci found. *H. dujardini* has been redescribed as *H. exemplaris*.



**Supplementary Figure S4: Comparisons of genic features between *H. exemplaris* and *R. varieornatus*.**

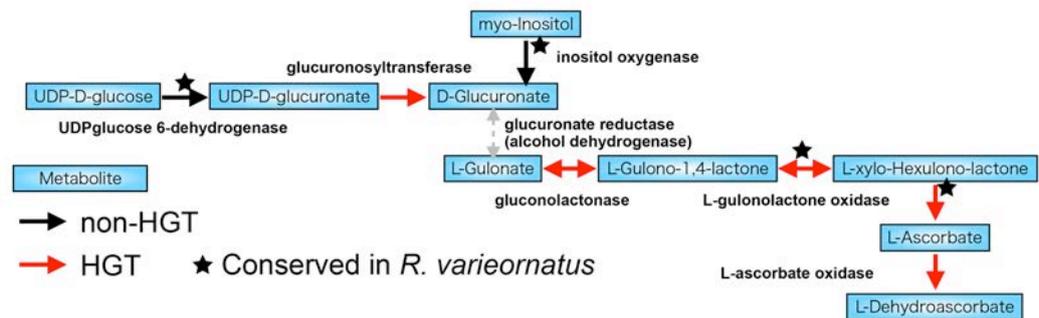
Comparisons of genes structures in 4,728 single-copy orthologues between *H. exemplaris* and *R. varieornatus*. Outliers are defined as genes in *H. exemplaris* which have CDS lengths 20% longer (long outliers; orange; 576 genes) or 20% shorter (short outliers; black; 294 genes) than their orthologues in *R. varieornatus*. Left panels show scatter plots, with *H. exemplaris* on the X and *R. varieornatus* on the Y axes. Right panels show frequency histograms of the ratios of genic feature lengths per gene. A positive  $\log_2$  ratio indicates a trend towards a larger count or span in *H. exemplaris*. *H. dujardini* has been redescribed as *H. exemplaris*.



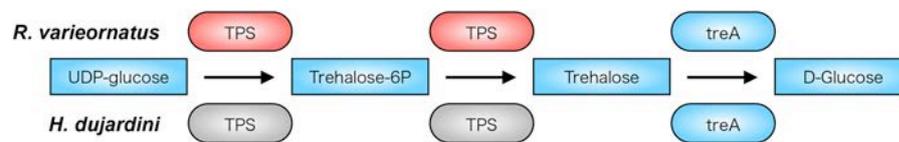
**Supplementary Figure S5: Clustered HGT loci in *H. exemplaris* and *R. varieornatus*.**

HGT candidates are clustered in the genomes of *H. exemplaris* and *R. varieornatus*. The number of genes between each HGT locus was calculated, and a frequency of separations calculated. A separation of 1 indicates that the genes are neighbours. Hd=*H. exemplaris*, Rv=*R. varieornatus*. *H. dujardini* has been rediscrined as *H. exemplaris*.

## A Ascorbate synthesis pathway

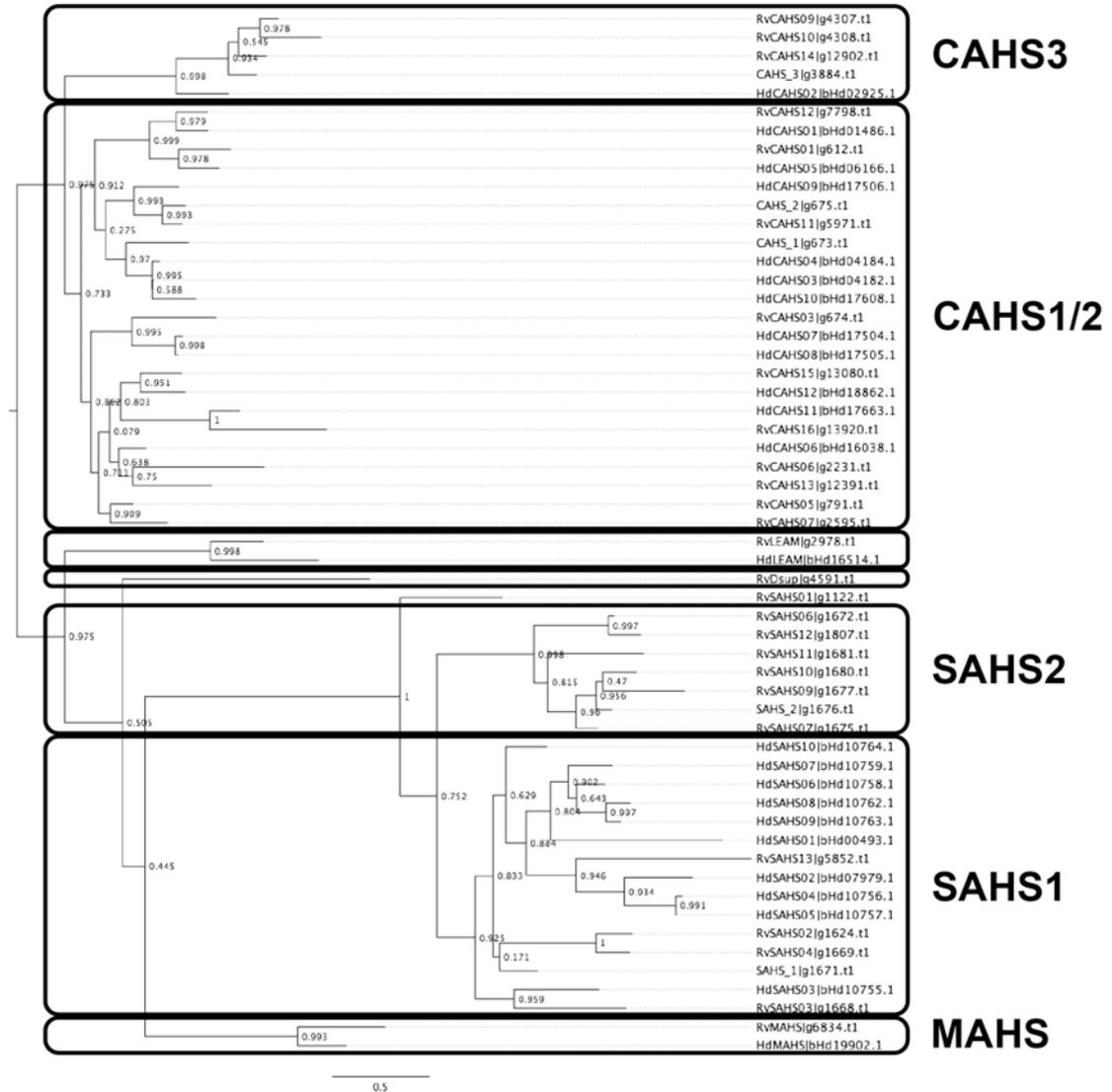


## B Trehalose synthesis via TPS



**Supplementary Figure S6: Biochemical pathways acquired or supplemented by HGT in *H. exemplaris* and *R. varieornatus*.**

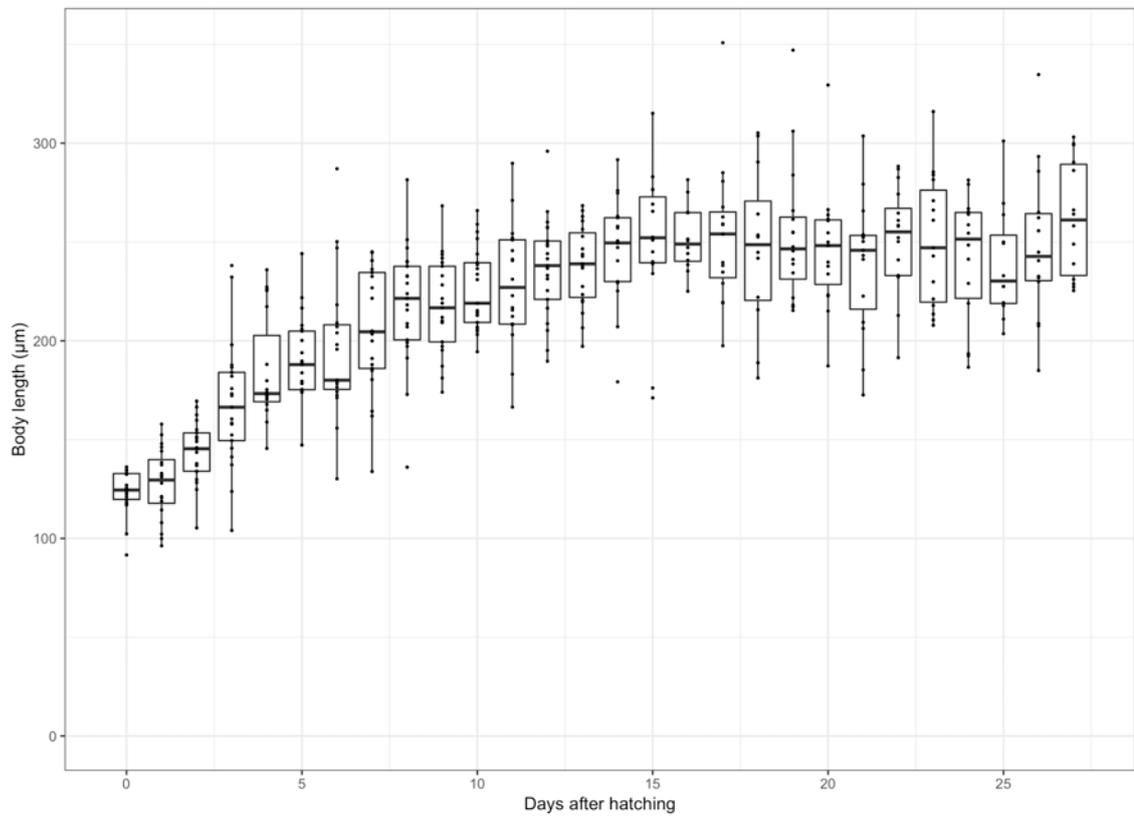
Trehalose and ascorbate synthesis pathways were reconstructed with KAAS and KEGG mapper. (A) Ascorbate synthesis. Many genes in this pathway were derived through HGT (red arrows). (B) Trehalose synthesis. *H. exemplaris* lacks TPS loci (grey lozenges) but contains *treA* trehalase loci (blue lozenges). Both enzymes are present in *R. varieornatus*. *H. dujardini* has been redescribed as *H. exemplaris*.



