Genetic architecture of behavioural resilience to ocean acidification

Robert Lehmann[1], Celia Schunter[2], Megan J. Welch[3], Stefan T. Arold[1,4,5], Göran E. Nilsson[6], Jesper N. Tegner[1,*], Philip L. Munday[3,*], and Timothy Ravasi[3,7,*]

[1] Bioscience Program, Biological and Environmental Sciences & Engineering Division, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia.

[2] Swire Institute of Marine Science, The School of Biological Sciences, The University of Hong Kong, Hong Kong SAR.

[3] Australian Research Council Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Queensland, Australia.

[4] Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia.

[5] Centre de Biologie Structurale, CNRS, INSERM, Université de Montpellier, Montpellier, France.

[6] Section for Physiology and Cell Biology, Department of Biosciences, University of Oslo, Oslo, Norway.

[7] Marine Climate Change Unit, Okinawa Institute of Science and Technology Graduate University (OIST), Okinawa, Japan.

* Corresponding authors

E-mail: jesper.tegner@kaust.edu.sa, philip.munday@jcu.edu.au, timothy.ravasi@oist.jp

Abstract

Genetic variation is essential for adaptation to rapid environmental changes. Identifying genetic variation associated with climate-change related phenotypes is therefore the necessary first step towards predictive models of genomic vulnerability.

Here we used a whole-genome scan to identify candidate genetic variants associated with differences in behavioural resilience to ocean acidification in a coral reef fish. We identified three genomic regions that differ between individuals that are behaviourally tolerant compared with behaviourally sensitive to elevated CO_2. These include a dopamine receptor (*drd4rs*), cadherin related family member 5-like (*cdhr5l*), Synapse-associated protein 1 (*syap1*), and GRB2 Associated Regulator of MAPK1 Subtype 2 (*garem2*), which have previously been found to modify behaviour related to boldness, novelty seeking, and learning in other species, and differ between behaviourally tolerant and sensitive individuals.

Consequently, the identified genes are promising candidates in the search of the genetic underpinnings and adaptive potential of behavioural resilience to ocean acidification in fishes.

Keywords

Climate Change, Ocean Acidification, Genome Scan, Genetic Variation, Adaptation, Selection.

Introduction

Anthropogenic stressors are impacting the physiology, ecology, and behaviour of marine and terrestrial animals at a global scale (Poloczanska et al. 2013; Buxton et al. 2017; Hendry et al. 2008). Yet the response of individual organisms or populations to different environments is not uniform, with significant variation in traits ranging from the camouflage of walking stick insects (Farkas et al. 2013), foraging behaviour of salamanders (Urban 2013) to the body composition of copepods (Charette and Derry 2016). Instead, there is intraspecific phenotypic variation conferring higher fitness of some individuals in altered environmental

conditions. In the dark-eyed Junco, for example, populations inhabiting regions with higher thermal heterogeneity also show increased flexibility in thermogenic capacity compared to populations in thermally homogenous regions (Stager et al. 2021). Indeed, the impact of such intraspecific phenotypic trait variation on a range of ecological response variables is expected to be similar to phenotypic variation across species (Des Roches et al. 2017). If intraspecific phenotypic variation has a large genetic component it can be the basis for adaptation (Falconer and Mackay 1996) to climate change (Bitter et al. 2019; Yang et al. 2021) or other anthropogenic stressors (Biro and Post 2008; Arlinghaus et al. 2017) via natural selection. Therefore, identifying the genetic variation associated with differences in individual fitness in different environments is crucial to making accurate predictions about the biological impacts of climate change and other anthropogenic stressors over the timeframes at which they are occurring (Hoffmann and Sgrò 2011; Razgour et al. 2019; Munday et al. 2013b). In the yellow warbler, genetic variation close to the genes *DRD4* and *DEAF1*, linked to exploratory and novelty-seeking behaviour in several species, was found to be important for successful climate adaptation and allowed the construction of a predictive model for this species (Bay et al. 2018). Modern genomics approaches are making it possible to directly identify candidate allelic variants of genes or loci that can be the raw material for genetic adaptation (Waldvogel et al. 2020) and, therefore, might enable species to adapt to rapid environmental change.

Ocean acidification, caused by the uptake of additional carbon dioxide from the atmosphere, has diverse effects on marine species, including decreased survivorship, altered metabolism, reduced calcification, growth and development (Wittmann and Pörtner 2013; Kroeker et al. 2013; Kelly and Hofmann 2013). In addition, elevated carbon dioxide partial pressure (pCO_2) has been linked to behavioural changes in some fish and invertebrates, with broad-ranging effects on sensory systems, learning and decision making (Paula et al. 2019; Cattano et al. 2018; Wang and Wang 2020; Munday et al. 2019; Heuer and Grosell 2014). These behavioural changes have been linked to impaired function of $GABA_A$ neurotransmitter receptors in the brain (Nilsson et al. 2012; Thomas et al. 2020; Schunter et al. 2019; Heuer et al. 2016; Chivers et al. 2014), as a consequence of acid-base regulation to defend tissue pH against the acidifying effects of high pCO_2.

One behavioural effect observed in fish exposed to elevated CO_2 is an altered response to olfactory cues, including an impaired response to the chemical cues of predators and chemical alarm cues (CAC) from conspecifics (Munday et al. 2010; Nilsson et al. 2012; Porteus et al. 2018; Williams et al. 2019; Ferrari et al. 2011; Ou et al. 2015). Alarm cues are chemicals released from the skin of injured prey, which reliably signal high predation risk to other individuals of the same species (Chivers and Smith 1998). Failure to respond appropriately to predator odor or CAC increases the risk of predation (Ferrari et al. 2010). Indeed, field-based experiments show that larval reef fishes exposed to elevated CO_2 for 4-5 days, which induced impaired responses to predator odor and CAC, exhibit markedly higher rates of mortality from predation in their natural habitat (Munday et al. 2010; Ferrari et al. 2011; Chivers et al. 2014). Nevertheless, variation in behavioural sensitivity to elevated CO_2 has been detected, with some individuals displaying greater tolerance to elevated CO_2 than others (Welch and Munday 2017; Schunter et al. 2016; Munday et al. 2013a). While the behavioural response of some individuals to predator odor or CAC is impaired at CO_2 levels predicted to occur by the end of the century (700-800 µatm), the behaviour of other individuals is unaffected (Munday et al. 2010, 2013b; Ferrari et al. 2011; Welch et al. 2014). In the spiny damselfish, *Acanthochromis polyacanthus*, variation in behavioural response to CAC under elevated pCO_2 is correlated in fathers and their offspring, suggesting a heritable genetic basis (Welch and Munday 2017), which might enable populations of this species to adapt to rising CO_2 levels in the ocean. The transcriptional brain response of juvenile *A. polyacanthus* to high CO_2 has been studied, showing differential expression patterns between the offspring of behaviourally tolerant or sensitive parents (Schunter et al. 2016, 2018), but the genetic variants associated with the observed variation in behavioral tolerance to elevated CO_2 remain unknown.

Here, we used a whole-genome scan to compare the genotype of *A. polyacanthus* individuals that have a strongly impaired behavioural response to conspecific alarm cue (i.e. are behaviourally sensitive) compared with the genotype of individuals that show a typical response (i.e. are behaviourally tolerant) to CAC under elevated CO_2 conditions. To do this we re-sequenced the genomes of fish reared in conjunction with the study of Welch and Munday (Welch and Munday 2017) to test the heritability of behavioural tolerance to

elevated CO_2 in *A. polyacanthus*. Briefly, wild-caught adult fish from the same population were exposed to elevated CO_2 (754 µatm, consistent with climate change projections) and their behavioural response to CAC was tested, assigning individuals retaining a strong natural avoidance of CAC to the tolerant group and individuals being attracted to CAC to the sensitive group. Breeding pairs were formed from similar-sized males and females in the behaviourally tolerant group and from similar-sized males and females in the behaviourally sensitive group. Offspring from these tolerant and sensitive breeding pairs were then reared under control or elevated CO_2 conditions for up to five months. In the current study, the genomes of adult fish (N=18) together with a large number of their offspring (N=210) were then re-sequenced to generate a high confidence set of genetic variants. Comparing this genetic variant set in the adult fish that had been categorized as either tolerant or sensitive to elevated CO_2 allowed us to identify genomic candidate regions and genes that differentiate individuals that are behaviourally tolerant or sensitive to elevated CO_2 (Fig. 1a).

Results & Discussion

A chromosome-scale reference genome for *Acanthochromis polyacanthus*

A whole-genome scan for genetic differentiation requires a high-quality reference genome. We generated a long-read dataset with 131x coverage of brain genome from an adult *Acanthochromis polyacanthus* using the SMRT sequencing platform of PacBio and assembled a high quality highly contiguous brain reference genome assembly, which was then placed in chromosome-scale scaffolds using a chromatin contact map from Hi-C data (see Tables S1, S2, S3, and Figures S2, S3). The resulting *A. polyacanthus* genome assembly consists of 25,468 annotated genes on 24 pseudo-chromosomes (Fig. 1b). It features an N50 of 41.7 Mb with 96 % of the initial assembly being placed in chromosomes and 57 unplaced contigs. The assembly includes more transposable element insertion sites, leading to an increased repeat content of 38 % (Fig. S4, Table S4), compared to a previous short-read based assembly (Schunter et al. 2016) as well as a higher assembly level completeness (BUSCO score 96.7 %) with reduced duplication (Table S5). Assembly of the mitochondrial genome and construction of a phylogeny using mitochondrial and nuclear marker genes also confirms the species of the sequenced individual (Fig. S5, Table S6).

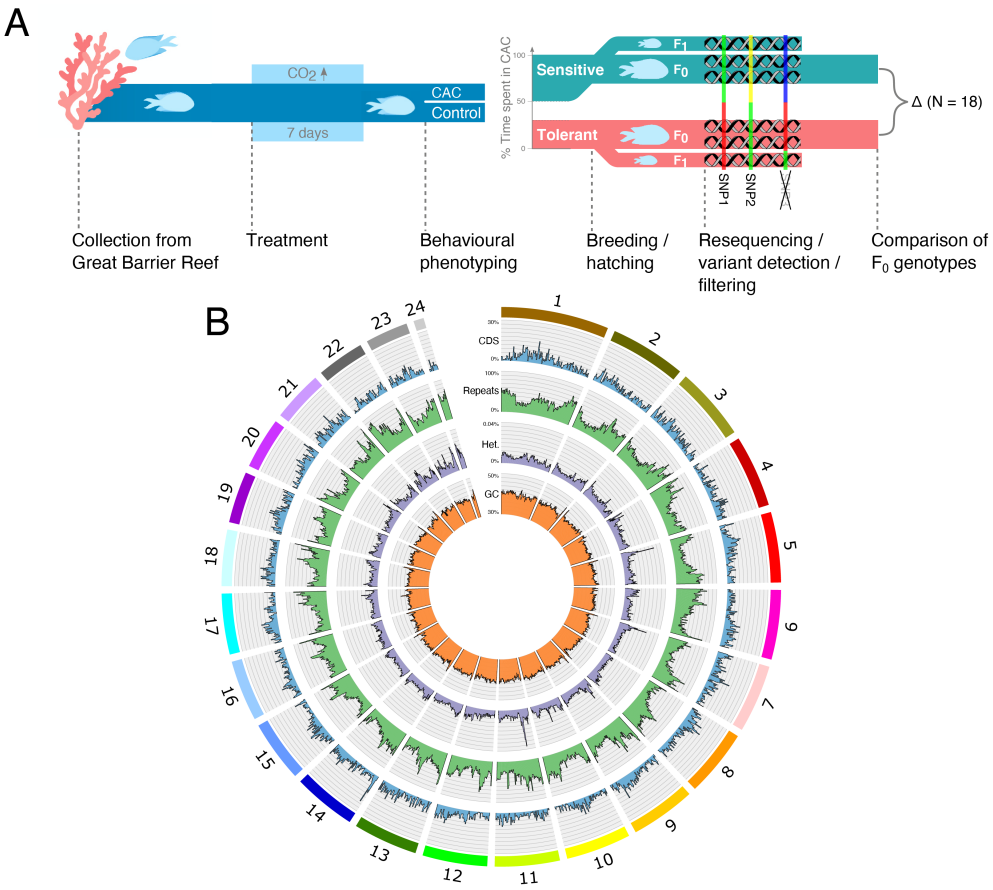

Figure 1: Experiment setup and the reference genome assembly. a) Experimental design with the F_1 rearing treatments elaborated in Fig. S1 **b)** The genome assembly of *A. polyacanthus*, consisting of 24 pseudo-chromosomes (outermost ring), with coding sequence density (blue), repetitive sequence density (green), heterozygosity in non-repetitive sequence (purple), and GC content (orange) shown for non-overlapping windows of 500 kb width.

