Long non-coding RNA as environmental regulator in a non-model fish

Jingliang Kang[1], Arthur Chung[1], Sneha Suresh[1], Lucrezia L. Bonzi[2], Jade M. Sourisse[1], Sandra Ramirez[1], Daniele Romeo[1], Natalia Petit-Marty[1], Cinta Pegueroles[3], Celia Schunter[1*]

[1]Swire Institute of Marine Science, School of Biological Sciences, The University of Hong Kong, Pokfulam Road, Hong Kong SAR

[2]Red Sea Research Center, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

[3]Department of Genetics, Microbiology and Statistics, Institute for Research on Biodiversity (IRBio), University of Barcelona, Barcelona, Spain

*Correspondence: celiaschunter@gmail.com

ABSTRACT

The majority of the transcribed genome does not have coding potential but is composed of non-coding transcripts that are involved in transcriptional and post-transcriptional regulation of protein-coding genes. Regulation of gene expression is important in determining the response of organisms to changes in the environment, and therefore their persistence as population or species under global change. However, long non-coding RNAs (lncRNAs) are scarcely studied especially in non-model organisms due to the lack of a reliable pipeline for their accurate identification and annotation. Here, we present a pipeline which uses a combination of alignment-dependent and independent methods for the identification of conserved and species-specific lncRNAs from RNA-Seq data. Validation of this pipeline was performed using existing RNA-Seq data from *Acanthochromis polyacanthus* brain tissue, identifying a total of 4,728 lncRNAs across the whole genome, the majority of which (3,272) are intergenic. To investigate the possible implications of these lncRNAs, we estimated the expression changes of lncRNAs and neighbouring coding genes in response to ocean acidification. We found that intergenic lncRNAs might adjust the expression of neighbouring coding genes related to pH regulation, neural signal transduction and ion transport, which are known to be important in the response to ocean acidification in fish. Overall, this pipeline enables the use of existing RNA sequencing technology to reveal additional underlying molecular mechanisms involved in the response to environmental changes in non-model species by integrating the study of lncRNAs with gene expression.

Keywords: epigenetic regulation, annotation, RNA sequencing, non-model organism, environmental change, ocean acidification

INTRODUCTION

Among pervasive genomic regions that can be transcribed, some encode long non-coding RNAs (lncRNAs), defined as RNAs longer than 200 nucleotides that are not translated into functional proteins (Statello et al., 2021). LncRNAs are a highly heterogeneous class of transcripts mainly transcribed by RNA polymerase II and, usually, inefficiently spliced. Some remain in the nucleus while others are polyadenylated and exported to the cytoplasm (Statello et al., 2021). Although initially considered to be products of transcriptional noise or spurious transcription (Struhl, 2007), lncRNAs are now known to play critical roles in diverse biological processes, including DNA repair, proliferation, and embryonic development (Fernandes et al., 2019; Li et al., 2019; Vance & Ponting, 2014). LncRNAs are involved at many levels, including chromatin modifications, pre-transcription, transcription, and post-transcription through regulation of the associated gene expression (Gardini & Shiekhattar, 2015; Kornfeld & Brüning, 2014; Necsulea et al., 2014; Ulitsky, 2016; H. Xu et al., 2019; Zhu et al., 2013). In addition, lncRNAs are involved in specific physiological processes and gene dysfunction and can be used as biomarkers of certain diseases (Beck et al., 2018; Fernandes et al., 2019; Jiang et al., 2016).

LncRNAs are not as conserved as protein-coding sequences due to their fast turn-over (Lopez-Ezquerra et al., 2017; Pegueroles et al., 2019), and they tend to be cell and tissue specific (Cabili et al., 2015; Derrien et al., 2012). In addition, many lncRNAs are lowly expressed (Guo et al., 2020; Hezroni et al., 2015; Necsulea et al., 2014; Quinn et al., 2016). As such, despite the growing interest in lncRNAs and the availability of RNA sequencing data in many species, lncRNAs remain unexplored in most species, particularly in non-model species. To date, several bioinformatic tools have been developed to detect lncRNAs by computing coding potential score (CPS) of transcripts, either by using sequence alignments (alignment-dependent) or detecting intrinsic features of the input RNA sequences (alignment-free). In the alignment-dependent methods, such as PhyloCSF (Lin et al., 2011) and CPC (Kong et al., 2007), sequences are either aligned between species or to protein databases, which may be biased toward misclassifying species-specific or lowly conserved coding and non-coding transcripts. In contrast, although the alignment-free methods like CPAT (Wang et al., 2013) and FEELnc (Wucher et al., 2017) are useful to discriminate species-specific lncRNA, these tools use different intrinsic features (such as the length and integrity of the longest open reading frame) of the input RNA sequences, which can result in differences in detecting lncRNAs. Nevertheless, most studies related to lncRNA identification used only one method (Boltaña et

al., 2016; Dettleff et al., 2020; Mu et al., 2016; Paneru et al., 2018; D. Quan et al., 2020; Ren et al., 2020; J. Xu et al., 2019). Hence, a pipeline combining both alignment-dependent and alignment-free strategies is essential to detect conserved and species-specific lncRNAs.

Environmental perturbations produce molecular responses needed to maintain cellular homeostasis and function, and lncRNAs may play a role in some of these responses by regulating gene expression (Paneru et al., 2018; Sarangdhar et al., 2018) through cis- and trans-regulation (Gil & Ulitsky, 2020; Luo et al., 2019; Nadal-Ribelles et al., 2014; D. Quan et al., 2020). A large proportion of lncRNAs with functional implications are in fact cis-regulatory and hence affect the regulation of the nearby/neighbouring protein-coding genes allowing for plasticity in gene networks across time (Engreitz et al., 2016; Gil & Ulitsky, 2020). In the model-plant genus *Arabidopsis*, for instance, a lncRNA tightly controls the expression of highly conserved transcription factors that promote cold tolerance in many plant species (Kindgren et al., 2018). In vertebrates, such as fish, lncRNAs identified in rainbow trout (*Oncorhynchus mykiss*) can potentially mediate regulation of the heat stress response (J. Quan et al., 2020) and one lncRNA affects the mucosal immunity of turbot (*Scophthalmus maximus*) in response to bacterial infection (N. Yang et al., 2020). However, more studies are needed to understand the importance that lncRNAs play in the responses to changing environments and their potential involvement in adaptive plasticity to environmental change also in non-model organisms.

To this end, we propose a comprehensive roadmap and guideline for the identification of lncRNAs in model as well as non-model organisms and use a case study to evaluate the involvement of lncRNAs and related coding genes in the response to an environmental change. For this, we use the spiny chromis *Acanthochromis polyacanthus*, a common coral reef fish in the Western Pacific, extensively studied for fish acclimation processes to environmental changes in the last decade. Generous RNA sequencing has been performed on this species (Bernal et al., 2022; Kang et al., 2022; Schunter et al., 2021, 2016) to interpret its transcriptional responses to ocean environmental changes. This extensive knowledge on the transcriptional processes to ocean acidification in *A. polyacanthus* allows for a broad evaluation of the involvement of lncRNAs and provides a guideline for future studies also in other systems. As lncRNAs can tune neighbouring coding gene expression by directly affecting nuclear architecture or indirectly affecting their transcription or translation (Ransohoff et al., 2018), we aim to combine the study of lncRNAs with coding gene expression to evaluate the

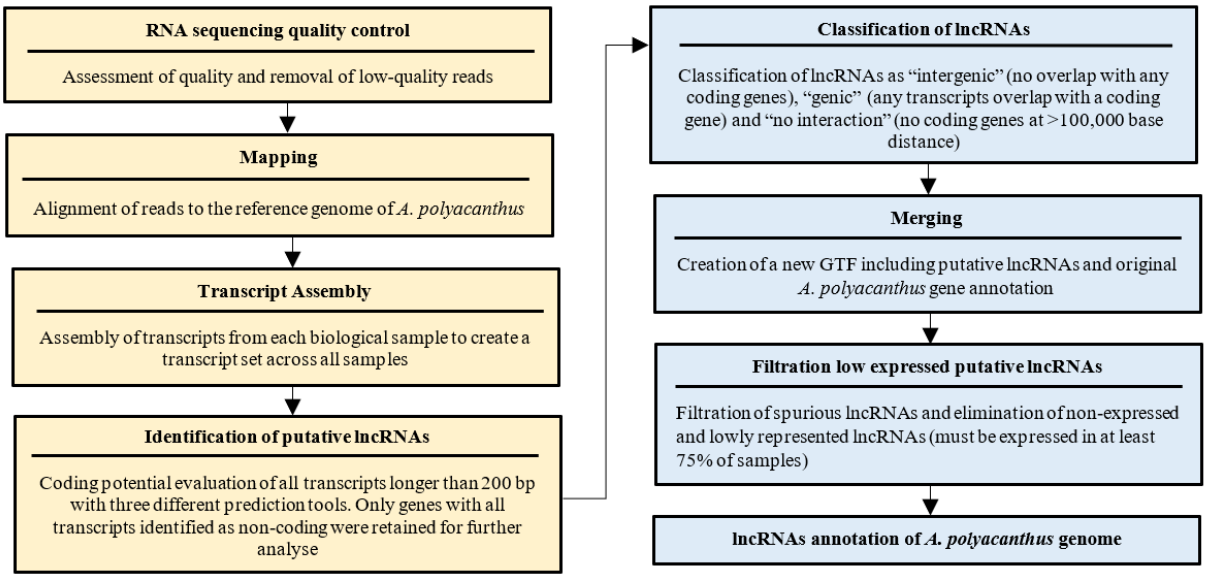
112 transcriptional and epigenetic factors that define the biological response of organisms to
113 environmental changes, providing information on the adaptive potential of the species.
114