**Supplementary Figure S7: Phylogeny of protection-related proteins.**

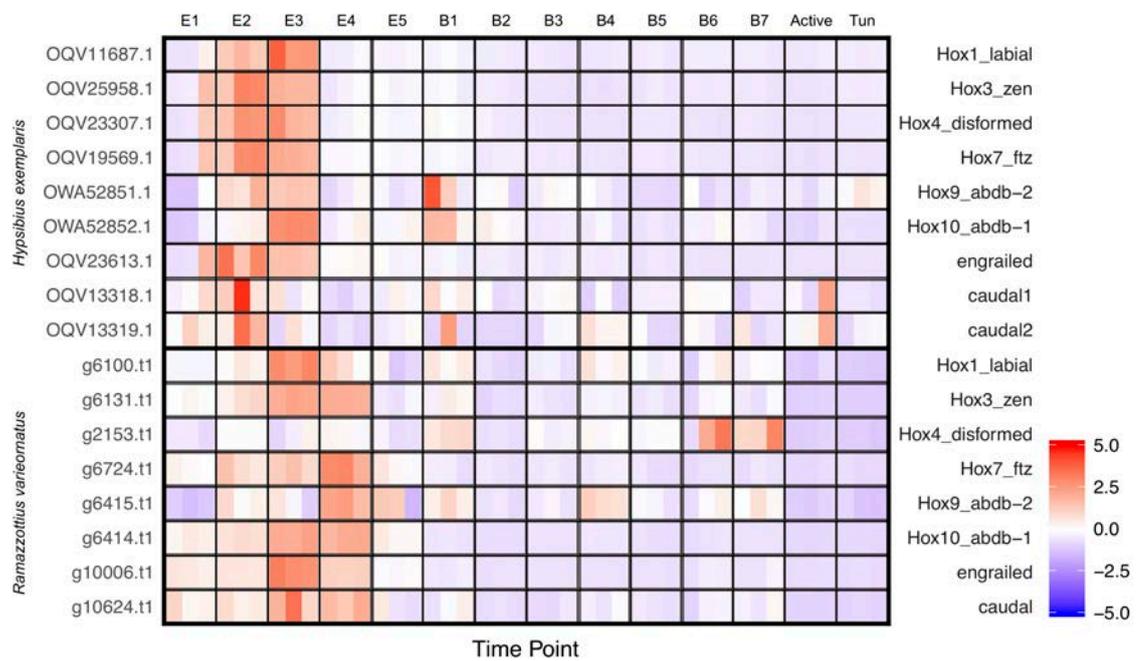
The amino acid sequences of *CAHS*, *SAHS*, *MAHS*, *RvLEAM* and *Dsup* genes in *H. exemplaris* and *R. varieornatus* were aligned with ClustalW2, and a phylogenetic tree was inferred with FastTree with 1,000 bootstraps. Each clade was annotated with the corresponding subtype as defined in *R. varieornatus* (*CAHS1*: g673, *CAHS2*: g675, *CAHS3*: g3884, *SAHS1*:g1671, *SAHS2*: g1676, *MAHS*: g6834, *RvLEAM*: g2978, *Dsup*: g4591). *H. dujardini* has been redescribed as *H. exemplaris*.

## Chapter 4



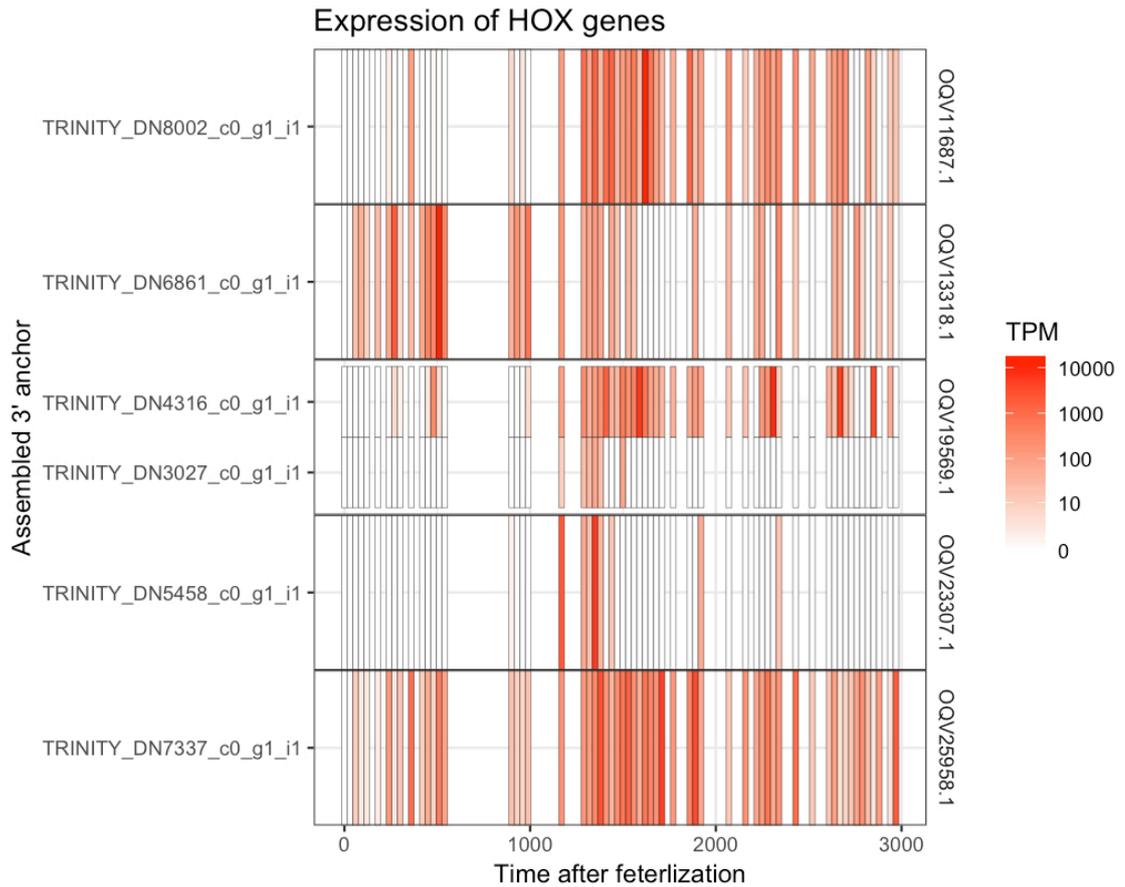
**Supplementary Figure S8: Body length in developing individuals of *H. exemplaris*.**

The body length was quantified for new hatchlings and observed for 28 days.



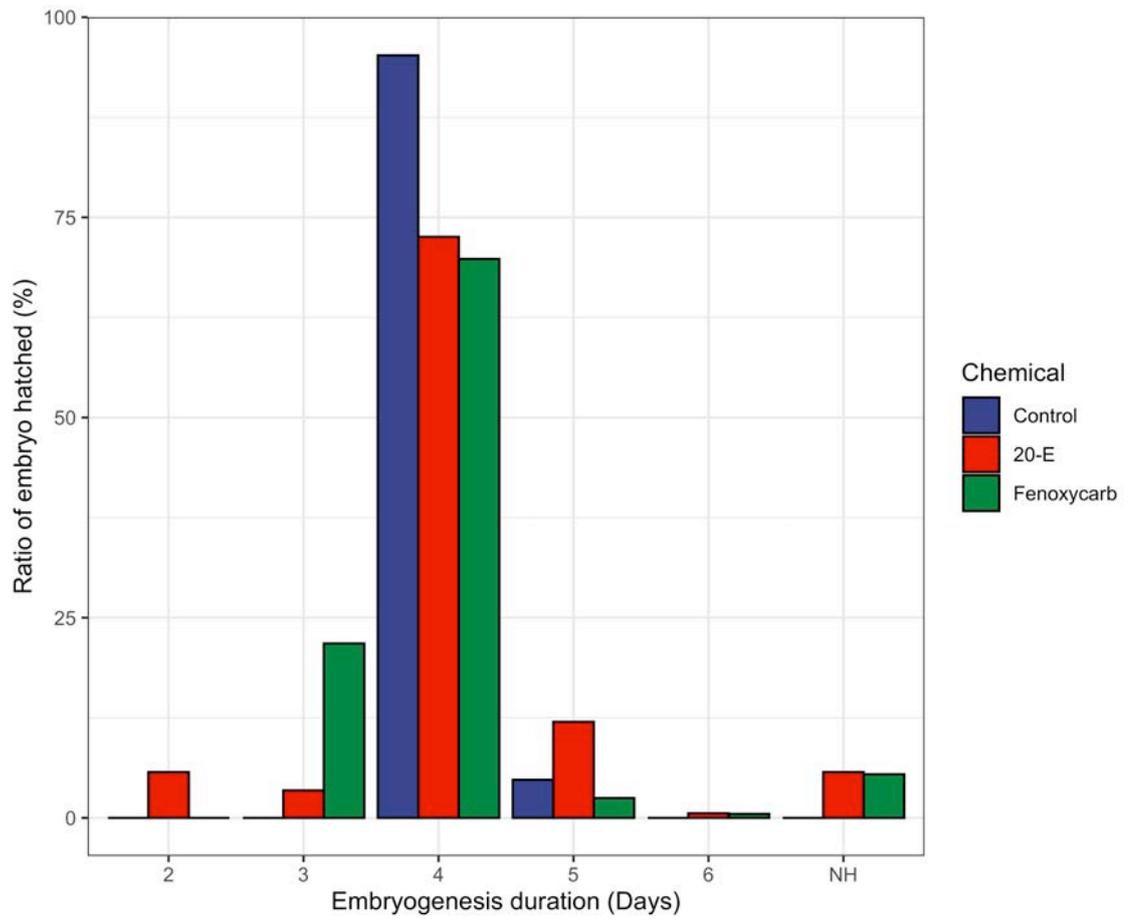
**Supplementary Figure S9: HOX genes were upregulated around Egg 2-3 d in both species.**

Z-scaled TPM values of HOX genes identified in (Yoshida *et al.*, 2017) were visualized as a heatmap. E: Egg, B: Juvenile, Active/Tun: Adult stages.



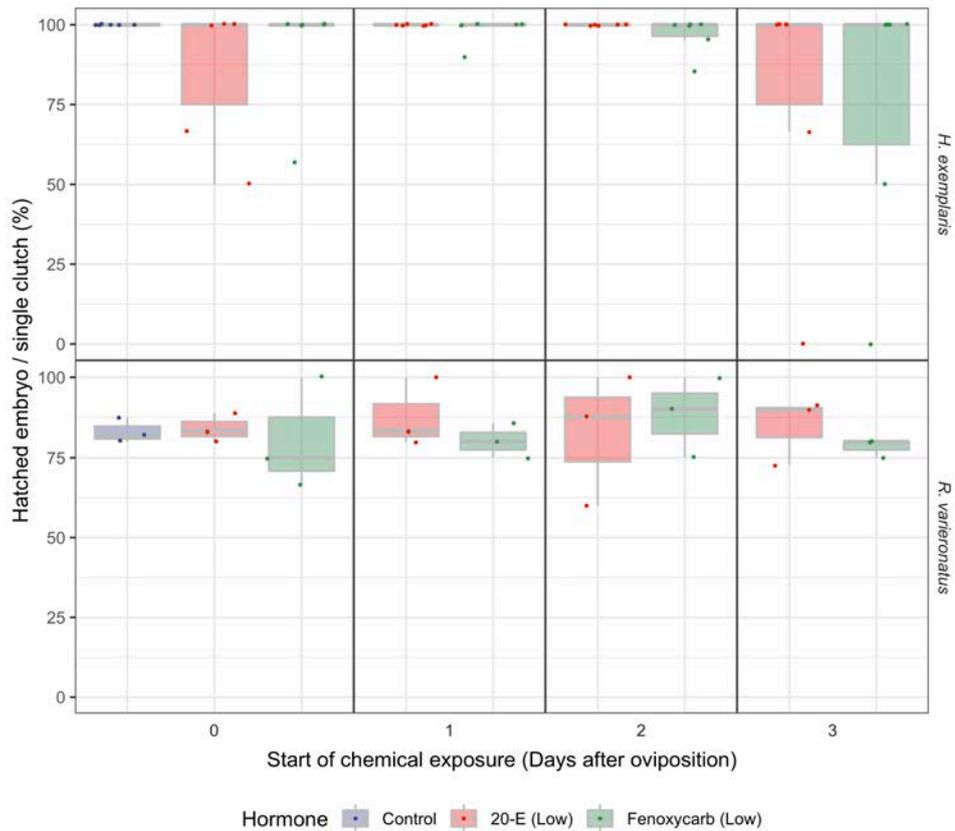
**Supplementary Figure S10: Expression of HOX genes in the CEL-Seq data set.**

CEL-Seq reads were assembled to construct a 3' strand biased transcriptome, used to identify HOX genes. The CEL-Seq reads were mapped against this assembly to calculate gene expression (TPM).