Genomic regions segregating with behavioural phenotype

To identify a set of polymorphic single nucleotide polymorphisms (SNPs) in the sample population of *A. polyacanthus* we re-sequenced the genomes of 18 behaviourally tested adult individuals. In addition, we re-sequenced 210 offspring of these adults to allow for pedigree-based filtering later on. Sequencing yielded an averaged 103.5 million raw reads per individual resulting in 32.7x coverage (Table S7). Comparison of all re-sequenced individuals to the reference genome assembly allowed for the identification of 16.4 million

polymorphic genomic sites. The pedigree information available for the 210 offspring allowed for rigorous quality filtering resulting in a final set of 5.88 million SNPs (Table S8). This high-quality set of variants was then used to identify genomic loci linked to the behavioural phenotype in the eighteen adults that had been behaviourally tested. For this we used the cross-population haplotype-based statistic XP-nSL (Szpiech et al. 2020), cross-population extended haplotype homozygosity XP-EHH (Sabeti et al. 2007), and the fixation index F_{ST} in a windowed genome-wide scan (Fig. 2, see Methods for details on window definition and filtering). Similar analyses with comparable number of samples have been used previously to identify genomic regions associated with tame behaviour in farm-bred red foxes (Kukekova et al. 2018), genomic signatures of speciation for in Lake Victoria cichlids (Nakamura et al. 2021) and North American songbirds (Termignoni-Garcia et al. 2022) amongst others. This analysis yielded three genomic outlier regions with consistently strong signal for all three measures (Table S9, Fig. 2). Region I located in the center of chromosome one contains the strongest differentiation signal found in coding sequences across the whole genome, with large positive values for XP-nSL and XP-EHH, specifically in the protocadherin gene *cdhr5l* and the dopamine receptor D4 (*drd4rs*) (Fig. 2 bottom). The positive value of XP-nSL and XP-EHH indicates longer homozygous haplotypes in the tolerant cohort, suggesting a shared origin. In contrast, Region II on chromosome five features significantly negative XP-nSL and XP-EHH values (Fig. S10), which suggests a shared origin of the sensitive haplotype that is centered around the gene coding for the Synapse-associated protein 1 (*syap1*). Region III on chromosome 16 also features negative XP-nSL and XP-EHH values and is centered around GRB2 associated regulator of MAPK1 subtype 2 (*garem2*) (Fig. S11). In summary, the behaviourally tolerant cohort features a determining haplotype at the *drd4rs*/*cdhr5l* locus while the behaviourally sensitive cohort is characterized by two determining haplotypes at the *syap1* and *garem2* loci.

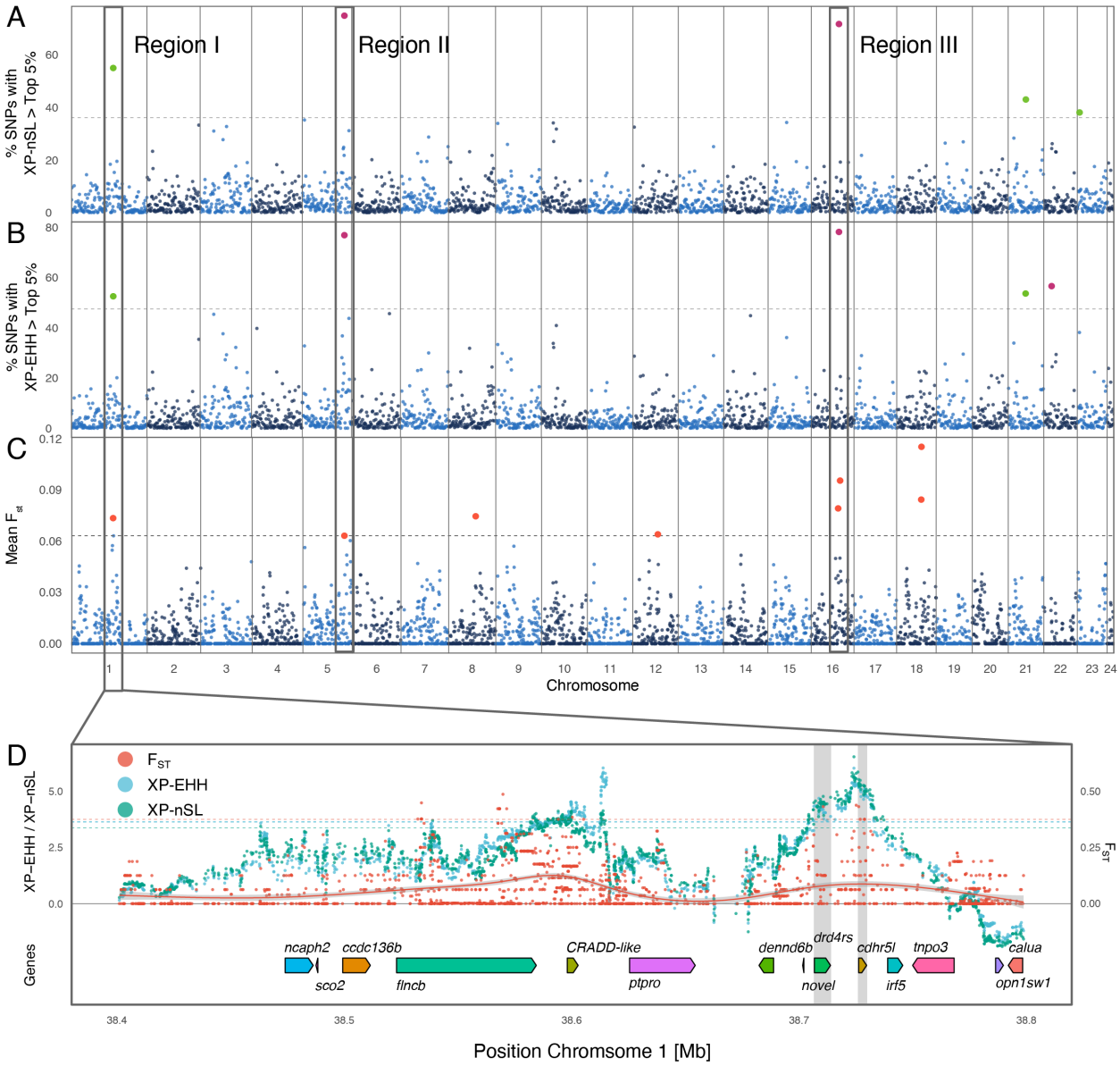

Figure 2: Genome-wide scan for genetic differentiation between behaviourally tolerant and sensitive fish to elevated CO_2. Density of SNPs with extreme a) cross-population nSL (XP-nSL) and b) cross-population extended haplotype homozygosity score (XP-EHH). Each point represents a genomic window of 400 kB width for which the percentage of SNPs among the top 5% most extreme values genome-wide is shown. c) Mean fixation index F_{ST} within 500 kB genomic windows. The dotted lines mark the top 0.2 % cutoff for each measure across all windows, outlier windows are marked in green for positive XP measures, purple for negative XP measures, and orange for F_{ST}. Three regions marked I to III are consistent outliers across measures. d) Detailed comparison of SNP-wise differentiation measures in genomic region I together with annotated genes and the genomic loci of *drd4rs* and *cdhr5l* are shaded in gray. A local polynomial fit of F_{ST} values is shown as a solid line.

174

175

176 **Dopamine receptor 4 (*drd4rs*) shows signature of genetic differentiation**

177 Region I (Fig. 2) contains the two genes with the strongest differentiation signal (Table S9) including one
178 of three dopamine D4 receptors (*drd4rs*, mean XP-EHH 4.21, mean XP-nSL 4.38) and its neighboring
179 gene cadherin-related family member 5-like (*cdhr5l*, mean XP-EHH 4.95, mean XP-nSL 5.13). The local
180 polynomial fit of the fixation index F_{ST} follows the pattern of the haplotype-based measures with two peaks
181 of positive value, suggesting the causal haplotype in the tolerant cohort. The G-protein coupled receptor
182 gene *drd4rs* of *A. polyacanthus* (1,238 bp) consists of 4 exons that code for 7 transmembrane domains and
183 one long intracellular loop (Fig. 3a). A SNP in the third exon, encoding the fifth transmembrane domain
184 (TM5) segregates with the behavioural phenotype (F_{ST} 0.43, top 0.0004 percentile) while inducing a
185 synonymous codon change.

186

187 In birds, synonymous SNPs in *drd4* are associated with behavioural traits such as novelty-seeking or escape
188 behaviour from a cage (Mueller et al. 2014; Kluen et al. 2012; Mueller et al. 2013). Specifically, two
189 synonymous SNPs in *drd4* exon 3 of invasive populations of the Yellow-crowned bishop explain 11% and
190 15% of the neophobic and neophilic behavioural phenotypes for two populations, respectively (Mueller et
191 al. 2014). In blue tit populations, a synonymous exon 3 SNP was associated with escape behaviour from a
192 cage (Kluen et al. 2012). In some great tit populations one synonymous SNP in *drd4* exon 3 was found to
193 be associated with behaviour in a novel environment chamber (Mueller et al. 2013). Furthermore, genetic
194 variation upstream of the *drd4* gene of the yellow warbler appears to be highly relevant to its ecology. This
195 locus was found to feature one of the highest associations between genotype and climate across a range of
196 populations, for which the genotype information was then used to predict genomic vulnerability to climate
197 change (Bay et al. 2018). Taken together, these results demonstrate that in wild bird populations, genetic
198 polymorphisms in the dopamine 4 receptor gene are at least partially responsible for setting the baseline
199 behavioural response to avoidance inducing stimuli, suggesting that the observed synonymous variant in

the *drd4rs* gene of *A. polyacanthus* might similarly modify the typically avoidant response to alarm cues when exposed to elevated CO_2.