115 METHODS

116 *Discovery and Annotation of lncRNAs*

117 To identify lncRNAs in the *Acanthochromis polyacanthus* genome, 226 RNA-seq samples
118 were compiled from published studies investigating the effect of elevated CO_2 on the brain
119 transcriptome of *A. polyacanthus* (Kang et al., 2022; Schunter et al., 2018, 2016;
120 PRJNA658203). This allows for a comprehensive annotation across the genome of lncRNAs
121 associated with the exposure to elevated CO_2 in laboratory and natural settings. We processed
122 all samples to identify a reliable set of lncRNAs across the genome (Figure 1 & S1). In our
123 pipeline, the raw reads quality was first assessed with FASTQCv0.11.9 (Andrews, 2010), and
124 then adapter sequences and low-quality sections of reads were removed by Trimmomatic v0.39
125 (Bolger et al., 2014) using the following parameters: "ILLUMINACLIP:2:30:10 LEADING:4
126 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:40". The resulting clean reads were aligned
127 to the reference genome of *A. polyacanthus* (NCBI database, accession number:
128 ASM210954v1) with HISAT2 v2.1.0 (Kim et al., 2019) with "*--known-splicesite-infile*" to
129 provide known splice sites.

130



131

132 *Figure 1: Flowchart illustrating the process for long non-coding RNA (lncRNA) identification*
133 *in* Acanthochromis polyacanthus *genome. In yellow are the steps for the initial identification*
134 *of putative lncRNAs, while in blue are the post-processing analyses for the creation of the final*
135 *high confidence lncRNA set.*

136

137 StringTie v2.1.5 (M. Pertea et al., 2015) was then applied to assemble transcripts. All resulting
138 transcripts among the 226 fish samples were merged for a unified set of non-redundant
139 transcripts (M. Pertea et al., 2016), which was then used for transcript abundance estimation
140 and lncRNAs detection. Transcript abundances were then re-estimated based on the merged
141 transcript assembly.
142
143 To estimate the coding potential of all transcripts, we extracted the transcript sequences from
144 the *A. polyacanthus* genome sequence using GffRead v0.12.7 (Pertea & Pertea, 2020) with the
145 parameters "*-w, -g*". Three tools, including one alignment-dependent (CPC v0.9-r2, Kong et
146 al., 2007) and two alignment-free methods (CPAT v1.2.4, Wang et al., 2013; FEELnc v0.2,
147 Wucher et al., 2017), were used to calculate the coding potential of all transcripts. CPC v0.9-
148 r2 (Kong et al., 2007) was used with default parameters to estimate coding potential based on
149 sequence similarity between transcripts and Uniref90 protein database, and the transcripts with
150 a coding potential score < 0 were considered as putative lncRNAs. For the two alignment-free
151 tools, CPAT v1.2.4 (Wang et al., 2013) was run with default parameters using the coding
152 mRNA sequences of zebrafish as training model, and transcripts with a coding probability <
153 0.38 were considered as putative lncRNAs, which is the cut-off value suggested by CPAT
154 (Wang et al., 2013). The second alignment-free tool we used was FEELnc v0.2 (Wucher et al.,
155 2017), which annotates lncRNAs based on a Random Forest model trained with general
156 features such as multi k-mer frequencies and relaxed open reading frames. We ran the program
157 with FEELnc_filter.pl to filter out spurious coding transcripts using "-b
158 transcript_biotype=protein_coding, --monoex=-1", and then FEELnc_codpot.pl with default
159 parameters to compute the coding potential score for each transcript by setting the mode of the
160 lncRNA simulation to intergenic. The putative lncRNAs were then identified based on the best
161 coding potential cutoff of receiver operating characteristic curve plot (Sing et al., 2005). For
162 all three methods, only genes with all transcripts identified as non-coding were retained for
163 further analyses.
164
165 For a more conservative approach in avoiding spurious lncRNA annotations we only retained
166 the transcripts that were identified as lncRNAs by at least two tools. These putative lncRNAs
167 were categorized using default parameters of FEELnc_classifier.pl (Wucher et al., 2017) with
168 respect to the location and the direction of transcription of proximal coding transcripts. A
169 lncRNA is classified as: a. "intergenic" (i.e., lincRNA) if all transcripts of this lncRNA have
170 no location overlap with any neighbouring coding genes; b. "genic" if any of its transcripts

171 overlaps with a coding gene; c. "no interaction" if no coding genes are detected within 100,000
172 base pairs on either side of the lncRNA. Due to the fact that intergenic lncRNA (lincRNA)
173 gene expression patterns, sequence conservation and perturbation outcomes are easier to
174 interpret than those of transcripts from other lncRNAs overlapping with coding genes (Ulitsky
175 & Bartel, 2013) only lncRNA genes classified as intergenic were used in further analyses to
176 investigate their potential involvement in the response to ocean acidification in fish.
177
178 The expression levels of the putative lincRNAs and mRNAs were quantified using
179 FeatureCounts v2.0.0 (Liao et al., 2014) allowing for multi-mapped reads to be counted
180 fractionally. Transcripts with reads number > 0 in at least 75% of the samples were retained to
181 remove potential spurious transcripts. The read numbers of remaining transcripts were
182 normalized using DESeq2 v1.32.0 (Love et al., 2014) and lincRNAs with a normalised
183 expression < 1 in 10% samples were removed to obtain the final list of candidate lincRNA
184 genes. This provided us with a final set of high confidence lincRNAs in both location and
185 expression. To investigate possible functional implications of highly expressed lincRNAs,
186 functional enrichment (Fisher's Exact Test with False Discovery Rate (FDR) adjusted *p*-value
187 < 0.05; Benjamini & Hochberg, 1995) was then performed in OmicsBox v2.0.36 (BioBam
188 Bioinformatics, 2019) for the neighbouring coding genes of highly expressed lincRNAs (> 500
189 normalized reads) using as a reference dataset all the annotated genes in *A. polyacanthus*
190 genome that had at least one read across all our samples (Benjamini & Hochberg, 1995).
191
192 *Case study: Expression patterns of lincRNAs in fish living in CO_2 seeps*
193 Using RNA sequencing data from wild *A. polyacanthus* samples collected in Papua New
194 Guinea (Kang et al., 2022; NCBI Bioproject PRJNA691990), we performed a case study to
195 evaluate the expression changes of lincRNAs in different environmental conditions. Seven
196 brain samples were collected from fish from a coral reef situated in a naturally bubbling CO_2
197 seep with elevated CO_2 levels close to the predicted levels for the end of this century due to
198 ocean acidification (pH= 7.77, pCO_2= 843 µatm; IPCC, 2022). Further eleven fish were
199 sampled from an adjacent control reef approximately 500 m away from the CO_2 seep (pH=
200 8.01, pCO_2= 443 µatm; Kang et al., 2022). There was no significant difference in temperature
201 and salinity between the CO_2 seep and the control site (Fabricius et al., 2011) allowing for the
202 evaluation of effects of long-term ocean acidification conditions on fish.
203

204 To evaluate differential gene expression of lincRNAs and coding genes between fish living at
205 different CO_2 levels, we performed differential gene expression analysis using DESeq2 v1.34
206 (Love et al., 2014) in R v.3.5.1. Between samples from control and CO_2 seep, lincRNAs and
207 coding genes were considered as differentially expressed (DE) with an FDR adjusted p-value
208 ≤ 0.05, and the average of the normalized count values (basemean) ≥ 10 as well as
209 Log2FoldChange ≥ 0.3. A principal component analysis (PCA) was performed using the log
210 2-fold normalized expression of the samples from the two different CO_2 level sites. To evaluate
211 potential functional implications of the DE lincRNAs, their neighbouring coding genes were
212 used for functional enrichment in OmicsBox v2.0.36 as previously described (BioBam
213 Bioinformatics, 2019).
214
215 To identify lincRNAs and coding gene co-expression modules with significant correlation with
216 CO_2 levels, a weighted gene co-expression network analysis (WGCNA; Langfelder & Horvath,
217 2008) was also applied on our fish samples. The co-expression similarity of gene modules was
218 obtained by selecting a *signed* network adjacency type with a soft thresholding power of 10
219 calculated according to the scale-free topology criteria. We established the expression
220 similarity between nodes of genes that are co-expressed by calculating the Topological Overlap
221 Measure (TOM). These co-expression networks were grouped into different coloured modules
222 based on their eigengenes values. Pearson correlation was then calculated to evaluate the
223 correlation between gene modules and samples from control and CO_2 seep. Genes of
224 significantly correlated gene module ($p < 0.01$) were analysed for functional enrichment in
225 OmicsBox v2.0.36 (BioBam Bioinformatics, 2019), with the Fisher's Exact Test (FDR < 0.05).
226 The enrichments were performed on the neighbouring coding genes of all the lincRNAs in the
227 module (the coding genes were not necessarily found within the module), as well as on the
228 neighbouring coding genes exclusively found within the same module of their lincRNAs.
229