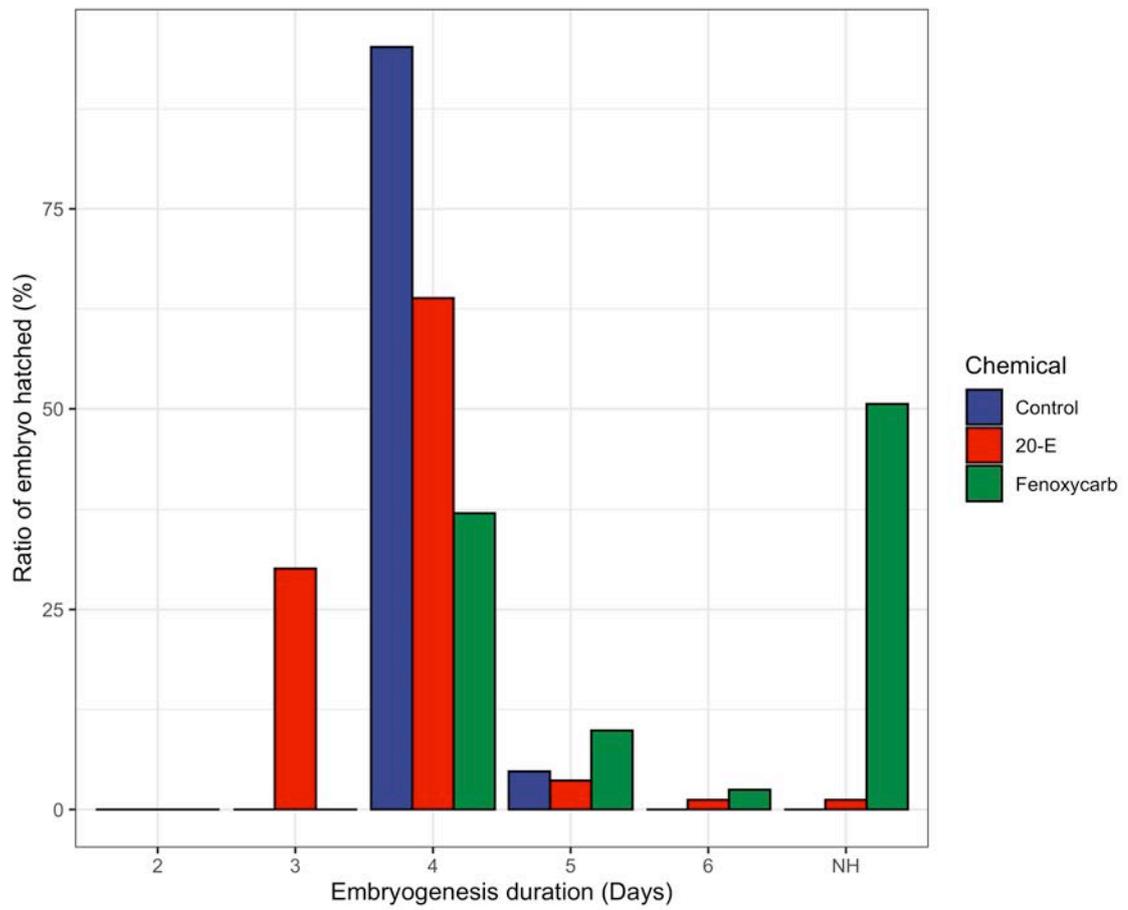
**Supplementary Figure S11: Duration of *H. exemplaris* embryo exposed to low concentration chemicals.**

Days required for hatching in embryo exposed to low concentrations. NH: Not hatched.



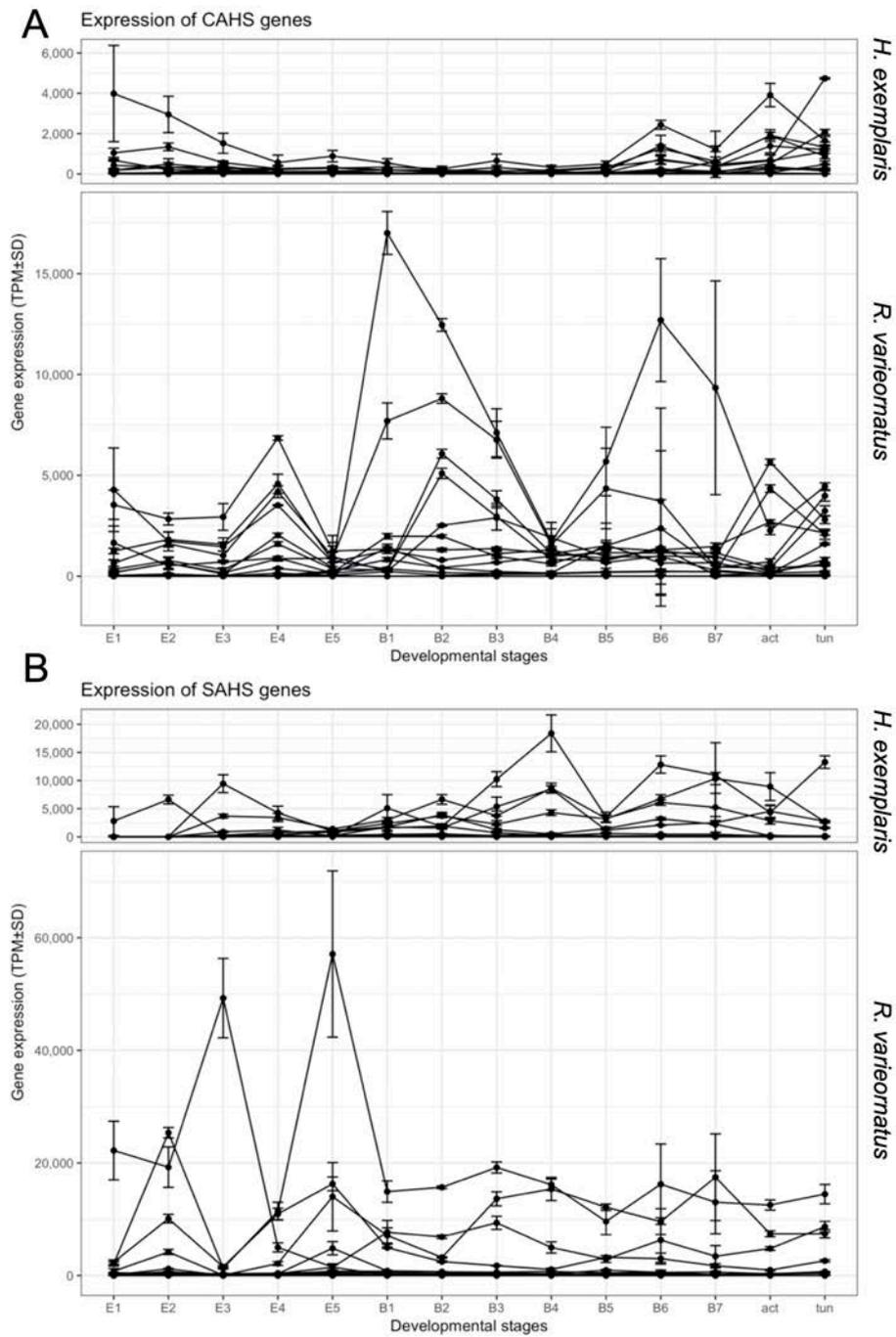
**Supplementary Figure S12: Exposure to low concentration chemicals in both tardigrades.**

Embryos were exposed to low concentrations of Fenoxycarb and 20-E at the indicated days, and the hatching ratio was recorded. Slightly higher variation was observed in the *H. exemplaris* embryos, but not in *R. varieronatus*.



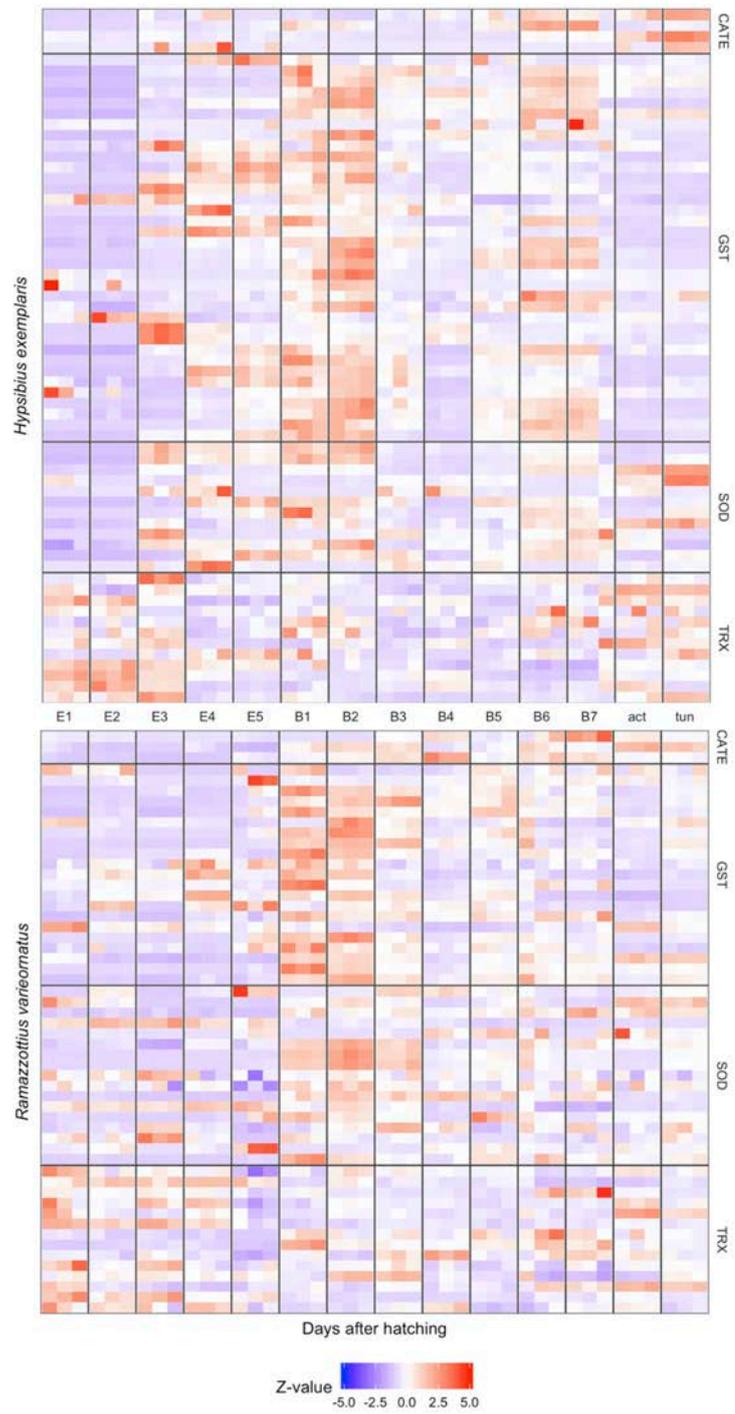
**Supplementary Figure S13: Duration of *H. exemplaris* embryogenesis exposed to high-concentration chemicals.**

Days required for hatching in embryo exposed to high concentrations. NH: Not hatched.