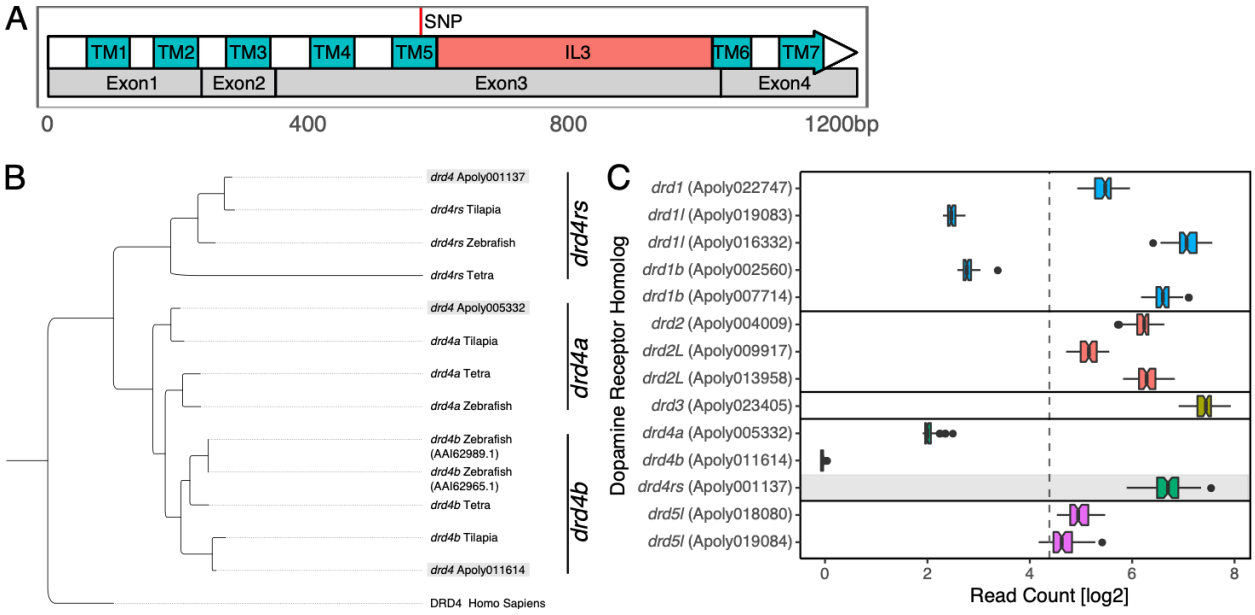

Figure 3: Domain structure of Dopamine Receptor D4 (*drd4rs*) and phylogenetic and transcriptional analysis of dopamine receptor genes. a) A SNP segregating with the behavioural phenotype is located in the third exon of *drd4rs*, encoding the fifth transmembrane domain (TM5) prior to the intracellular loop 3 (IL3). **b)** Phylogenetic tree of *drd4* homologs from our study species, *A. polyacanthus*, and *Oreochromis niloticus*, *Astyanax mexicanus*, and *Danio rerio*, using the human D4 receptor gene as outgroup. **c)** Gene expression levels of all members of the dopamine receptor family (mean normalized read count including all samples from all CO_2 treatments). The mean dataset-wide expression is marked as gray line while gene boxes are color-coded by receptor subgroup 1 to 5.

The family of dopamine receptor genes is separated into two major groups (group 1: *drd1* and *5* and group 2: *drd2*, *3*, and *4*) (Opazo et al. 2018). While mammals and birds generally only have one *drd4* gene, teleosts feature three homologous D4 receptor genes a, b, and rs. Through phylogenetic analysis we ascertained that it is the drd4rs which is carrying the identified genetic variation (Fig. 3b). To evaluate if there is active transcription of this or other candidate genes in the brain of *A. polyacanthus*, we re-analyzed previously published transcriptomics data (Schunter et al. 2018) of 72 offspring exposed to elevated CO_2 conditions for various lengths of time from behaviourally tolerant and sensitive breeding parents (Table S12) . Briefly,

these CO_2 conditions included transgenerational exposure of parents and offspring to elevated CO_2, developmental exposure of offspring to elevated CO_2 from hatching, and acute exposure to elevated CO_2 for four days in five months old offspring (see Fig. S1 and Methods for more details). Comparison of dopamine receptor family members shows *drd4a* and *b* to be extremely low or not expressed while *drd4rs* is expressed constitutively across all individuals and treatments (Fig. 3c), which is consistent with the possible activity of *drd4rs* in the brain.

D4 receptors might modulate aversive behavioural response signaling via indirect pathway in D2 receptors

In order to respond to alarm cue stimulus different environmental inputs (Leahy et al. 2011) are integrated before inducing the avoidance behaviour. This points towards the involvement of the basal ganglia, a part of the forebrain that plays a central role in motivational and cognitive learning, where the so-called direct and indirect pathways mediate the response to positive and negative stimuli (Albin et al. 1989). An aversion stimulus, such as a chemical alarm cue, decreases dopaminergic signaling which reduces the induction of direct pathway neurons, the D1 receptors, and de-repression of indirect pathway neurons, which are the dopamine receptors of the D2 family to which D4 receptors belong (Albin et al. 1989). It was demonstrated experimentally that suppressing firing of dopaminergic neurons in the ventral tegmental area of mice is sufficient to activate the indirect pathway via D2 receptors, inducing aversive behaviour and learning (Danjo et al. 2014). Dopamine receptors of type D4 were shown to form heterodimers with D2 receptors, thereby modulating their output in the striatum in humans (Borroto-Escuela et al. 2011; González et al. 2012). Specifically, dopaminergic signaling is increased upon reduced D2-D4 heteromerization, in this case due to a structural variation in the D4 receptor gene (Bonaventura et al. 2017; Sánchez-Soto et al. 2018). The comparison of the protein structure of the *A. polyacanthus drd4rs* gene to the human D4 receptor protein reveals a well conserved dopamine binding pocket structure (Fig. 4a, S13) and a long unstructured third intracellular loop (IL3) mediating the heteromerization of the D4 receptor with other receptors (Woods 2010), supporting functional homology between both. While the observed genetic differences between

245 behaviourally tolerant and sensitive individuals do not reflect an amino acid change in the D4 protein in *A.*
246 *polyacanthus*, this difference might still have functional consequences. Synonymous SNPs in the D2 gene
247 were found to modify the expression and stability of the mRNA and thus influence the protein abundance
248 (Duan et al. 2003) by changing the ribosomal pausing propensity (McCarthy et al. 2017), explaining the
249 observed association of seemingly silent SNPs to a psychiatric disorder in humans. Although gene
250 expression based on whole-brain tissue of this gene did not differ in our experiment on *A. polyacanthus*, it
251 is possible that the observed non-coding SNP in the *drd4rs* gene differentiating tolerant and sensitive
252 individuals might alter the protein levels in certain brain regions. This could lead to an alteration of D2-D4
253 receptor heterodimers and a modified D2-dependent indirect pathway activation. Behavioural phenotypes
254 fitting this hypothesis were found in experimental pharmacological perturbations of dopaminergic signaling
255 in *Danio rerio*. Amongst the observed phenotypes were swimming distance (Ek et al. 2016), occurrence of
256 swimming episodes (Thirumalai and Cline 2008), activity (Tran et al. 2015), dark-phase activity (Shontz et
257 al. 2018), and notably hypoactivity in response to a high affinity D4 receptor antagonist (Boehmler et al.
258 2007). Similarly, an increased expression of the D2 receptor resulted in bolder behaviour in *Danio rerio*
259 when placed in a novel tank (Thörnqvist et al. 2019). The involvement of the direct/indirect pathway
260 circuitry in fish is supported by the extensive homology of brain regions between *Danio rerio* and mammals
261 (Parker et al. 2013). Furthermore, a study of cleaner wrasse (*Labroides dimidiatus*) and client fish (*Naso*
262 *elegans*) interactions under elevated CO_2 conditions found a large decrease in dopamine in the midbrain of
263 the cleaner wrasse, which was associated with decreased cleaning interactions (Paula et al. 2019). Hence,
264 there is evidence from a range of model systems, including fish, that alterations in dopaminergic signaling
265 can lead to similar boldness-related behavioural phenotypes as observed in *A. polyacanthus*. Furthermore,
266 synonymous SNPs can also lead to alterations in dopamine receptor abundance, potentially causing an
267 imbalance between dopamine receptor homologs and thereby disturbing dopaminergic signaling.
268
269 Prior research has demonstrated the involvement of $GABA_A$ receptors in the behavioural response to
270 olfactory stimuli under elevated CO_2 in coral reef and other marine fishes (Nilsson et al. 2012; Chivers et

al. 2014; Williams et al. 2019; Schunter et al. 2019). The synaptic homeostasis of both D1 and D2 receptor-expressing neurons is maintained by extra-synaptic $GABA_A$ receptors in mice, which modulate their excitability by induction of a tonic current in response to prolonged activation (Maguire et al. 2014). Moreover, the activity of the ventral tegmental area has been found to be modulated via dense GABAergic projections (Barrot et al. 2012) in response to stimuli and three distinct fear and anxiety-related behaviours were reduced upon introduction of lesions in this area (Jhou et al. 2009). This interweaving of dopaminergic and GABAergic signaling is consistent with the observation that perturbation of GABAergic signaling with gabazine can reverse the high CO_2-induced behavioural phenotype (Nilsson et al. 2012; Chivers et al. 2014). It is thus tempting to suggest that the genetic variation we see in the dopamine D4 receptor *drd4rs* of *A. polyacanthus* modifies the behavioural response to a stimulus by altering the sensitivity balance between the direct and indirect pathway. This variation gives rise to bolder individuals unlikely to respond aversely to the alarm cue stimulus under elevated CO_2 and less bold individuals who are more likely to respond aversely to alarm cue.

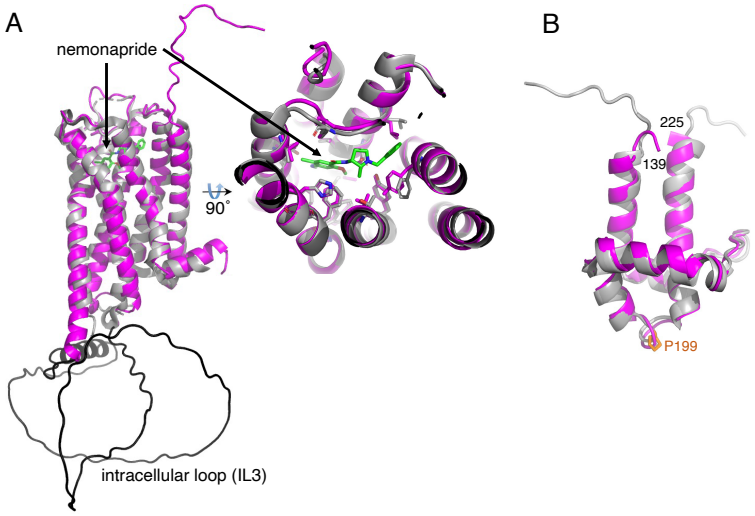

Figure 4: Protein structure prediction of candidate genes segregating behavioural phenotypes. a) The AlphaFold predicted structure of the *A. polyacanthus drd4rs* gene (pink) superimposes with an RMSD of 0.8 Å to the human D4 receptor gene (grey, PDB Id: 5wiu). The selective antagonist nemonapride is shown as a green stick model. The long predicted intracellular loop IL3 is shown in black. On the right, the view into the dopamine binding pocket reveals ligand-binding residues, which are nearly all strictly conserved (key residues and nemonapride are shown as stick models). **b)** Structural AlphaFold model of the BSD

domain of the *syap1* Leu199 variant (magenta; only residues 139-225 are shown), superimposed onto the human Synapse-associated protein 1 BSD domain structure (grey; RMSD is 0.7 Å; PDB accession 1x3a) featuring a highly conserved proline in this position (shown as orange stick model).