RESULTS

LncRNA annotation pipeline

Brain *Acanthochromis polyacanthus* RNA sequencing samples had on average 32,7 million high-quality paired-end reads (Table S1) and on average, 87.5% of these reads mapped to the reference genome (Table S2). The assembly steps resulted in a total of 116,237 transcripts across the genome. We used different algorithms to predict the coding potential score (CPS) of these transcripts. This resulted in 29,127 non-coding transcripts with CPAT, 28,865 with CPC and 11,897 with FEELnc, belonging to 17,434, 7,970 and 3,633 putative long non-coding RNA genes (lncRNAs) respectively. Of these, 9,209 putative lncRNAs were identified by at least two programs (Figure 2A). After filtering out lowly expressed putative lncRNAs, we obtained a final set of 4,728 lncRNAs, consisting of 1,313 genic, 3,272 intergenic and 143 non-neighbouring lncRNAs.

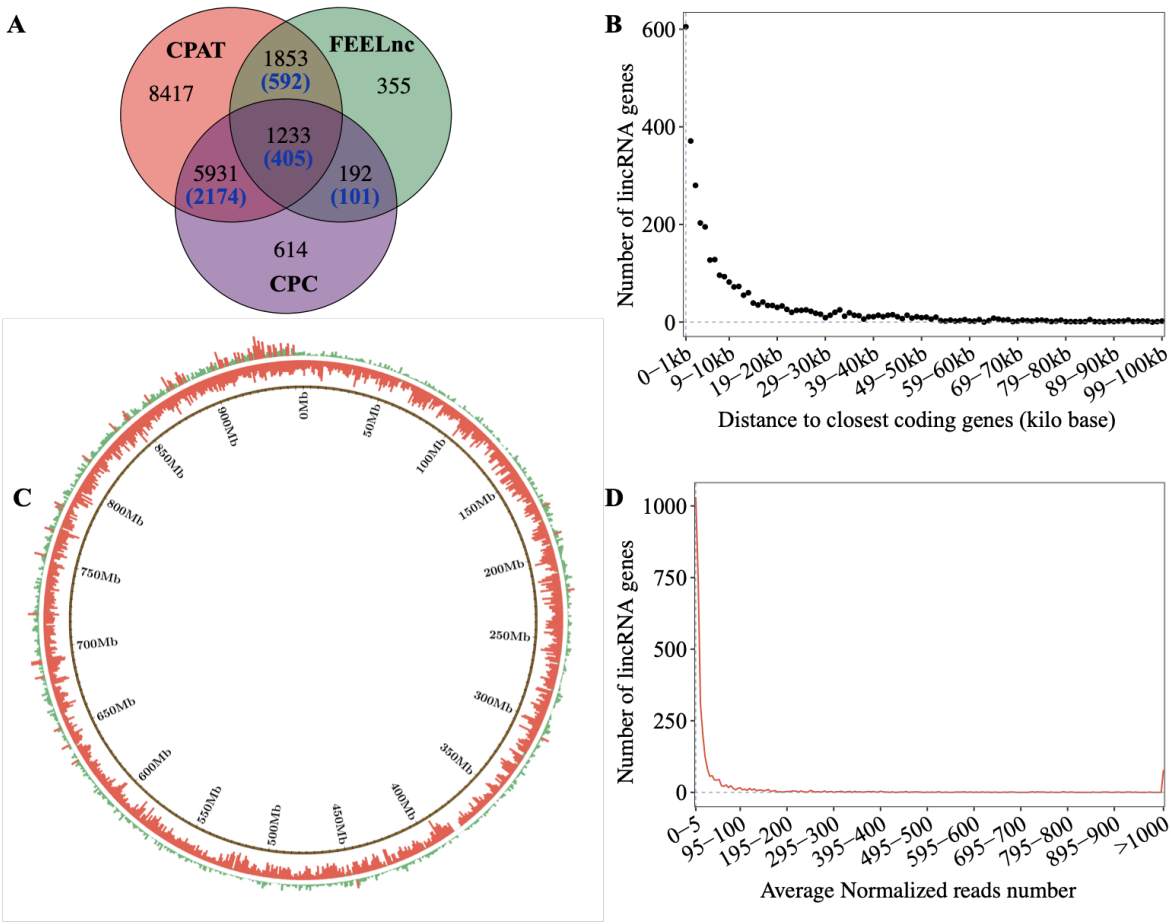


*Figure 2: Intergenic lncRNA genes (lincRNAs) along the Acanthochromis polyacanthus genome. **A**. Venn diagram of putative lncRNAs identified by CPAT, CPC and FEELnc*

247 *programs. Blue numbers in brackets indicate overlapping lincRNAs among the total lncRNAs.*
248 ***B**. Distance ranges between the identified 3,272 lincRNAs and their neighbouring coding genes.*
249 ***C**. Distribution of lincRNAs (first outer circle) and coding genes (second outer circle) along*
250 *the whole genome (concatenating all scaffolds together from longest to shortest). Genome*
251 *sliding windows are indicated in the inner black circle. Bar height and color are indicative of*
252 *the number of genes found in each 1Mb sliding window, with red and green bars indicating*
253 *more and less than 10 lincRNAs or coding genes per window, respectively.* ***D.*** *LincRNAs levels*
254 *of expression across different expression level ranges.*
255
256 The intergenic lncRNAs (lincRNAs; Table S3) were mostly located within a short distance to
257 their neighbouring coding genes (Figure 2B) with 67% located within 0-1kb from their
258 neighbouring coding gene. Despite lincRNAs being dispersed throughout the whole genome
259 (Figure 2C), the lincRNAs density of short scaffolds (length < 1Mb) was significantly higher
260 compared to long scaffolds. Out of 614 scaffolds with at least one lincRNA, the lincRNAs
261 density in 392 short scaffolds was significantly higher (two-sample Wilcoxon test, $p = 0.00026$)
262 compared to the 222 long scaffolds (Figure S2). The overall lincRNAs density was
263 significantly lower than in protein coding genes. Among 955 genomic sliding windows of 1Mb
264 in the *A. polyacanthus* genome, 829 regions included more than one lincRNA and 53 regions
265 included more than ten lincRNAs. Protein coding genes, by contrast, displayed a higher density
266 (934 regions with more than 10 coding genes). In addition, most lincRNAs exhibited
267 expression with less than 500 normalized reads, and only 125 lincRNAs had elevated
268 expression levels (average normalized reads > 500; Figure 2D; Table S4). The neighbouring
269 coding genes of the 125 most expressed lincRNAs were involved in signal transduction,
270 GABA-A receptor, circadian rhythm, and ion transport (Figure 3A; Table S4). The highest
271 expressed lincRNA (MSTRG.9528; expression > 190,000 normalized reads) neighbours the
272 coding gene peripherin-2 (PRPH2).
273

