Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Decoding the constraints acting on a coastal fish using landscape transcriptomics

Emma Gairin

emma.gairin@hotmail.fr

Okinawa Institute of Science and Technology Graduate University https://orcid.org/0000-0001-6898-8084

Saori Miura

Okinawa Institute of Science and Technology Graduate University

Zoé Chamot

Okinawa Institute of Science and Technology Graduate University

Camille Sautereau

Okinawa Institute of Science and Technology Graduate University

Jann Zwahlen

Okinawa Institute of Science and Technology Graduate University

Hiroki Takamiyagi

Okinawa Institute of Science and Technology Graduate University

Yann Gibert

Indiana University School of Medicine

Marcela Herrera

Okinawa Institute of Science and Technology Graduate University

Vincent Laudet

Okinawa Institute of Science and Technology

Article

Keywords: Landscape transcriptomics, Coastal fish, Nursery, Damselfish, Environmental monitoring

Posted Date: November 25th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7230643/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Version of Record: A version of this preprint was published at Nature Communications on October 9th, 2026. See the published version at https://doi.org/10.1038/s41467-026-77751-2.

Decoding the constraints acting on a coastal fish using landscape transcriptomics

Emma Gairin[1*], Saori Miura[1], Zoé Chamot[1†], Camille A. Sautereau[2,3], Jann Zwahlen[1], Hiroki Takamiyagi[1], Yann Gibert[4], Marcela Herrera[1‡], Vincent Laudet[1,3,5]

[1] Marine Eco-Evo-Devo Unit, Okinawa Institute of Science and Technology, 904-0495 Onna-son, Okinawa, Japan.

[2] Sorbonne Université, CNRS, Biologie Intégrative des Organismes Marins, BIOM, Observatoire Océanologique, 66650 Banyuls-sur-Mer, France

[3] CNRS IRL 2028 "Eco-Evo-Devo of Coral Reef Fish Life Cycle" (EARLY)

[4] Department of Medical and Molecular Genetics, Indiana University School of Medicine, 975 W. Walnut Street IB 454C, Indianapolis, IN 46202, USA

[5] Marine Research Station, Institute of Cellular and Organismic Biology (ICOB), Academia Sinica, 23-10, Dah-Uen Rd, Jiaoxi, Yilan 262, Taiwan

†: Present address: Department of Biology and Marine Program, Boston University, Boston, MA, USA

‡: Present address: Institut des Sciences de l'Evolution de Montpellier, ISEM, University of Montpellier, CNRS, IRD, EPHE, 34095 Montpellier, France

*Corresponding author. Email: emma.gairin@oist.jp

Keywords

Landscape transcriptomics; Coastal fish; Nursery; Damselfish; Environmental monitoring

Abstract

Understanding how wild organisms respond to complex mixtures of environmental stressors remains a major challenge in ecology[1,2]. To move beyond laboratory-based studies of the effect of individual stressors on model species, landscape transcriptomics – the study of how genome-wide expression patterns reveal environmental conditions – has been suggested[3]. Here, we apply landscape transcriptomics to the damselfish *Chrysiptera cyanea*, a sentinel species from Okinawa's coastal ecosystems, by sampling juveniles and adults across 18 sites from natural to urbanized habitats. We uncover a striking molecular signature of urbanization: fish from human-impacted sites exhibit elevated expression of genes involved in inflammation and immune responses. Remarkably, their transcriptomes resemble those of overfed individuals from laboratory feeding experiments – suggesting that urban reefs act as "fast food" environments, offering energy-rich diets while imposing chronic stress. This pattern is illustrated by the altered expression of key genes in metabolic and stress pathways (*e.g.*, downregulation of *ppargc1a* and *insrb*, indicative of dietary intake surplus and insulin resistance; upregulation of interleukins, chemokines, toll-like receptors and caspase linked to the immune response and inflammation). Our results demonstrate that wild transcriptomes capture the invisible biological costs of environmental degradation, positioning fish as living biosensors for coastal ecosystem health.

Main

Human activities alter natural habitats through local and global stressors[1], yet current methods fall short of detecting their compounded consequences in organisms. Researchers commonly rely on parallel approaches: monitoring changes in environmental parameters, such as water quality or biological communities[4–6] and exposing model organisms to a single or few stressors in controlled laboratory conditions to understand their impacts[7,8]. Despite their usefulness, these approaches often fail to capture the full effect of the cocktail of stressors in natural environments. With the rise of high-throughput sequencing, landscape transcriptomics – the study of how genome-wide expression patterns reflect dynamic, landscape-scale environmental drivers – has been proposed as a comprehensive method to detect multiple stressors[3,9]. Unlike targeted approaches, transcriptomics provides unbiased and hypothesis-free gene expression data[3,10–12]. Importantly, variations in gene expression can precede physiological changes, serving as an early warning signal for later-stage ecological damage such as disease outbreaks and population declines[12]. Moreover, transcriptomics can provide sensitive mechanistic insights about the consequences of environmental change[13]. Lastly, by focusing on conserved genes and pathways, transcriptomic responses can ultimately be compared across diverse taxa, enabling the construction of a common conceptual framework[14]. Landscape transcriptomics therefore shows promise as a universal tool to monitor environmental change both in wildlife and human populations. However, the application of transcriptomics at ecological scales remains in its infancy and awaits broader empirical validation.

To test the value of landscape transcriptomics under field conditions, we selected the blue damselfish *Chrysiptera cyanea,* a territorial reef species which, through its presence in a variety of shallow coastal environments, provides access to a gradient of urbanisation on the subtropical island of Okinawa, Japan[15]. These coastal zones offer contrasting conditions for

68 reef fish; after a period of larval dispersal in the open ocean, different species may prefer
69 specific habitat features – such as higher nutrient availability or stable temperatures – when
70 settling on a reef[16,17]. As such, natural variability influences fish growth, survival, and
71 potentially gene expression[18]. However, many coastal ecosystems globally are degraded due to
72 human activities[19,20]. In Okinawa, the northern part of the island remains relatively pristine,
73 while more than a million people live within the southern half of the 1,200 km^2 island, with
74 housing, industries, and agricultural activities taking place along the shore[21,22]. This is an ideal
75 setting to study the relationship between human presence, coastal environments, and the
76 organisms inhabiting them.

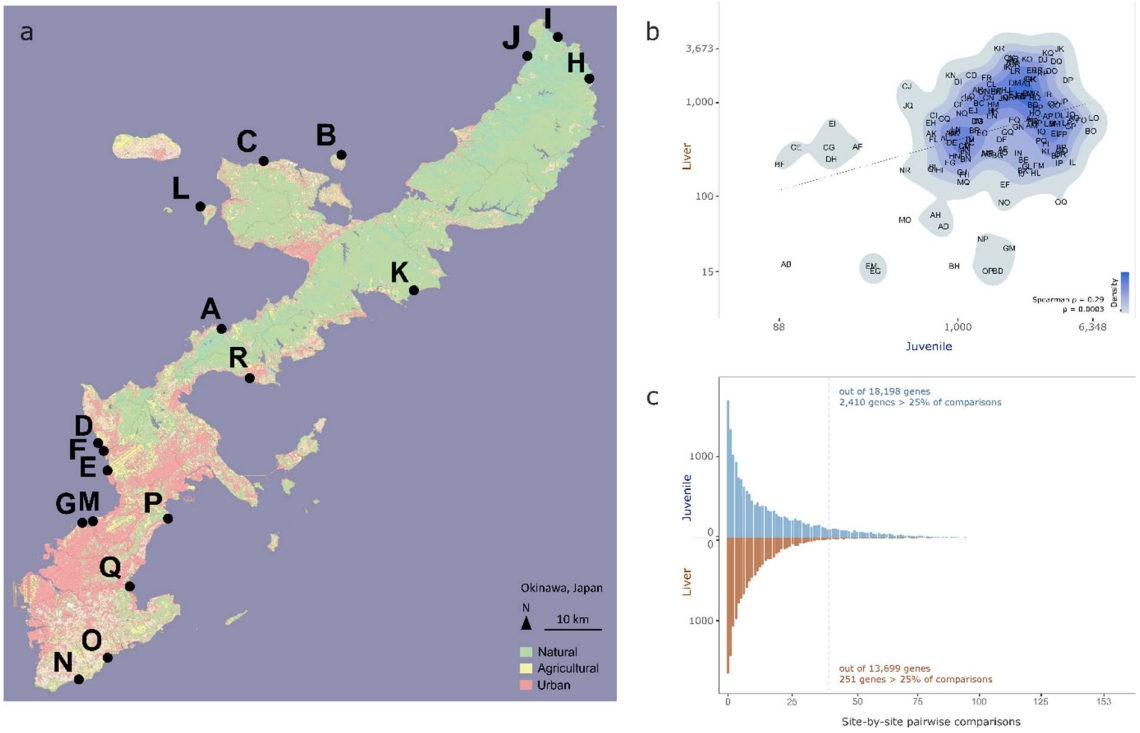
77 In this study, we predicted that urbanisation leads to detectable shifts in gene expression
78 in juvenile and adult *C. cyanea*. Gene expression is however known to be influenced by factors
79 such as sampling time, season, developmental stage, or reproductive status, which have been
80 suggested to obscure patterns of interest in field samples[3]. To address the co-occurrence of
81 multiple environmental factors, our sampling design drew from strategies developed for
82 experiments with multiple factors. Indeed, the "combinatorial explosion" frequently
83 encountered by experimentalists can be countered by assessing subsets of randomised
84 combinations of factors, which can gather more significant results than experiments
85 constrained to a controllable number of parameters and that may lack a key explanatory
86 factor[7,23,24]. We adopted a semi-randomised sampling approach by targeting 18 sites along a
87 defined gradient of human presence, and collecting five juvenile fish of varying sizes and four
88 to five adult fish of both sexes under a range of environmental conditions (different water
89 temperature, salinity, chlorophyll-α, and sampling time). We focused on the gene expression of
90 the whole body of juveniles to detect organism-wide changes in gene expression, and of the
91 liver in adults to screen long-term consequences of human activities on energy metabolism and
92 detoxifying processes. Using this extensive dataset, we detected the impact of individual and

combinations of parameters on gene expression and found that human presence profoundly influenced immunity, cell cycle regulation, stress response, and energy metabolism. Experimental validation to extract the transcriptomic signature of food availability identified a 'fast food' effect of urbanisation, with adults showing both a lipid-rich diet and physiological stress. Using dimensional analyses, we summarised the relative expression pattern of pre-defined groups of genes to a single value for each sample, which we visualised with radar plots; this novel approach, tailored for landscape transcriptomics, facilitated data analysis and stands as a powerful and simple way to extract a diagnosis about organismal status.

Non-random variability in gene expression captured in transcriptomes

We sampled *Chrysiptera cyanea* across 18 coastal sites in Okinawa, Japan, along a gradient of human presence (from 2 to 94% of natural zones within 1km of each site). As a first approach to detect variations across the sites (Fig. 1a), we found that, across the 28,173 genes in the genome of *C. cyanea*[25], 88 to 6,348 genes significantly differed in expression level in pairwise site comparisons (*e.g.,* site A *vs.* B) for the juvenile whole-body samples, and 15 to 3,763 for the adult livers (Fig. 1b). Interestingly, the number of differentially expressed genes was significantly correlated in juvenile whole-body and adult liver samples for the same pairs of sites (Fig. 1b): there was a coherent effect of local conditions at both life stages. As only one organ was considered for adults, the number of genes and range of physiological responses expressed in the samples was lower than for juvenile whole-body samples. Lastly, juveniles – likely still acclimating to their new environment – showed numerous frequently differentially expressed genes (2,410 genes in >25% of the pairwise site comparisons; Fig. 1c) while adults – which have resided on their sites for months to years – showed site-specific profiles (251 genes retrieved in >25% comparisons); this is indicative of a transition from broad acclimation responses to long-term fine-tuned mechanisms.

Fig. 1: Environmental context and overview of the differential gene expression analysis for juvenile whole body and adult liver samples of *C. cyanea* across 18 coastal sites in Okinawa, Japan.



a, Map of the eighteen field sites in Okinawa, overlaid on land-use data from JAXA (2023). **b**, Number of differentially expressed genes (logarithm-10 scale) in juvenile whole body and adult liver samples resulting from pairwise comparisons of 18 sites (absolute fold-change >1.5, $p < 0.01$). **c,** Number of pairwise site comparisons in which each gene was differentially expressed.

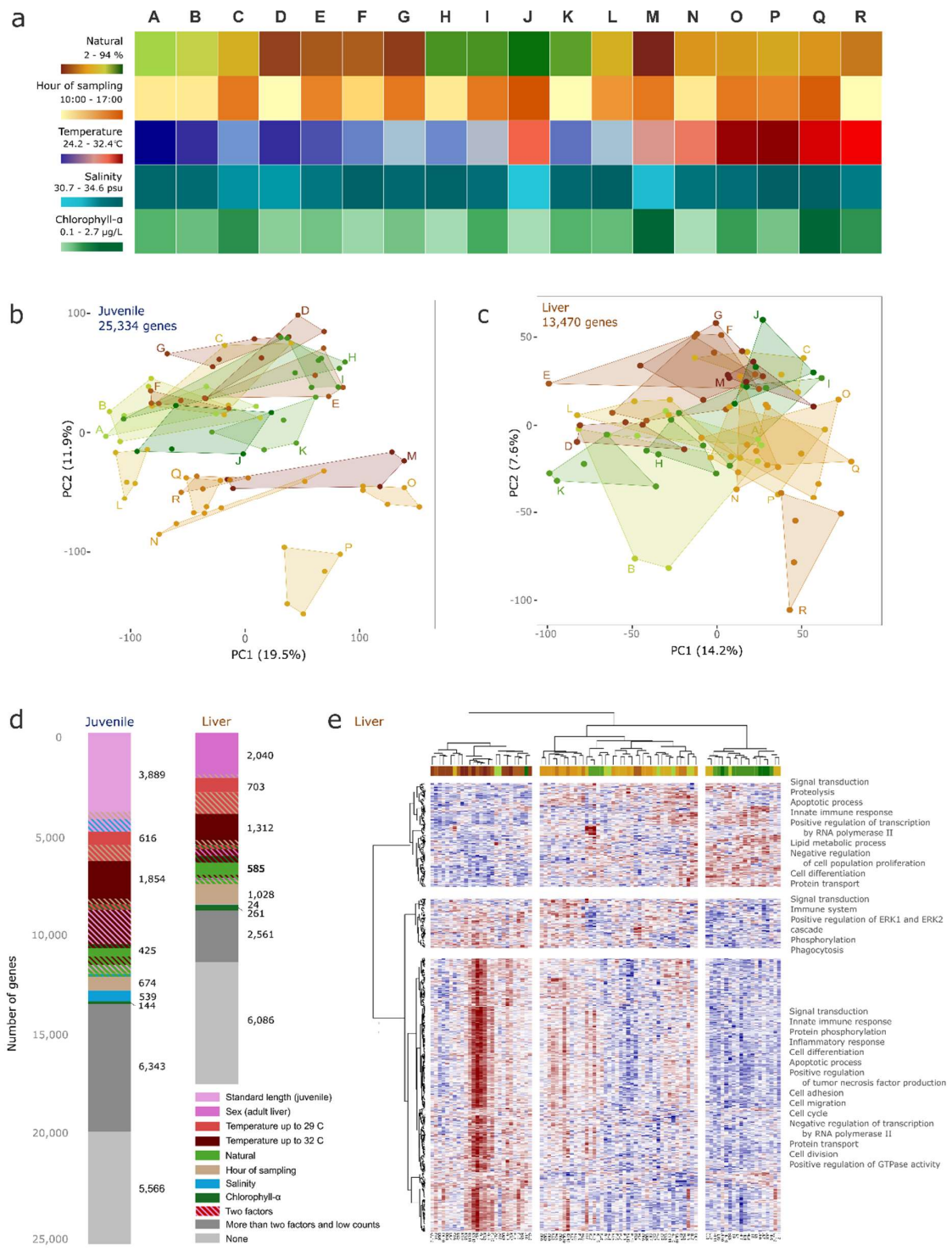
Identifying gene expression drivers through sampling heterogeneity

Dimensional reduction using Principal Component Analyses (PCAs) showed that, although fish size and sex influenced gene expression, samples from the same sites conveyed a coherent gene expression message by clustering closely in response to local environmental drivers (Fig. 2b,c; Extended Data Fig. 1). Taking advantage of our semi-randomised sampling approach, we fitted linear and quadratic models between the expression level of each gene, the

133 percentage of natural zones within a kilometre of the site, environmental factors (temperature,
134 salinity, chlorophyll-α, sampling time), and fish characteristics (juveniles: standard length;
135 adults: sex). A significant relationship with at least one of the parameters was found in >78%
136 of the genes for juvenile whole-body and >65% for adult livers (Fig. 2d). Interestingly, 1,854
137 genes in juvenile whole-body and 1,312 genes in adult livers responded exclusively to
138 temperature, which varied extensively in our sampling. When the four warmest sites (29.6-
139 32.4ºC) were excluded from the fit, the number of temperature-associated genes decreased by
140 almost two-thirds in juvenile whole-body and by half in adult livers (Fig. 2d). Sampling time
141 – due to circadian cycles –, fish size and sex also had a strong footprint on gene expression. A
142 similar proportion of genes in juvenile whole-body and adult liver samples responded to the
143 same factors, highlighting a parallel response of different tissues across developmental stages
144 (Pearson's Chi-squared test, χ^2=3548, df=19, $p<10^{-15}$).