Possible association of fish-specific protocadherin to behavioural variation

Next to the dopamine receptor, the gene with the strongest differentiation signal between tolerant and sensitive behavioural phenotypes under elevated CO_2 in *A. polyacanthus* is cadherin related family member 5-like (*cdhr5l,* Apoly001138) (Fig. 2). It is part of the superfamily of cadherin genes which is comprised of several large gene families (classical cadherins, desmosomal cadherins and protocadherins) with diverse functions in cell-cell interaction during development, differentiation, migration, axon outgrowth, dendrite arborization, synapse formation, stabilization, and plasticity (Seong et al. 2015). *CDHR5* is classified as non-clustered protocadherin group epsilon according to Kim *et al.* (Kim et al. 2011) due to the four cadherin repeat domains (human *CDHR5* as well as *A. polyacanthus*' *cdhr5l* Apoly001138). We found a significantly differentiated synonymous SNP featuring the highest fixation index (F_{ST} 0.43, top 0.0004 percentile), while one non-synonymous SNP (F_{ST} 0.31, top 0.0024 percentile) and two successive SNPs close to the splice acceptor site of exon 10 also feature significant differentiation (F_{ST} 0.375, top 0.001 percentile). The non-synonymous SNP Met348Thr is located within the third cadherin repeat domain that is part of the interface to cadherin repeat 4 (Fig. S9), where it codes either for a methionine or threonine amino acid. Cadherins are transmembrane proteins which mediate cell-cell adhesion via their extracellular cadherin domains. These cadherin domains need to be stabilised in an extended position by calcium ions that are lodging between successive cadherin domains to stabilise their linker regions. The substitution of the long hydrophobic methionine with a shorter and partly hydrophilic threonine is expected to affect the associations of this position with the core of the cadherin repeat 3, and thus the position of the loop containing arginines 89 and 90. Such a conformational change in one of the extracellular cadherin repeat domains may influence the responsiveness and dynamics of the cadherin to calcium, and hence affect its capability and dynamics of cell-cell interaction mediation between neurons in the brain.

The differentiating *cdhr5l* Apoly001138 as well as its homolog Apoly011615 are highly expressed in the *A. polyacanthus* transgenerational elevated CO_2 treatment (Fig. S12). Furthermore, various members of this gene family are expressed in the brain and differentially regulated in response to different CO_2 treatments. In mammalian model organisms non-clustered *PCDH* are predominantly expressed in the nervous system and are important for the establishment of selective synaptic connections between the cerebral cortex and other brain regions, such as the thalamus (Kim et al. 2007), and the maintenance and plasticity of the adult hippocampus (Kim et al. 2010). Similarly, protocadherins exhibit complex expression patterns in the brain of *Danio rerio*, where they are required for normal functioning and maintenance (Liu et al. 2015). A recent study showed that cadherin mediates the stabilization and long term potentiation of excitatory synapses at dopaminergic neurons in the ventral tegmental area (Mills et al. 2017). Perhaps unsurprisingly then, mutations in various protocadherins have been linked to a wide range of behavioural disorders in humans (Tsai and Huber 2017). These results suggest a potential role of this still unstudied protocadherin homolog *cdhr5l* in the maintenance and plasticity of brain regions upstream of the basal ganglia of *A. polyacanthus,* allowing for the possibility that the observed genetic variation might impact the behavioural response to alarm cue under elevated CO_2.

Synapse associated protein 1 of region II is associated with synaptic plasticity-related behavioural phenotypes in various model systems

We found three paralogs of Synapse-associated protein 1 (*SYAP1),* expressed at low levels in the brain of offspring *A. polyacanthus* (Fig. S10), with one harboring a non-synonymous SNP at amino acid position 199 (either a proline or a leucine) in the functionally relevant BSD domain, which segregates with the behaviourally tolerant and sensitive phenotypes (F_{st} 0.28, top 0.005 percentile). Similar to the tolerant cohort, the experimental structure of the 80% identical BSD domain from human Syap1 features a proline in *cis* conformation in position 199 (PDB accession 1x3a). This *cis*-bonded proline allows the α4-α4 loop

342 to connect both helices while sealing the small hydrophobic core (Fig. 4b). Prolines are energetically more
343 favorable and hence >100 times more likely to form *cis* bonds than all other residues (Joseph et al. 2012).
344 Therefore, substituting Pro199 with a leucine as found in the sensitive cohort is expected to alter the α4-α4
345 loop conformation and to destabilize the BSD domain. The function of BSD domains is unknown, but
346 prolines are highly conserved at this position (Doerks et al. 2002), suggesting that this *cis*-proline is
347 functionally relevant. *Cis-trans* changes are associated with the evolution of new functions (Joseph et al.
348 2012), supporting a role of this variant in fish behavior.
349
350 Support for a possible role of *syap1* in the response to alarm cue is found in fruit fly and mouse model
351 systems. Firstly, larvae of Sap47 knockout fruit flies exhibit deficiencies in short-term plasticity involved
352 in olfactory associative memory processing (Saumweber et al. 2011). Specifically, these larvae show a
353 ~50% reduction in the ability to learn and/or remember the association of an odorant with a rewarding
354 tastant. Furthermore, recent work demonstrated that Sap47 knockout reduces the lifespan, impairs climbing
355 proficiency, and reduces the plasticity in circadian rhythm and sleep (Blanco-Redondo et al. 2019).
356 Secondly, in mice *SYAP1* is prominently expressed in the nervous system where a knockout reduces
357 locomotor activity in early phases when voluntary movement is initiated (Von Collenberg et al. 2019).
358 Accordingly, the observed genetic variation in *A. polyacanthus* might modulate the protein functional
359 efficiency and thereby olfactory learning and/or voluntary movement initiation in response to alarm cue in
360 elevated CO_2.
361
362 ***garem2* in region III is a modifier of anxiety-like behavior in mice**
363 The third region significantly differentiating between the two cohorts contains a member of the Grb2-
364 associated and regulator of Erk/MAPK gene family (*GAREM*) (Tashiro et al. 2009). *A. polyacanthus*
365 features one copy of *garem2*, which is expressed in the brain of offspring from both phenotypes, with a
366 mean-variance stabilized expression level of 8.2 across all samples. Furthermore, we observe significant

367 expression induction under developmentally elevated CO_2 treatment of the offspring (log2-fold-change
368 0.34, $p < 2.08E-6$). Two SNPs in the first intron exhibit a high F_{ST} of 0.38 (top 0.0009 percentile) as well
369 as a set of SNPs upstream (Fig. S11), while six variants within the coding sequence do not segregate with
370 the phenotype ($F_{ST} < 0.0625$, below top 0.12 percentile). Prior work in mice has shown the involvement of
371 *GAREM2* with behaviour similar to the alarm cue response. Knocking out *GAREM2* in mice leads to a
372 reduction of anxiety-like behavior, increased social approaching and exploratory behavior, and a reduction
373 in novelty-induced anxiety (Nishino et al. 2019).
374

375 **Candidate loci for genetic predisposition of behavioural tolerance to elevated CO_2 stressor**

376 Currently, the genetic structure responsible for the phenotypic variation in behavioral resilience to ocean
377 acidification is not known. Here we present a whole-genome scan to detect genomic candidate loci that
378 might be responsible for this variation. We find variation in the genes coding for dopamine receptor D4
379 (*drd4rs*), cadherin related family member 5-like (*cdhr5l*), Synapse-associated protein 1 (*syap1*), and GRB-
380 associated and regulator of ERK/Mapk subtype 2 (*garem2*) that correlate with behaviourally tolerant and
381 sensitive phenotypes expressed under elevated CO_2 conditions. Each of these genes has been connected to
382 modifications of boldness, exploratory, and anxiety-related behavioural patterns in a range of model
383 species, supporting their role in the impairment of behavioural response to risk cues (e.g. alarm cues and
384 predator cues) in *A. polyacanthus* and other fishes (Welch et al. 2014; Nagelkerken and Munday 2016;
385 Cattano et al. 2018; Williams et al. 2019; Munday et al. 2019; Porteus et al. 2018). The family of dopamine
386 receptors is mechanistically well studied, leading to the hypothesis that the observed genetic variation in
387 behaviourally tolerant *A. polyacanthus* modifies the aversive behavior signaling in response to negative
388 stimuli in the basal ganglia, thereby expanding the acclimatization range compared to sensitive individuals.
389 The genetic variation in *cdhr5l*, *syap1* and *garem2* of sensitive individuals might have the opposite effect
390 upstream of the basal ganglia, collectively modifying the baseline behavioural response of an individual to
391 a negative stimulus, such as conspecific alarm cues. Elevated environmental CO_2 modifies the genetically

predisposed typical behavior, rendering sensitive individuals less responsive to alarm cue, while tolerant individuals manage to respond appropriately, as they would under normal CO_2 conditions.

Our results suggest a link between the alarm cue avoidance behaviour to well-described signaling mechanisms during aversive behaviour in several model organisms and furthermore suggest that there is standing genetic variation in key behaviour-associated genes that would provide the raw material for adaptation of behavioural responses in *A. polyacanthus*, and probably other fishes, to rising CO_2 levels in the ocean. Our results also constitute the ideal starting point for further validation using quantitative genetics, pharmacological perturbation, or genome editing to elucidate the cellular mechanisms responsible for altered behavioural responses to elevated CO_2 and the genetic variation that could foster genetic adaptation of marine animal behaviour to ocean acidification.