**Supplementary Figure S14: Expression profiles of CAHS and SAHS orthologs.**

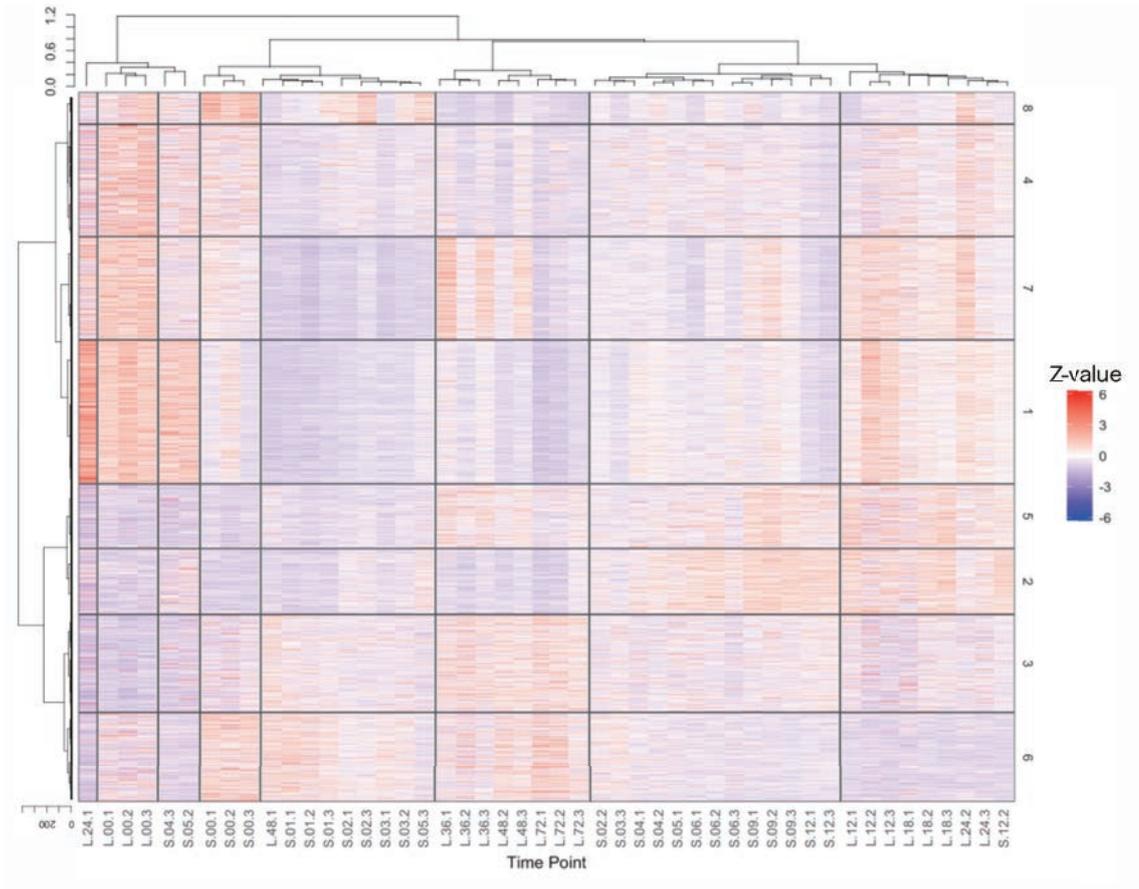
TPM values of (A) CAHS and (B) SAHS genes were plotted with standard deviation as error bars.



**Supplementary Figure S15: Profiles of anti-oxidative stress related genes.**

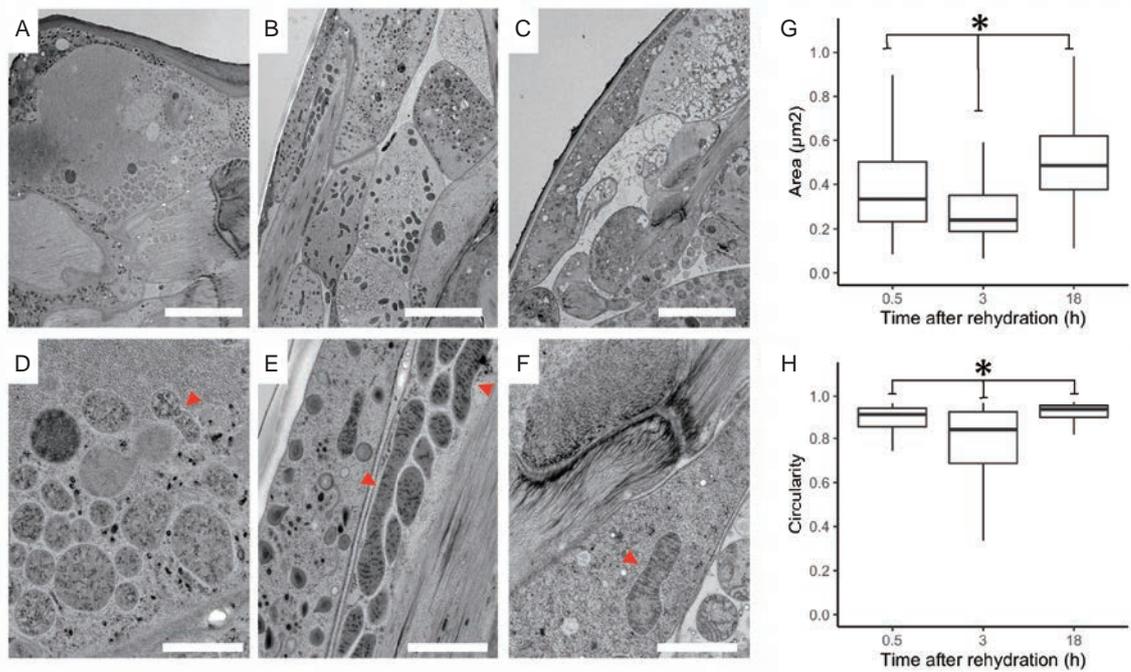
Z-scaled TPM values of catalase (CATE), glutathione S-transferase (GST), superoxide dismutase (SOD), and Thioredoxin reductase (TRX) of both species were plotted as a heatmap.

## Chapter 5



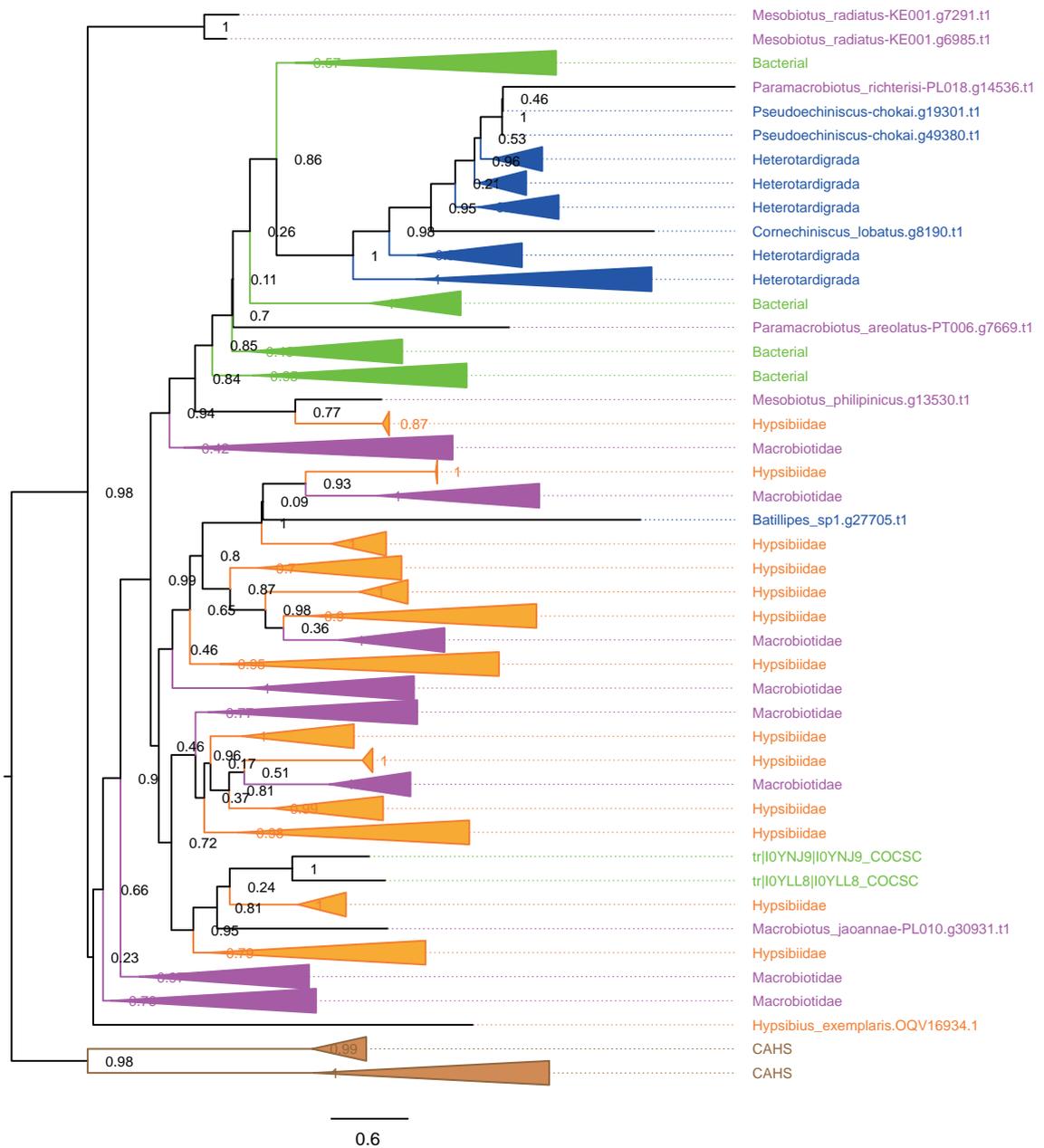
**Supplementary Figure S16: Expression profiles of UVC exposed *R. varieornatus*.**

Gene expression profiles of *R. varieornatus* specimens exposed to 2.5 kJ/m<sup>2</sup> UVC shown as a heatmap. TPM values were Z-scaled to between samples. Expression and sample profiles were clustered with by Ward's method based on Spearman correlation.