145 We identified 425 genes in juvenile whole-body and 585 genes in adult livers uniquely
146 linked to human activities. These conveyed generalised stress at multiple biological levels, with
147 modifications of the expression level of genes involved in apoptosis, immunity, and signal
148 transduction (genes such as interleukins, chemokines, caspases; KEGG pathways including
149 antigen processing and presentation, virus infection, T cell, toll-like, and NOD-like receptor
150 signalling, chemokine and NF-kappa B signalling). We also found a decrease in the expression
151 of genes related to cell differentiation, tissue homeostasis, protein transport and proteolysis,
152 and transcription in favour of inflammation, carcinogenesis, reactive oxygen species, protein
153 phosphorylation, and cell cycle regulation near human activities (Fig. 2e). These genes are
154 potential markers of human-related stress for environmental monitoring in *C. cyanea* as they
155 do not respond to other environmental or organismal factors.

156 **Fig. 2: Relationship between the environment and gene expression in *Chrysiptera cyanea*.**



a, Percentage of natural zones within a kilometre of each site, sampling time, temperature, salinity, and chlorophyll-α at each site (full dataset in Supplementary Materials). **b**, Principal Component Analysis of the expression level of all genes from the transcriptomes of juvenile

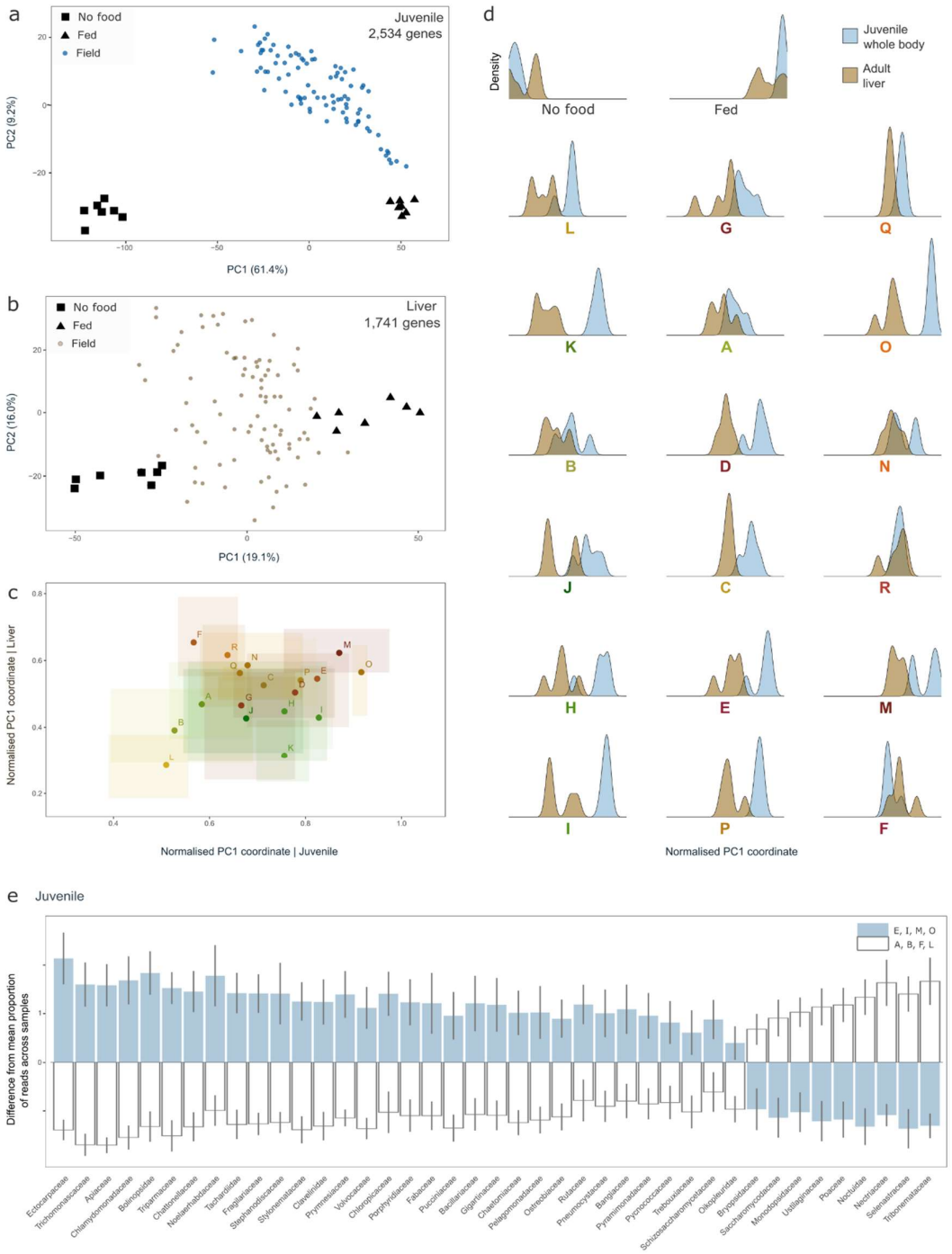
whole-body samples from 18 sites. Colour-coding based on the proportion of natural zones within a kilometre of the site. **c**, Principal Component Analysis of the expression level of selected genes from the transcriptomes of adult liver samples from 18 sites. 3,973 genes responding exclusively to the sex of the adult were removed from the analysis. Colour-coding based on the proportion of natural zones within a kilometre of the site. **d**, Bar plot of the number of genes with a significant linear or quadratic difference in expression level in response to fish characteristics or environmental parameters, from DRomics 2.6-2. **e,** Heatmap of the 585 genes with significant linear or quadratic differences in expression level with respect to the percentage of natural zones within one kilometre in the adult liver samples, and Gene Ontology families corresponding to the 3 main gene clusters, presented in order of prevalence within each gene cluster.

Integrating experimental and field transcriptomes to infer feeding status

Landscape transcriptomes can be assessed through the lens of local parameters (Fig. 2) but can also be compared to reference datasets from controlled conditions[26]. To test the value of combined field and laboratory approaches, we explored the impact of food availability on gene expression in juvenile whole-body and adult liver samples using two conditions: fasted or fed. Following 3-day and 14-day experiments for juveniles and adults respectively (durations tailored to their tolerance to fasting), PCAs showed a clear separation of the fasted and fed transcriptome samples along PC1 (Extended Data Fig. 2). The top 10% of genes driving PC1 in juvenile whole-body and adult liver transcriptomes were related to cell cycle and division, DNA transcription, protein modification, lipid and carbohydrate metabolism, and glucose homeostasis (Extended Data Fig. 3). We used these genes to generate secondary PCAs grouping experimental and field samples (Fig. 3a,b), from which we extracted the feeding status of each field sample using its PC1 coordinate as a proxy.

186 Strikingly, adults showed higher feeding status near urban zones (linear mixed-effect
187 model, site as a random effect, p=0.008; Fig. 3c). Among diet-related genes with a modified
188 expression in livers from urban zones, we found Insulin Receptor b *insrb* and Peroxisome
189 Proliferator Activated Receptor Gamma Coactivator 1 Alpha *ppargc1a*, markers of insulin
190 resistance and energy surplus. Coupled with an upregulated inflammation and immune
191 response (Fig. 2e), this suggests that human activities lead to "fast-food-like" habitats, where
192 macronutrients are plentiful but fish are exposed to immune challenges. A similar trade-off was
193 reported in transcriptomic studies of two bird species, where an upregulation of the immune
194 response was coupled with a lipid-rich diet in urban zones[27,28]. This potential commonality in
195 transcriptome signatures in response to human activities must be explored in more ecosystems
196 and taxa.
197 In juveniles, size was positively linked to feeding status (linear mixed-effect model, site
198 as random effect, $p < 0.001$). Although we found significant differences in PC1 coordinates
199 between 39 pairs of sites (Kruskal-Wallis $\chi^2=65.7$, df=17, $p<10^{-6}$ followed by Dunn test), none
200 of the environmental parameters measured in the study was significant. We thus analysed
201 juvenile diet using a meta-transcriptomic approach; by taxonomically classifying the
202 sequencing reads which did not align to *C. cyanea*'s reference transcriptome, we retrieved
203 significantly more tunicates (*Clavelinidae, Oikopleuridae*), ctenophores (*Bolinopsidae*), brown
204 algae (*Ectocarpaceae*), and red algae (*Gigartinaceae*) reads in juveniles from sites associated
205 with a higher feeding status (Fig. 3e).
206 Lastly, the mean PC1 coordinate was greater (higher feeding status) for juveniles than
207 adults, with the exception of sites F and R (Fig. 3d); the fasted treatment could have had a
208 stronger impact on the juveniles over three days than on the adults over two weeks, leading to
209 a more extreme response in the fasted juvenile group. Alternatively, Okinawa's coastal zones
210 may provide more adequate food to juveniles than to adults.

Fig. 3: Feeding status of juvenile and adult fish from the field through a comparison with a fasted-fed laboratory experiment, and meta-transcriptomic juvenile diet analysis.



a, PCA of the expression level of 2,534 genes in the field and laboratory experiment samples, for juvenile whole-body samples. These 2,534 genes are the top 10% contributors to the PC1

variance of a preliminary PCA using the laboratory samples. **b**, PCA of the expression level of 1,741 genes in the field and laboratory experiment samples, for adult liver samples. These 1,741 genes are the top 10% contributors to the PC1 variance of a preliminary PCA using the laboratory samples. **c**, Mean normalised PC1 coordinate of the juvenile and liver samples from each site (dot) as well as minimum and maximum coordinates (shaded rectangles). The colour coding corresponds to the percentage of natural zones within a kilometre of each field site. **d**, Density spectra of the juvenile and liver normalised PC1 coordinate on each site; sites presented in order of the mean PC1 coordinate for the adult liver samples, from low to high. **e**, Proportion of taxonomically assigned reads in juvenile samples deviating from the mean in the sites with highest (E, I, M, O) and lowest (A, B, F, L) feeding status.

Radar plots link transcriptomes to local stressors

Okinawa's *C. cyanea* transcriptomes revealed extensive differences linked to human presence, environmental parameters, and feeding (Figs. 1,2,3): they can serve as indicators of fish health and local conditions. However, the use of landscape transcriptomics by environmental managers may be impeded by the complexity of grouping and analysing numerous samples, environmental conditions, and biological processes. To tackle this multidimensionality, novel data summarisation and visualisation tools are required. Building upon manually curated lists of genes related to energy metabolism, neurotransmission, endocrine regulation, and stress response[14], we generated PCAs for each gene group, calculated the average coordinates of each site for the top PC dimensions, and identified the axis of maximum variance; we then projected the coordinates of each sample onto this axis and set lower values to correspond to samples with overall lower gene expression (Table S7; Extended Data Fig. 5). This yielded a single value per sample and gene group representative of the position of the sample compared to all others. We then visualised these values on radar plots,

241 enabling the comparison of samples and sites and identification of co-occurring processes and
242 site-specific stress signatures.

243 These radar plot profiles were consistent in fish exposed to similar conditions, and intra-
244 site variability could be attributed to size or sex in many cases (Fig. 4). Fish from similar sites
245 showed closely related profiles (*e.g.,* for juveniles, A, B, and L with medium urbanisation and
246 low food availability, N and R with similar urbanisation and water temperature; for adults, H
247 and I, among the most natural, as well as E, G, and M, among the most urbanised). Genes
248 related to a single and strong environmental driver directly relayed local characteristics, such
249 as heat-shock markers in fish exposed to >31.9ºC (O and P) or the bimodal phototransduction
250 signature in juveniles depending on morning or afternoon sample collection.

251 Transcriptomes are complex datasets, mixing genes that can be snapshots of stress with
252 those conveying messages about long-term conditions. The comparison of juvenile whole-body
253 and adult liver profiles allowed us to discern timescales at which environmental factors affect
254 gene expression. For instance, the two most urban sites, G and M, had similar radars for adult
255 livers but not for juvenile whole-body samples, with differences in corticoids,
256 neurotransmission, and energy metabolism (also see the respective PCAs on Fig. 2b,c). We
257 hypothesise that juveniles from G benefited from its more open reef morphology with frequent
258 water renewal, that can wash away waterborne stressors. In contrast, adults that have been
259 exposed to local conditions for longer durations could have integrated chronic stress,
260 particularly in terms of immune response and neurotransmission. Landscape transcriptomes
261 can thus yield time-integrated differences across environments that may be deemed similar,
262 thereby making the 'invisible' visible.
263

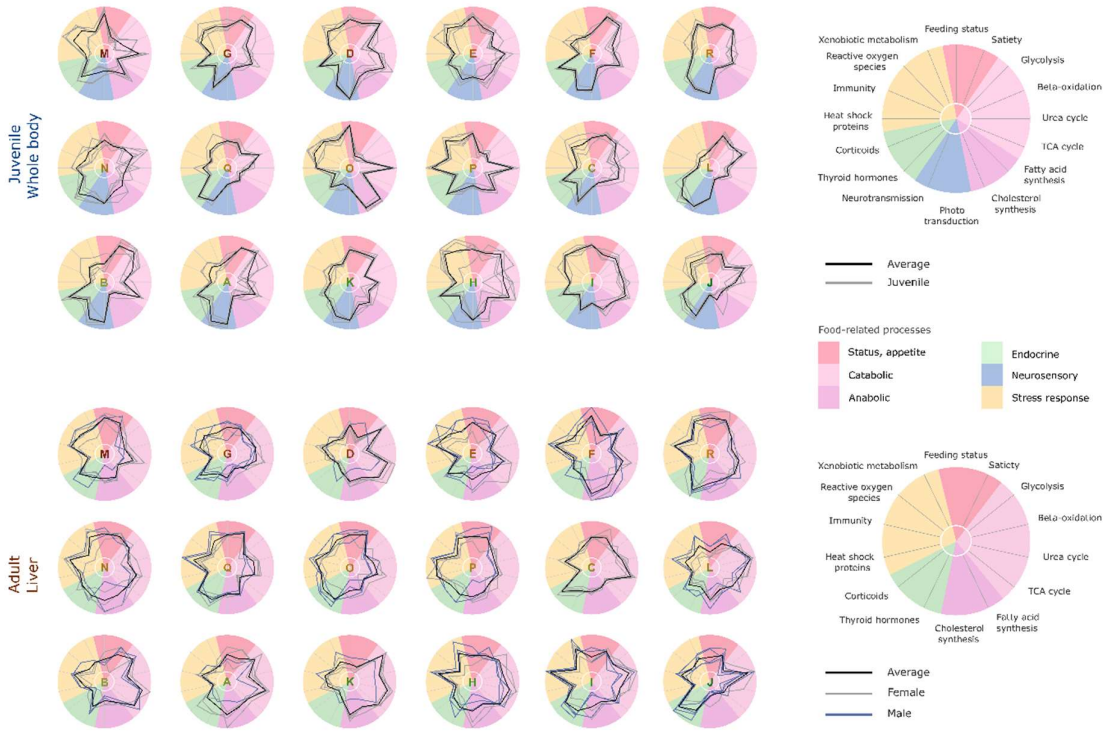


Fig. 4: Radar plots presenting the status of juvenile whole-body (top) and adult liver (bottom) across key biological processes. The coordinates were calculated through an orthogonal projection of each sample onto the main axis of site-wise variance based on the top dimensions representing at least 50% of the variance of PCA performed for the following subsets of genes: satiety, glycolysis, beta-oxidation, urea cycle, TCA cycle, fatty acid synthesis, cholesterol synthesis, neurotransmission (juvenile only), photoreception (juvenile only), thyroid hormones, corticoids, heat shock proteins, immunity, reactive oxygen species, xenobiotic metabolism. The feeding status of each sample, as derived from the analysis from Fig. 3, is also presented.