Methods

Specimen collection and experimental design

A total of 121 adult spiny damselfish *Acanthochromis polyacanthus* collected in the central Great Barrier Reef (GBR), Australia (18°38'24,3" S, 146°29'31,8" E) were exposed to 754 ± 92 µatm CO_2, a level projected to occur by the end of this century (Collins et al. 2013), for 7 days as previously described (Welch and Munday 2017). The 7 day exposure duration was chosen because previous studies show that fish display impaired responses to chemical stimuli after four or more days of exposure to elevated CO_2 (Ferrari et al. 2011; Heuer et al. 2016; Munday et al. 2010). A two-chamber flume was used to determine the behavioural phenotype of these 121 individuals in response to conspecific chemical alarm cue (CAC). A ratio of one donor fish per test fish was used in the preparation of CAC. The CAC was obtained by euthanizing the donor fish with a quick blow to the head, making superficial incisions on both sides of the body, rinsing the body with 60 ml of control water, and adding that to 10 l of seawater with elevated CO_2. The flume was fed with control and CAC water at a constant rate of 450 ml/min from two header tanks (one with 10 l control water and one with 10 l CAC treated water). A fresh preparation of CAC treated water was used for each test fish. Individuals were subjected to nine minute long behavioural trials consisting of two minutes

habituation, a two minute recording period, one minute switching of CAC and control sides, followed by another two minutes habituation and two minutes recording period. The position of the fish in the flume was noted every five seconds during the two x two minute recording periods. Large variations in the behavioural response to CAC were observed, with some individuals avoiding the CAC as they normally do in ambient conditions, whereas other individuals preferred CAC over control water. The former were termed behaviourally 'tolerant' (< 30 % of the time in CAC) while the latter were termed behaviourally 'sensitive' (> 50 % of time in CAC) to elevated CO_2. Since the experiment was designed to answer a range of complementary questions pertaining to the heritability of the observed behavioural phenotype, including epigenetic effects, it included generating a generation of offspring from the phenotyped adult individuals. The offspring from this experiment that were collected for molecular analyses were not behaviourally phenotyped, and are thus not informative for the comparison of tolerant vs. sensitive individuals, but their genotypes are useful to filter the SNP panel which is why they were included here. Breeding pairs were then assigned based on sex and size into four different groups: (1) tolerant male and tolerant female, (2) sensitive male and sensitive female, (3) tolerant male and sensitive female, and (4) sensitive male and tolerant female. Only the first two groupings, comprising both tolerant or both sensitive individuals are considered in this study. The latter two "mixed phenotype" groups were used in (Welch and Munday 2017), but are not included in the current study. Half of the tolerant and sensitive breeding pairs in each group were allowed to acclimate to control condition and half to elevated CO_2 conditions for three months prior to the start of the breeding season. A total of five tolerant and five sensitive pairs bred in control conditions, and a total of four tolerant and five sensitive pairs bred in elevated CO_2 conditions. To reduce any potential family related bias, two parental pairs from each behavioural phenotype were allowed to breed in control condition and then moved to elevated CO_2, acclimated for three months and bred in elevated CO_2 (included in counts above). Following breeding, five randomly selected tolerant pairs and five randomly selected sensitive pairs were sacrificed for extraction of brain tissue (Table S1). This sums to a total of 20 parental fish, but the total number of sequenced adults was 18 due to DNA degradation observed in four samples and the addition of two available phenotyped samples not selected for breeding pairs. The average

443 percentage time in CAC of the sampled/sequenced tolerant adults was 3.6 % (SD 10.3) and the average
444 percentage time in CAC of the sampled/sequenced tolerant adults was 87 % (SD 18.2).
445 On hatching, offspring clutches from the breeding pairs were immediately moved to separate tanks
446 maintaining the CO_2 conditions of the respective parental pair, where they were reared for up to five months.
447 In addition, some offspring reared under control conditions were exposed to elevated CO_2 levels for four
448 days prior to sampling. These combinations resulted in four different treatments: 1. (Control) parents and
449 offspring under control CO_2 level, 2. (Transgenerational) parents and offspring under elevated CO_2 level,
450 3. (Developmental) parents under control CO_2 level while offspring are exposed to elevated CO_2
451 immediately after hatching, and 4. (Acute) parents and offspring are exposed to control CO_2 levels, but with
452 a four-day long exposure of offspring to elevated CO2 levels before sampling (Fig. S1). Offspring
453 environmental condition exposures were maintained for two durations, five weeks and five months, after
454 which individuals were euthanized. This yielded 228 individuals (18 adults and 210 offspring) suitable for
455 DNA sequencing: F_0 adult nine tolerant / nine sensitive, F_1 five weeks old: 102, F_1 five months old: 108.
456 From this experiment brains of 72 five months old offspring were previously extracted for RNA and
457 sequenced (see Schunter et al. 2018) and re-analyzed in this study. The experiment was conducted under
458 James Cook University ethics approval A1828.

459 **DNA extraction, library preparation, and sequencing for *de novo* genome assembly**

460 To build a reference genome sequence, we collected one large adult fish from the wild at Bramble reef on
461 the Great Barrier Reef, Australia. This individual was not phenotyped since the reference does not influence
462 the result of a genome scan contrasting the sensitive and the tolerant group, as neither are required to match
463 the reference. The whole brain tissue was dissected, snap-frozen in liquid nitrogen and kept at -80°C prior
464 to processing. High molecular weight DNA was then extracted from this tissue using the Qiagen Genomic-
465 tip 100/G extraction kit. Briefly, after homogenization of the whole brain tissue using sterile beads and lysis
466 buffer G2 supplemented with 200 µg/mL RNase A for 30 sec, proteinase K was added followed by
467 overnight incubation at 50°C. DNA was then extracted following the protocol resulting in a final elution

volume of 200 µl. Pulsed-field gel electrophoresis was used to assess DNA fragment size and quality. The extracted DNA was sheared with a g-TUBE (Covaris, MA, USA) to a target size of 20 kb prior to the preparation of SMRTbell libraries according to the protocol provided by Pacific Biosciences, CA, USA. BluePippin pulse-field gel electrophoresis (Sage Science, MA, USA) was used to perform a fragment size selection to obtain one library with a minimum size of 10 kb and one with 5kb, which were then sequenced using a PacBio RS II instrument at the King Abdullah University of Science and Technology (KAUST) Bioscience Core Laboratory using P6-C4 chemistry and 109 SMRT cells.

From our controlled aquarium experiment, the whole brain tissue was dissected from offspring, snap-frozen with liquid nitrogen and kept at -80°C prior to processing. For the adult individuals, fin clips were taken from the dorsal fins and kept in ethanol for further processing. For genome resequencing, the DNA and RNA of the 228 samples were extracted using a Qiagen AllPrep DNA/RNA Mini Kit. For DNA extraction from the 18 adult finclips a Qiagen Blood and Tissue kit was used. Approximately 10 RNase and DNase free one-use silica beads (Daintree Scientific, Australia) were placed into Eppendorf tubes together with the tissue samples, which were then homogenized for 30 seconds in a pre-frozen metal tray with a Thermo Fisher Scientific bead beater. DNA and total RNA were then purified according to the manufacturers' protocol and stored at -80 ºC. Illumina sequencing libraries for 49 samples were produced with a TruSeq DNA library preparation kits and sequenced on the HiSeq4000 platform by Macrogen (Macrogen South Korea). An additional 179 samples were sequenced using the same procedure at the King Abdullah University of Science and Technology (KAUST) Bioscience Core Laboratory.

Genome assembly and proximity guided assembly scaffolding

The PacBio reads (Table S2) were assembled using FALCON v0.4.0 (Chin et al. 2016) varying the parameters generating 11 candidate assemblies (Table S3). These candidate assemblies were compared with respect to contiguity and the best was selected for phasing with FALCON_Unzip and initial polishing with quiver. A chromatin contact map was assembled by Phase Genomics (Seattle, WA, USA). For this procedure, an adult fish was dissected, the flash-frozen brain tissue was fixed and sent to Phase Genomics,

where the chromatin was isolated and a library prepared for 80 bp paired-end sequencing. Chromosome-scale scaffolds were then obtained by mapping the read data to the reference assembly, then clustering, ordering, and orienting contigs into 24 clusters (Fig. S2) using Proximo (Bickhart et al. 2017; Burton et al. 2013) as previously described (Peichel et al. 2017). The scaffolded assembly was then polished over three rounds with Arrow, which resulted in the final assembly (Table S3). To confirm the agreement of a short read based genome assembly (GCA_002109545.1) (Schunter et al. 2016) with the new assembly, a whole-genome alignment was performed with Mummer (Kurtz et al. 2004) using default parameters and the result visualized with dotplotly (Fig. S3). This resulted in an alignment of 97.35 % of the short-read assembly to 93.6 % of the new assembly with 99 % average identity.

Repeat element annotation

A repeat library was constructed with RepeatModeler v1.08 (Smit and Hubley 2008). A second library was constructed with LtrHarvest (Ellinghaus et al. 2008) and LTRdigest (Steinbiss et al. 2009), from the genometools suite 1.5.6 (Gremme et al. 2013) (parameters: -seed 76 -xdrop 7 -mat 2 -mis -2 -ins -3 -del -3 -mintsd 4 -maxtsd 20 -minlenltr 100 -maxlenltr 6000 -maxdistltr 25000 -mindistltr 1500 -similar 90). The combined results were deduplicated via clustering with USEARCH (Edgar 2010) (>90% sequence identity) retaining only cluster representatives. The repeat library was then classified by RepeatClassifier. A total of 38% (Table S4) of the genome assembly was masked by RepeatMasker (Smit et al. 2010) using the *de novo* library and the Repbase v22.05 (Bao et al. 2015) *Danio rerio* repeat library. The comparison of transposable element (TE) content between the short-read assembly (annotated using the same *de novo* library, resulting in 25.2 % of sequence masked) and the new long-read assembly reveals that particularly recently inserted TE copies were assembled more successfully (Fig. S4).

Structural and functional gene annotation

After mapping the RNA-seq data from this previously published experiment (Schunter et al. 2016, 2018) to the final assembly using STAR v2.5.2b (Dobin et al. 2013), BRAKER1 v1.9 (Hoff et al. 2016) was used to perform an *ab-initio* annotation of the soft-masked reference genome assembly, providing the RNA-seq

hints and protein sequences of the short read assembly and the closest related fish species with a high-quality reference genome available, the orange clownfish (Lehmann et al. 2019) (GCA_003047355.1) as evidence. This annotation was filtered to obtain a high-quality gene set to train the Augustus (Stanke et al. 2006) gene prediction pipeline. A subsequent BRAKER1 run with similar hints then identified 41,975 genes. The MAKER2 v2.31.8 (Holt and Yandell 2011) gene annotation pipeline was used to annotate the AED score and only genes with an AED < 0.7 were retained for the final annotation (Table S5). InterProScan 5 was executed to obtain the Pfam protein domain. The current UniProtKB/Swiss-Prot, TrEMBL, and NCBI non-redundant database were obtained at 09/2019 and blast 2.6.0 was used to align the annotated protein sequences to these databases, retaining the best hits when falling below an e-value of $1*e^{-5}$. Annotations in the regions of interest were furthermore validated by mapping protein sequences annotated by Ensembl to the short-read assembly to the new assembly using genomethreader v1.7.0.

Assembly of mitochondrial genome

The obtained PacBio data were aligned to the mitochondrial genome sequence of the closest related fish species, the orange clownfish *Amphiprion percula* (Lehmann et al. 2019) (CM011763.1). This filtering step retained 466 mapping reads with an N50 of 12,531 bp and 3,997,697 bp total length which corresponds to 240x expected coverage of the assembled mitochondrial genome. This read dataset was then assembled with Organelle_PBA (Soorni et al. 2017) and annotated with MitoAnnotator (Iwasaki et al. 2013) (Fig. S5a). A phylogeny based on the sequences of the ATP synthase 8/6, Cytochrome B (*Cyt b*), and V(D)J recombination-activating protein 1 (*RAG1*) genes from eight damselfish species including another sample of *A. polyacanthus* (Table S6) was then constructed to confirm the species of the sequenced individual. After aligning the concatenated sequences with ClustalW 2.1 (Stamatakis 2006), a maximum-likelihood tree was obtained using RAxML (Larkin et al. 2007) with the GTRGAMMA model and 500 rounds of bootstrapping, affirming the identity of the sequenced individual as *A. polyacanthus* (Fig. S5b).