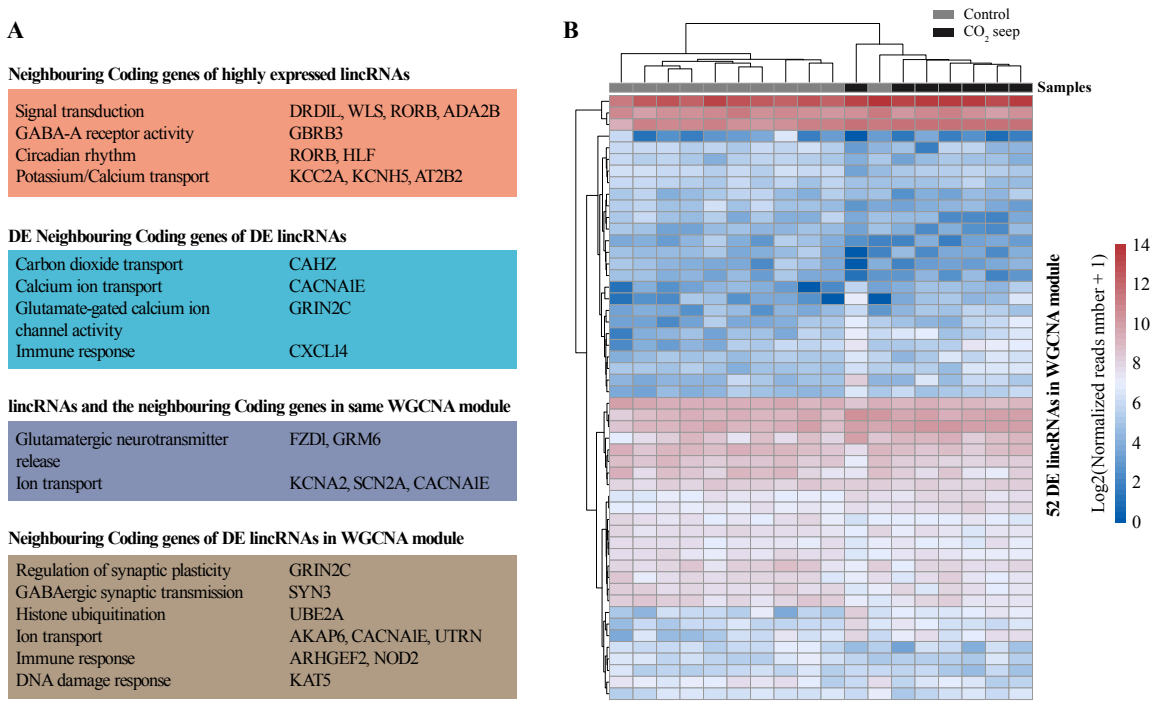


Figure 3: Functions of neighbouring coding genes of lincRNAs (A) and gene expression pattern of 52 differential expressed lincRNAs which were also related to pH level in module turquoise by WGCNA analysis (B).

<u>Case study: response to ocean acidification in fish</u>

By applying our pipeline to already existing transcriptomic data from wild *A. polyacanthus* collected from a CO_2 seep and a control reef site in Papua New Guinea, we were able to evaluate the involvement of genomic regulations by lincRNAs in fish brain in response to an environmental factor. Fish from the CO_2 seep displayed a distinct brain expression pattern, both when the analysis was run using all genes (coding genes and lincRNAs; Figure 4A) as well as only lincRNAs (Figure 4B).

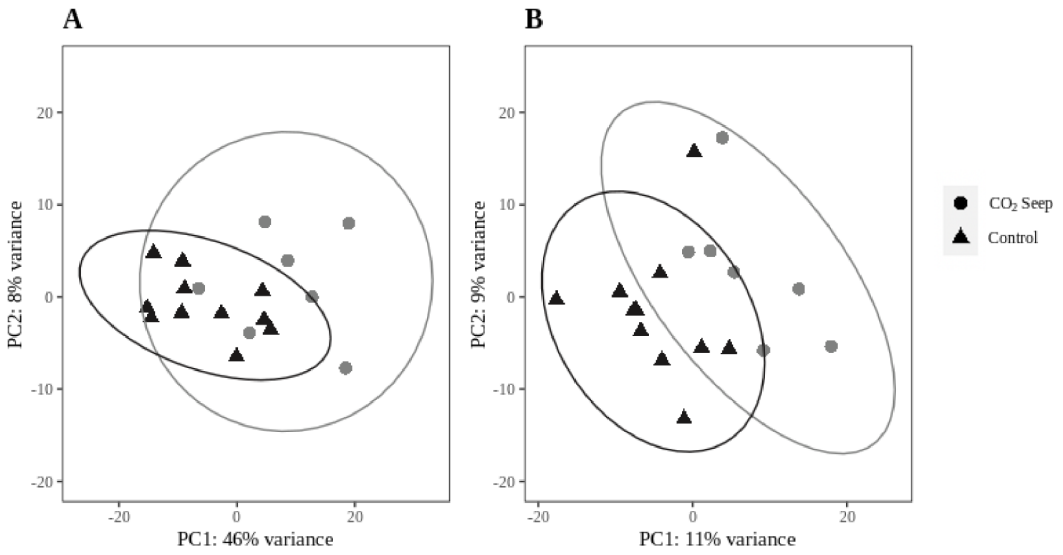


Figure 4: Principal component analyses (PCAs) of log2-fold normalized expression values of (A) lincRNA and coding genes or (B) lincRNAs only in the brain of wild collected Acanthochromis polyacanthus from a CO_2 seep (blue-green) and an adjacent control reef (red). Ellipse areas represent a 95% confidence level.

Between samples from CO_2 seep and control site, 3,431 coding genes (Figure S3A) and 97 lincRNAs (Table 1; Table S5) were significantly differentially expressed (DE). The DE coding genes with neighbouring lincRNAs had significantly higher expression changes (Pearson's Chi-squared test, p = 3.425e-05). 3.5% (7/198) of the DE coding genes with neighbouring lincRNA exhibit high expression difference (|log2FoldChange| > 2), whereas only 0.7% (23/3233) of the DE coding genes without neighbouring lincRNA display elevated expression (Figure S3B). Among the neighbouring coding genes of the 97 DE lincRNAs, 13 coding genes also displayed significant differential expression (Table S6). These DE neighbouring coding genes are related to carbon dioxide transport and hypotonic salinity response (CAHZ: carbonic anhydrase), calcium ion transport (CACNA1E: calcium voltage-gated channel subunit alpha1 E), glutamate-gated calcium ion channel activity (GRIN2C: glutamate ionotropic receptor NMDA type subunit 2C) or immune system response (CXCL14: C-X-C motif chemokine 14; Figure 3A).

Three weighted correlated gene network modules were found to be significantly correlated with environmental CO_2 levels ("turquoise", p = 0.001, Table S7; "sky-blue", p= 0.009, Table S8; "dark-green", p= 0.006, Table S9). Turquoise and sky-blue modules were negatively correlated with CO_2 levels and included 329 lincRNAs and 6,259 coding genes (Table 1; Table

310 S7), and 26 lincRNAs and 136 coding genes (Table 1; Table S8), respectively. Dark green
311 module was conversely positively correlated with CO_2 levels and consisted of 6 lincRNAs and
312 231 coding genes (Table 1; Table S9). Within the turquoise module only we found 68
313 lincRNAs whose neighbouring coding genes were also included in the same co-expression
314 module (Tables S7 & S10). The functional implications of these neighbouring coding genes
315 show involvement in, for instance, glutamatergic neurotransmitter release (FZD1: frizzled
316 class receptor 1; GRM6: glutamate metabotropic receptor 6) and ion transport (KCNA2:
317 potassium voltage-gated channel subfamily A member 2; SCN2A: sodium voltage-gated
318 channel alpha subunit 2; CACNA1E: calcium voltage-gated channel subunit alpha1 E; Figure
319 3A).
320
321 52 lincRNAs in the turquoise module were also significantly differentially expressed showing
322 an overlap between the two gene expression analyses (Figure 3; Tables S7 & S11). The
323 neighbouring coding genes of these lincRNAs were related to the regulation of synaptic
324 plasticity (GRIN2C: glutamate receptor ionotropic NMDA 2C), modulation on GABAergic
325 synaptic transmission (SYN3: synapsin-3), histone ubiquitination (UBE2A: ubiquitin-
326 conjugating enzyme E2 A), regulation of ion transport (AKAP6: A-kinase anchor protein 6;
327 CACNA1E: calcium voltage-gated channel subunit alpha1 E; UTRN: utrophin), innate
328 immune response (ARHGEF2: rho guanine nucleotide exchange factor 2; NOD2: nucleotide-
329 binding oligomerization domain-containing protein 2) and DNA damage response (KAT5:
330 histone acetyltransferase KAT5; Figure 3A). Of the other two significantly correlated modules,
331 sky-blue and dark green, only two and one lincRNAs, respectively, were also found to be
332 differentially expressed (Table S8 & S9).
333

334

335 *Table 1. Differentially expressed lincRNAs and coding genes identified through differential*
336 *expression analysis (DESeq2) and weighted gene co-expression network analysis (WGCNA).*
337 *In the "lincRNAs with neighbouring coding gene" row are the numbers of lincRNAs for which*
338 *the neighbouring coding genes were also differentially expressed or found in the same co-*
339 *expression module.*

Gene type	**Differential Expression**	**Weighted co-expression network**		
		turquoise	sky-blue	dark-green
lincRNAs	97	329	26	6
coding genes	3431	6,259	136	231
lincRNAs with neighbouring coding gene	13	65	0	0

340

341 # DISCUSSION

342

343 *lncRNAs discovery pipeline for non-model organisms*

344

345 The boost in RNA sequencing studies including a large variety of non-model organisms
346 provides an excellent resource that could be (re-)used to obtain a more complete picture of
347 lncRNAs regulations on coding genes. Here we present a solid pipeline that can make use of
348 already existing data to detect conserved as well as species-specific lncRNAs. This pipeline
349 can be easily applied to any non-model species with an available reference genome. Most
350 studies related to lncRNA identification rely on only one method of coding potential detection
351 (Boltaña et al., 2016; Dettleff et al., 2020; Mu et al., 2016; Paneru et al., 2018; J. Quan et al.,
352 2020; Ren et al., 2020; H. Xu et al., 2019). In contrast, our results based on an alignment-
353 dependent (CPC) and two alignment-free methods (CPAT, FEELnc), indicate that 92.0%, 51.7%
354 and 90.2% lncRNAs detected by CPC, CPAT and FEELnc respectively were called as
355 lncRNAs by at least two tools. Different tools revealed to have varying performances
356 (Schneider et al., 2017; Wucher et al., 2017) and the three tools used in our study perform well
357 in predicting lncRNAs also in other species including both model and non-model organisms
358 (Duan et al., 2021). Thus, lncRNAs predicted using our pipeline could be reliable resources for
359 future studies to investigate lncRNAs functions.