Divergent strategies detected in organisms across environments

Grouping multiple biological processes in the same visualisation enabled the detection of distinct strategies depending on developmental stage, environmental conditions, and feeding. For instance, well-fed juveniles suppressed satiety-signalling genes; this can boost food intake

and thus growth (Extended Data Fig. 6). Similarly, energy metabolism shifted from catabolic to anabolic processes with increasing feeding status: juveniles and adults with low-food status (*e.g.,* juveniles from A, B, F, L; adults from A, B, C, H, I, J; Fig. 3) privileged glycolysis, which allows for efficient energy conversion from sugars; in some sites, β-oxidation, which can use fatty reserves to maintain energy homeostasis, was also upregulated (juveniles from B, L; adults from C, H, I, J). Inversely, well-fed fish showed high fatty acid and cholesterol synthesis (highest in O in juveniles, also in D, E, H, I, K, M, N, P; for adults, in F, N, R). A switch in energy investment strategy was detected in high-stress environments: well-fed adult fish showed low metabolism-related gene expression in a trade-off for high immune response in M, and with thermal stress in O and P. We therefore show that landscape transcriptome data provide a rich framework for testing the ecological relevance of physiological regulations identified in experimental animal models.

Typical pollution marker genes show limited reliability in Okinawa's *C. cyanea*

As a last step in our elucidation of environmental and human impacts on fish transcriptomes, we assessed common marker genes of pollution, including immunity, endocrine processes, xenobiotic metabolism (involved in the removal of chemical pollutants), and neurotransmission. While we found a close association between immune response genes and human presence (Fig. 2e, 4), most marker genes responded to other drivers.

Starting with endocrine processes, we identified an upregulation of nuclear thyroid hormone receptors under high temperature, overriding any other environmental driver; this is coherent with previous data from coral reef fish[29]. Other thyroid and corticoid-related genes also responded mostly to temperature, although, for juveniles, *dio3b* – which acts by deactivating thyroid hormones in target tissues – and genes in the neuroendocrine cascade leading to cortisol release, notably corticotropin-releasing hormone receptors, pro-

opiomelanocortin B, and ferredoxin, were upregulated under low food conditions (Extended Data Figs. 7, 8). Interestingly, *hsd11b1*, a gene involved in the conversion of the inactive cortisone to the active cortisol, was upregulated in juveniles from the natural sites H and I. This indicates an acute stress response, potentially linked to the strain of capture on the fish, which may be dampened elsewhere due to chronic stress.

We found feeding status, rather than pollution, to be linked to xenobiotic metabolism at both stages (Fig. 4; Extended Data Fig. 9). This could be because many of these genes, involved in the modification, conjugation, and excretion of compounds, have a pleiotropic role and intervene in endogenous processes dependent on food-derived energy. Lastly, neurotransmission and immune defence mechanisms, which are energetically costly[30,31], were suppressed under low food conditions in juveniles (Extended Data Fig. 9).

The tolerance of *C. cyanea* to a large range of conditions and its ubiquitous presence along Okinawa's coastline allowed us to extract environmental signals in juvenile whole-body and adult liver transcriptomes, revealing intricate patterns of gene expression depending on local conditions, developmental stage, tissue, sex and size. We identified which genes and biological processes respond to human activities, and importantly, which do not. While thermal stress and food availability could be confounding factors, the absence of clear signal from 'classic' pollution marker genes may also be due to the choice of species: *C. cyanea* are omnivorous and planktivorous fish near the base of the trophic food web, and whilst territorial and living in shallow waters, they are not benthic fish. They might therefore not bioaccumulate toxicants to levels leading to a significant transcriptomic signature of chemical stress in typical marker genes, notably those involved in xenobiotic metabolism. Comparing the signals obtained with ubiquitous species higher in the food chain or with benthic feeders could be of interest.

Discussion and outlook

Our study of the localised gene expression response of a common fish species in Okinawa is a step towards building a generalisable framework for environmental state awareness, or 'environmental phenotyping', using transcriptomes, with implications for global ecosystem monitoring. In the case of *C. cyanea*, further studies will be required to further elucidate temporal and seasonal variability. In addition, the analysis of multiple tissues at different developmental stages would provide insight into more biological processes as well as help to identify carry-over effects throughout development. Future studies may benefit from detailed environmental characterisations, including long-term water quality data series and measurements of chemical pollution. Lastly, generating additional laboratory-based transcriptomes (*e.g.,* temperature, salinity, chemical compounds) could be an efficient method to isolate the footprint of different stressors in the field.

Landscape transcriptomics provided an unparalleled depth of information about the status of the blue damselfish *C. cyanea* on Okinawa's coastal zones. Using semi-randomised sampling and complementary laboratory experiments, we identified the combined effects of environmental parameters and the human footprint on transcriptomes and found biological trade-offs in the choice of a habitat. With sufficient genomic resources and environmental information for baseline analyses, our methodology can be generalised to sentinel species across any ecosystem and holds huge potential for environmental monitoring.

References

1. Gissi, E. *et al.* A review of the combined effects of climate change and other local human stressors on the marine environment. *Science of The Total Environment* **755**, 142564 (2021).
2. Pirotta, E. *et al.* Understanding the combined effects of multiple stressors: A new perspective on a longstanding challenge. *Science of The Total Environment* **821**, 153322 (2022).
3. Keagy, J. *et al.* Landscape transcriptomics as a tool for addressing global change effects across diverse species. *Molecular Ecology Resources* **0**, 1–16 (2023).
4. Forio, M. A. E. & Goethals, P. L. M. An Integrated Approach of Multi-Community Monitoring and Assessment of Aquatic Ecosystems to Support Sustainable Development. *Sustainability* **12**, 5603 (2020).
5. Karydis, M. & Kitsiou, D. Marine water quality monitoring: A review. *Marine Pollution Bulletin* **77**, 23–36 (2013).
6. Martinez-Haro, M. *et al.* A review on the ecological quality status assessment in aquatic systems using community based indicators and ecotoxicological tools: what might be the added value of their combination? *Ecological Indicators* **48**, 8–16 (2015).
7. Rillig, M. C. *et al.* The role of multiple global change factors in driving soil functions and microbial biodiversity. *Science* **366**, 886–890 (2019).
8. Walker, C. H., Sibly, R. M., Sibly, R. M. & Peakall, D. B. *Principles of Ecotoxicology*. (CRC Press, Boca Raton, 2016). doi:10.1201/b11767.
9. Martyniuk, C. J. Are we closer to the vision? A proposed framework for incorporating omics into environmental assessments. *Environmental Toxicology and Pharmacology* **59**, 87–93 (2018).

10. Evans, T. G. Considerations for the use of transcriptomics in identifying the 'genes that matter' for environmental adaptation. *Journal of Experimental Biology* **218**, 1925–1935 (2015).
11. Johnson, K. J. *et al.* A Transformative Vision for an Omics-Based Regulatory Chemical Testing Paradigm. *Toxicological Sciences* **190**, 127–132 (2022).
12. Page, T. M. & Lawley, J. W. The Next Generation Is Here: A Review of Transcriptomic Approaches in Marine Ecology. *Front. Mar. Sci.* **9**, (2022).
13. Snape, J. R., Maund, S. J., Pickford, D. B. & Hutchinson, T. H. Ecotoxicogenomics: the challenge of integrating genomics into aquatic and terrestrial ecotoxicology. *Aquatic Toxicology* **67**, 143–154 (2004).
14. Herrera, M. *et al.* From Genes to Pathways: A Curated Gene Approach to Accurate Pathway Reconstruction in Teleost Fish Transcriptomics. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **n/a**, (2025).
15. Lecchini, D., Adjeroud, M., Pratchett, M. S., Cadoret, L. & Galzin, R. Spatial structure of coral reef fish communities in the Ryukyu Islands, southern Japan. *Oceanologica Acta* **26**, 537–547 (2003).
16. Komyakova, V. & Swearer, S. E. Contrasting patterns in habitat selection and recruitment of temperate reef fishes among natural and artificial reefs. *Marine Environmental Research* **143**, 71–81 (2019).
17. Pratchett, M. S., Coker, D. J., Jones, G. P. & Munday, P. L. Specialization in habitat use by coral reef damselfishes and their susceptibility to habitat loss. *Ecology and Evolution* **2**, 2168–2180 (2012).
18. Evans, T. G. & Hofmann, G. E. Defining the limits of physiological plasticity: how gene expression can assess and predict the consequences of ocean change. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 1733–1745 (2012).

19. Halpern, B. S. *et al.* Recent pace of change in human impact on the world's ocean. *Sci Rep* **9**, 11609 (2019).
20. Halpern, B. S. *et al.* A Global Map of Human Impact on Marine Ecosystems. *Science* **319**, 948–952 (2008).
21. Japan Aerospace Exploration Agency (JAXA). High-Resolution Land Use and Land Cover Map of Okinawa Island. (2023).
22. Japan Statistics Bureau. *POPULATION AND HOUSEHOLDS OF JAPAN 2020*. https://www.stat.go.jp/english/data/kokusei/2020/summary.html (2020).
23. Katzir, I., Cokol, M., Aldridge, B. B. & Alon, U. Prediction of ultra-high-order antibiotic combinations based on pairwise interactions. *PLOS Computational Biology* **15**, e1006774 (2019).
24. Lundstedt, T. *et al.* Experimental design and optimization. *Chemometrics and Intelligent Laboratory Systems* **42**, 3–40 (1998).
25. Gairin, E. *et al.* The genome of the sapphire damselfish Chrysiptera cyanea: a new resource to support further investigation of the evolution of Pomacentrids. *Gigabyte* **2024**, 0–0 (2024).
26. Oomen, R. A. & Hutchings, J. A. Transcriptomic responses to environmental change in fishes: Insights from RNA sequencing. *FACETS* **2**, 610–641 (2017).
27. Watson, H., Videvall, E., Andersson, M. N. & Isaksson, C. Transcriptome analysis of a wild bird reveals physiological responses to the urban environment. *Sci Rep* **7**, 44180 (2017).
28. Damiani, G., Sebastiano, M., Dell'Omo, G. & Costantini, D. Blood transcriptome analysis of common kestrel nestlings living in urban and non-urban environments. *Science of The Total Environment* **928**, 172585 (2024).

29. Besson, M. *et al.* Anthropogenic stressors impact fish sensory development and survival via thyroid disruption. *Nat Commun* **11**, 3614 (2020).
30. Bajgar, A. *et al.* Extracellular Adenosine Mediates a Systemic Metabolic Switch during Immune Response. *PLOS Biology* **13**, e1002135 (2015).
31. Adamo, S. The Integrated Defense System: Optimizing Defense against Predators, Pathogens, and Poisons. *Integr Comp Biol* **62**, 1536–1546 (2022).
32. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* **428**, 726–731 (2016).

Figure legends

Fig. 1: Environmental context and overview of the differential gene expression analysis for juvenile whole body and adult liver samples of *C. cyanea* across 18 coastal sites in Okinawa, Japan. **a**, Map of the eighteen field sites in Okinawa, overlaid on land-use data from JAXA (2023). **b**, Number of differentially expressed genes (logarithm-10 scale) in juvenile whole body and adult liver samples resulting from pairwise comparisons of 18 sites (absolute fold-change >1.5, $p < 0.01$). **c,** Number of pairwise site comparisons in which each gene was differentially expressed.

Fig. 2: Relationship between the environment and gene expression in *Chrysiptera cyanea*. **a**, Percentage of natural zones within a kilometre of each site, sampling time, temperature, salinity, and chlorophyll-α at each site (full dataset in Supplementary Materials). **b**, Principal Component Analysis of the expression level of all genes from the transcriptomes of juvenile whole-body samples from 18 sites. Colour-coding based on the proportion of natural zones within a kilometre of the site. **c**, Principal Component Analysis of the expression level of

445 selected genes from the transcriptomes of adult liver samples from 18 sites. 3,973 genes
446 responding exclusively to the sex of the adult were removed from the analysis. Colour-coding
447 based on the proportion of natural zones within a kilometre of the site. **d**, Bar plot of the number
448 of genes with a significant linear or quadratic difference in expression level in response to fish
449 characteristics or environmental parameters, from DRomics 2.6-2. **e,** Heatmap of the 585 genes
450 with significant linear or quadratic differences in expression level with respect to the
451 percentage of natural zones within one kilometre in the adult liver samples, and Gene Ontology
452 families corresponding to the 3 main gene clusters, presented in order of prevalence within
453 each gene cluster.
454
455 Fig. 3: Feeding status of juvenile and adult fish from the field through a comparison with a
456 fasted-fed laboratory experiment, and meta-transcriptomic juvenile diet analysis. **a**, PCA of the
457 expression level of 2,534 genes in the field and laboratory experiment samples, for juvenile
458 whole-body samples. These 2,534 genes are the top 10% contributors to the PC1 variance of a
459 preliminary PCA using the laboratory samples. **b**, PCA of the expression level of 1,741 genes
460 in the field and laboratory experiment samples, for adult liver samples. These 1,741 genes are
461 the top 10% contributors to the PC1 variance of a preliminary PCA using the laboratory
462 samples. **c**, Mean normalised PC1 coordinate of the juvenile and liver samples from each site
463 (dot) as well as minimum and maximum coordinates (shaded rectangles). The colour coding
464 corresponds to the percentage of natural zones within a kilometre of each field site. **d**, Density
465 spectra of the juvenile and liver normalised PC1 coordinate on each site; sites presented in
466 order of the mean PC1 coordinate for the adult liver samples, from low to high. **e**, Density
467 spectra of the juvenile and liver normalised PC1 coordinate on each site; sites presented in
468 order of the mean PC1 coordinate for the adult liver samples, from low to high. **e**, Proportion

of taxonomically-assigned reads in juvenile samples deviating from the mean in the sites with highest (E, I, M, O) and lowest (A, B, F, L) feeding status.

Fig. 4: Radar plots presenting the status of juvenile whole-body (top) and adult liver (bottom) across key biological processes. The coordinates were calculated through an orthogonal projection of each sample onto the main axis of site-wise variance based on the top dimensions representing at least 50% of the variance of PCA performed for the following subsets of genes: satiety, glycolysis, beta-oxidation, urea cycle, TCA cycle, fatty acid synthesis, cholesterol synthesis, neurotransmission (juvenile only), photoreception (juvenile only), thyroid hormones, corticoids, heat shock proteins, immunity, reactive oxygen species, xenobiotic metabolism. The feeding status of each sample, as derived from the analysis from Fig. 3, is also presented.

Materials and Methods

Study location

The subtropical island of Okinawa, Japan, is the largest island of the Ryukyu archipelago. It has an area of approximately 1,200 km² and a population of 1.38 million [1]. Most inhabitants live in the southern half of the island. The average ocean temperature is 25°C through the action of the warm Kuroshio current, and the island is surrounded by a fringing and patch reef system, which has a length of approximately 380 km. Coral communities represent an area of 69.8 km^2 [1].

Okinawa is an ideal location for this study, with coastal ecosystems heavily influenced by nearby human activities and a range of contrasted coastal ecosystems along different zones, from dense urban areas near the city of Naha in the south, to the Yanbaru forest in the north. The water quality in Okinawa is affected by human activities. Rivers in urban zones were found to have high ammonium and chemical compound concentrations, notably of pharmaceuticals, with compounds such as bezafibrate, caffeine, clarithromycin, and sulfamethoxazole found at higher levels than the 90th percentile across 109 major Japanese rivers [2]. Endocrine disruptors such as nonylphenol, Bisphenol A, and per- and polyfluoroalkyl substances have been measured in rivers and reefs with a relationship with the density of human activity [3,4]. Agricultural zones of Okinawa are linked with nitrite, nitrate, ammonium, and red soil runoff [5]. Human presence is also associated with habitat degradation and coastal hardening to protect shores from erosion and reclaim fringing reefs [6,7].