Whole-genome resequencing data processing

Illumina short-read sequences obtained from the individuals from the controlled aquarium experiment were assessed with FastQC (Andrews) and low quality regions were trimmed with Trimmomatic v0.33 (Bolger et al. 2014) using parameters: 2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:40, leaving an average of 97 million reads per sample for analysis (Table S7). Trimmed reads were mapped to the long read assembly and sorted with BWA mem (Li and Durbin 2010), duplicates were removed with sambamba (Tarasov et al. 2015), and read groups added with Picard tools (Broad Institute 2018). Single nucleotide polymorphisms were jointly called in all samples with GATK HaplotypeCaller (McKenna et al. 2010) 3.8.0. An initial high confidence SNP set was generated by filtering the raw variants according to these criteria: QD < 2.0, FS > 60.0, MQ < 40.0, MQRankSum < -12.5, ReadPosRankSum < -8.0, retaining only bi-allelic sites with minimum allele frequency > 0.01, DP > 3, and at least 40 samples with genotype calls. Furthermore, only SNPs conforming to Mendelian inheritance according to our pedigree were retained, yielding 4,296,736 sites (S7 Table, High confidence set). The obtained high confidence variant set was then used as training set for Variant Quality Score Recalibration (VQSR) using these annotations: QD MQRankSum ReadPosRankSum FS MQ SOR DP. The final variant set was then obtained by filtering the VQSR result for an average genotype quality value above 35 and less than 10 violations of Mendelian inheritance per locus. This procedure resulted in a total of 5,882,916 variants (S7 Table, VQSR filtered), which were then phased with shapeit v4.1.3 (Delaneau et al. 2019) using an increased number of iterations with option 10b,1p,1b,1p,1b,1p,1b,1p,10m. Genome-wide distribution of heterozygosity (Fig. 1b) was assessed by identifying repetitive regions with more than three times the expected coverage in each sample, excluding SNPs located in these repetitive regions, and calculating the fraction of remaining heterozygous SNPs in 500 kb non-overlapping windows. PCA (Fig. S6) and Admixture (Alexander et al. 2009) were used to confirm the absence of a population structure signal. The normalized cross-population extended haplotype homozygosity (XP-EHH) and XP-nSL are then calculated between nine sensitive parent individuals and nine tolerant ones using selscan v1.2.0 (Szpiech and Hernandez 2014) setting the sensitive cohort as reference. To define genomic regions that segregate with the behavioural phenotype, the top 5%

of SNPs with the most extreme score genome-wide were then defined as outliers as implemented in selscan's normalization procedure. This resulted in an upper and lower threshold for XP-EHH of 1.95 and -2.04, respectively. Similarly, the upper and lower thresholds for XP-nSL were 1.96 and -2.01, respectively. The genome-wide as well as per-chromosome distributions for SNP-wise XP-EHH and XP-nSL scores are shown in Fig. S7 A-F together with quantile-quantile plots illustrating the defined outlier thresholds. Furthermore, the selscan normalization procedure includes the division of the genome into non-overlapping windows, here set to 400 kb width, where the percentage of outlier SNPs within each window indicates differentiated regions. We selected the top 0.2 % windows with the highest outlier SNP percentage as candidate regions in each differentiation measure, corresponding to 47% and 36% of outlier SNPs per window as thresholds for XP-EHH and XP-nSL, respectively (Fig. S8, Fig. 2 A and B).

The Weir-Cockerham fixation index F_{ST} was calculated between the parental individuals of the same phenotype with vcftools v0.1.16 (Danecek et al. 2011) (see Fig. S7 G for genome-wide distribution), again followed by a windowing into 500kb wide regions with 250kb overlap and calculating the mean F_{st} (Fig. 2 C. Again, the top 0.2% of windows with the highest mean F_{st} were selected as candidate regions, corresponding to a threshold of 0.063 (Fig S8 C). Finally, only regions detected as differentiated by all three measures were retained for detailed evaluation to ensure rigorous filtering and avoid false positives. The application of all three genomic differentiation measures was limited to the parental individuals since phenotypes for offspring are not available making their genomic information not informative beyond the SNP filtering step.

The effect of SNPs was estimated with SNPeff (Cingolani et al. 2012).

Phylogeny construction

A multiple amino acid sequence alignment was constructed using the three *drd4* homologs found in the long-read genome assembly as well as the three *drd4* genes of the Nile tilapia (*Oreochromis niloticus*), Mexican tetra (*Astyanax mexicanus*), and Zebrafish (*Danio rerio*) (Table S11), and lastly the human DRD4 gene (MAFFT using auto setting). A maximum-likelihood phylogenetic tree was obtained from the multiple

sequence alignment using RAxML (Larkin et al. 2007) with the model PROTGAMMAAUTO and 500 rounds of bootstrapping.

RNA-seq data processing

For the RNA-seq read data for 72 samples from the controlled aquarium experiment (Schunter et al. 2016, 2018), data integrity was assessed with FastQC (Andrews), trimmed with Trimmomatic v0.33 (Bolger et al. 2014) (2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:40) and mapped to the long read assembly with STAR v2.5.2b (Dobin et al. 2013) (Table S12) using the MAKER2 annotation as well as the genomethreader annotation. Differential expression analysis and visualization was then performed with R and the DESeq2 package 1.22.2 (Love et al. 2014). Pairwise differential expression analyses were performed between the acute, developmental, and transgenerational samples against control individuals, separating by parental phenotype.

Protein structure prediction

Structural models were produced by AlfaFold (Jumper et al. 2021) using the AlphaFold2_advanced.ipynb colab implementation with default values, except for the Filter options (cov: 90; qid: 30). Importantly, this colab version does not use structural templates, and hence is not biased by known PDB structures. Predicted LDDT per-residue scores were above 80 for the secondary structure regions in the transmembrane domain of Drd4 and for the BSD domain of Syap1 (Fig. S13).

Data Access

All generated data have been deposited at the NCBI under BioProject ID PRJNA671567. Supplementary figures and tables are available at https://doi.org/10.5281/zenodo.7219978.

Competing Interests

The authors declare no competing interests.

Acknowledgements

This work was supported by the Office of Competitive Research Funds (OSR-2015-CRG4-2541 and FCC/1/1976-25 to T.R., P.L.M., S.T.A) from the King Abdullah University of Science and Technology,

and the Australian Research Council (ARC) as well as the ARC Centre of Excellence for Coral Reef Studies to P.L.M.

Contributions

M.J.W. and P.L.M. designed and managed the fish rearing experiments. M.J.W. performed the behavioural phenotyping of the adult fish. C.S. prepared the DNA and RNA samples for sequencing, R.L. designed and performed the analyses, S.T.A. performed the protein structure modeling, R.L. wrote the paper with input from C.S., P.L.M., G.E.N., T.R., and J.N.T.

References

Albin RL, Young AB, Penney JB. 1989. The functional anatomy of basal ganglia disorders. *Trends Neurosci* **12**: 366–375.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**: 1655–64.

Andrews S. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Accessed April 27, 2020).

Arlinghaus R, Laskowski KL, Alós J, Klefoth T, Monk CT, Nakayama S, Schröder A. 2017. Passive gear-induced timidity syndrome in wild fish populations and its potential ecological and managerial implications. *Fish Fish* **18**: 360–373.

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* **6**: 11.

Barrot M, Sesack SR, Georges F, Pistis M, Hong S, Jhou TC. 2012. Braking dopamine systems: a new GABA master structure for mesolimbic and nigrostriatal functions. *J Neurosci* **32**: 14094–101.

Bay RA, Harrigan RJ, Underwood V Le, Gibbs HL, Smith TB, Ruegg K. 2018. Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science (80-)* **359**: 83–86.

Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, Lee J, Lam ET, Liachko I, Sullivan ST, et al. 2017. Single-molecule sequencing and chromatin conformation capture enable *de novo* reference assembly of the domestic goat genome. *Nat Genet* **49**: 643–650.

Biro PA, Post JR. 2008. Rapid depletion of genotypes with fast growth and bold personality

651 traits from harvested fish populations. *Proc Natl Acad Sci* **105**: 2919–2922.
652 Bitter MC, Kapsenberg L, Gattuso J-P, Pfister CA. 2019. Standing genetic variation fuels rapid
653 adaptation to ocean acidification. *Nat Commun* **10**: 5821.
654 Blanco-Redondo B, Nuwal N, Kneitz S, Nuwal T, Halder P, Liu Y, Ehmann N, Scholz N, Mayer
655 A, Kleber J, et al. 2019. Implications of the Sap47 null mutation for synapsin
656 phosphorylation, longevity, climbing proficiency and behavioural plasticity in adult
657 Drosophila. *J Exp Biol* **222**.
658 Boehmler W, Carr T, Thisse C, Thisse B, Canfield VA, Levenson R. 2007. D4 Dopamine
659 receptor genes of zebrafish and effects of the antipsychotic clozapine on larval swimming
660 behaviour. *Genes, Brain Behav* **6**: 155–166.
661 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence
662 data. *Bioinformatics* **30**: 2114–20.
663 Bonaventura J, Quiroz C, Cai N-S, Rubinstein M, Tanda G, Ferré S. 2017. Key role of the
664 dopamine D $_4$ receptor in the modulation of corticostriatal glutamatergic neurotransmission.
665 *Sci Adv* **3**: e1601631.
666 Borroto-Escuela DO, Craenenbroeck K Van, Romero-Fernandez W, Guidolin D, Woods AS,
667 Rivera A, Haegeman G, Agnati LF, Tarakanov AO, Fuxe K. 2011. Dopamine D2 and D4
668 receptor heteromerization and its allosteric receptor–receptor interactions. *Biochem*
669 *Biophys Res Commun* **404**: 928–934.
670 Broad Institute. 2018. Picard Tools. *Broad Institute, GitHub Repos*.
671 Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. 2013. Chromosome-scale
672 scaffolding of de novo genome assemblies based on chromatin interactions. *Nat*
673 *Biotechnol* **31**: 1119–1125.
674 Buxton RT, McKenna MF, Mennitt D, Fristrup K, Crooks K, Angeloni L, Wittemyer G. 2017.
675 Noise pollution is pervasive in U.S. protected areas. *Science (80-)* **356**: 531–533.
676 Cattano C, Claudet J, Domenici P, Milazzo M. 2018. Living in a high CO $_2$ world: a global meta-
677 analysis shows multiple trait-mediated fish responses to ocean acidification. *Ecol Monogr*
678 **88**: 320–335.
679 Charette C, Derry AM. 2016. Climate alters intraspecific variation in copepod effect traits
680 through pond food webs. *Ecology* **97**: 1239–1250.
681 Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R,
682 Figueroa-Balderas R, Morales-Cruz A, et al. 2016. Phased diploid genome assembly with
683 single-molecule real-time sequencing. *Nat Methods* **13**: 1050–1054.

Chivers DP, McCormick MI, Nilsson GE, Munday PL, Watson SA, Meekan MG, Mitchell MD, Corkill KC, Ferrari MCO. 2014. Impaired learning of predators and lower prey survival under elevated CO2: A consequence of neurotransmitter interference. *Glob Chang Biol* **20**: 515–522.

Chivers DP, Smith RJF. 1998. Chemical alarm signalling in aquatic predator-prey systems: A review and prospectus. *Écoscience* **5**: 338–352.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)* **6**: 80–92.