360

361 We identified a total of 9,209 putative lncRNAs, of which 49% were expressed in at least 75%
362 of our *A. polyacanthus* brain samples. While the numbers found in mammal species are higher
363 (6,010 - 96,411 from the rhesus macaque to human; NONCODE v5.0, Fang et al., 2018) it has
364 to be noted that in our case we only investigate one tissue, and the inclusion of more tissues is
365 likely to increase the number of lncRNAs. Nonetheless, teleost model species show similar
366 numbers of expressed lncRNAs. Zebrafish lncRNAs identified so far are 3,503 (Zhao et al.,
367 2021) and between three- and six thousand lncRNAs have been found to be expressed across
368 different tissues also in rainbow trout (Al-Tobasei et al., 2016), coho (Leiva et al., 2020) and
369 Atlantic salmon (Boltaña et al., 2016). Other studies found higher numbers of lncRNAs in non-
370 model teleost species, such as 31,984 in the brain of *Larimichthys crocea* (Liu et al., 2018)
371 using CPC or 14,614 across several tissues of *Genypterus chilensis* (Dettleff et al., 2020) using
372 CPAT. However, these studies only used one method to evaluate coding potential and did not

373 ensure that all transcripts of a lncRNA were non-coding, which are less stringent criteria
374 compared to what was applied here.
375
376 Overall, the majority of lncRNAs we identified throughout the genome of *A. polyacanthus* had
377 relatively low expression. This observation is in accordance with many previous studies that
378 report a typical low expression of lncRNAs across species and tissues (Guo et al., 2020;
379 Hezroni et al., 2015; Necsulea et al., 2014; Quinn et al., 2016), as lncRNAs are constantly
380 submitted to various regulation mechanisms and they are typically short-lived (Necsulea et al.,
381 2014; Wu et al., 2014). Despite low expression levels, these lncRNAs can be functional since
382 some regulatory mechanisms do not require high concentration of effector molecules (Aprea
383 & Calegari, 2015). However, some lncRNAs can also be expressed at elevated levels and
384 highly expressed lincRNAs, for instance, are associated with diseases (Wan et al., 2013; F.
385 Yang et al., 2011). In the brains of our coral reef fish, we also found some highly expressed
386 lincRNAs with neighbouring coding genes involved in fundamental functions such as a
387 $GABA_A$ receptor, circadian rhythm genes and genes related to signal transduction (such as
388 DRD1L and ADA2B), important for normal brain function and neuronal activity (Bhat et al.,
389 2010; Logan & McClung, 2019). In humans, rodents and zebrafish, a lncRNA promotes the
390 expression of homeobox transcription factors required for the development of GABAergic
391 neurons Y-Aminobutyric acid (or GABA) which is the main inhibitory neurotransmitter in the
392 vertebrate brain (Feng et al., 2006). Furthermore, lncRNAs have been shown to regulate core
393 circadian rhythm genes in mammals (Mosig & Kojima, 2021). With a substantial fraction of
394 lncRNAs affecting the gene expression of their neighbouring coding genes (Engreitz et al.,
395 2016), it is no surprise to see elevated expression in lncRNAs neighbouring $GABA_A$ receptor
396 and core circadian rhythm genes as these are known to play important roles in the
397 transcriptional response to ocean acidification (Kang et al., 2022; Schunter et al., 2021, 2016).
398 As our RNAseq data is from ocean acidification experiments, this now hints at a regulatory
399 involvement of lncRNAs in these key functions. Hence, our pipeline of lncRNA annotation
400 and expression analysis recovers conserved lncRNAs found to be involved in essential
401 functions also in other species, but it also allows for the discovery of new lncRNAs involved
402 in the response to an environmental change in a wild coral reef fish.
403
404 *Functional responses of lncRNAs to environmental change*
405

406 Our pipeline facilitates the investigation of lncRNAs as a potential regulatory mechanism in
407 response to environmental change and opens doors to fields like environmental science and
408 ecology. In our case study, we made use of a published RNA sequencing set of *Acanthochromis*
409 *polyacanthus* brains from a volcanic CO_2 seep, representing future ocean acidification
410 conditions, and a control site. We detected differentially expressed lincRNAs and mRNAs in
411 the brain of *A. polyacanthus* as a response to natural environmental differences (CO_2 levels).
412 Interestingly, for the coding genes that have a neighbouring lincRNA in close proximity, we
413 found larger expression differences between fish from CO_2 seeps and control sites suggesting
414 a regulation of gene expression by neighbouring lncRNAs with the exposure to elevated CO_2.
415 For the differentially expressed lincRNAs, the set of neighbouring protein coding genes that
416 were also differentially expressed are involved in functions related to pH regulation in fish.
417 When fish respond to ocean acidification, carbonic anhydrase (CAHZ) is often upregulated, as
418 in our data here, to catalyze the hydration of CO_2 and play an essential role in acid-base and
419 ion regulatory functions (Heuer & Grosell, 2014). A further reduction of intracellular pH in
420 fish is necessary to prevent acidosis when faced with elevated CO_2 levels in fish (Schmidt,
421 2019), and this process performed by coding gene GRIN2C may be regulated by a differentially
422 expressed lincRNA.
423
424 Further biological processes are known to be involved in the response to ocean acidification in
425 fish brains. One of them is synaptic transmission and the down regulation of a lincRNA and
426 the neighbouring coding CACNA1E (gene voltage-gated calcium channel complex) can
427 contribute to synaptic transmission (Berecki et al., 2014), which was also found to behave
428 similarly to a previous RNAseq *A. polyacanthus* data (Schunter et al., 2018). In addition,
429 immune responses are commonly associated with ocean acidification in fish including our
430 study species (De Souza et al., 2014; Kang et al., 2022; Machado et al., 2020), and here we
431 find differential expression of lincRNAs neighbouring a variety of immune response genes that
432 show co-expression in a gene network significantly correlated with CO_2 levels. This reveals
433 that lincRNAs annotated in our coral reef fish show differential expression in response to an
434 environmental factor, CO_2 levels, and that some of these differentially expressed lincRNAs are
435 in close proximity to coding genes that are known to have functional implications in the
436 response to ocean acidification in fishes.
437
438 We identified candidate lincRNAs to be involved in acclimation to elevated CO_2 levels and
439 future genetic manipulation experiments would allow to verify the direct effect of a lincRNA

440 on the nearby coding gene (Akay et al., 2019; Rodriguez-Lopez et al., 2022). CRISPR/Cas9
441 based genome editing is a promising technique to experimentally evaluate the role of
442 lincRNAs, although it is not readily available for most non-model organisms. Our study is a
443 start to understand regulatory elements potentially involved in the regulation of the gene
444 expression patterns observed with an environmental change. This case study shows a pipeline
445 that can be applied by using RNA sequencing reads, now frequently incorporated into
446 experimental studies, and adds an epigenetic mechanistic aspect with no additional sequencing
447 effort needed. We demonstrate that lncRNAs can be annotated and evaluated in a non-model
448 organism and encourage more molecular ecologists to discover lncRNAs in their study species
449 and investigate the mechanistic role of lncRNA in response to environmental change.
450

ACKNOWLEDGMENTS

We would like to thank Timothy Ravasi and Philip L. Munday for for helping with the collection and sequencing of the large RNA sequencing datasets produced in previous studies. We are grateful to Helen Leung, Ho Wu Cheuck and Kam Yan Chit for their support with this project.