Environmental data collection

Eighteen sites were visited from May to July 2023 along the coastline of Okinawa (Supplementary Table 1,2), during the reproductive season, which is when juvenile *C. cyanea*

504 can be collected. On each site, using a multiparameter probe (RINKO-Profiler CTD logger),
505 water temperature, salinity, dissolved oxygen, turbidity, and chlorophyll-α were measured.
506 Three replicate 25m long x 4m wide transects were performed to record the fish
507 community. These transects were positioned parallel to the shore, ten meters apart, in at least
508 one meter of water. Transects could not be conducted at two sites, Uka (J) and Kayo (K), due
509 to limited water depth or strong currents. Two passes were performed per transect; mobile,
510 more visible fishes were recorded during the first pass and more cryptic fishes during the
511 second pass [8]. All fish were identified to the species level. The average fish density (number of
512 fish per m^2) and species richness (number of species per m^2) were calculated. Additionally, the
513 substrate cover on each site was assessed across the three transects using a point-intercept
514 method, with the type of substrate (live coral, dead coral, macroalgae, seagrass, rubble, hard
515 substrate, sand) recorded every meter.
516 Land use in the watersheds near each site was assessed based on a publicly available
517 dataset released in January 2023 [9]. The percentage of urban, agricultural, and natural zones
518 within a radius of 1km of the sites (excluding water) was calculated for each site.
519 Environmental parameters (salinity, temperature, dissolved oxygen, turbidity,
520 chlorophyll-alpha, substrate cover, fish density, number of species, and Shannon diversity
521 index) were normalised across sites using *scale* of the *base* package in R for all analyses.
522 The full dataset is provided in the Supplementary Information.
523
524 **Study species**
525 The blue damselfish *Chrysiptera cyanea* is a coral reef damselfish characterised by its
526 bright blue pigmentation (Fig. 1A). It occurs on shallow (<10 meters) subtropical and tropical
527 reefs of the Indo-Pacific and can most generally be found around rubble and corals in subtidal
528 reef flats and tide pools (Fig. 1B). It is one of the most abundant fish along the coastline of the

529 island of Okinawa, Japan, and can be found from urban river mouths to isolated beaches [10]. *C.*
530 *cyanea* is territorial and lives in small to large schools that typically consist of a few males and
531 several females [11,12]. The species displays protogynous hermaphroditism, with females turning
532 into males. In Okinawa, during the reproduction season (typically May-August; *49*, *50*), large
533 numbers of juveniles can be found in clusters with or without adults.
534
535 **Fish collection and sample preparation**
536 On 18 sites along the coast of Okinawa, 5 juvenile and 5 adult *C. cyanea* (4 on site B)
537 were collected using hand nets and snorkelling or SCUBA diving. In total, 90 juveniles, 49
538 females, and 40 males were collected. The fish were euthanised in a 200mg/L Tricaine
539 Methanesulfonate (MS222) solution following the guidelines for animal use issued by the
540 Animal Resources Section of OIST Graduate University. Juvenile fish were photographed and
541 directly placed in RNAlater for tissue preservation. They could not be weighed due to field
542 logistics. The average size of the juveniles was 9.6 ± 0.7 mm SL [range of 7.9-1.2 mm SL].
543 There was no significance difference in fish standard length across sites. Adult fish were
544 photographed, weighed, and dissected on-site. The livers were preserved using RNAlater. The
545 average size of the females was 34.5 ± 5.3 mm SL and of males was 43.8 ± 4.6 mm SL [23.5 –
546 56.0 mm SL]. The Fulton's body condition factor (total length in cm divided by cube of the
547 weight in grams) of the females was 1.73 ± 0.31 and of the males was 1.66 ± 0.23 [1.14 – 2.60].
548 There was no significance difference in Fulton's body condition factor across sites for females
549 and males. The characteristics of each fish from the field are listed in Supplementary Tables 3
550 and 4.
551
552 **Feeding experiment**

553 The effect of food availability on the transcriptomic profile of *C. cyanea* was assessed
554 with two experiments, for juveniles and adults. The experiment lasted 3 days in juveniles and
555 14 days in adults due to the difference in tolerance to fasting at the distinct developmental
556 stages.

557 For juvenile fish, juveniles collected in Apogama, Okinawa, Japan (26.50003 N,
558 127.84261 E), on July 30th, 2024 were kept at the OIST Marine Science Station. After one night
559 of acclimation, they were placed in two tanks with 14 juvenile fish per tank. Water input came
560 from pumped seawater. Two treatments were used: fasted and fed. In one tank, the fish were
561 fed normal amounts of food for three days (dry pellets twice per day, ASIN number:
562 B0BFJ9MY8P); in the other tank, the fish were not fed and a mesh filter was used on the water
563 inlet to remove particles and plankton. After three days, each fish was weighed, measured,
564 euthanised with MS222, and dissected. The whole body of each individual was preserved in
565 RNAlater. Of these samples, seven per treatment (14 in total) were used for RNA sequencing
566 (Supplementary Table 5). The average weight of the sequenced fish at the end of the experiment
567 was 33 ± 9 mg for the fasted group and 77 ± 22 mg for the fed group. Fulton's condition index
568 was 1.36 ± 0.20 for the fasted group and 2.00 ± 0.21 for the fed group.

569 For adult fish, female *C. cyanea* collected in Sesoko, Okinawa, Japan (26.65056 N,
570 127.8549 E), on September 30th, 2022 were kept at the OIST Marine Science Station. After an
571 acclimation period, on November 23rd, 2022, they were placed in two tanks with eight fish per
572 tank. Water input came from pumped seawater. The experiment had two groups of fish: fasted
573 and fed (control). In one tank, the fish were fed normal amounts of food for two weeks (dry
574 pellets twice per day, ASIN number: B0BFJ9MY8P); in the other tank, the fish were not fed
575 for two weeks. The average weight of the fish at the start of the experiment was 2.3 ± 0.7 g for
576 the fasted group and 2.3 ± 0.9 g for the fed group. After one week, each fish was weighed. The
577 average weight of the fish after one week was 2.1 ± 0.8 g for the fasted group and 2.4 ± 1.1 g

578 for the fed group. After two weeks, on December 6th, 2022, each fish was weighed, measured,
579 euthanised with MS222, and dissected. The livers were preserved in RNAlater. Of these
580 samples, seven per treatment (14 in total) were used for RNA sequencing (Supplementary Table
581 6). The average weight of the sequenced fish after two weeks, at the end of the experiment,
582 was 2.0 ± 0.7 g for the fasted group and 2.4 ± 0.8 g for the fed group. Fulton's condition index
583 was 1.70 ± 0.09 for the fasted group and 2.04 ± 0.22 for the fed group. The Hepatosomatic
584 index (100 x liver weight over the total weight of the fish) was 0.80 ± 0.14 for the fasted group
585 and 1.55 ± 0.27 for the fed group.
586
587 **RNA extraction, library preparation, sequencing, and data processing**
588 RNA extractions were performed using the Maxwell® RSC simplyRNA tissue kits and
589 an automated Maxwell® RSC instrument based on the manufacturer's recommendations
590 (Promega, Cat. No. AS2000, Wisconsin, USA). The quality and concentration of RNA were
591 assessed with an Invitrogen Qubit Flex benchtop fluorometer and an Agilent 4200 TapeStation.
592 Library preparation was performed using the NEBNext Ultra II Directional RNA library prep
593 kit for Illumina (New England BioLabs, USA). Sequencing was done on an Illumina
594 NovaSeq6000 platform at the OIST Sequencing section. Paired end reads (2 x 150 bp) were
595 obtained.
596 Quality control of the RNA sequencing data was performed with FastQC [15]. Following
597 this quality check, the data was processed by trimming the adapters used by Illumina on each
598 transcript as well as dropping low quality sections with Trimmomatic 0.39 [16]. The sequences
599 were aligned to the reference transcriptome for this species [17] and the number of transcripts per
600 gene were quantified using Kallisto 0.46.2 [18]. The percentage of reads mapping against the
601 reference transcriptome was 68.1 ± 3.8 (mean ± sd) for field juveniles, 81.7 ± 7.1 for field
602 livers, 58.5 ± 3.5 for feeding experiment juveniles, and 73.3 ± 1.7 for feeding experiment livers.

603

604 **RNAseq data analysis**

605 All analyses were performed on the transcriptomes of juvenile whole-body and adult
606 liver samples separately. All data analysis and figure generation scripts are available on
607 FigShare. R 4.3.3 was used. Plots were generated using *ggplot2* 3.5.1*, ggtern* 3.5.0*, ggrepel*
608 0.9.5*, ggradar* 0.2*, pheatmap* 1.0.12 in R. A threshold of $p < 0.01$ was used throughout the
609 study unless otherwise specified. Principal Component Analyses (PCA) were obtained with the
610 *prcomp* function from *stats* 4.3.3. Statistical tests included the non-parametric Kruskal-Wallis
611 Rank Sum test and linear mixed effect models with environmental variables, fish size or sex as
612 fixed effects and site as a random effect, using the top parameters identified with the *regsubset*
613 model selection function of *leaps* 3.2.

614 Using all field samples as input, gene counts were normalised using *DESeq2* 1.36.0 [19].
615 Genes with more than 10 counts in at least two sites were kept. Following variance stabilising
616 transformation (vsd) with *DESeq2,* the matrix of the gene expression level for each gene in the
617 field samples was used as the basis for the analysis. A second vsd-normalised matrix combining
618 the field and feeding experiment samples together was also used, keeping genes with more than
619 10 counts in at least two sites or feeding treatments.

620 Using *DESeq2,* the number of differentially expressed genes in pairwise site
621 comparisons was estimated at the threshold of $p < 0.01$ and absolute fold change of 1.5 (logFC
622 of 0.58). The number of genes responding to sex in the adult liver samples was also calculated.
623 Using *DRomics* 2.6-2 [20], the number of genes responding linearly or quadratically to various
624 environmental factors (percentage of natural zones near each site, temperature, sampling time,
625 salinity, chlorophyll-α) as well as size for juvenile fish was calculated at a significance
626 threshold of $p < 0.01$.

627 An annotated genome is available for *Chrysiptera cyanea* with gene names identified
628 for 26,300 genes using BLAST [17]. Manually curated sets of genes corresponding to key
629 biological processes – appetite regulation, glycolysis, beta-oxidation, urea cycle, TCA cycle,
630 fatty acid synthesis, cholesterol biosynthesis, phototransduction, neurotransmission, thyroid
631 hormones, corticoids, reactive oxygen species – were used [21]. In addition, genes associated
632 with the immune response, heat shock, and xenobiotic metabolism were identified for the
633 analysis. Module analysis based on subsets of genes was performed with *WGCNA* 1.72-5 [22].
634 GO terms (*org.Dr.eg.db*) were analysed semantically using *GoSemSim* 2.28.1 [23]. KEGG terms
635 were annotated using BlastKOALA [32].
636
637 **Meta-transcriptomic analysis**
638 To estimate diet in juvenile *C. cyanea*, using Kaiju 1.10.0 [24], juvenile whole-body reads
639 that did not map to the reference *C. cyanea* transcriptome were compared with 2,462 reference
640 NCBI protein sequences available for the following taxa: Artemia (NCBI ID: 6660),
641 Chlorophyta (3041), Ctenophora (10197), Cyanobacteria (1117), Dinoflagellata (2864),
642 Embrophyta (3193), Fungi (4751), Haptophyta (2830), Ochrophyta (2696291), Pancrustacea
643 (197562), Rhodophyta (2763), Tunicata (7712). The proportion of reads in each sample
644 corresponding to different taxonomic families was identified.
645
646 **Feeding and field experiment analysis**
647 To compare the feeding and field samples, a PCA was performed on the feeding
648 experiment samples only (*i.e.,* fish that were fasted or fed over 3 days for the juvenile
649 experiment, and 14 days for the adult experiment). After identifying the top 10% of genes with
650 the highest loading along PC1, a second PCA was generated based on the subset of genes for
651 the feeding experiment and field samples together. As the field samples fell between the feeding

652 samples along PC1, the PC1 coordinate of each sample was obtained as an indicator of each
653 sample's feeding status. This yielded five datapoints per site and sample type (four for adult
654 livers from site B).
655
656 **Radar plots**
657 Radar plots were produced to summarise the dimensionality in vsd-transformed gene
658 expression levels for various subsets of genes as obtained using Principal Component Analysis
659 (PCA). For each PCA, the minimum number of dimensions accounting for at least to 50% of
660 the total variance was calculated. The average coordinates of each site were calculated for these
661 top dimensions, and the main axis of variance along these coordinates was obtained. The
662 orthogonal projection of all samples along this main axis of variance was then calculated,
663 yielding a segment bounded by the two most extreme samples. The gene expression levels of
664 the two sites which had the most extreme average projections along the segment was compared.
665 After determining which site had the most genes showing lower expression levels, the sample
666 at the extremity of the vector closest to this selected site was set to have a value of 0, while the
667 opposite extremity was set to have a value of 1 (sketch of the method provided in
668 Supplementary Materials). For most biological processes screened to obtain the radar plots,
669 there was a linear correlation between the calculated coordinate of each sample and the mean
670 expression level of the genes within the set (Supplementary Materials), although exceptions
671 were noted (*e.g.,* phototransduction in juveniles, where the set of genes showed a clear
672 difference in expression pattern between morning and afternoon sampling on the PCA, but
673 some genes were upregulated while others were downregulated). The relationship between
674 radar plot coordinates and gene expression was also checked with heatmaps (examples in the
675 Supplementary Materials). If expression patterns with various possible environmental drivers

were found, the data structure was further dissected with module analysis tools (*e.g.,* for xenobiotic metabolism and environmental stress, Supplementary Materials).

Population genetics

For juvenile whole-body transcriptomes, in which a higher number of genes was expressed than in the adult liver samples, the raw sequencing reads were mapped to the reference genome using 2-pass mapping using STAR 2.7.9a followed by data cleaning with Picard (AddOrReplaceReadGroups, MarkDuplicates). GATK 4.4.0.0 was then used for variant calling following the pipeline from Genomics Core at NYU CGSB [25]. SNPs were filtered to exclude low quality variants. Principal Component Analysis was performed using PLINK, removing samples with over 25% of missing genotype data (11 samples out of 90) and SNPs with missing genotype rates higher than 25% (9,040 variants retained). Discriminant Analysis of Principal Components (DAPC) was performed in R using *adegenet 2.1.10*. No significant population structure was found across five zones around the island – northwest, north, east, south, west (Supplementary Materials).

Methods references

1. Japanese Ministry of the Environment, Nature Conservation Bureau, Natural Environment Planning Division. Coral Reefs of Japan. 14 (2020).
2. Mano, H. *et al.* Water quality comparison of secondary effluent and reclaimed water to ambient river water of southern Okinawa Island via biological evaluation. *Environ Monit Assess* **189**, 442 (2017).
3. Camacho, C. G. *et al.* PFAS surveillance within a highly militarized island: a case study of Okinawa, Japan. *Environmental Science: Processes & Impacts* **27**, 46–51 (2025).

4. Kawahata, H., Ohta, H., Inoue, M. & Suzuki, A. Endocrine disrupter nonylphenol and bisphenol A contamination in Okinawa and Ishigaki Islands, Japan—within coral reefs and adjacent river mouths. *Chemosphere* **55**, 1519–1527 (2004).
5. Shilla, D. J., Mimura, I., Takagi, K. K. & Tsuchiya, M. Preliminary survey of the nutrient discharge characteristics of Okinawa Rivers, and their potential effects on inshore coral reefs. *Galaxea, Journal of Coral Reef Studies* **15**, 172–181 (2013).
6. Masucci, G. D. & Reimer, J. D. Expanding walls and shrinking beaches: loss of natural coastline in Okinawa Island, Japan [PeerJ]. *PeerJ* **7**, e7520 (2019).
7. Nakajima, R., Masucci, G. D., Kuba, R., Wee, H. B. & Reimer, J. D. Son of a beach: Coastal development and the loss of natural beaches over time (1919 to 2018) on Okinawa Island, southern Japan. *Marine Pollution Bulletin* **212**, 117459 (2025).
8. Lecchini, D. & Galzin, R. Spatial repartition and ontogenetic shifts in habitat use by coral reef fishes (Moorea, French Polynesia). *Marine Biology* **147**, 47–58 (2005).
9. Japan Aerospace Exploration Agency (JAXA). High-Resolution Land Use and Land Cover Map of Okinawa Island. (2023).
10. Lecchini, D., Adjeroud, M., Pratchett, M. S., Cadoret, L. & Galzin, R. Spatial structure of coral reef fish communities in the Ryukyu Islands, southern Japan. *Oceanologica Acta* **26**, 537–547 (2003).
11. Tamilmani, G. & Gopakumar, G. Chrysiptera cyanea (Quoy & Gaimard, 1825). in 301–305 (ICAR - Central Marine Fisheries Research Institute, Kochi, 2017).
12. Wacker, S., Ness, M. H., Östlund-Nilsson, S. & Amundsen, T. Social structure affects mating competition in a damselfish. *Coral Reefs* **36**, 1279–1289 (2017).
13. Bapary, J. M. A., Nurul Amin, Md. & Takemura, A. Food availability as a possible determinant for initiation and termination of reproductive activity in the tropical damselfish Chrysiptera cyanea. *Marine Biology Research* **8**, 154–162 (2012).