Collins M, Knutti R, Arblaster J, Dufresne J-L, Fichefet T, Friedlingstein P, Xuejie G. 2013. Long-term climate change: Projections, commitments and irreversibility. In *Climate Change 2013: The Physical Science Basis* (eds. T.F. Stocker, D. Qin, G.-K. Plattner, M.M.B. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P.M. Midgley), Cambridge Univ. Press, Cambridge.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

Danjo T, Yoshimi K, Funabiki K, Yawata S, Nakanishi S. 2014. Aversive behavior induced by optogenetic inactivation of ventral tegmental area dopamine neurons is mediated by dopamine D2 receptors in the nucleus accumbens. *Proc Natl Acad Sci U S A* **111**: 6455–60.

Delaneau O, Zagury JF, Robinson MR, Marchini JL, Dermitzakis ET. 2019. Accurate, scalable and integrative haplotype estimation. *Nat Commun* **10**: 1–10.

Des Roches S, Post DM, Turley NE, Bailey JK, Hendry AP, Kinnison MT, Schweitzer JA, Palkovacs EP. 2017. The ecological importance of intraspecific variation. *Nat Ecol Evol 2017 21* **2**: 57–64.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.

Doerks T, Huber S, Buchner E, Bork P. 2002. BSD: A novel domain in transcription factors and synapse-associated proteins. *Trends Biochem Sci* **27**: 168–170.

Duan J, Wainwright MS, Comeron JM, Saitou N, Sanders AR, Gelernter J, Gejman P V. 2003. Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Hum Mol Genet* **12**: 205–216.

717 Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*
718 **26**: 2460–2461.
719 Ek F, Malo M, Åberg Andersson M, Wedding C, Kronborg J, Svensson P, Waters S, Petersson
720 P, Olsson R. 2016. Behavioral Analysis of Dopaminergic Activation in Zebrafish and Rats
721 Reveals Similar Phenotypes. *ACS Chem Neurosci* **7**: 633–646.
722 Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for *de*
723 *novo* detection of LTR retrotransposons. *BMC Bioinformatics* **9**: 18.
724 Falconer D, Mackay T. 1996. *Introduction to quantitative genetics*. Longman Scientifc &
725 Technical, Essex, UK.
726 Farkas TE, Mononen T, Comeault AA, Hanski I, Nosil P. 2013. Evolution of Camouflage Drives
727 Rapid Ecological Change in an Insect Community. *Curr Biol* **23**: 1835–1843.
728 Ferrari MCO, Dixson DL, Munday PL, McCormick MI, Meekan MG, Sih A, Chivers DP. 2011.
729 Intrageneric variation in antipredator responses of coral reef fishes affected by ocean
730 acidification: implications for climate change projections on marine communities. *Glob*
731 *Chang Biol* **17**: 2980–2986.
732 Ferrari MCO, Wisenden BD, Chivers DP. 2010. Chemical ecology of predator–prey interactions
733 in aquatic ecosystems: a review and prospectusThe present review is one in the special
734 series of reviews on animal–plant interactions. *https://doi.org/101139/Z10-029* **88**: 698–
735 724.
736 González S, Rangel-Barajas C, Peper M, Lorenzo R, Moreno E, Ciruela F, Borycz J, Ortiz J,
737 Lluís C, Franco R, et al. 2012. Dopamine D4 receptor, but not the ADHD-associated D4.7
738 variant, forms functional heteromers with the dopamine D2S receptor in the brain. *Mol*
739 *Psychiatry* **17**: 650–62.
740 Gremme G, Steinbiss S, Kurtz S. 2013. GenomeTools: a comprehensive software library for
741 efficient processing of structured genome annotations. *IEEE/ACM Trans Comput Biol*
742 *Bioinforma* **10**: 645–56.
743 Hendry AP, Farrugia TJ, Kinnison MT. 2008. Human influences on rates of phenotypic change
744 in wild animal populations. *Mol Ecol* **17**: 20–29.
745 Heuer RM, Grosell M. 2014. Physiological impacts of elevated carbon dioxide and ocean
746 acidification on fish. *https://doi.org/101152/ajpregu000642014* **307**: R1061–R1084.
747 Heuer RM, Welch MJ, Rummer JL, Munday PL, Grosell M. 2016. Altered brain ion gradients
748 following compensation for elevated CO2 are linked to behavioural alterations in a coral
749 reef fish. *Sci Reports 2016 61* **6**: 1–10.

750 Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: Unsupervised
751 RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*
752 **32**: 767–9.
753 Hoffmann AA, Sgrò CM. 2011. Climate change and evolutionary adaptation. *Nat 2011 4707335*
754 **470**: 479–485.
755 Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management
756 tool for second-generation genome projects. *BMC Bioinformatics* **12**: 491.
757 Iwasaki W, Fukunaga T, Isagozawa R, Yamada K, Maeda Y, Satoh TP, Sado T, Mabuchi K,
758 Takeshima H, Miya M, et al. 2013. Mitofish and mitoannotator: A mitochondrial genome
759 database of fish with an accurate and automatic annotation pipeline. *Mol Biol Evol* **30**:
760 2531–2540.
761 Jhou TC, Fields HL, Baxter MG, Saper CB, Holland PC. 2009. The Rostromedial Tegmental
762 Nucleus (RMTg), a GABAergic Afferent to Midbrain Dopamine Neurons, Encodes Aversive
763 Stimuli and Inhibits Motor Responses. *Neuron* **61**: 786–800.
764 Joseph AP, Srinivasan N, De Brevern AG. 2012. Cis-trans peptide variations in structurally
765 similar proteins. *Amino Acids* **43**: 1369–1381.
766 Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates
767 R, Žídek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with
768 AlphaFold. *Nat 2021 5967873* **596**: 583–589.
769 Kelly MW, Hofmann GE. 2013. Adaptation and the physiology of ocean acidification. *Funct Ecol*
770 **27**: 980–990.
771 Kim S-Y, Chung HS, Sun W, Kim H. 2007. Spatiotemporal expression pattern of non-clustered
772 protocadherin family members in the developing rat brain. *Neuroscience* **147**: 996–1021.
773 Kim S-Y, Yasuda S, Tanaka H, Yamagata K, Kim H. 2011. Non-clustered protocadherin. *Cell*
774 *Adh Migr* **5**: 97–105.
775 Kim SY, Mo JW, Han S, Choi SY, Han SB, Moon BH, Rhyu IJ, Sun W, Kim H. 2010. The
776 expression of non-clustered protocadherins in adult rat hippocampal formation and the
777 connecting brain regions. *Neuroscience* **170**: 189–199.
778 Kluen E, Kuhn S, Kempenaers B, Brommer JE. 2012. A simple cage test captures intrinsic
779 differences in aspects of personality across individuals in a passerine bird. *Anim Behav* **84**:
780 279–287.
781 Kroeker KJ, Kordas RL, Crim R, Hendriks IE, Ramajo L, Singh GS, Duarte CM, Gattuso JP.
782 2013. Impacts of ocean acidification on marine organisms: Quantifying sensitivities and

783 interaction with warming. *Glob Chang Biol* **19**: 1884–1896.
784 Kukekova A V., Johnson JL, Xiang X, Feng S, Liu S, Rando HM, Kharlamova A V., Herbeck Y,
785 Serdyukova NA, Xiong Z, et al. 2018. Red fox genome assembly identifies genomic
786 regions associated with tame and aggressive behaviours. *Nat Ecol Evol* **2**: 1479–1491.
787 Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004.
788 Versatile and open software for comparing large genomes. *Genome Biol* **5**.
789 Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F,
790 Wallace IM, Wilm A, Lopez R, et al. 2007. Clustal W and Clustal X version 2.0.
791 *Bioinformatics* **23**: 2947–2948.
792 Leahy SM, McCormick MI, Mitchell MD, Ferrari MCO. 2011. To fear or to feed: the effects of
793 turbidity on perception of risk by a marine fish. *Biol Lett* **7**: 811–813.
794 Lehmann R, Lightfoot DJ, Schunter C, Michell CT, Ohyanagi H, Mineta K, Foret S, Berumen
795 ML, Miller DJ, Aranda M, et al. 2019. Finding Nemo's Genes: A chromosome-scale
796 reference assembly of the genome of the orange clownfish *Amphiprion percula*. *Mol Ecol*
797 *Resour* **19**: 570–585.
798 Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform.
799 *Bioinformatics* **26**: 589–95.
800 Liu Q, Bhattarai S, Wang N, Sochacka-Marlowe A. 2015. Differential expression of
801 protocadherin-19, protocadherin-17, and cadherin-6 in adult zebrafish brain. *J Comp*
802 *Neurol* **523**: 1419–1442.
803 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for
804 RNA-seq data with DESeq2. *Genome Biol* **15**: 550.
805 Maguire EP, Macpherson T, Swinny JD, Dixon CI, Herd MB, Belelli D, Stephens DN, King SL,
806 Lambert JJ. 2014. Tonic inhibition of accumbal spiny neurons by extrasynaptic α4βδ
807 GABAA receptors modulates the actions of psychostimulants. *J Neurosci* **34**: 823–38.
808 McCarthy C, Carrea A, Diambra L. 2017. Bicodon bias can determine the role of synonymous
809 SNPs in human diseases. *BMC Genomics* **18**: 227.
810 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler
811 D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework
812 for analyzing next-generation DNA sequencing data. *Genome Res* **20**: 1297–303.
813 Mills F, Globa AK, Liu S, Cowan CM, Mobasser M, Phillips AG, Borgland SL, Bamji SX. 2017.
814 Cadherins mediate cocaine-induced synaptic plasticity and behavioral conditioning. *Nat*
815 *Neurosci* **20**: 540–549.