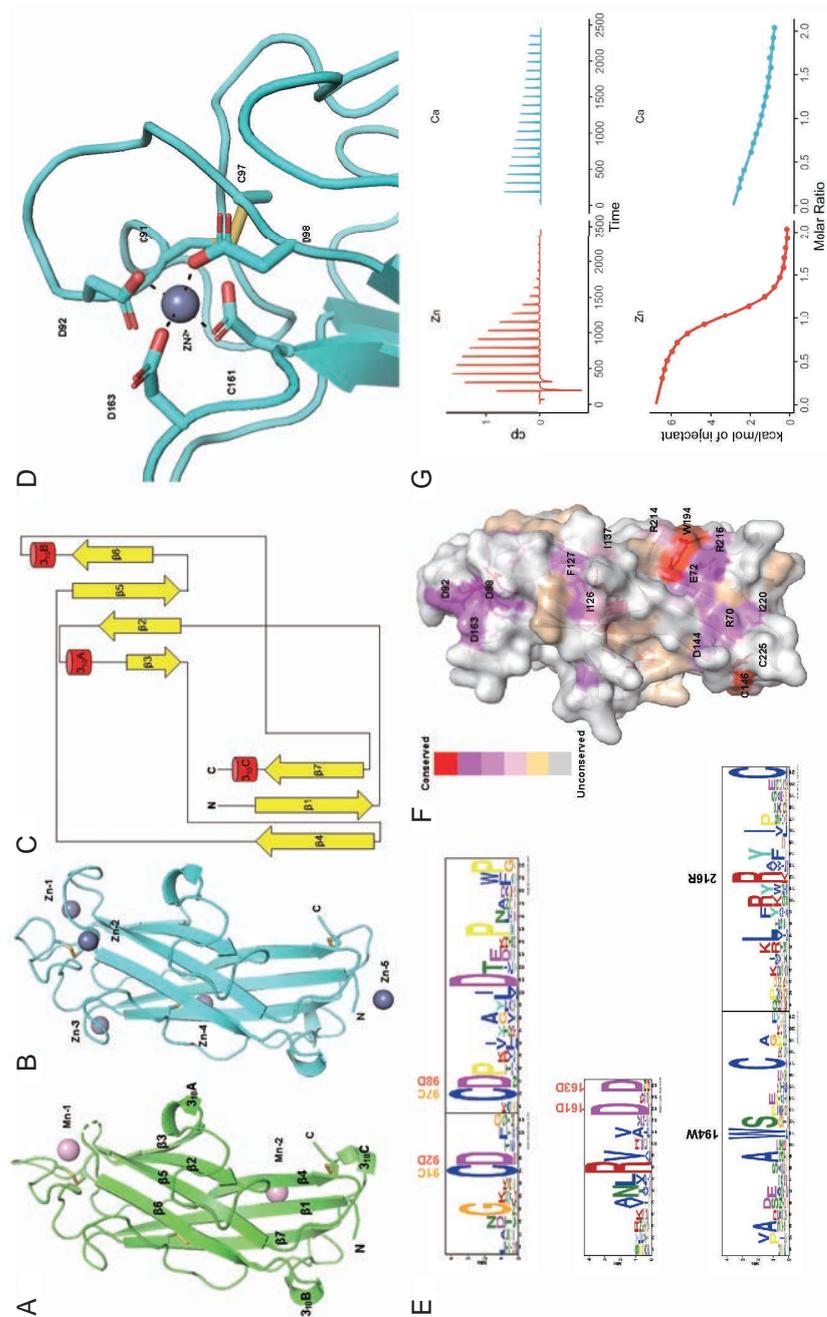
**Supplementary Figure S17: Abnormal mitochondrial structures in anhydrobiosis recovering *R. varieornatus*.**

Transmission electron microscopy images of *R. varieornatus* specimens during recovery from anhydrobiosis. (a,d): 0.5 h (b,e): 3 h, (c,f): 18 h post recovery. (a,b,c): 4,200x magnification of mitochondria. Scale bar 5  $\mu\text{m}$ . (d,e,f): 21,000 magnification of mitochondria. Abnormal mitochondria structures can be observed (Red arrows). Scale bar 1  $\mu\text{m}$ . (g,h) Decrease in area (g) and circularity (h) values in individuals recovering from anhydrobiosis. Mitochondria structures were outlined in ImageJ (0.5 h: 159, 3 h: 355, 18 h: 154 structures) and was statistically tested (ANOVA + Tukey HSD, \* FDR<0.05).



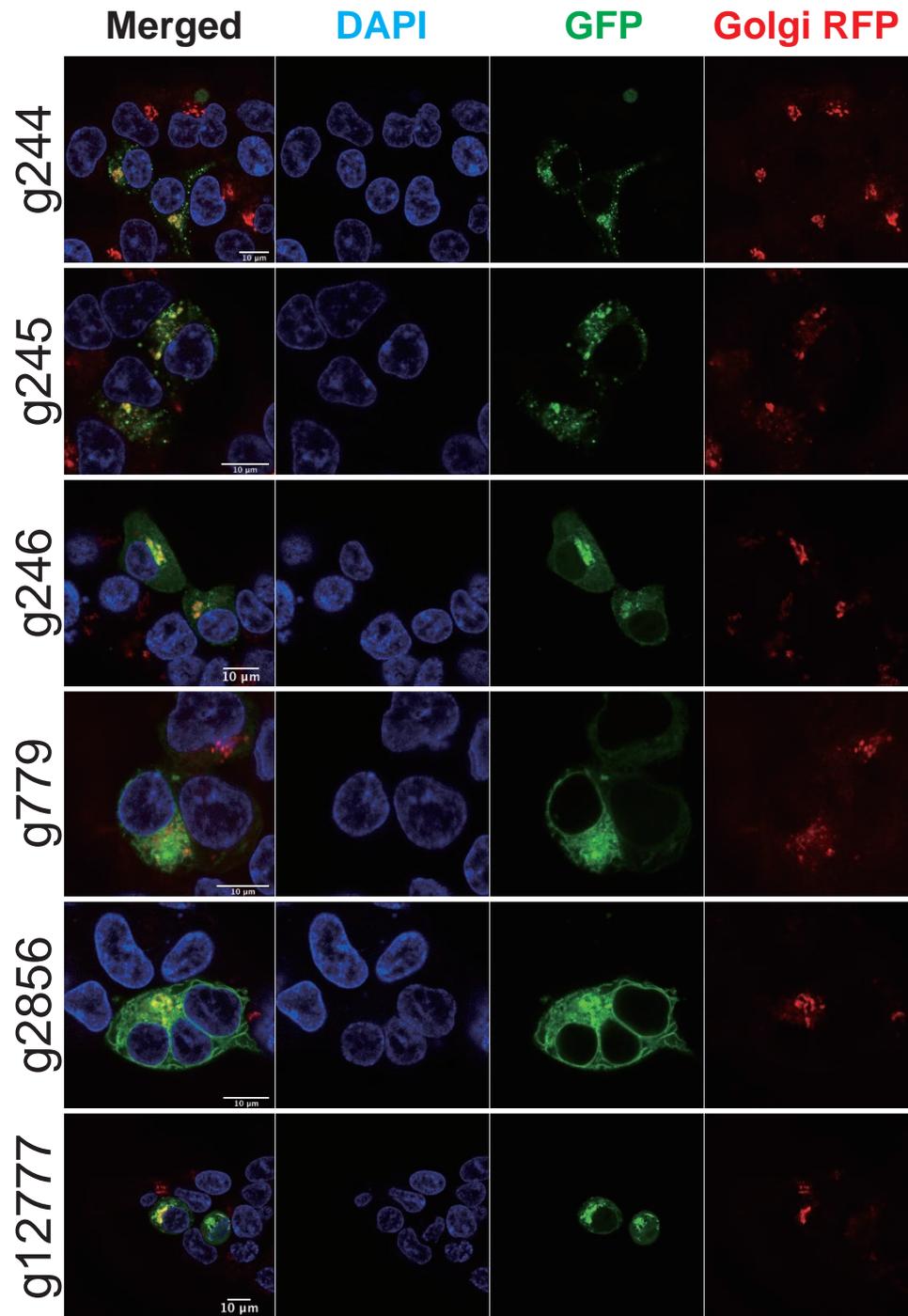
### Supplementary Figure S18: Phylogenetic analysis g241 orthologs within Tardigrada.

Phylogenetic tree of tardigrade g241 orthologs and bacterial orthologs. Amino acid sequences were aligned with MAFFT and the phylogenetic tree was constructed with FastTree with 1,000 bootstraps (-boot, -gamma). CAHS genes were used as an outgroup.



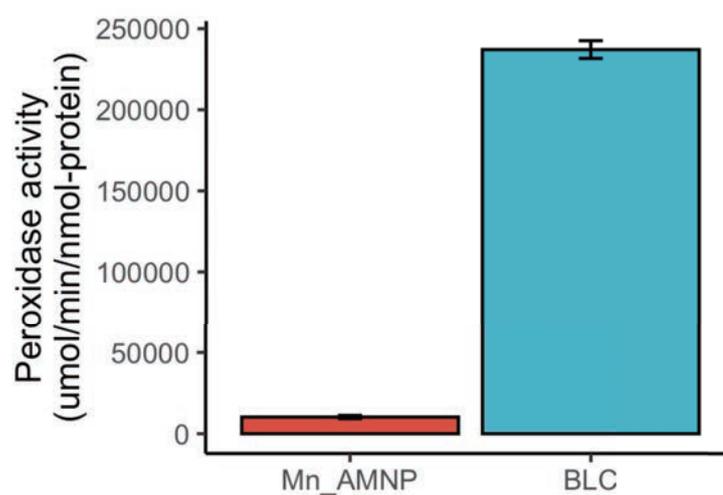
### Supplementary Figure S19: Metal-binding property and structural detail of g12777 globular domain.

(a,b) The crystal structures of catalytic domain of g12777 complexed with  $\text{Mn}^{2+}$  (a) and  $\text{Zn}^{2+}$  (b). All crystallographically observed metal ion binding sites are indicated. (c) Topology diagram of g12777 catalytic domain. (d) Close-up view of the  $\text{Zn}^{2+}$ -binding site. The residues comprising binding site 1 (D92, D98, D161, and D163) and the disulfide bond (C91 and C97) are indicated. (e) Conservation of residues. All g12777 orthologs were submitted to MEME search. (f) Conserved residues of g12777. The conserved residues are colored according to the bit scores obtained from MEME analysis. (g)  $\text{Ca}^{2+}$  and  $\text{Zn}^{2+}$  ion binding affinity measured by isothermal titration calorimetry. The upper two panels indicate the raw data, while the bottom two panels represent the integrated heat values corrected for the heat of dilution and fit to a one-site binding model (solid line); red:  $\text{Ca}^{2+}$ , cyan:  $\text{Zn}^{2+}$ .