14. Bapary, M. A. J., Fainuulelei, P. & Takemura, A. Environmental control of gonadal development in the tropical damselfish Chrysiptera cyanea. *Marine Biology Research* **5**, 462–469 (2009).
15. Andrews, S., Biggins, L., Inglesfield, S. & Montgomery, J. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. *FastQC. A quality control tool for high throughput sequence data.* https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2019).
16. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
17. Gairin, E. *et al.* The genome of the sapphire damselfish Chrysiptera cyanea: a new resource to support further investigation of the evolution of Pomacentrids. *Gigabyte* **2024**, 0–0 (2024).
18. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525–527 (2016).
19. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
20. Larras, F. *et al.* DRomics: A Turnkey Tool to Support the Use of the Dose–Response Framework for Omics Data in Ecological Risk Assessment. *Environ. Sci. Technol.* **52**, 14461–14468 (2018).
21. Herrera, M. *et al.* From Genes to Pathways: A Curated Gene Approach to Accurate Pathway Reconstruction in Teleost Fish Transcriptomics. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **n/a**, (2025).
22. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

749 23. Yu, G. *et al.* GOSemSim: an R package for measuring semantic similarity among GO terms
750 and gene products. *Bioinformatics* **26**, 976–978 (2010).
751 24. Menzel, P., Ng, K. L. & Krogh, A. Fast and sensitive taxonomic classification for
752 metagenomics with Kaiju. *Nat Commun* **7**, 11257 (2016).
753 25. Khalfan, M. Variant Calling Pipeline using GATK4 – Genomics Core at NYU CGSB.
754 https://gencore.bio.nyu.edu/variant-calling-pipeline-gatk4/ (2020).
755

756 **Acknowledgements**

757 We thank Elise Billoir for data analysis support with DRomics, Manon Mercader and Erina
758 Kawai for fieldwork support, James Hutasoit for experimental support, Brandon Conlon for
759 scientific writing advice, the OIST Sequencing Section for library preparation and sequencing,
760 and the Scientific Computing and Data Analysis section of Core Facilities at OIST. We also
761 thank James Reimer, Manon Mercader, Eugene Myers, and Daniel Rokhsar for fruitful
762 discussions about the project.
763

764 **Author contributions**

765 Conceptualization: EG, VL

766 Methodology: EG, SM, MH, VL

767 Investigation: EG, SM, ZC, CAS, JZ, HT, YG, MH, VL

768 Visualization: EG

769 Funding acquisition: EG, VL

770 Project administration: SM, VL

771 Supervision: SM, MH, VL

772 Writing – original draft: EG

Writing – review & editing: EG, YG, MH, VL

Competing interests

Authors declare that they have no competing interests.

Data and code availability

The transcriptome sequencing reads are deposited in the NCBI GenBank database under the BioProject IDs: PRJNA1254330 (field samples), PRJNA1271234 (juvenile feeding samples), PRJNA1168266 (adult feeding samples). All files and codes used in this manuscript for data analysis and visualisation are available at Figshare (doi: 10.6084/m9.figshare.27966972).

Ethics

Ethical approval for the feeding experiment was granted by the Okinawa Institute of Science and Technology (ACUP-2023-014).

Funding

Funding for this research was provided by the Japanese Society of the Promotion of Science KAKENHI 24KJ2183, the Okinawa Institute of Science and Technology Graduate University SHINKA Grant, and the 49th Iwatani Science and Technology Research Grant FY2023.

Additional information

Supplementary Information is available for this paper. Correspondence and requests for materials should be addressed to Emma Gairin (emma.gairin@hotmail.fr).

Supplementary Information

Table S1: Site name, coordinates, date and hour of sampling, percentage of urban, agricultural, and natural zones within a kilometre of each site, and environmental parameters (temperature, salinity, chlorophyll-α, turbidity, dissolved oxygen).

Table S2: Substrate characteristics (percentage of live coral, dead coral, macroalgae, seagrass, rubble, sand, rock) and fish community description (density of juveniles and adults of all species, density of *Chrysiptera cyanea* juveniles and adults, number of species, and Shannon diversity index).

Table S3: Sample description of the juveniles from the field.

Table S4: Sample description of the adults from the field.

Table S5: Sample description of the juveniles from the feeding experiment.

Table S6: Sample description of the adults from the feeding experiment.

Table S7: Spearman correlation coefficient between the projected coordinate of each sample along the main axis of variance across the top dimensions of Principal Component Analyses for subsets of genes, and the mean expression level of the genes.

Extended Data Fig. 1: Drivers of the overall data structure in two Principal Component Analyses (PCA), using all genes with counts above 10 in at least two sites, for juvenile and liver samples separately. For juveniles (**a**) and adults (**b**), bar charts of the percentage of contribution to the variance of the top 6 PC dimensions, corresponding to over 50% of the total variance. Tile plots presenting the significance of the linear relationship between various environmental parameters and PC coordinates. The tests were linear mixed effect model with site as a random effect and variables selected from preliminary linear models with lowest BIC value using *regsubset* (*leaps* package in R). **c**, Relationship between PC1 (PCA with all genes with counts above 10 in at least two sites for juvenile samples) and the standard length of the juvenile fish. **d**, Principal Component Analysis of the expression of all genes with normalised

counts (from DESEQ2) above an average of 10 in at least two sites, for adult liver samples, *i.e.,* for 17,868 genes. Triangles correspond to female livers and circles to male livers. Sex was assigned based on the tail colour of the fish at sampling (females have transparent caudal fins while males have opaque blue caudal fins). The bottom panel is a violin plot indicating the distribution of the male and female samples along PC1. There is a significant difference in PC1 values depending on sex (Kruskal-Wallis rank sum test, $\chi^2 = 40.4$, df = 1, $p < 10^{-9}$).

Extended Data Fig. 2: Feeding experiment results for the juvenile and liver samples. a, Volcano plot displaying the number of significantly differently expressed genes in the whole body (absolute Log2 fold change > 0.58, $p < 0.01$) from DESEQ2 for the juvenile feeding experiment (no food *vs.* fed). **b**, Volcano plot displaying the number of significantly differently expressed genes in the liver (absolute Log2 fold change > 0.58, $p < 0.01$) from DESEQ2 for the adult feeding experiment (no food *vs.* fed). **c**, Venn Diagram displaying the number of differentially expressed genes (absolute Log2 fold change > 0.58, adjusted $p < 0.01$) from DESEQ2 in the juvenile and liver samples, with the genes in common shown based on the intersection of the different diagram cells. **d**, Principal Component Analysis of the gene expression level across all genes (with mean expression above 10 counts) for the juvenile feeding experiment samples (7 juvenile fish with no food, 7 fed for 3 days). **e**, Principal Component Analysis of the gene expression level across all genes (with mean expression above 10 counts) for the liver feeding experiment samples (7 adult fish with no food, 7 fed for 3 days).

Extended Data Fig. 3: Most common GO terms across the top 10 percent genes driving the loading of dimension 1 on the Principal Component Analysis (PCAs) of the feeding experiment samples for juveniles (a) and adult liver (b). The PCAs are presented on Fig. 3. GO terms were trimmed based on semantic similarity at a threshold of 0.7 using GOSemSim and the most unique terms are presented here.

Extended Data Fig. 4: Most common GO terms across the top 10 percent genes driving the loading of dimension 2 on the Principal Component Analysis of the feeding and field samples (Fig. 3b from the main text) in juveniles. GO terms were trimmed based on semantic similarity at a threshold of 0.7 using GOSemSim and the most unique terms are presented here.

Extended Data Fig. 5: Sketch of the method to obtain the radar plots presented in Fig. 4. This is an example where the first two PC dimensions are used. The analysis includes all PC dimensions representing up to 50% of the total variance. **a**, First, the Principal Component Analysis, here based on the normalized expression level of 16 genes linked to corticoids, across 18 sites in Okinawa, with 5 samples per site, is obtained. **b**, The mean PC1/PC2 coordinate of the samples from each of the 18 sites is calculated. **c**, The direction of the main variance in the mean coordinate of the sites is then calculated, here presented as a dashed axis. **d**, The orthogonal projection of all mean site coordinates as well as the projection of each sample onto the main axis of variance is calculated. The two furthermost samples across the main variance axis are set to have values of 0 and 1. Based on this method, the orthogonal projection of all samples and sites was calculated for multiple gene sets. The results are then summarized as a radar plot in the main manuscript.

Extended Data Fig. 6: Heatmap of the vsd-normalised expression level of 45 appetite-related genes in juvenile field samples. Rows are clustered based on correlation using the Ward D2 method with *pheatmap*. Environmental (temperature, hour of sampling) and intrinsic fish variables (feeding status, standard length) are indicated above the heatmap. Increasingly dark colours indicate higher values for the radar plot coordinates, feeding status, standard length, and hour of sampling. Temperature is indicated from a blue (low temperature) to red (high temperature) scale.

Extended Data Fig. 7: Heatmap of the vsd-normalised expression level of 27 thyroid-related genes in juvenile field samples. Rows are clustered based on correlation using the

Ward D2 method with *pheatmap*. Environmental (temperature, hour of sampling) and intrinsic fish variables (feeding status, standard length) are indicated above the heatmap. Increasingly dark colours indicate higher values for the radar plot coordinates, feeding status, standard length, and hour of sampling. Temperature is indicated from a blue (low temperature) to red (high temperature) scale.

Extended Data Fig. 8: Heatmap of the vsd-normalised expression level of 25 corticoid-related genes in juvenile field samples. Rows are clustered based on correlation using the Ward D2 method with *pheatmap*. Environmental (temperature, hour of sampling) and intrinsic fish variables (feeding status, standard length) are indicated above the heatmap. Increasingly dark colours indicate higher values for the radar plot coordinates, feeding status, standard length, and hour of sampling. Temperature is indicated from a blue (low temperature) to red (high temperature) scale.

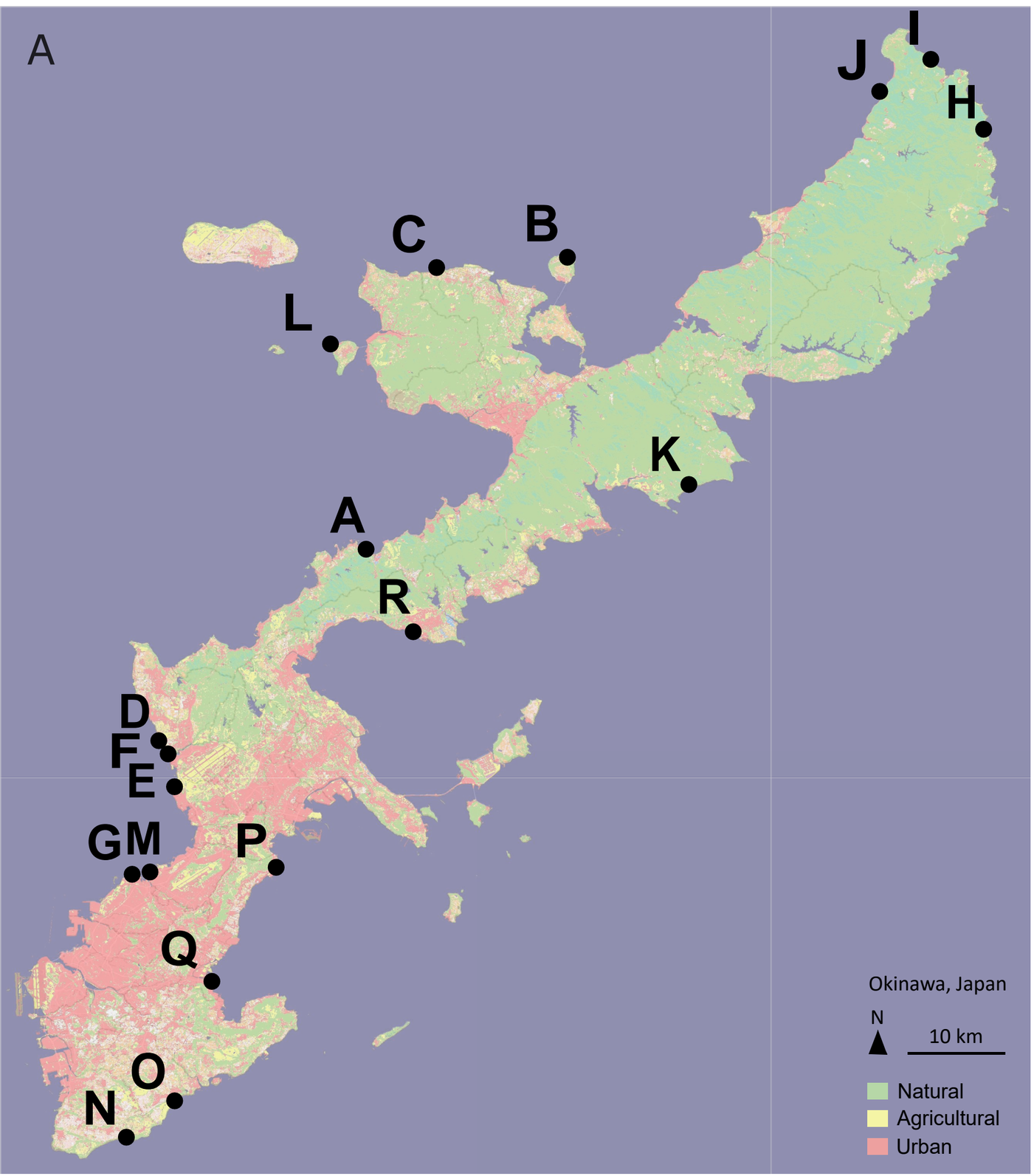
Extended Data Fig. 9: Results from a module analysis for the juvenile whole-body and adult liver gene expression. Analysis performed using WGCNA (Weighted Correlation Network Analysis) with a minimum module size of 20 genes for juveniles and 13 genes for adult liver samples (corresponding to the lowest number of genes within a category) and a soft-thresholding power of 9 for juveniles and 6 for livers. 13 modules of genes with similar expression patterns across the samples were retrieved for the juveniles and 8 for the livers. The following categories of genes were used, with 974 genes retrieved in total: heat shock proteins, reactive oxygen species, corticoids, thyroid hormones, neurotransmission, immunity, and xenobiotic metabolism. Environmental and intrinsic fish variables (percentage of natural zones within a kilometre of the site, feeding status as derived from Fig. 3 of the main text, standard length for juveniles, sex for livers, temperature, and hour of sampling) were used in a linear-mixed effect model (with site as a random effect and a selection of fixed effects based on model dredging) to determine the relationship between the gene expression patterns in each module

and environmental or intrinsic fish drivers. The resulting t-values are presented. Higher t-values (in red) indicate positive relationships between the gene expression level within the module and the given variable, and inversely (in blue). Top: Percentage of genes of each category retrieved within a given module. Bottom: t-value of a linear mixed-effect model retrieving significant relationships between environmental or intrinsic fish characteristics and gene expression patterns within the modules.