816 Mueller JC, Edelaar P, Carrete M, Serrano D, Potti J, Blas J, Dingemanse NJ, Kempenaers B,
817 Tella JL. 2014. Behaviour-related *DRD4* polymorphisms in invasive bird populations. *Mol*
818 *Ecol* **23**: 2876–2885.
819 Mueller JC, Korsten P, Hermannstaedter C, Feulner T, Dingemanse NJ, Matthysen E, van Oers
820 K, van Overveld T, Patrick SC, Quinn JL, et al. 2013. Haplotype structure, adaptive history
821 and associations with exploratory behaviour of the *DRD4* gene region in four great tit (
822 *Parus major*) populations. *Mol Ecol* **22**: 2797–2809.
823 Munday PL, Dixson DL, McCormick MI, Meekan M, Ferrari MCO, Chivers DP. 2010.
824 Replenishment of fish populations is threatened by ocean acidification. *Proc Natl Acad Sci*
825 **107**: 12930–12934.
826 Munday PL, Jarrold MD, Nagelkerken I. 2019. Ecological effects of elevated CO2 on marine and
827 freshwater fishes: From individual to community effects. In *Fish Physiology*, Vol. 37 of, pp.
828 323–368, Elsevier Inc.
829 Munday PL, McCormick MI, Meekan M, Dixson DL, Watson S-A, Chivers DP, Ferrari MCO.
830 2013a. Selective mortality associated with variation in CO2 tolerance in a marine fish.
831 *Ocean Acidif* **1**: 1–5.
832 Munday PL, Warner RR, Monro K, Pandolfi JM, Marshall DJ. 2013b. Predicting evolutionary
833 responses to climate change in the sea. *Ecol Lett* **16**: 1488–1500.
834 Nagelkerken I, Munday PL. 2016. Animal behaviour shapes the ecological effects of ocean
835 acidification and warming: Moving from individual to community-level responses. *Glob*
836 *Chang Biol* **22**: 974–989.
837 Nakamura H, Aibara M, Kajitani R, Mrosso HDJ, Mzighani SI, Toyoda A, Itoh T, Okada N,
838 Nikaido M. 2021. Genomic Signatures for Species-Specific Adaptation in Lake Victoria
839 Cichlids Derived from Large-Scale Standing Genetic Variation. *Mol Biol Evol* **38**: 3111–
840 3125.
841 Nilsson GE, Dixson DL, Domenici P, McCormick MI, Sørensen C, Watson S-A, Munday PL.
842 2012. Near-future carbon dioxide levels alter fish behaviour by interfering with
843 neurotransmitter function. *Nat Clim Chang* **2**: 201–204.
844 Nishino T, Tamada K, Maeda A, Abe T, Kiyonari H, Funahashi Y, Kaibuchi K, Takumi T, Konishi
845 H. 2019. Behavioral analysis in mice deficient for GAREM2 (Grb2-associated regulator of
846 Erk/MAPK subtype2) that is a subtype of highly expressing in the brain. *Mol Brain* **12**: 94.
847 Opazo JC, Zavala K, Miranda-Rottmann S, Araya R. 2018. Evolution of dopamine receptors:
848 phylogenetic evidence suggests a later origin of the DRD $_{2l}$ and DRD $_{4rs}$ dopamine receptor

849 gene lineages. *PeerJ* **6**: e4593.
850 Ou M, Hamilton TJ, Eom J, Lyall EM, Gallup J, Jiang A, Lee J, Close DA, Yun SS, Brauner CJ.
851 2015. Responses of pink salmon to CO2-induced aquatic acidification. *Nat Clim Chang* **5**:
852 950–957.
853 Parker MO, Brock AJ, Walton RT, Brennan CH. 2013. The role of zebrafish (Danio rerio) in
854 dissecting the genetics and neural circuits of executive function. *Front Neural Circuits* **7**:
855 63.
856 Paula JR, Repolho T, Pegado MR, Thörnqvist PO, Bispo R, Winberg S, Munday PL, Rosa R.
857 2019. Neurobiological and behavioural responses of cleaning mutualisms to ocean
858 warming and acidification. *Sci Rep* **9**: 1–10.
859 Peichel CL, Sullivan ST, Liachko I, White MA. 2017. Improvement of the Threespine Stickleback
860 genome using a Hi-C-Based proximity-guided assembly. *J Hered* **108**: 693–700.
861 Poloczanska ES, Brown CJ, Sydeman WJ, Kiessling W, Schoeman DS, Moore PJ, Brander K,
862 Bruno JF, Buckley LB, Burrows MT, et al. 2013. Global imprint of climate change on marine
863 life. *Nat Clim Chang 2013 310* **3**: 919–925.
864 Porteus CS, Hubbard PC, Uren Webster TM, van Aerle R, Canário AVM, Santos EM, Wilson
865 RW. 2018. Near-future CO2 levels impair the olfactory system of a marine fish. *Nat Clim*
866 *Chang* **8**: 737–743.
867 Razgour O, Forester B, Taggart JB, Bekaert M, Juste J, Ibáñez C, Puechmaille SJ, Novella-
868 Fernandez R, Alberdi A, Manel S. 2019. Considering adaptive genetic variation in climate
869 change vulnerability assessment reduces species range loss projections. *Proc Natl Acad*
870 *Sci U S A* **116**: 10418–10423.
871 Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll
872 SA, Gaudet R, et al. 2007. Genome-wide detection and characterization of positive
873 selection in human populations. *Nature* **449**: 913–918.
874 Sánchez-Soto M, Yano H, Cai N-S, Casadó-Anguera V, Moreno E, Casadó V, Ferré S. 2018.
875 Revisiting the Functional Role of Dopamine D4 Receptor Gene Polymorphisms:
876 Heteromerization-Dependent Gain of Function of the D4.7 Receptor Variant. *Mol Neurobiol*
877 1–8.
878 Saumweber T, Weyhersmüller A, Hallermann S, Diegelmann S, Michels B, Bucher D, Funk N,
879 Reisch D, Krohne G, Wegener S, et al. 2011. Behavioral and synaptic plasticity are
880 impaired upon lack of the synaptic protein SAP47. *J Neurosci* **31**: 3508–3518.
881 Schunter C, Ravasi T, Munday PL, Nilsson GE. 2019. Neural effects of elevated CO2 in fishmay

882 be amplified by a vicious cycle. *Conserv Physiol* **7**.
883 Schunter C, Welch MJ, Nilsson GE, Rummer JL, Munday PL, Ravasi T. 2018. An interplay
884 between plasticity and parental phenotype determines impacts of ocean acidification on a
885 reef fish. *Nat Ecol Evol* **2**: 334–342.
886 Schunter C, Welch MJ, Ryu T, Zhang H, Berumen ML, Nilsson GE, Munday PL, Ravasi T.
887 2016. Molecular signatures of transgenerational response to ocean acidification in a
888 species of reef fish. *Nat Clim Chang*.
889 Seong E, Yuan L, Arikkath J. 2015. Cadherins and catenins in dendrite and synapse
890 morphogenesis. *Cell Adh Migr* **9**: 202–13.
891 Shontz EC, Souders CL, Schmidt JT, Martyniuk CJ. 2018. Domperidone upregulates dopamine
892 receptor expression and stimulates locomotor activity in larval zebrafish (*Danio rerio*).
893 *Genes, Brain Behav* **17**: e12460.
894 Smit AFA, Hubley R. 2008. RepeatModeler Open-1.0.
895 Smit AFA, Hubley R, Green P. 2010. RepeatMasker Open-4.0.
896 Soorni A, Haak D, Zaitlin D, Bombarely A. 2017. Organelle_PBA, a pipeline for assembling
897 chloroplast and mitochondrial genomes from PacBio DNA sequencing data. *BMC*
898 *Genomics* **18**: 49.
899 Stager M, Senner NR, Swanson DL, Carling MD, Eddy DK, Greives TJ, Cheviron ZA. 2021.
900 Temperature heterogeneity correlates with intraspecific variation in physiological flexibility
901 in a small endotherm. *Nat Commun 2021 121* **12**: 1–11.
902 Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with
903 thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.
904 Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: *ab initio*
905 prediction of alternative transcripts. *Nucleic Acids Res* **34**: W435–W439.
906 Steinbiss S, Willhoeft U, Gremme G, Kurtz S. 2009. Fine-grained annotation and classification
907 of *de novo* predicted LTR retrotransposons. *Nucleic Acids Res* **37**: 7002–13.
908 Szpiech ZA, Hernandez RD. 2014. selscan: An Efficient Multithreaded Program to Perform
909 EHH-Based Scans for Positive Selection. *Mol Biol Evol* **31**: 2824–2827.
910 Szpiech ZA, Novak TE, Bailey NP, Stevison LS. 2020. High-altitude adaptation in rhesus
911 macaques. *bioRxiv Evol Biol* 2020.05.19.104380.
912 Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. 2015. Sambamba: fast processing of NGS
913 alignment formats. *Bioinformatics* **31**: 2032–2034.
914 Tashiro K, Tsunematsu T, Okubo H, Ohta T, Sano E, Yamauchi E, Taniguchi H, Konishi H.

2009. GAREM, a novel adaptor protein for growth factor receptor-bound protein 2, contributes to cellular transformation through the activation of extracellular signal-regulated kinase signaling. *J Biol Chem* **284**: 20206–20214.

Termignoni-Garcia F, Kirchman JJ, Clark J, Edwards S V. 2022. Comparative Population Genomics of Cryptic Speciation and Adaptive Divergence in Bicknell's and Gray-Cheeked Thrushes (Aves: Catharus bicknelli and Catharus minimus). *Genome Biol Evol* **14**.

Thirumalai V, Cline HT. 2008. Endogenous Dopamine Suppresses Initiation of Swimming in Prefeeding Zebrafish Larvae. *J Neurophysiol* **100**: 1635–1648.

Thomas JT, Munday PL, Watson SA. 2020. Toward a Mechanistic Understanding of Marine Invertebrate Behavior at Elevated CO2. *Front Mar Sci* **7**: 345.

Thörnqvist P-O, McCarrick S, Ericsson M, Roman E, Winberg S. 2019. Bold zebrafish (Danio rerio) express higher levels of delta opioid and dopamine D2 receptors in the brain compared to shy fish. *Behav Brain Res* **359**: 927–934.

Tran S, Nowicki M, Muraleetharan A, Gerlai R. 2015. Differential effects of dopamine D1 and D2/3 receptor antagonism on motor responses. *Psychopharmacology (Berl)* **232**: 795–806.

Tsai NP, Huber KM. 2017. Protocadherins and the Social Brain. *Biol Psychiatry* **81**: 173–174.

Urban MC. 2013. Evolution mediates the effects of apex predation on aquatic food webs. *Proc R Soc B Biol Sci* **280**.

Von Collenberg CR, Schmitt D, Rülicke T, Sendtner M, Blum R, Buchner E. 2019. An essential role of the mouse synapse-associated protein Syap1 in circuits for spontaneous motor activity and rotarod balance. *Biol Open* **8**.

Waldvogel A, Feldmeyer B, Rolshausen G, Exposito-Alonso M, Rellstab C, Kofler R, Mock T, Schmid K, Schmitt I, Bataillon T, et al. 2020. Evolutionary genomics can improve prediction of species' responses to climate change. *Evol Lett* **4**.

Wang T, Wang Y. 2020. Behavioral responses to ocean acidification in marine invertebrates: new insights and future directions. *J Oceanol Limnol* **38**: 759–772.

Welch MJ, Munday PL. 2017. Heritability of behavioural tolerance to high CO_2 in a coral reef fish is masked by nonadaptive phenotypic plasticity. *Evol Appl* **10**: 682–693.

Welch MJ, Watson S-A, Welsh JQ, McCormick MI, Munday PL. 2014. Effects of elevated CO2 on fish behaviour undiminished by transgenerational acclimation. *Nat Clim Chang* **4**: 1086–1089.

Williams CR, Dittman AH, McElhany P, Busch DS, Maher MT, Bammler TK, MacDonald JW, Gallagher EP. 2019. Elevated CO_2 impairs olfactory-mediated neural and behavioral

948 responses and gene expression in ocean-phase coho salmon (*Oncorhynchus kisutch*).
949 *Glob Chang Biol* **25**: 963–977.
950 Wittmann AC, Pörtner HO. 2013. Sensitivities of extant animal taxa to ocean acidification. *Nat*
951 *Clim Chang* **3**: 995–1001.
952 Woods AS. 2010. The dopamine D(4) receptor, the ultimate disordered protein. *J Recept Signal*
953 *Transduct Res* **30**: 331–6.
954 Yang Z, Sun F, Liao H, Zhang Z, Dou Z, Xing Q, Hu J, Huang X, Bao Z. 2021. Genome-wide
955 association study reveals genetic variations associated with ocean acidification resilience
956 in Yesso scallop Patinopecten yessoensis. *Aquat Toxicol* 105963.
957