**Supplementary Figure S20: Localization of other g12777 orthologs.**

The localization of top six high expressed g12777 orthologs were validated using the same method. All of these orthologs showed localization to the Golgi apparatus.



**Supplementary Figure S21: Comparison of peroxidase function with bovine liver catalase.**

The peroxidase function of g12777 (Mn\_AMNP) was compared to those of bovine liver catalase (BLC).

## Supplementary Data

### List of Supplementary Data

S1	Phylogenetic analyses of potential horizontal gene transfer candidates. . . . .	196
S2	463 putative HGTs in <i>H. exemplaris</i> . . . . .	196
S3	Input proteome FASTA files for orthology analyses. . . . .	196
S4	Orthology analyses—OrthoFinder clustering output. . . . .	197
S5	Orthology analyses—KinFin input and output file. . . . .	197
S6	<i>H. exemplaris</i> miRNA data. . . . .	197
S7	Orthology analyses: Analysis of overrepresentation of protein families in tardigrades. . . . .	198
S8	List of tardigrade specific cluster domains. . . . .	198
S9	Physical clustering of horizontal gene transfer candidates in the genome. . . . .	199
S10	Genes differentially expressed in anhydrobiosis. . . . .	199
S11	Functional annotation of genes differentially expressed in anhydrobiosis. . . . .	200
S12	Dayhoff recoded phylogenetic analyses in NEWICK format. . . . .	200
S13	Expression profiles of genes in group 1 and 2 of SOM clustering in <i>H. exemplaris</i> . . . . .	201
S14	Excel file containing the expression profiles of genes focused in this analysis. . . . .	201
S15	The shells of embryo exposed to more than 24 hours to high concentration of Fenoxycarb were peeled off, and movement were observed. . . . .	201
S16	Annotations of genes overlapping piRNA loci predicted in both PILFER and proTRAC. . . . .	201
S17	Data collection and refinement statistics for g12777. . . . .	202
S18	Statistics of RNA-Seq data obtained in this study. . . . .	202
S19	Highly expressed or regulated genes in Short time-course. . . . .	202
S20	Highly expressed or regulated genes in Long time-course. . . . .	202
S21	Abnormal mitochondrial structure in recovering specimens. . . . .	202
S22	List of differentially expressed genes. . . . .	203
S23	Enrichment analysis of DEGs. . . . .	203
S24	Details on identification of novel stress-responsive gene families. . . . .	203
S25	Expression of g241 orthologs. . . . .	203
S26	Expression of g12777 orthologs. . . . .	203
S27	g12777 and g241 orthologs within Tardigrada identified by BLASTP. . . . .	204
S28	g12777 copy numbers within Tardigrada identified by OrthoFinder. . . . .	204
S29	g241 copy numbers within Tardigrada identified by OrthoFinder. . . . .	204
S30	Refinement of g12777 ortholog gene regions. . . . .	204
S31	Categorization of g241 orthologs. . . . .	204
S32	HGT statistics of g241 orthologs from previous studies. . . . .	204
S33	g241.t1 orthologs identified from NCBI nr database. . . . .	205
S34	g241.t1 orthologs identified from RefSeq database. . . . .	205
S35	Syntenic information of g2856 and g241 orthologs in <i>R. varieornatus</i> . . . . .	205
S36	Details on the crystal structure of g12777 protein. . . . .	205
S37	Metal ion affinity of recombinant proteins. . . . .	205

## Chapter 2

### Supplementary Data S1: Phylogenetic analyses of potential horizontal gene transfer candidates.

[https://github.com/abs-yy/Hypsibius\\_dujardini\\_manuscript/blob/master/Supplementary\\_datas/Supp\\_data\\_S5\\_HGT\\_candidates\\_trees.tar.gz](https://github.com/abs-yy/Hypsibius_dujardini_manuscript/blob/master/Supplementary_datas/Supp_data_S5_HGT_candidates_trees.tar.gz)

These data are the phylogenies from which putative HGTs were classified by taxon-of-origin. This compressed file (tar-gzipped) contains a directory, that holds five subdirectories, “complex”, “metazoan”, “non-metazoan”, “prokaryotes”, “viruses”. In each subdirectory is the set of text files containing NEWICK format trees for putative HGT loci generated by RAxML search, as described in the Methods. Each tree has a name of the form “RAxML\_bestTree.bHd00409.1.faa.aln.ml.newick”: RAxML best tree for gene bHd00409.1, amino acid fasta alignment, generated with maximum likelihood, NEWICK format.

### Supplementary Data S2: 463 putative HGTs in *H. exemplaris*.

[https://github.com/abs-yy/Hypsibius\\_dujardini\\_manuscript/blob/master/Supplementary\\_datas/Supp\\_data\\_S4\\_463\\_putative\\_HGTs.xlsx](https://github.com/abs-yy/Hypsibius_dujardini_manuscript/blob/master/Supplementary_datas/Supp_data_S4_463_putative_HGTs.xlsx)

This excel file lists the properties of the 463 loci identified as potential HGTs in *H. exemplaris* using the HGT index approach. The columns in the table are as follows: (1) protein: name of the *H. exemplaris* HGT candidate; (2) best\_metazoan\_hit: identifier for the best hit in Metazoa (as identifier|database|classification); (3) metazoan\_hitscore: bit score for the best metazoan hit; (4) best\_nonmetazoan\_hit: identifier for the best hit in non-metazoan organisms (as identifier|database|classification); (5) nonmetazoan\_hitscore: bit score for the best non-metazoan hit; (6) hits\_to\_metazoan\_nonmetazoan: summary of whether the protein has hits to one or both of Metazoa and non-Metazoa; (7) score\_difference: difference in BLAST bitscore between metazoan and non-metazoan hits, the HGT index; (8) HGT\_index\_gt30: is the difference between best Metazoan and best non-metazoan hit scores greater than 30 — the HGT index score; (9) has\_metazoan\_neighbour\_on\_same\_scaffold: Does the locus have neighbours on the same scaffold with uncomplicated metazoan similarities? (10) intron\_count: number of introns predicted in the gene; (11) expr\_level\_in\_any\_lib\_gt\_1tpm: Does the gene have evidence of expression from RNA-Seq data? hiexp = greater than 1 TPM, lowexp = less than 1 TPM; (12) tree\_classification: the taxonomic classification of the locus based on RAxML phylogenetic analyses; (13) OrthogroupName: the OrthoFinder orthogroup the locus is a member of; (14) High\_likelihood\_HGTs: classification of likelihood of being a true HGT, with evidence supporting HighLikelihood calls; (15) Ramazzottius\_HGT\_homologues: orthologues and homiologues in *R. varieornatus* (names are *R. varieornatus* gene names); (16) Short\_InterproScan\_Names: summary of InterProScan-derived domain annotations for the locus. Where more than one domain is found, the different domains are separated by “|” characters.

### Supplementary Data S3: Input proteome FASTA files for orthology analyses.

[https://github.com/abs-yy/Hypsibius\\_dujardini\\_manuscript/tree/master/Supplementary\\_datas/Supp\\_data\\_S12\\_proteome\\_fastas](https://github.com/abs-yy/Hypsibius_dujardini_manuscript/tree/master/Supplementary_datas/Supp_data_S12_proteome_fastas)

A tarred and gzipped archive of the protein fasta format files used in orthology analyses, including data from 29 species. In these files only the longest isoform has been retained from groups of sequences derived from the same locus.

**Supplementary Data S4: Orthology analyses—OrthoFinder clustering output.**

[https://github.com/abs-yy/Hypsibius\\_dujardini\\_manuscript/blob/master/Supplementary\\_datas/Supp\\_data\\_S10\\_Orthofinder\\_clustering.tar.gz](https://github.com/abs-yy/Hypsibius_dujardini_manuscript/blob/master/Supplementary_datas/Supp_data_S10_Orthofinder_clustering.tar.gz)

This file, a tarred and gzipped archive, contains the outputs from OrthoFinder analyses conducted under different inflation values.

**Supplementary Data S5: Orthology analyses—KinFin input and output file.**

[https://github.com/abs-yy/Hypsibius\\_dujardini\\_manuscript/blob/master/Supplementary\\_datas/Supp\\_data\\_S11\\_KinFin\\_input.tar.gz](https://github.com/abs-yy/Hypsibius_dujardini_manuscript/blob/master/Supplementary_datas/Supp_data_S11_KinFin_input.tar.gz)

This tarred and gzipped archive contains the input files for KinFin and the output from the analyses. The six input files include two alternative tree hypotheses, a file giving the functional annotation of the proteins, a species classification file and files giving the links between species and sequences. The output files are given for each inflation value applied in OrthoFinder, and include analyses under both hypotheses of tardigrade relationships.

**Supplementary Data S6: *H. exemplaris* miRNA data.**

[https://github.com/abs-yy/Hypsibius\\_dujardini\\_manuscript/blob/master/Supplementary\\_datas/Supp\\_data\\_S1\\_miRNA\\_data.xlsx](https://github.com/abs-yy/Hypsibius_dujardini_manuscript/blob/master/Supplementary_datas/Supp_data_S1_miRNA_data.xlsx)

(1) Provisional id: identifier for miRNA sequence, based on scaffold of origin. (2) miRDeep2 score: score for miRNA as given by miRDeep2. (3) Total read count: total number of reads mapping to this miRNA model. (4) Mature read count: number of reads mapping to majority strand of predicted hairpin model. (5) Loop read count: number of reads mapping to predicted loop region of model. (6) Star read count: number of reads mapping to minority strand of predicted hairpin model. (7) Significant randfold *p*-value. (8) Example miRBase miRNA with the same seed: an example from miRBase of this miRNA family. (9) Consensus mature sequence. (10) Consensus star sequence. (11) Consensus precursor sequence. (12) Precursor coordinate: coordinates of the predicted miRNA on the scaffold.