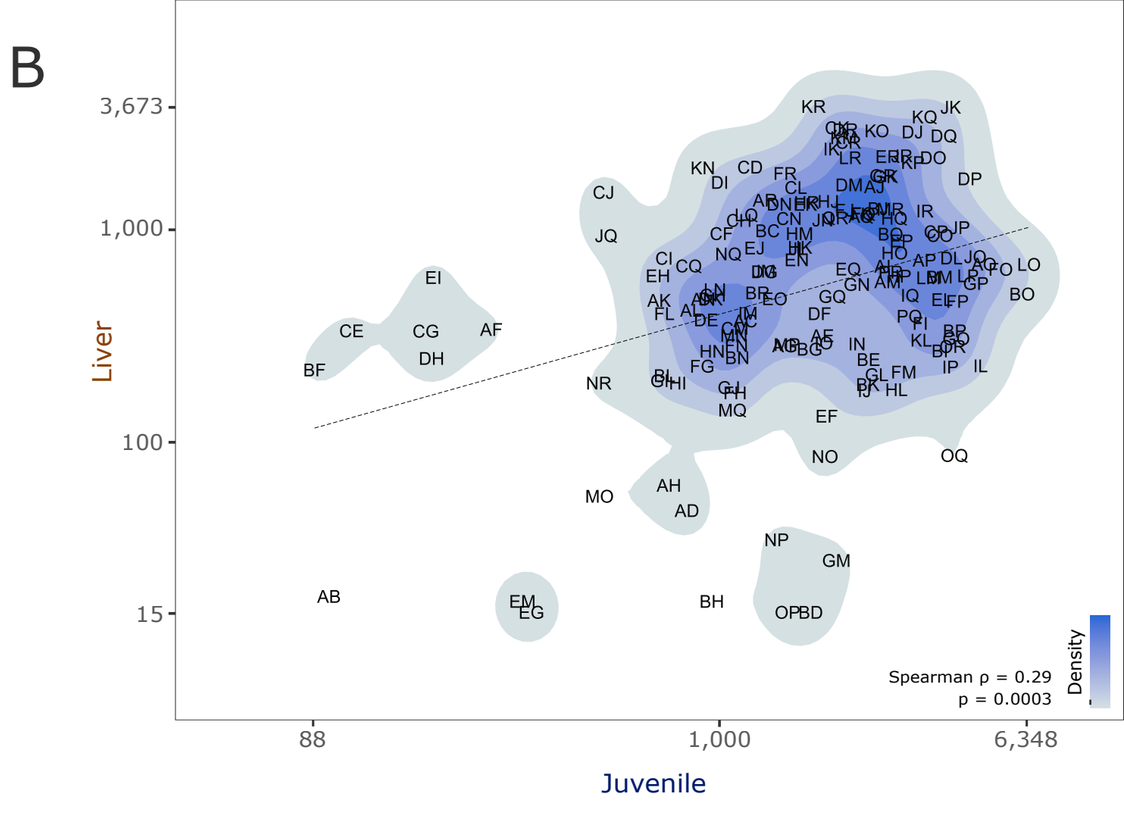
Extended Data Fig. 10: Environmental characterisation of the sites. a, Ternary plot presenting the percentage of natural, agricultural, and urban zones within 1km of the sites (excluding water). **b**, Substrate cover on three 25m transects performed parallel to shore near the sampling (maximum depth: 2m). **c**, NMDS analysis of the juvenile fish community structure and contributing environmental variables based on *vegan* function envfit ($p < 0.1$). **d**, NMDS analysis of the adult fish community structure and contributing environmental variables based on *vegan* function envfit ($p < 0.1$). No fish community structure was recorded on sites J and K due to logistical constraints.

Extended Data Fig. 11: First two discriminant functions of a Discriminant Analysis of Principal Components (DAPC) of *C. cyanea* juveniles. Colours indicate 5 zones: northwest, north, east, south, west. The analysis is based on filtered SNPs (mind = 0.25, geno = 0.25) following variant calling using the Genome Analysis Toolkit (GATK) pipeline.

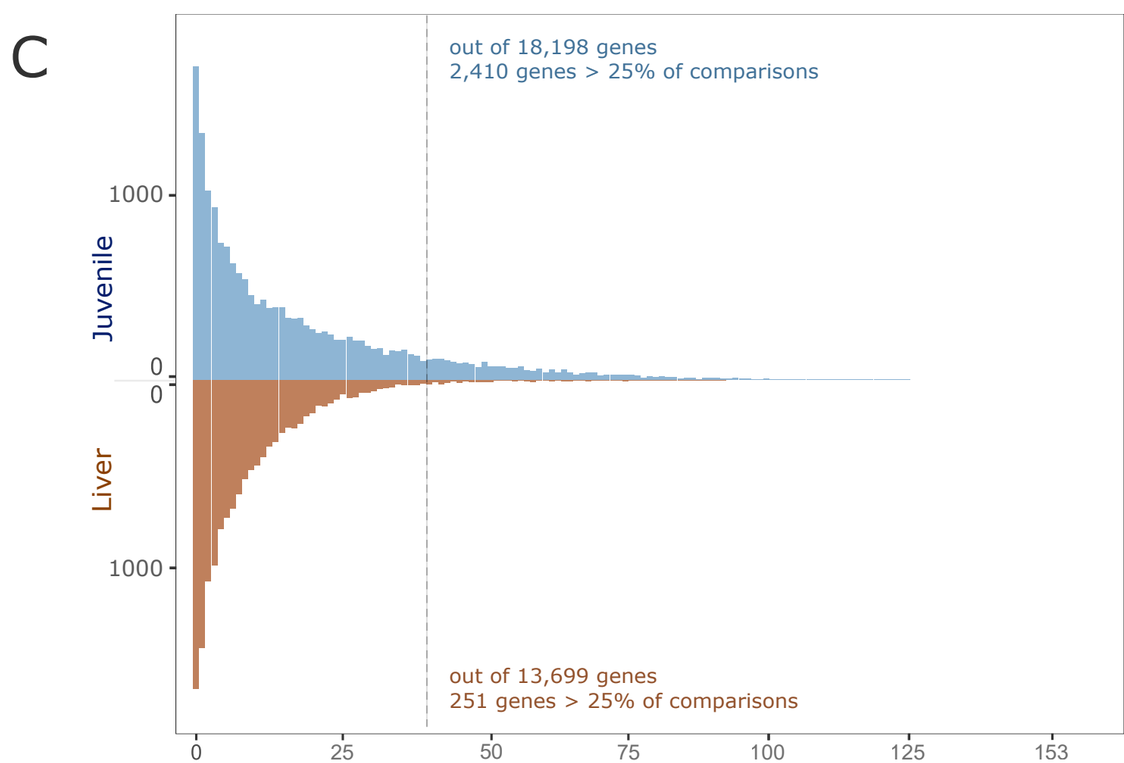
A


A
B
C
D
E
F
G
M
H
I
J
K
L
N
O
P
Q
R
Okinawa, Japan
N
10 km
Natural
Agricultural
Urban

B


3,673
1,000
100
15
Liver
88
1,000
6,348
Juvenile
Density
Spearman ρ = 0.29
p = 0.0003

C


out of 18,198 genes
2,410 genes > 25% of comparisons
1000
0
0
1000
Juvenile
Liver
out of 13,699 genes
251 genes > 25% of comparisons
0
25
50
75
100
125
153
Site-by-site pairwise comparisons

A

A B C D E F G H I J K L M N O P Q R

Natural
2 - 94 %

Hour of sampling
10:00 - 17:00

Temperature
24.2 - 32.4℃

Salinity
30.7 - 34.6 psu

Chlorophyll-α
0.1 - 2.7 μg/L

B

Juvenile
25,334 genes

PC2 (11.9%)

PC1 (19.5%)

C

Liver
13,470 genes

PC2 (7.6%)

PC1 (14.2%)

D

Juvenile

Liver

Number of genes

3,889
616
1,854
425
674
539
144
6,343
5,566
2,040
703
1,312
585
1,028
24
261
2,561
6,086

Standard length (juvenile)
Sex (adult liver)
Temperature up to 29 C
Temperature up to 32 C
Natural
Hour of sampling
Salinity
Chlorophyll-α
Two factors
More than two factors and low counts
None

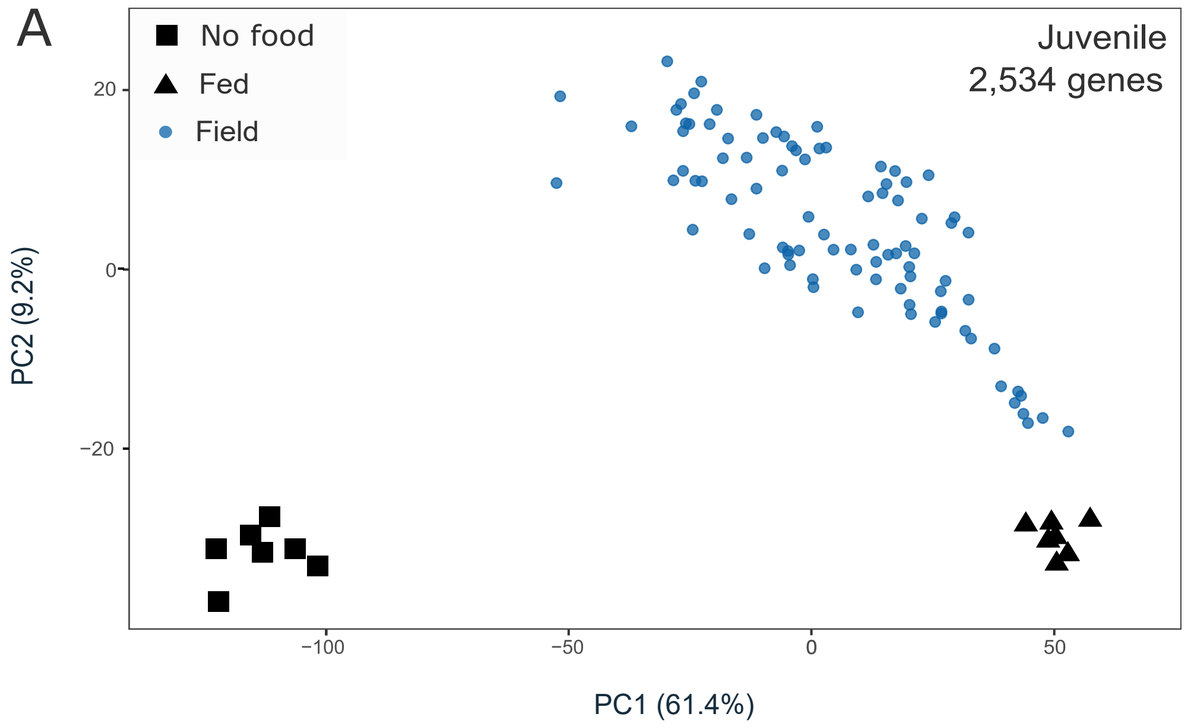
E Liver

Signal transduction
Proteolysis
Apoptotic process
Innate immune response
Positive regulation of transcription by RNA polymerase II
Lipid metabolic process
Negative regulation of cell population proliferation
Cell differentiation
Protein transport

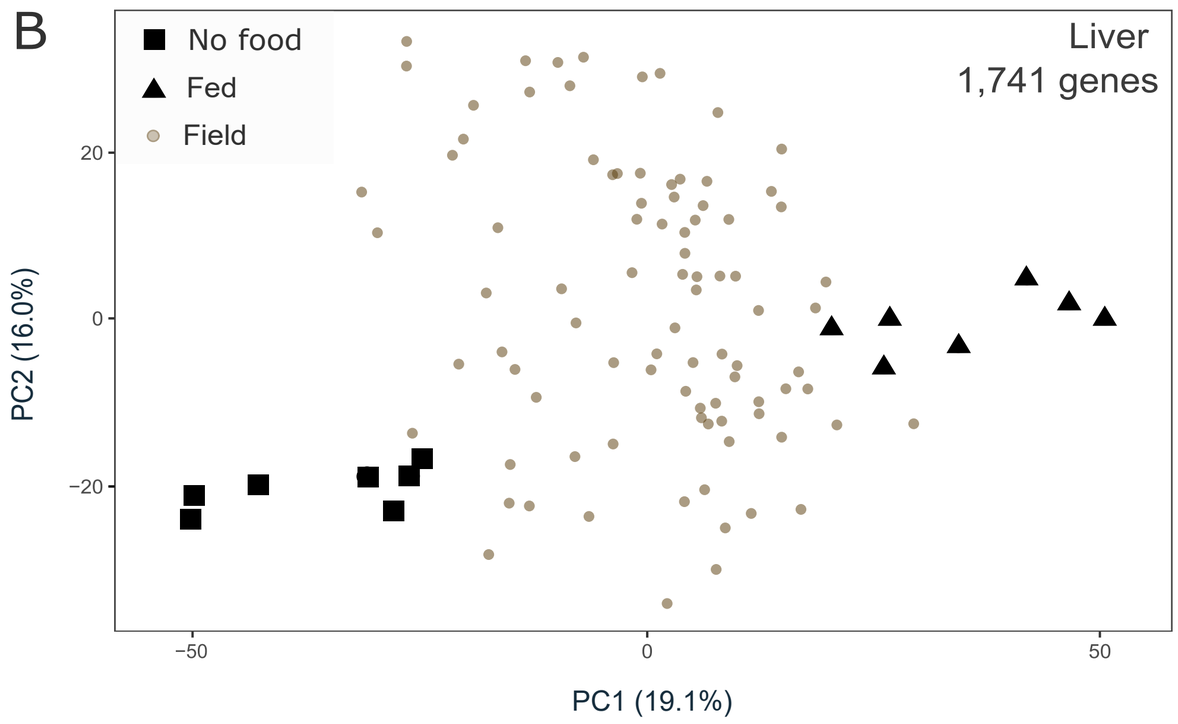
Signal transduction
Immune system
Positive regulation of ERK1 and ERK2 cascade
Phosphorylation
Phagocytosis

Signal transduction
Innate immune response
Protein phosphorylation
Inflammatory response
Cell differentiation
Apoptotic process
Positive regulation of tumor necrosis factor production
Cell adhesion
Cell migration
Cell cycle
Negative regulation of transcription by RNA polymerase II
Protein transport
Cell division
Positive regulation of GTPase activity