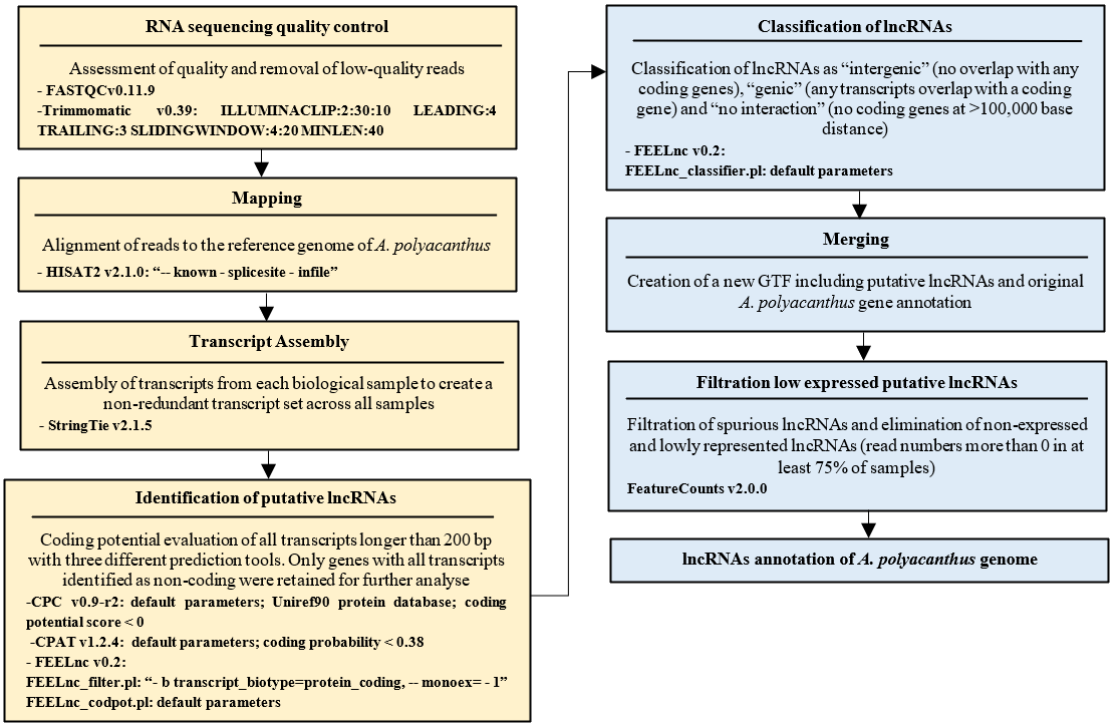
DATA AVAILABILITY

The RNA sequencing raw data used in this study is found in the Bioprojects: PRJNA691990; PRJNA311159; PRJNA658203 (Reviewer link as the last is not published yet: https://dataview.ncbi.nlm.nih.gov/object/PRJNA658203?reviewer=47r5c4kubkjn124c770t5vvt60)

The lncRNA annotation of *Acanthochromis polyacanthus* is available here: 10.6084/m9.figshare.20045780. Reviewer link: https://figshare.com/s/33802daaa28dc7ccd876

466 SUPPLEMENTARY FIGURES

467



468

469 Figure S1: Detailed flowchart illustrating the process for long non-coding RNA (lncRNA)

470 identification including the programs and parameters used.

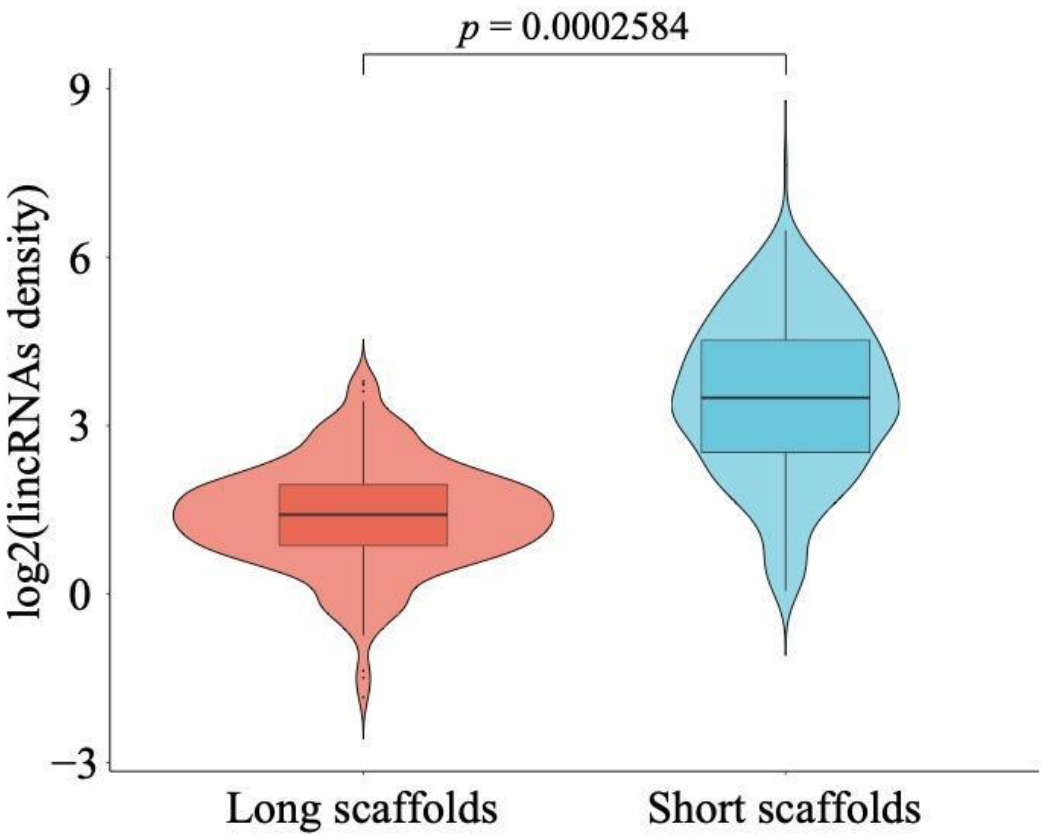


Figure S2: Density of lincRNAs (per million bases) in *Acanthochromis polyacanthus* genome scaffolds with at least one lincRNAs. The density of lincRNAs in 392 short scaffolds (length < 1Mb) was significantly higher (two-sample Wilcoxon test, $p = 0.0002584$) than in 222 long scaffolds (length >= 1Mb).

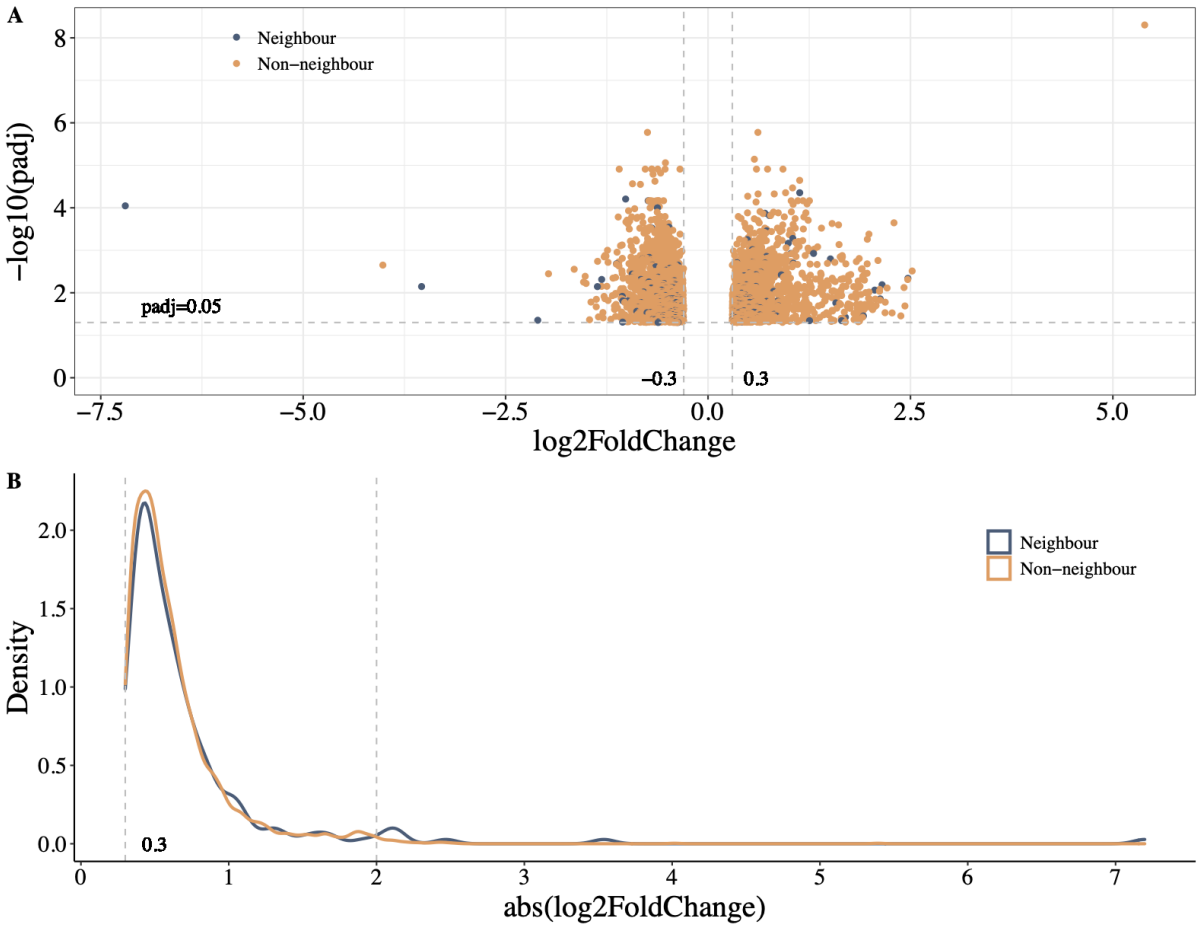


Figure S3: Expression pattern comparison between coding genes with (blue) and without (orange) neighbouring lincRNAs. A. Volcano plot of all expressed coding genes. The gene expression log2 fold change and log10(adjusted P-value) between samples from CO_2 seep and control site are reported on the X and the Y axes, respectively. B. Density of log2FoldChange absolute value of gene expression between samples from CO_2 seep and control site.