### Supplementary Data S7: Orthology analyses: Analysis of overrepresentation of protein families in tardigrades.

[https://github.com/abs-yy/Hypsibius\\_dujardini\\_manuscript/blob/master/Supplementary\\_datas/Supp\\_data\\_S2\\_Tardigrade\\_counts\\_representation\\_tests.xlsx](https://github.com/abs-yy/Hypsibius_dujardini_manuscript/blob/master/Supplementary_datas/Supp_data_S2_Tardigrade_counts_representation_tests.xlsx)

Supplementary comments on protein families with four-fold overrepresentation in tardigrades. Family OG0000104 had 284 members, 276 of which derived from tardigrades, and was annotated as having a receptor ligand binding domain, but was not otherwise distinguished. OG004022 had 5 members in *H. exemplaris* and 9 members in *R. variornatus*, but a mode of 1 (and maximum of 4 in *Octopus bimaculatus*) in other species. It is a member of a deeply conserved, otherwise uncharacterized, transmembrane protein family of unknown function. OG0001636 gathers a deeply conserved ATP-binding cassette family, and while the 27 other species had a mode of 1 (and a maximum of 2), *R. variornatus* had 4 and *H. exemplaris* had 9 copies. OG0002927 encodes protein kinases, present in 23 of the 29 species, with 6 in *H. exemplaris*, 5 in *R. variornatus* and a mode of 1 elsewhere. OG0004228 is annotated as a relish-like transcription factor, and has 1 copy in the non-tardigrade species (except for two insects with 2) and 5 copies in each tardigrade. OG0001359, with 1 copy in most species, 8 in *H. exemplaris*, 8 in *R. variornatus*, and 4 in *Solenopsis invictus*, is likely to be a SAM-dependent methyltransferase (type 11), possibly involved in coenzyme biosynthesis. OG001949 had 1 copy in most species but 6 in *H. exemplaris* and 4 in *R. variornatus*, and is annotated as a RAB GTP hydrolase. OG0003870 was unannotated (containing only matches to domain of unknown function DUF1151), and elevated in *R. variornatus* (9 copies) compared to other species (mode of 1; *H. exemplaris* had 2). The three clusters with depletion in the tardigrades were OG0000604, encoding an exoribonuclease (1 copy in each tardigrade, but an average of three copies in the other 27 species), OG0000950, a 3'5'-cyclic nucleotide phosphodiesterase (1 in tardigrades versus 2.3 elsewhere) and OG00001138, an EF-hand protein (1 in tardigrades versus 2 elsewhere). In the Excel-format file the columns are as follows: (1) cluster\_id: protein family identifier; (2) cluster\_status: "present" indicates present in either or both tardigrades; (3) cluster\_type: "shared" indicates present in both tardigrades; (4) cluster\_protein\_count: the number of proteins in the cluster; (5) cluster\_proteome\_count: the number of proteomes contributing a member to the cluster; (6) Tardigrade\_protein\_count: the number of members that come from the tardigrade species; (7) Tardigrade\_mean\_count: the mean number of members per tardigrade; (8) non\_Tardigrade\_mean\_count: the mean number of members across the other species; (9) representation: the direction of the difference found; "enriched" = enriched in tardigrades; "depleted" = depleted in tardigrades; (10) log2\_mean(Tardigrade/others): the ratio of mean per-species membership in tardigrades to others; (11) mwu\_pvalue(Tardigrade vs. others): Mann-Whitney U-test for difference in representation in the tardigrades; (12) Tardigrade\_proteome\_coverage: the proportion of tardigrade proteomes that contributed to the cluster; (13) Tardigrade\_proteomes\_present\_count: the number of tardigrade proteomes in the clustering analyses; (14) Tardigrade\_proteomes\_present: which tardigrade proteomes contributed members to the cluster; (15) GO: GO terms associated with the cluster (based on InterPro matches; given as GO identifier:number of members with that term); (16) IPR: InterPro domains associated with the cluster (given as IPR identifier:number of members with that domain); (17) Pfam: Pfam domains associated with the cluster (given as Pfam identifier:number of members with that domain); (18) SignalP\_EUK: Presence of eukaryotic signal peptides in members of the cluster (given as number of members with each pattern of signal peptide (SP) and transmembrane (TM) domains).

### Supplementary Data S8: List of tardigrade specific cluster domains.

[https://github.com/abs-yy/Hypsibius\\_dujardini\\_manuscript/blob/master/Supplementary\\_datas/Supp\\_data\\_S3\\_cluster\\_domains\\_by\\_node.Tardigrada.xlsx](https://github.com/abs-yy/Hypsibius_dujardini_manuscript/blob/master/Supplementary_datas/Supp_data_S3_cluster_domains_by_node.Tardigrada.xlsx)

### Supplementary Data S9: Physical clustering of horizontal gene transfer candidates in the genome.

[https://github.com/abs-yy/Hypsibius\\_dujardini\\_manuscript/blob/master/Supplementary\\_datas/Supp\\_data\\_S6\\_HGT\\_cluster\\_matrix.xlsx](https://github.com/abs-yy/Hypsibius_dujardini_manuscript/blob/master/Supplementary_datas/Supp_data_S6_HGT_cluster_matrix.xlsx)

This table summarises the gene neighbourhood analysis of *H. exemplaris* HGT candidates. Highlighted rows indicate putative HGT gene sets that are very close on the *H. exemplaris* genome. For this analysis, the scaffolds were concatenated into one linear sequence, and thus one step in the assessment was determining the scaffold origin of loci (columns 13, 14). The columns are as follows: (1) Gene: putative HGT candidate gene name; (2) Count: index count; (3) Scaffold: scaffold from which the gene is predicted; (4) Number: number of gene on scaffold; (5) PrevNum: number of the closest HGT candidate to the left of the gene (set to -1,000,000 if the gene is the first on the scaffold); (6) PrevGene: putative downstream HGT candidate gene name; (7) Number: number of gene on scaffold; (8) NextNum: number of the closest HGT candidate to the right of the gene (set to 1,000,000 if the gene is the last on the scaffold); (9) NextGene: putative upstream HGT candidate gene name; (10) DifPrev: number of genes between focal gene and next HGT candidate to left; (11) DifNext: number of genes between focal gene and next HGT candidate to right; (12) DiffClose: number of genes between focal gene and next closest HGT candidate (minimum of columns 10 and 11); (13) ScaffClose: the scaffold on which the closest next HGT gene resides; (14) SameScaf: whether this scaffold is the same as the scaffold on which the focal gene resides; (15) DiffClose: number of genes between focal gene and next closest HGT candidate on the same scaffold (minimum of columns 10 and 11); (16) Phylo: phylogenetic classification of the taxon-of-origin of the focal locus; (17) TrEMBL: best BLAST match of the focal locus in TrEMBL; (18) Swiss-Prot: best BLAST match of the focal locus in SwissProt.

### Supplementary Data S10: Genes differentially expressed in anhydrobiosis.

[https://github.com/abs-yy/Hypsibius\\_dujardini\\_manuscript/blob/master/Supplementary\\_datas/Supp\\_data\\_S7\\_DEG\\_list.xlsx](https://github.com/abs-yy/Hypsibius_dujardini_manuscript/blob/master/Supplementary_datas/Supp_data_S7_DEG_list.xlsx)

This table describes the analysis of differential gene expression under anhydrobiosis of *H. exemplaris*, based on mRNA-Seq of active and anhydrobiotic, “tun” animals. TPM: transcripts per million mapped. The columns are as follows: (1) Gene: gene name; (2) Active-1: TPM in sample of active animals 1, assessed by Kallisto; (3) Active-2: TPM in sample of active animals 2, assessed by Kallisto; (4) Active-3: TPM in sample of active animals 3, assessed by Kallisto; (5) Tun-1: TPM in sample of tun animals 1, assessed by Kallisto; (6) Tun-2: TPM in sample of tun animals 2, assessed by Kallisto; (7) Tun-3: TPM in sample of tun animals 3, assessed by Kallisto; (8) FoldChange: fold change in tun animals (mean of tun TPM/mean of active TPM); (9) Active-1: fragments mapped in sample of active animals 1 using BWA; (10) Active-2: fragments mapped in sample of active animals 2 using BWA; (11) Active-3: fragments mapped in sample of active animals 3 using BWA; (12) Tun-1: fragments mapped in sample of tun animals 1 using BWA; (13) Tun-2: fragments mapped in sample of tun animals 2 using BWA; (14) Tun-3: fragments mapped in sample of tun animals 3 using BWA; (15) *p*-value: *p*-value of differential expression between tun and active animals estimated using DESeq; (16) FDR: false discovery rate estimate; (17) Database Search: BLAST matches for locus based on BLAST search of SwissProt/TrEMBL. Blank entries indicate no matches with *E*-values less than 1E-5.

**Supplementary Data S11: Functional annotation of genes differentially expressed in anhydrobiosis.**

[https://github.com/abs-yy/Hypsibius\\_dujardini\\_manuscript/blob/master/Supplementary\\_datas/Supp\\_data\\_S9\\_Tardigrade\\_DEGs\\_functional\\_annotation.xlsx](https://github.com/abs-yy/Hypsibius_dujardini_manuscript/blob/master/Supplementary_datas/Supp_data_S9_Tardigrade_DEGs_functional_annotation.xlsx)

Functional annotations of the *H. exemplaris* genes found to be differentially expressed under anhydrobiosis, based on InterPro domain matches. The columns in the table are: (1) cluster\_id: cluster identifier summary; (2) cluster\_id: cluster identifier (with one line for each domain matched); (3) domain\_source: database source of domain model (IPR = InterProScan); (4) domain\_id: database identifier for domain model; (5) domain\_description: summary description for domain model; (6) protein\_count: total number of proteins in the cluster; (7) protein\_count\_with\_domain: number of proteins in the cluster annotated with this domain; (8) HDUJA-DEG: number of *H. exemplaris* members of the cluster that are differentially expressed under anhydrobiosis; (9) HDUJA-DEG-DOMAINS: number of *H. exemplaris* members of the cluster that are differentially expressed under anhydrobiosis that have this domain, and how many copies of the domain each has (e.g., 4x1 means four members have 1 copy each); (10) HDUJA-HGT: number of *H. exemplaris* members of the cluster that are potential HGT candidates; (11) RVARI-DEG: number of *R. varieornatus* members of the cluster that are differentially expressed under anhydrobiosis; (12) RVARI-DEG-DOMAINS: number of *R. varieornatus* members of the cluster that are differentially expressed under anhydrobiosis that have this domain, and how many copies of the domain each has (e.g., 4x1 means four members have 1 copy each); (13) proteomes\_with\_domain\_fraction: proportion of all proteomes represented in the cluster where a member has the domain; (14) proteomes\_with\_domain: proteomes with members in the cluster where that member has the domain.

**Supplementary Data S12: Dayhoff recoded phylogenetic analyses in NEWICK format.**

[https://github.com/abs-yy/Hypsibius\\_dujardini\\_manuscript/blob/master/Supplementary\\_datas/Supp\\_data\\_S8\\_Dayhoff\\_recoded\\_phylogenetic\\_analyses\\_in\\_NEWICK\\_format.txt](https://github.com/abs-yy/Hypsibius_dujardini_manuscript/blob/master/Supplementary_datas/Supp_data_S8_Dayhoff_recoded_phylogenetic_analyses_in_NEWICK_format.txt)

## Chapter 4

**Supplementary Data S13: Expression profiles of genes in group 1 and 2 of SOM clustering in *H. exemplaris*.**

[https://github.com/abs-yy/Yoshida\\_and\\_Sugiura\\_etal\\_2018/blob/master/Sdata\\_1\\_hypsibius\\_som\\_group1ANND2.xlsx](https://github.com/abs-yy/Yoshida_and_Sugiura_etal_2018/blob/master/Sdata_1_hypsibius_som_group1ANND2.xlsx)

**Supplementary Data S14: Excel file containing the expression profiles of genes focused in this analysis.**

[https://github.com/abs-yy/Yoshida\\_and\\_Sugiura\\_etal\\_2018/blob/master/Sdata\\_2\\_geneexpression.xlsx](https://github.com/abs-yy/Yoshida_and_Sugiura_etal_2018/blob/master/Sdata_2_geneexpression.xlsx)

**Supplementary Data S15: The shells of embryo exposed to more than 24 hours to high concentration of Fenoxycarb were peeled off, and movement were observed.**

[https://github.com/abs-yy/Yoshida\\_and\\_Sugiura\\_etal\\_2018/blob/master/Sdata\\_3\\_piRNA\\_clusters\\_proTRAC.bed](https://github.com/abs-yy/Yoshida_and_Sugiura_etal_2018/blob/master/Sdata_3_piRNA_clusters_proTRAC.bed)

**Supplementary Data S16: Annotations of genes overlapping piRNA loci predicted in both PILFER and proTRAC.**

[https://github.com/abs-yy/Yoshida\\_and\\_Sugiura\\_etal\\_2018/blob/master/Sdata\\_5\\_piRNA\\_clusters\\_intersect.bed](https://github.com/abs-yy/Yoshida_and_Sugiura_etal_2018/blob/master/Sdata_5_piRNA_clusters_intersect.bed)

## Chapter 5

Supplementary data for this chapter can be accessed from the bioRxiv preprint (<https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.supplementary-material>).