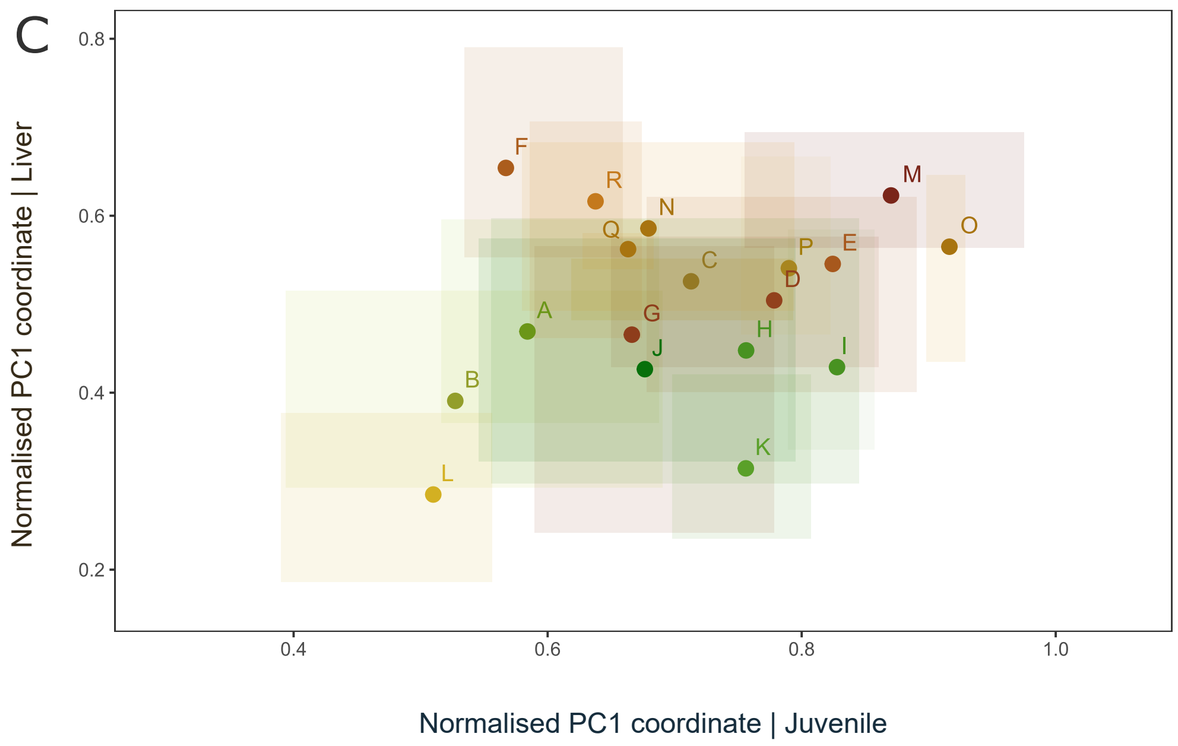
A



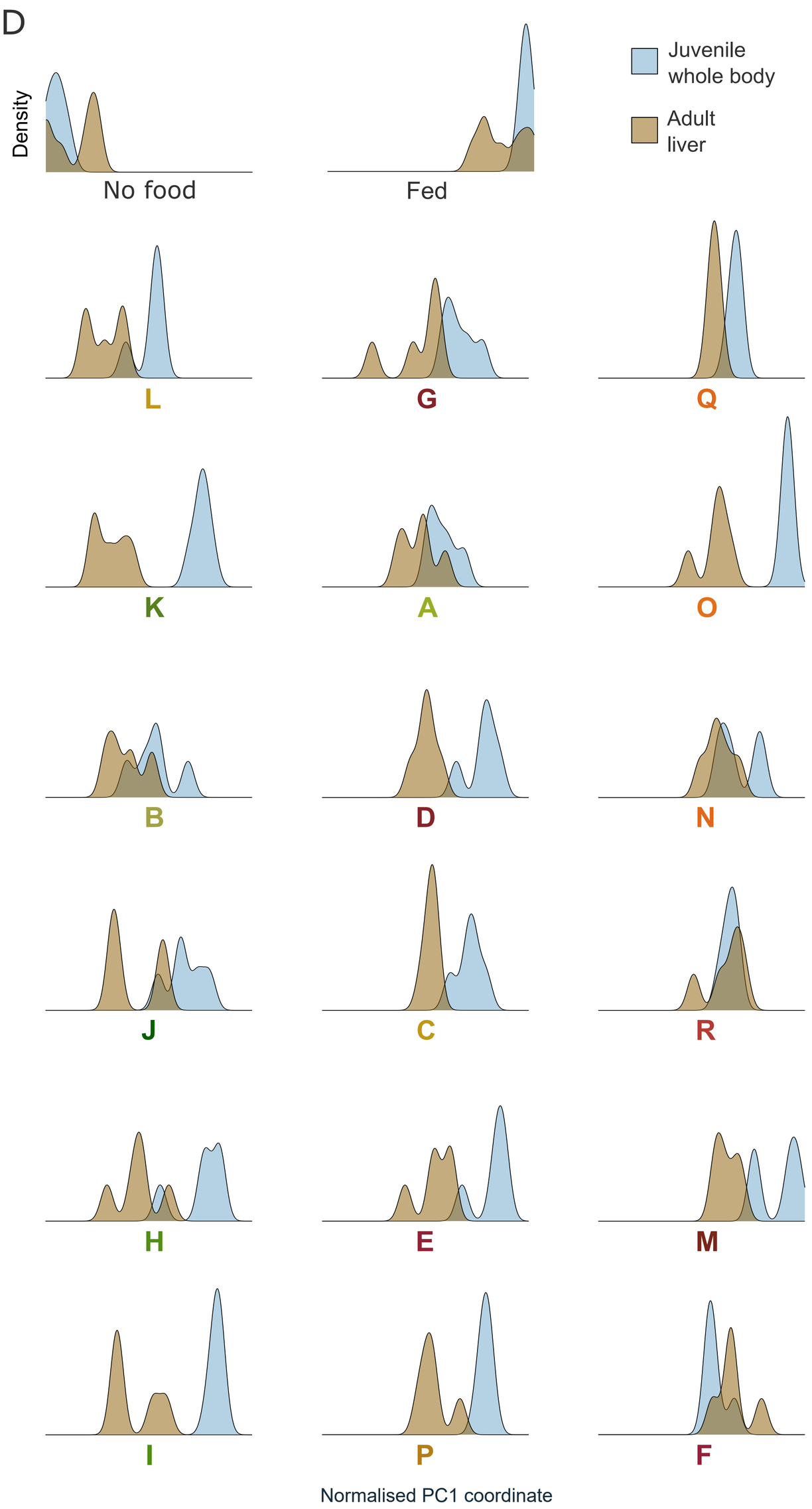
B



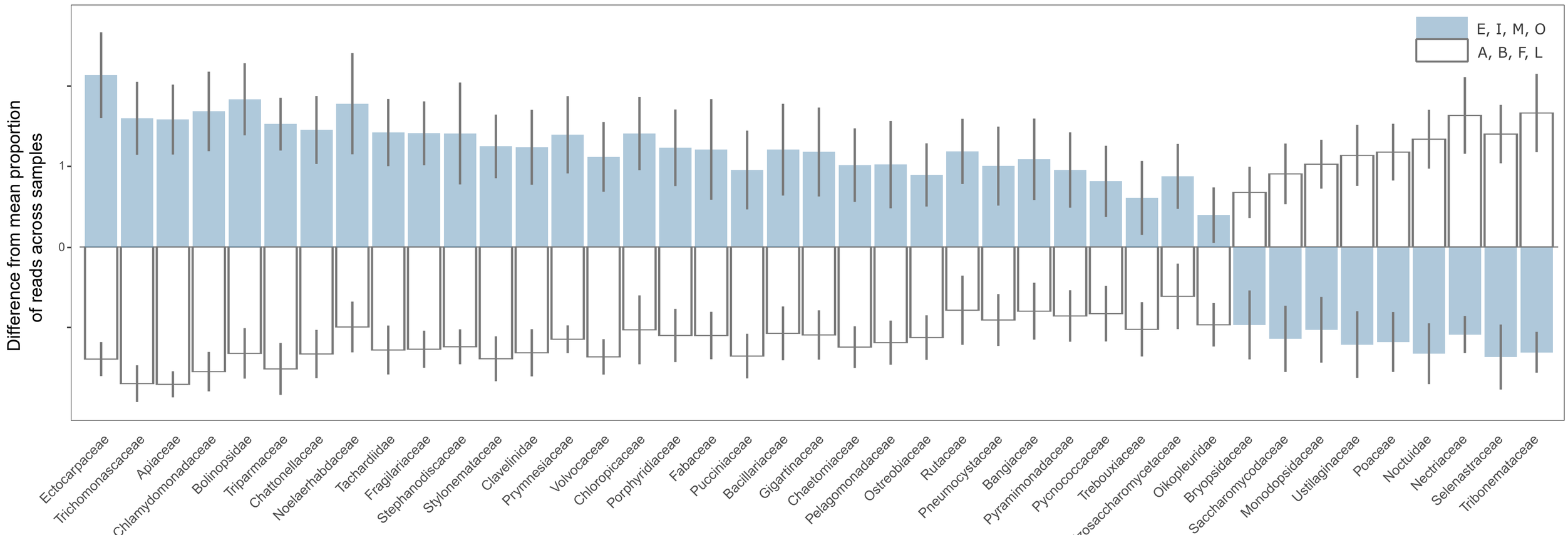
C



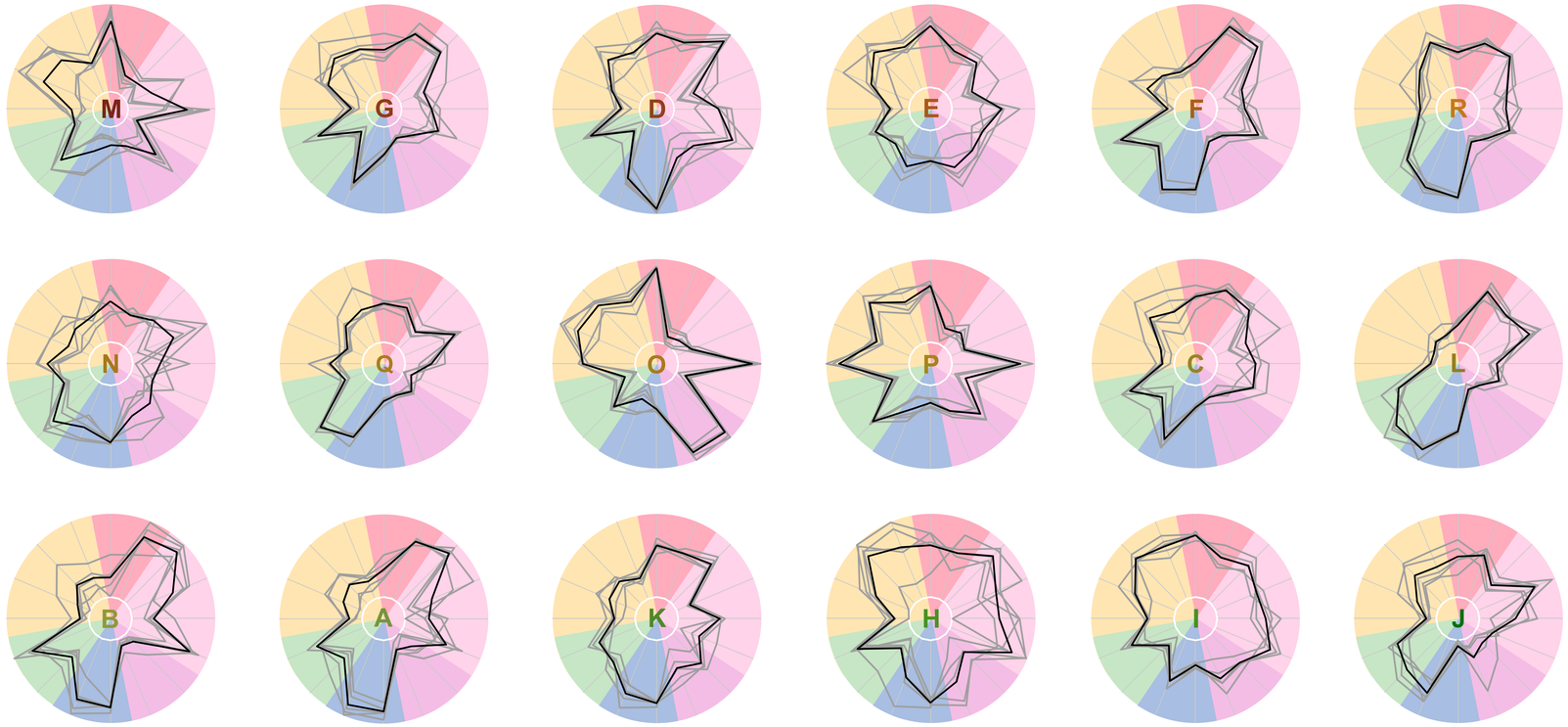
D

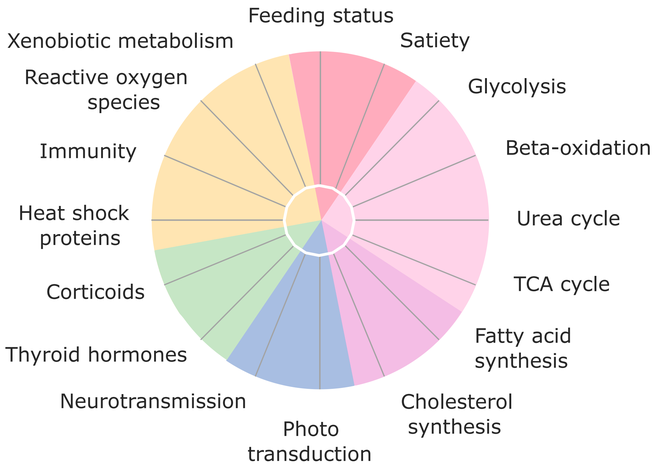


E Juvenile



Juvenile
Whole body



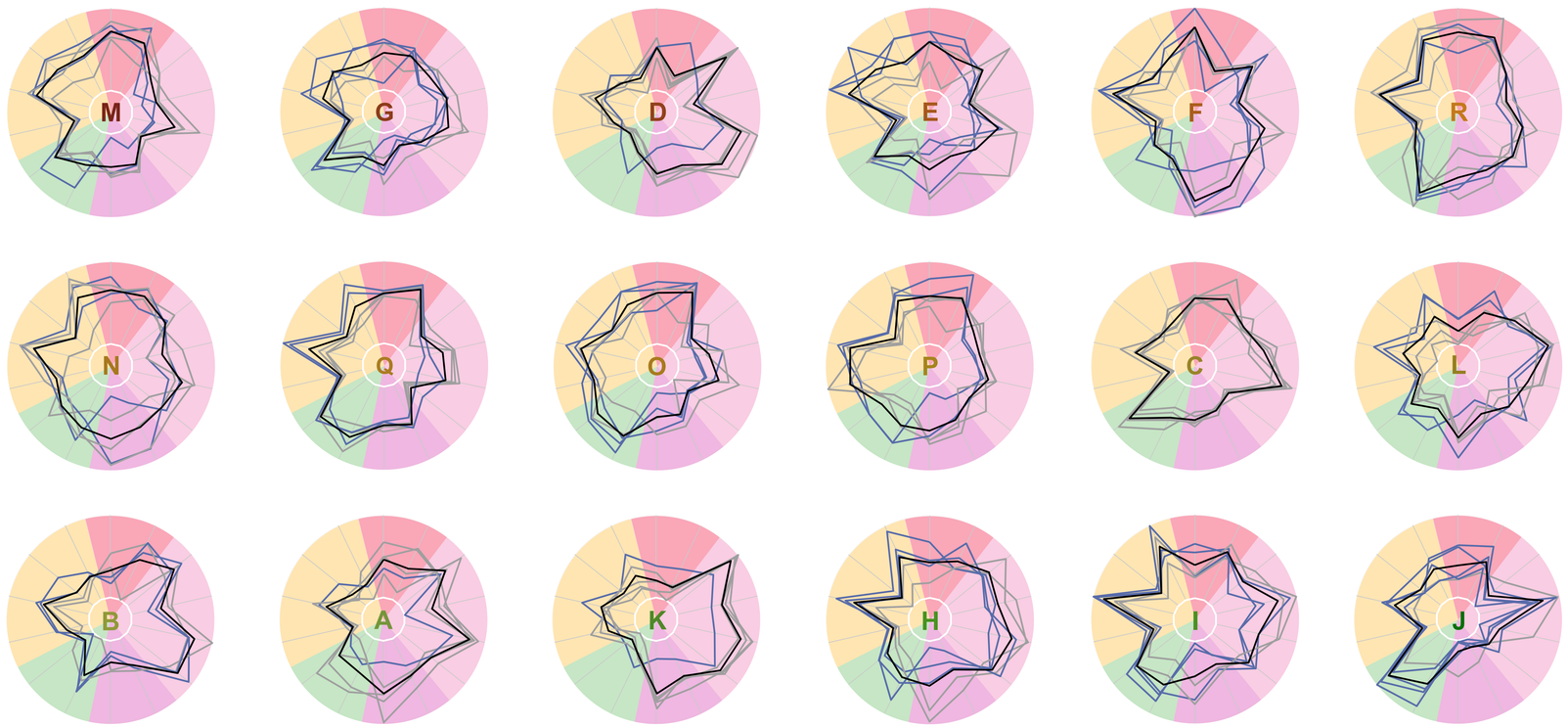


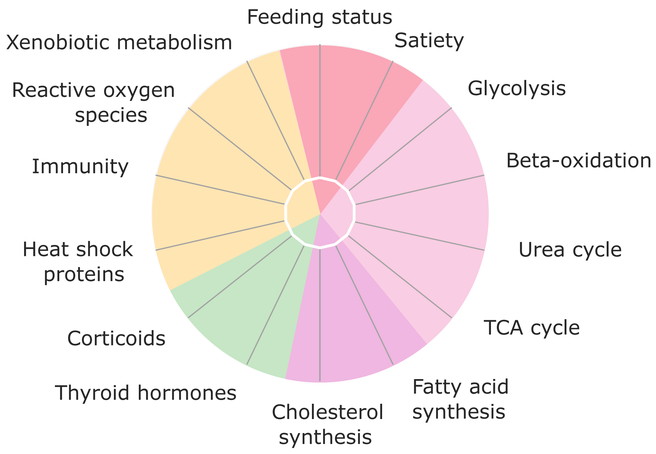
Average
Juvenile

Food-related processes

Status, appetite
Catabolic
Anabolic
Endocrine
Neurosensory
Stress response

Adult
Liver





Average
Female
Male

Supplementary Information

Table S1: Site name, coordinates, date and hour of sampling, percentage of urban, agricultural, and natural zones within a kilometre of each site, and environmental parameters (temperature, salinity, chlorophyll-α, turbidity, dissolved oxygen).