REFERENCES

Akay, A., Jordan, D., Navarro, I. C., Wrzesinski, T., Ponting, C. P., Miska, E. A., & Haerty, W. (2019). Identification of functional long non-coding RNAs in *C. elegans*. *BMC Biology*, *17*(1), 1–14.

Al-Tobasei, R., Paneru, B., & Salem, M. (2016). Genome-wide discovery of long non-coding RNAs in rainbow trout. *PLoS ONE*, *11*(2).

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Retrieved May 24, 2022, from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Aprea, J., & Calegari, F. (2015). Long non-coding RNA s in corticogenesis: deciphering the non-coding code of the brain. *The EMBO Journal*, *34*(23), 2865–2884.

Beck, D., Thoms, J. A. I., Palu, C., Herold, T., Shah, A., Olivier, J., … Pimanda, J. E. (2018). A four-gene LincRNA expression signature predicts risk in multiple cohorts of acute myeloid leukemia patients. *Leukemia*, *32*(2), 263–272.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300.

Berecki, G., McArthur, J. R., Cuny, H., Clark, R. J., & Adams, D. J. (2014). Differential Cav2.1 and Cav2.3 channel inhibition by baclofen and α-conotoxin Vc1.1 via $GABA_B$ receptor activation. *Journal of General Physiology*, *143*(4), 465–479.

Bernal, M. A., Ravasi, T., Rodgers, G. G., Munday, P. L., & Donelson, J. M. (2022). Plasticity to ocean warming is influenced by transgenerational, reproductive, and developmental exposure in a coral reef fish. *Evolutionary Applications*, *15*(2), 249–261.

Bhat, R., Axtell, R., Mitra, A., Miranda, M., Lock, C., Tsien, R. W., & Steinman, L. (2010). Inhibitory role for GABA in autoimmune inflammation. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(6), 2580–2585.

BioBam Bioinformatics. (2019). OmicsBox – Bioinformatics made easy. *BioBam Bioinformatics*. Retrieved from https://www.biobam.com/omicsbox/

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Boltaña, S., Valenzuela-Miranda, D., Aguilar, A., Mackenzie, S., & Gallardo-Escárate, C. (2016). Long noncoding RNAs (lncRNAs) dynamics evidence immunomodulation during ISAV-Infected Atlantic salmon (*Salmo salar*). *Scientific Reports*, *6*(1), 1–13.

Cabili, M. N., Dunagin, M. C., McClanahan, P. D., Biaesch, A., Padovan-Merhar, O., Regev, A., … Raj, A. (2015). Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biology*, *16*(1).

De Souza, K. B., Jutfelt, F., Kling, P., Förlin, L., & Sturve, J. (2014). Effects of increased CO_2 on fish gill and plasma proteome. *PLoS ONE*, *9*(7), e102901.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., … Guigó, R. (2012). The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Research*, *22*(9), 1775–1789.

Dettleff, P., Hormazabal, E., Aedo, J., Fuentes, M., Meneses, C., Molina, A., & Valdes, J. A. (2020). Identification and Evaluation of Long Noncoding RNAs in Response to Handling Stress in Red Cusk-Eel (*Genypterus chilensis*) via RNA-seq. *Marine Biotechnology*, *22*(1), 94–108.

Duan, Y., Zhang, W., Cheng, Y., Shi, M., & Xia, X. Q. (2021). A systematic evaluation of bioinformatics tools for identification of long noncoding RNAs. *RNA*, *27*(1), 80–98.

Engreitz, J. M., Haines, J. E., Perez, E. M., Munson, G., Chen, J., Kane, M., … Lander, E. S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*, *539*(7629), 452–455.

Fabricius, K. E., Langdon, C., Uthicke, S., Humphrey, C., Noonan, S., De'ath, G., … Lough, J. M. (2011). Losers and winners in coral reefs acclimatized to elevated carbon dioxide concentrations. *Nature Climate Change*, *1*(3), 165–169.

Fang, S., Zhang, L., Guo, J., Niu, Y., Wu, Y., Li, H., … Zhao, Y. (2018). NONCODEV5: A comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Research*, *46*(D1), D308–D314.

Feng, J., Bi, C., Clark, B. S., Mady, R., Shah, P., & Kohtz, J. D. (2006). The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes and Development*, *20*(11), 1470–1484.

Fernandes, J. C. R., Acuña, S. M., Aoki, J. I., Floeter-Winter, L. M., & Muxel, S. M. (2019). Long non-coding RNAs in the regulation of gene expression: Physiology and disease. *Non-Coding RNA, 5*(1), 17.

Gardini, A., & Shiekhattar, R. (2015). The many faces of long noncoding RNAs. *FEBS Journal*, *282*(9), 1647–1657.

Gil, N., & Ulitsky, I. (2020). Regulation of gene expression by cis-acting long non-coding RNAs. *Nature Reviews Genetics, 21*, 102–117.

Guo, C. J., Ma, X. K., Xing, Y. H., Zheng, C. C., Xu, Y. F., Shan, L., … Chen, L. L. (2020). Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells. *Cell*, *181*(3), 621-636.e22.

Heuer, R. M., & Grosell, M. (2014). Physiological impacts of elevated carbon dioxide and ocean acidification on fish. *American Journal of Physiology - Regulatory Integrative and Comparative Physiology*, *307*(9), R1061–R1084.

Hezroni, H., Koppstein, D., Schwartz, M. G., Avrutin, A., Bartel, D. P., & Ulitsky, I. (2015). Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. *Cell Reports*, *11*(7), 1110–1122.

IPCC. (2022). *Summary for Policymakers. In: Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Retrieved from www.ipcc.ch

Jiang, X., Lei, R., & Ning, Q. (2016). Circulating long noncoding RNAs as novel biomarkers of human diseases. *Biomarkers in Medicine, 10*(7), 757–769.

Kang, J., Nagelkerken, I., Rummer, J. L., Rodolfo-Metalpa, R., Munday, P. L., Ravasi, T., & Schunter, C. (2022). Rapid evolution fuels transcriptional plasticity to ocean acidification. *Global Change Biology*, *00*, 1–16.

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915.

Kindgren, P., Ard, R., Ivanov, M., & Marquardt, S. (2018). Transcriptional read-through of the long non-coding RNA SVALKA governs plant cold acclimation. *Nature Communications*, *9*(1).

Kong, L., Zhang, Y., Ye, Z. Q., Liu, X. Q., Zhao, S. Q., Wei, L., & Gao, G. (2007). CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research*, *35*(SUPPL.2), 345–349.

Kornfeld, J. W., & Brüning, J. C. (2014). Regulation of metabolism by long, non-coding RNAs. *Frontiers in Genetics, 5,* 57.

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics, 9*(1), 1–13.

Leiva, F., Rojas-Herrera, M., Reyes, D., Bravo, S., Garcia, K. K., Moya, J., & Vidal, R. (2020). Identification and characterization of miRNAs and lncRNAs of coho salmon (*Oncorhynchus kisutch*) in normal immune organs. *Genomics*, *112*(1), 45–54.

Li, Z., Zhao, W., Wang, M., & Zhou, X. (2019). The Role of Long Noncoding RNAs in Gene