### **Supplementary Data S17: Data collection and refinement statistics for g12777.**

<https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.supplementary-material>

Crystal parameters and refinement statistics for Mn<sup>2+</sup>- and Zn<sup>2+</sup>-bound g12777 protein catalytic domain are summarized. One crystal for each structure was used for diffraction data collection.

### **Supplementary Data S18: Statistics of RNA-Seq data obtained in this study.**

<https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.supplementary-material>

Statistics of each RNA-Seq data sequenced in this study. Mapping ratio for BWA mapping are shown.

### **Supplementary Data S19: Highly expressed or regulated genes in Short time-course.**

<https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.supplementary-material>

Genes with TPM values > 1,000 and fold change > 4 are highlighted in purple and orange, respectively.

### **Supplementary Data S20: Highly expressed or regulated genes in Long time-course.**

<https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.supplementary-material>

Genes with TPM values > 1,000 and fold change > 4 are highlighted in purple and orange, respectively.

### **Supplementary Data S21: Abnormal mitochondrial structure in recovering specimens.**

We found mitochondria-related genes, in particular, the mitochondrial chaperone BCS1 and the mitophagy related gene Sequestosome, to have a high fold change. We have found similar inductions during *H. exemplaris* anhydrobiosis, suggesting the existence of mitochondria stresses during desiccation or UVC exposure-response. Oxidative stress caused during desiccation (or UVC exposure) would also induce extensive stress to the mitochondria. We conducted TEM observations of *R. varieornatus* specimens recovering from desiccation. We observed abnormal mitochondrial morphology in specimens 3 hours after rehydration, not in 30 minutes (Supplementary Figure S17). This suggests that mitochondrial stress may be occurring during the initial 3 hours, resulting in morphological changes at 3 hours after rehydration. Induction of Sequestosome may be related to mitophagy of damaged mitochondria. A previous study using *H. exemplaris* specimens also has observed mitochondrial morphology, however, they have seen an increase in mitochondrial size, not a decrease as we have observed (Richaud *et al.*, 2020). These inconsistencies may reflect the differences in mitochondrial damage that occurred during anhydrobiosis in the two species.

**Supplementary Data S22: List of differentially expressed genes.**

[https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.  
supplementary-material](https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.supplementary-material)

List of genes differentially expressed in both UVC response and slow-dry desiccation. Known anhydrobiosis genes are highlighted in orange.

**Supplementary Data S23: Enrichment analysis of DEGs.**

[https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.  
supplementary-material](https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.supplementary-material)

Enrichment analysis of KEGG pathway, Pfam-A and Gene ontology term of differentially expressed genes. Terms colored in red are terms suggested to be related to anhydrobiosis in previous studies.

**Supplementary Data S24: Details on identification of novel stress-responsive gene families.**

To validate the conservation within Tardigrada, we submitted gene sequences to a BLAST search against our in-house and publicly accessible genomes/transcriptomes. We first predicted approximately 10-90 thousand genes in various tardigrade lineages and then identified orthologs by BLASTP search or OrthoFinder clustering (Supplementary Data S27, S28, S29). Most of the lineages showed conservation of g2856 (excluding Halechiniscidae family in Arthrotardigrada), indicating that this gene family is highly conserved in the phylum (Supplementary Data S27, Supplementary Data S28). On the contrary, we found that g241 was lost in the Apochela, Echiniscoididae and Arthrotardigrada lineages (Supplementary Data S27, S29). Additionally, initial TBLASTX searches against the publicly accessible *Echiniscoides sigismundi* and *Richtersius coronifer* transcriptome assemblies indicated the loss of g2856 orthologs both species. We determined several raw RNA-Seq reads that showed homology against g2856 coding sequences, which implied the existence of g2856 orthologs. Therefore, we re-assembled both the *E. sigismundi* and *R. coronifer* transcriptome using previously sequenced RNA-Seq data with Bridger. BLASTX searches against this re-assembly found approximately fifteen g2856 orthologs but no g241 orthologs in *E. sigismundi*. On the other hand, we detected twenty three g241 and sixty four g2856 orthologs in *R. coronifer*. The lack of g241 or g2856 orthologs in previous transcriptome assemblies may have occurred during the filtering stage. Our in-house transcriptome sequencing data supported the existence of both gene families in *R. coronifer* (fifty six g2856 and seventeen g241 orthologs) and twelve g2856 orthologs in *E. sigismundi*. Together, these results support that the g241 is lost in the Echiniscoididae, Arthropoda, and Apochela lineages. Three of the g2856 orthologs (including g2856) were found to be mispredicted in our gene set (Supplementary Data S30).

**Supplementary Data S25: Expression of g241 orthologs.**

[https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.  
supplementary-material](https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.supplementary-material)

TPM values of g241 orthologs in *R. varieornatus* UVC response.

**Supplementary Data S26: Expression of g12777 orthologs.**

[https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.  
supplementary-material](https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.supplementary-material)

TPM values of g12777 orthologs in *R. varieornatus* UVC response.

**Supplementary Data S27: g12777 and g241 orthologs within Tardigrada identified by BLASTP.**

[https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.  
supplementary-material](https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.supplementary-material)

Coding sequences of g241.t1 and g12777.t1 were submitted to BLASTP search against predicted proteome sequences of indicated species in our in-house genome database and additional transcriptome assemblies.

**Supplementary Data S28: g12777 copy numbers within Tardigrada identified by OrthoFinder.**

[https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.  
supplementary-material](https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.supplementary-material)

Number of genes classified as g12777 orthologs by OrthoFinder.

**Supplementary Data S29: g241 copy numbers within Tardigrada identified by OrthoFinder.**

[https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.  
supplementary-material](https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.supplementary-material)

Number of genes classified as g241 orthologs by OrthoFinder.

**Supplementary Data S30: Refinement of g12777 ortholog gene regions.**

[https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.  
supplementary-material](https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.supplementary-material)

Mispredicted g12777 orthologs in *R. varieornatus* genome.

**Supplementary Data S31: Categorization of g241 orthologs.**

The g241 orthologs in *H. exemplaris* were identified as a high-confidence horizontally transferred gene candidate in our previous study (Supplementary Data S32). BLASTP searches against TrEMBL and NCBI nr databases indicated that the majority of homologs originates from Gammaproteobacteria species (Supplementary Data S33, S34). This suggests that this gene family may have been integrated into the tardigrade genome before the divergence of these Eutardigrada and Heterotardigrada. We also have found several orthologs in Ciliophora, Chlorophyta, and Acanthamoeba, however, the conservation patterns suggest that these genes may also be results of horizontal transfer into these organisms. Several of these bacterial orthologs were found to be fused to the C-terminal end of haem peroxidases. These regions do not have any functional domains and are not predicted to be a disordered region, suggesting that this gene itself may not have anti-oxidative stress.

**Supplementary Data S32: HGT statistics of g241 orthologs from previous studies.**

[https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.  
supplementary-material](https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.supplementary-material)

HGT statistics calculated in our previous study for g241 orthologs.

**Supplementary Data S33: g241.t1 orthologs identified from NCBI nr database.**

[https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.  
supplementary-material](https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.supplementary-material)

List of orthologs identified by Diamond BLASTP search (E-value < 1E-5) against NCBI nr database. Tardigrade, non-tardigrade eukaryotic, and bacterial hits were each colored in gray, yellow, and green. *H. exemplaris* orthologs colored in red had gene annotations, but validation by RNA-Seq data mapping indicated that these genes may be mis-predicted.

**Supplementary Data S34: g241.t1 orthologs identified from RefSeq database.**

[https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.  
supplementary-material](https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.supplementary-material)

List of orthologs identified by Diamond BLASTP search (E-value < 1E-5) against NCBI Bacteria RefSeq complete genome sequences. Genes before and after g241 orthologs are shown.

**Supplementary Data S35: Syntenic information of g2856 and g241 orthologs in *R. varieornatus*.**

*varieornatus*.

[https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.  
supplementary-material](https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.supplementary-material)

Annotation of five genes prior and after g2856 orthologs in *R. varieornatus*. g2856 and g241 orthologs are highlighted in red.

**Supplementary Data S36: Details on the crystal structure of g12777 protein.**

In the Zn<sup>2+</sup>-bound crystal structure, five Zn<sup>2+</sup> ions (Zn-1-5) originated from the crystallization buffer containing 300 mM zinc acetate were observed, while Ca<sup>2+</sup> originated from the protein buffer (2 mM calcium chloride) was not. We prepared Mn<sup>2+</sup>-bound g12777 crystals by soaking method using Zn<sup>2+</sup>-bound crystals. Upon soaking of excess amount of Mn<sup>2+</sup> (50 mM), Zn<sup>2+</sup> was replaced with Mn<sup>2+</sup> in Mn-1 and newly appeared in Mn-2 (Supplementary Figure S19a). On the other hand, Zn<sup>2+</sup> in Zn-2-5 disappeared in the Mn<sup>2+</sup>-bound crystal structure. As predicted from the bioinformatics analysis, the crystal structure of putative catalytic domain (residues Gly63–Leu231) of the g12777 protein displayed a characteristic  $\beta$ -sandwich fold comprising seven  $\beta$ -strands and three  $3_{10}$  helices with three disulphide bridges (C91–C97, C146–C225, and C176–C200) (Figure 5.2a, Supplementary Figure S19b,c). In the crystal structure of Mn<sup>2+</sup>-bound form, a part of  $\beta$ 5– $\beta$ 6 loop (residues 163–168) around the Mn<sup>2+</sup>-binding site was disordered, suggesting its flexible nature. Comparison of the structure of the  $\beta$ -sandwich domain of g12777 with known protein structures revealed that the g12777 protein structure has very weak similarities with structures of calcium-binding C2 domain involved in lipid interaction (Z-score = 5.1–5.5; RMSD = 2.5–3.2 Å; identity = 5–13%; PDB codes: 2JGZ, 1WFJ, and 4IHB). In addition, it has subtle similarity with a Cu/Zn-SOD monomeric circular permutant (Z-score = 4.1; RMSD = 3.4 Å; identity = 5%; PDB code: 5J0C).

**Supplementary Data S37: Metal ion affinity of recombinant proteins.**

[https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.  
supplementary-material](https://www.biorxiv.org/content/10.1101/2020.11.06.370643v1.supplementary-material)

Affinity values for metal ions with g12777 proteins.