Expression Regulation. In *Gene Expression Profiling in Cancer*, 1–17.
Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930.
Lin, M. F., Jungreis, I., & Kellis, M. (2011). PhyloCSF: A comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics*, *27*(13), 275–282.
Liu, W., Liu, X., Wu, C., & Jiang, L. (2018). Transcriptome analysis demonstrates that long noncoding RNA is involved in the hypoxic response in *Larimichthys crocea*. *Fish Physiology and Biochemistry*, *44*(5), 1333–1347.
Logan, R. W., & McClung, C. A. (2019). Rhythms of life: circadian disruption and brain disorders across the lifespan. *Nature Reviews Neuroscience, 20*(1), 49–65.
Lopez-Ezquerra, A., Harrison, M. C., & Bornberg-Bauer, E. (2017). Comparative analysis of lincRNA in insect species. *BMC Evolutionary Biology*, *17*(1), 1–11.
Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21.
Luo, Y., Chen, J. J., Lv, Q., Qin, J., Huang, Y. Z., Yu, M. H., & Zhong, M. (2019). Long non-coding RNA NEAT1 promotes colorectal cancer progression by competitively binding miR-34a with SIRT1 and enhancing the Wnt/β-catenin signaling pathway. *Cancer Letters*, *440–441*, 11–22.
Machado, M., Arenas, F., Svendsen, J. C., Azeredo, R., Pfeifer, L. J., Wilson, J. M., & Costas, B. (2020). Effects of Water Acidification on Senegalese Sole *Solea senegalensis* Health Status and Metabolic Rate: Implications for Immune Responses and Energy Use. *Frontiers in Physiology*, *11*, 26.
Mosig, R. A., & Kojima, S. (2021). Timing without coding: How do long non-coding RNAs regulate circadian rhythms? *Seminars in Cell and Developmental Biology, 126*, 79–96.
Mu, C., Wang, R., Li, T., Li, Y., Tian, M., Jiao, W., … Bao, Z. (2016). Long Non-Coding RNAs (lncRNAs) of Sea Cucumber: Large-Scale Prediction, Expression Profiling, Non-Coding Network Construction, and lncRNA-microRNA-Gene Interaction Analysis of lncRNAs in *Apostichopus japonicus* and *Holothuria glaberrima* During LPS Challenge. *Marine Biotechnology*, *18*(4), 485–499.
Nadal-Ribelles, M., Solé, C., Xu, Z., Steinmetz, L. M., DeNadal, E., & Posas, F. (2014). Control of Cdc28 CDK1 by a Stress-Induced lncRNA. *Molecular Cell*, *53*(4), 549–561.
Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., … Kaessmann, H. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*, *505*(7485), 635–640.
Paneru, B., Ali, A., Al-Tobasei, R., Kenney, B., & Salem, M. (2018). Crosstalk among lncRNAs, microRNAs and mRNAs in the muscle "degradome" of rainbow trout. *Scientific Reports*, *8*(1), 8416.
Pegueroles, C., Iraola-Guzmán, S., Chorostecki, U., Ksiezopolska, E., Saus, E., & Gabaldón, T. (2019). Transcriptomic analyses reveal groups of co-expressed, syntenic lncRNAs in four species of the genus *Caenorhabditis*. *RNA Biology*, *16*(3), 320–329.
Pertea, G., & Pertea, M. (2020). GFF Utilities: GffRead and GffCompare. *F1000Research*, *9*.
Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, *11*(9), 1650–1667.
Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3), 290–295.
Quan, D., Chen, K., Zhang, J., Guan, Y., Yang, D., Wu, H., … Lv, L. (2020). Identification of lncRNA NEAT1/miR-21/RRM2 axis as a novel biomarker in breast cancer. *Journal of Cellular Physiology*, *235*(4), 3372–3381.

Quan, J., Kang, Y., Luo, Z., Zhao, G., Ma, F., Li, L., & Liu, Z. (2020). Identification and characterization of long noncoding RNAs provide insight into the regulation of gene expression in response to heat stress in rainbow trout (*Oncorhynchus mykiss*). *Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics*, *36*, 100707.

Quinn, J. J., Zhang, Q. C., Georgiev, P., Ilik, I. A., Akhtar, A., & Chang, H. Y. (2016). Rapid evolutionary turnover underlies conserved lncRNA-genome interactions. *Genes and Development*, *30*(2), 191–207.

Ransohoff, J. D., Wei, Y., & Khavari, P. A. (2018). The functions and unique features of long intergenic non-coding RNA. *Nature Reviews Molecular Cell Biology 19*(3), 143-157.

Ren, X., Chen, C., Luo, Y., Liu, M., Li, Y., Zheng, S., … Chen, R. (2020). LncRNA-PLACT1 sustains activation of NF-κB pathway through a positive feedback loop with IκBα/E2F1 axis in pancreatic cancer. *Molecular Cancer*, *19*(1), 1–19.

Rodriguez-Lopez, M., Anver, S., Cotobal, C., Kamrad, S., Malecki, M., Correia-Melo, C., … Bähler, J. (2022). Functional profiling of long intergenic non-coding RNAs in fission yeast. *ELife*, *11*.

Sarangdhar, M. A., Chaubey, D., Srikakulam, N., & Pillai, B. (2018). Parentally inherited long non-coding RNA Cyrano is involved in zebrafish neurodevelopment. *Nucleic Acids Research*, *46*(18), 9726–9735.

Schmidt, M. (2019). *Behavioural disturbances and underlying neurophysiological mechanisms during ocean acidification and warming in Gadus morhua and Boreogadus saida* (Doctoral dissertation, Universität Bremen). Retrieved from http://nbn-resolving.de/urn:nbn:de:gbv:46-00107505-11

Schneider, H. W., Raiol, T., Brigido, M. M., Walter, M. E. M. T., & Stadler, P. F. (2017). A Support Vector Machine based method to distinguish long non-coding RNAs from protein coding transcripts. *BMC Genomics*, *18*(1), 804.

Schunter, C., Jarrold, M. D., Munday, P. L., & Ravasi, T. (2021). Diel pCO_2 fluctuations alter the molecular response of coral reef fishes to ocean acidification conditions. *Molecular Ecology*, *30*(20), 5105–5118.

Schunter, C., Welch, M. J., Nilsson, G. E., Rummer, J. L., Munday, P. L., & Ravasi, T. (2018). An interplay between plasticity and parental phenotype determines impacts of ocean acidification on a reef fish. *Nature Ecology and Evolution*, *2*(2), 334–342.

Schunter, C., Welch, M. J., Ryu, T., Zhang, H., Berumen, M. L., Nilsson, G. E., … Ravasi, T. (2016). Molecular signatures of transgenerational response to ocean acidification in a species of reef fish. *Nature Climate Change*, *6*(11), 1014–1018.

Sing, T., Sander, O., Beerenwinkel, N., & Lengauer, T. (2005). ROCR: Visualizing classifier performance in R. *Bioinformatics*, *21*(20), 3940–3941.

Statello, L., Guo, C. J., Chen, L. L., & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology 22*(2), 96–118.

Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nature Structural and Molecular Biology 14*(2), 103–105.

Ulitsky, I. (2016). Evolution to the rescue: Using comparative genomics to understand long non-coding RNAs. *Nature Reviews Genetics 17*(10), 601–614.

Ulitsky, I., & Bartel, D.P. (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell* 154, 26–46.

Vance, K. W., & Ponting, C. P. (2014). Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends in Genetics 30*(8), 348–355.

Wan, G., Hu, X., Liu, Y., Han, C., Sood, A. K., Calin, G. A., … Lu, X. (2013). A novel non-coding RNA lncRNA-JADE connects DNA damage signalling to histone H4 acetylation. *EMBO Journal*, *32*(21), 2833–2847.

683 Wang, L., Park, H. J., Dasari, S., Wang, S., Kocher, J. P., & Li, W. (2013). CPAT: Coding-
684 potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids*
685 *Research*, *41*(6), e74–e74.
686 Wu, Z., Liu, X., Liu, L., Deng, H., Zhang, J., Xu, Q., … Ji, A. (2014). Regulation of lncRNA
687 expression. *Cellular and Molecular Biology Letters*, *19*(4), 561–575.
688 Wucher, V., Legeai, F., Hédan, B., Rizk, G., Lagoutte, L., Leeb, T., … Derrien, T. (2017).
689 FEELnc: A tool for long non-coding RNA annotation and its application to the dog
690 transcriptome. *Nucleic Acids Research*, *45*(8), e57–e57.
691 Xu, H., Cao, L., Sun, B., Wei, Y., & Liang, M. (2019). Transcriptomic analysis of potential
692 "lncRNA-mRNA" interactions in liver of the marine teleost *Cynoglossus semilaevis* fed
693 diets with different DHA/EPA ratios. *Frontiers in Physiology*, *10*(APR).
694 Xu, J., Meng, Q., Li, X., Yang, H., Xu, J., Gao, N., … Chen, R. (2019). Long noncoding RNA
695 MIR17HG promotes colorectal cancer progression via miR-17-5p. *Cancer Research*,
696 *79*(19), 4882–4895.
697 Yang, F., Zhang, L., Huo, X. S., Yuan, J. H., Xu, D., Yuan, S. X., … Sun, S. H. (2011). Long
698 noncoding RNA high expression in hepatocellular carcinoma facilitates tumor growth
699 through enhancer of zeste homolog 2 in humans. *Hepatology*, *54*(5), 1679–1689.
700 Yang, N., Wang, B., Yu, Z., Liu, X., Fu, Q., Cao, M., … Li, C. (2020). Characterization of a
701 novel lncRNA (SETD3-OT) in turbot (*Scophthalmus maximus L.*). *Fish and Shellfish*
702 *Immunology*, *102*, 145–151.
703 Zhao, L., Wang, J., Li, Y., Song, T., Wu, Y., Fang, S., … He, S. (2021). NONCODEV6: An
704 updated database dedicated to long non-coding RNA annotation in both animals and
705 plants. *Nucleic Acids Research*, *49*(D1), D165–D171.
706 Zhu, J. J., Fu, H. J., Wu, Y. G., & Zheng, X. F. (2013). Function of lncRNAs and approaches
707 to lncRNA-protein interactions. *Science China Life Sciences, 56*(10), 876–885.