Table S2: Substrate characteristics (percentage of live coral, dead coral, macroalgae, seagrass, rubble, sand, rock) and fish community description (density of juveniles and adults of all species, density of *Chrysiptera cyanea* juveniles and adults, number of species, and Shannon diversity index).

Table S3: Sample description of the juveniles from the field.

Table S4: Sample description of the adults from the field.

Table S5: Sample description of the juveniles from the feeding experiment.

Table S6: Sample description of the adults from the feeding experiment.

Table S7: Spearman correlation coefficient between the projected coordinate of each sample along the main axis of variance across the top dimensions of Principal Component Analyses for subsets of genes, and the mean expression level of the genes.

Extended Data Fig. 1: Drivers of the overall data structure in two Principal Component Analyses (PCA), using all genes with counts above 10 in at least two sites, for juvenile and liver samples separately. For juveniles (**a**) and adults (**b**), bar charts of the percentage of contribution to the variance of the top 6 PC dimensions, corresponding to over 50% of the total variance. Tile plots presenting the significance of the linear relationship between various environmental parameters and PC coordinates. The tests were linear mixed effect model with site as a random effect and variables selected from preliminary linear models with lowest BIC value using *regsubset* (*leaps* package in R). **c**, Relationship between PC1 (PCA with all genes with counts above 10 in at least two sites for juvenile samples) and the standard length of the juvenile fish. **d**, Principal Component Analysis of the expression of all genes with normalised counts (from DESEQ2) above an average of 10 in at least two sites, for adult liver samples, *i.e.,* for 17,868 genes. Triangles correspond to female livers and circles to male livers. Sex was assigned based on the tail colour of the fish at sampling (females have transparent caudal fins while males have opaque blue caudal fins). The bottom panel is a violin plot indicating the distribution of the male and female samples along PC1. There is a significant difference in PC1 values depending on sex (Kruskal-Wallis rank sum test, $\chi^2 = 40.4$, df = 1, $p < 10^{-9}$).

Extended Data Fig. 2: Feeding experiment results for the juvenile and liver samples. a, Volcano plot displaying the number of significantly differently expressed genes in the whole body (absolute Log2 fold change > 0.58, $p < 0.01$) from DESEQ2 for the juvenile feeding experiment (no food *vs.* fed). **b**, Volcano plot displaying the number of significantly differently expressed genes in the liver (absolute Log2 fold change > 0.58, $p < 0.01$) from DESEQ2 for the adult feeding experiment (no food *vs.* fed). **c**, Venn Diagram displaying the number of

differentially expressed genes (absolute Log2 fold change > 0.58, adjusted p < 0.01) from DESEQ2 in the juvenile and liver samples, with the genes in common shown based on the intersection of the different diagram cells. **d**, Principal Component Analysis of the gene expression level across all genes (with mean expression above 10 counts) for the juvenile feeding experiment samples (7 juvenile fish with no food, 7 fed for 3 days). **e**, Principal Component Analysis of the gene expression level across all genes (with mean expression above 10 counts) for the liver feeding experiment samples (7 adult fish with no food, 7 fed for 3 days).

Extended Data Fig. 3: Most common GO terms across the top 10 percent genes driving the loading of dimension 1 on the Principal Component Analysis (PCAs) of the feeding experiment samples for juveniles (a) and adult liver (b). The PCAs are presented on Fig. 3. GO terms were trimmed based on semantic similarity at a threshold of 0.7 using GOSemSim and the most unique terms are presented here.

Extended Data Fig. 4: Most common GO terms across the top 10 percent genes driving the loading of dimension 2 on the Principal Component Analysis of the feeding and field samples (Fig. 3b from the main text) in juveniles. GO terms were trimmed based on semantic similarity at a threshold of 0.7 using GOSemSim and the most unique terms are presented here.

Extended Data Fig. 5: Sketch of the method to obtain the radar plots presented in Fig. 4. This is an example where the first two PC dimensions are used. The analysis includes all PC dimensions representing up to 50% of the total variance. **a**, First, the Principal Component Analysis, here based on the normalized expression level of 16 genes linked to corticoids, across 18 sites in Okinawa, with 5 samples per site, is obtained. **b**, The mean PC1/PC2 coordinate of the samples from each of the 18 sites is calculated. **c**, The direction of the main variance in the mean coordinate of the sites is then calculated, here presented as a dashed axis. **d**, The orthogonal projection of all mean site coordinates as well as the projection of each sample onto the main axis of variance is calculated. The two furthermost samples across the main variance axis are set to have values of 0 and 1. Based on this method, the orthogonal projection of all samples and sites was calculated for multiple gene sets. The results are then summarized as a radar plot in the main manuscript.

Extended Data Fig. 6: Heatmap of the vsd-normalised expression level of 45 appetite-related genes in juvenile field samples. Rows are clustered based on correlation using the Ward D2 method with *pheatmap*. Environmental (temperature, hour of sampling) and intrinsic fish variables (feeding status, standard length) are indicated above the heatmap. Increasingly dark colours indicate higher values for the radar plot coordinates, feeding status, standard length, and hour of sampling. Temperature is indicated from a blue (low temperature) to red (high temperature) scale.

Extended Data Fig. 7: Heatmap of the vsd-normalised expression level of 27 thyroid-related genes in juvenile field samples. Rows are clustered based on correlation using the Ward D2 method with *pheatmap*. Environmental (temperature, hour of sampling) and intrinsic fish variables (feeding status, standard length) are indicated above the heatmap. Increasingly dark colours indicate higher values for the radar plot coordinates, feeding status, standard

length, and hour of sampling. Temperature is indicated from a blue (low temperature) to red (high temperature) scale.

Extended Data Fig. 8: Heatmap of the vsd-normalised expression level of 25 corticoid-related genes in juvenile field samples. Rows are clustered based on correlation using the Ward D2 method with *pheatmap*. Environmental (temperature, hour of sampling) and intrinsic fish variables (feeding status, standard length) are indicated above the heatmap. Increasingly dark colours indicate higher values for the radar plot coordinates, feeding status, standard length, and hour of sampling. Temperature is indicated from a blue (low temperature) to red (high temperature) scale.

Extended Data Fig. 9: Results from a module analysis for the juvenile whole-body and adult liver gene expression. Analysis performed using WGCNA (Weighted Correlation Network Analysis) with a minimum module size of 20 genes for juveniles and 13 genes for adult liver samples (corresponding to the lowest number of genes within a category) and a soft-thresholding power of 9 for juveniles and 6 for livers. 13 modules of genes with similar expression patterns across the samples were retrieved for the juveniles and 8 for the livers. The following categories of genes were used, with 974 genes retrieved in total: heat shock proteins, reactive oxygen species, corticoids, thyroid hormones, neurotransmission, immunity, and xenobiotic metabolism. Environmental and intrinsic fish variables (percentage of natural zones within a kilometre of the site, feeding status as derived from Fig. 3 of the main text, standard length for juveniles, sex for livers, temperature, and hour of sampling) were used in a linear-mixed effect model (with site as a random effect and a selection of fixed effects based on model dredging) to determine the relationship between the gene expression patterns in each module and environmental or intrinsic fish drivers. The resulting t-values are presented. Higher t-values (in red) indicate positive relationships between the gene expression level within the module and the given variable, and inversely (in blue). Top: Percentage of genes of each category retrieved within a given module. Bottom: t-value of a linear mixed-effect model retrieving significant relationships between environmental or intrinsic fish characteristics and gene expression patterns within the modules.

Extended Data Fig. 10: Environmental characterisation of the sites. a, Ternary plot presenting the percentage of natural, agricultural, and urban zones within 1km of the sites (excluding water). **b**, Substrate cover on three 25m transects performed parallel to shore near the sampling (maximum depth: 2m). **c**, NMDS analysis of the juvenile fish community structure and contributing environmental variables based on *vegan* function envfit ($p < 0.1$). **d**, NMDS analysis of the adult fish community structure and contributing environmental variables based on *vegan* function envfit ($p < 0.1$). No fish community structure was recorded on sites J and K due to logistical constraints.

Extended Data Fig. 11: First two discriminant functions of a Discriminant Analysis of Principal Components (DAPC) of *C. cyanea* juveniles. Colours indicate 5 zones: northwest, north, east, south, west. The analysis is based on filtered SNPs (mind = 0.25, geno = 0.25) following variant calling using the Genome Analysis Toolkit (GATK) pipeline.

Table S1: Site name, coordinates, date and hour of sampling, percentage of urban, agricultural, and natural zones within a kilometre of each site, and environmental parameters (temperature, salinity, chlorophyll-α, turbidity, dissolved oxygen).

Site	Site name	Latitude (N)	Longitude (E)	Nearby features	Date	Hour of sampling	Urban (%)	Agricultural (%)	Natural (%)	Temperature (oC)	Salinity (psu)	Chlorophyll-α (µg/L)	Turbidity (FTU)	Dissolved Oxygen (%)
A	Seragaki	26.50514	127.8801	Village, forest	24/05/2023	11:00	21	10	69	24.2	34.5	0.3	0.7	101
B	Kouri	26.71363	128.0197	Agriculture	25/05/2023	11:00	5	39	56	24.7	34.4	0.2	0.8	107
C	Nakijin	26.70474	127.9302	Village, agriculture	25/05/2023	15:00	26	35	38	26.0	34.0	0.5	1.2	123
D	Toguchi	26.36946	127.7332	Village, agriculture, military installation	26/05/2023	10:00	22	70	8	24.6	33.9	0.2	0.1	102
E	Junkyard	26.33682	127.7444	City, industry, military installation	26/05/2023	14:30	42	45	13	25.1	34.3	0.2	0.2	103
F	Mizugama	26.36009	127.7398	City, large river	29/05/2023	12:00	61	25	14	25.7	34.5	0.1	0.1	120
G	Minatogawa	26.27478	127.7143	Industrial zone, military installation	29/05/2023	15:00	80	13	7	26.9	34.4	0.2	0.1	137
H	Adan	26.80111	128.3195	Forest	07/06/2023	11:00	3	11	86	25.8	34.5	0.1	0.0	110
I	Oku	26.85067	128.2821	Forest, pig farm	07/06/2023	15:00	1	13	86	27.1	34.4	0.3	0.8	148
J	Uka	26.83113	128.2459	Forest, main road	07/06/2023	17:00	5	0	94	28.6	30.7	0.1	0.0	162
K	Kayo	26.54939	128.1106	Forest, village	09/06/2023	10:00	3	14	83	25.5	34.5	0.3	1.4	101
L	Sesoko	26.65056	127.8549	Hotel, agriculture	09/06/2023	14:00	23	39	38	26.9	34.6	0.2	0.3	122
M	Ginowan	26.27643	127.727	City	12/06/2023	15:00	84	13	2	27.8	33.4	1.1	0.7	120
N	Oodo	26.08846	127.7102	Village, agriculture	11/07/2023	11:00	11	60	28	28.3	34.4	0.1	0.0	106
O	Hanashiro	26.11425	127.7444	Golf course	11/07/2023	15:00	23	48	29	32.0	34.4	0.4	0.6	166

P	Kuba	26.27971	127.8164	Village, agriculture, power plant	12/07/2023	14:30	43	21	35	32.4	34.6	0.2	0.4	140
Q	Agarihama	26.19888	127.7708	City	12/07/2023	16:30	55	17	28	30.4	34.5	2.7	36.2	125
R	Kin	26.44667	127.9136	City, military installation	14/07/2023	10:00	66	14	20	29.7	34.2	0.5	1.4	102

Table S2: Substrate characteristics (percentage of live coral, dead coral, macroalgae, seagrass, rubble, sand, rock) and fish community description (density of juveniles and adults of all species, density of *Chrysiptera cyanea* juveniles and adults, number of species, and Shannon diversity index).

Site	Live coral (%)	Dead coral (%)	Macroalgae (%)	Seagrass (%)	Rubble (%)	Sand (%)	Rock (%)	Juvenile density (number of fish per 100 m^2)	Adult density (number of fish per 100 m^2)	Juvenile C. cyanea density (number of fish per 100 m^2)	Adult C. cyanea density (number of fish per 100 m^2)	Number of species (juveniles)	Number of species (adults)	Number of species (all stages)	Shannon index (juveniles)	Shannon index (adults)	Shannon index (all stages)
A	4	4	1	19	45	20	7	272	68	74	13	12	10	14	0.8	1.3	1.0
B	7	4	0	0	23	10	56	85	225	11	39	11	18	19	1.7	1.6	1.8
C	3	4	5	19	32	19	18	42	18	1	0	7	7	9	1.0	1.3	1.6
D	1	0	15	8	27	26	23	186	66	34	7	11	13	15	1.3	1.7	1.7
E	0	0	12	0	47	20	21	319	131	72	26	12	11	15	1.0	1.2	1.3
F	0	0	0	0	16	19	65	544	220	152	21	17	24	28	0.8	2.0	1.4
G	7	1	7	0	19	4	62	182	227	39	1	10	21	22	1.0	1.9	2.1
H	24	13	1	0	16	11	35	201	142	5	10	19	23	29	1.5	2.4	2.3
I	19	9	3	0	15	8	46	506	153	47	3	15	24	30	1.1	2.5	1.9
J	0	0	0	0	0	0	100	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
K	0	0	12	0	0	68	20	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
L	4	1	24	1	24	31	15	168	92	27	2	10	19	20	1.3	2.4	2.2
M	0	0	67	0	16	17	0	164	113	41	13	10	6	9	0.8	0.7	1.1

N	3	12	4	0	0	48	33	332	187	52	20	17	25	31	1.6	2.3	2.1
O	15	4	16	0	13	7	45	315	354	71	60	21	26	32	1.3	1.7	1.7
P	0	0	21	4	24	33	18	100	32	1	2	12	7	10	1.5	1.3	1.8
Q	0	0	0	0	32	20	48	84	78	1	18	6	9	9	1.1	1.3	1.6
R	17	1	11	0	17	39	15	477	371	38	18	34	19	38	2.2	1.7	2.2

Table S3: Sample description of the juveniles from the field.

Sample	Site	Standard length (cm)	Total length (cm)
AJ1	A	1.10	1.55
AJ2	A	1.01	1.35
AJ3	A	1.03	1.37
AJ4	A	0.93	1.14
AJ5	A	0.97	1.31
BJ1	B	1.06	1.44
BJ2	B	0.92	1.24
BJ3	B	0.91	1.23
BJ4	B	0.97	1.30
BJ6	B	0.92	1.28
CJ1	C	0.97	1.39
CJ2	C	0.96	1.39
CJ3	C	0.96	1.26
CJ4	C	1.03	1.40
CJ5	C	0.88	1.23
DJ1	D	1.00	1.34
DJ2	D	0.99	1.33
DJ3	D	0.91	1.19
DJ4	D	0.89	1.22
DJ5	D	0.93	1.31
EJ1	E	1.10	1.43
EJ2	E	0.99	1.34
EJ3	E	1.13	1.54
EJ4	E	0.97	1.37
EJ5	E	0.99	1.33
FJ1	F	1.04	1.48
FJ3	F	0.98	1.35
FJ4	F	0.91	1.22
FJ5	F	0.85	1.18
FJ6	F	0.80	1.14
GJ1	G	0.94	1.30
GJ3	G	0.90	1.28
GJ4	G	0.87	1.17
GJ6	G	0.96	1.31
GJ7	G	0.87	1.18
HJ1	H	1.03	1.44
HJ2	H	0.96	1.36
HJ3	H	1.24	1.76
HJ4	H	1.07	1.50
HJ5	H	0.90	1.34
IJ1	I	0.90	1.30
IJ2	I	1.00	1.34

IJ3	I	0.96	1.40
IJ4	I	1.00	1.41
IJ5	I	0.98	1.31
JJ1	J	0.89	1.24
JJ2	J	1.03	1.41
JJ3	J	1.00	1.38
JJ5	J	0.96	1.20
JJ6	J	0.94	1.35
KJ1	K	1.00	1.40
KJ2	K	0.96	1.39
KJ3	K	0.92	1.29
KJ4	K	1.00	1.28
KJ5	K	1.00	1.26
LJ1	L	0.86	1.18
LJ2	L	0.92	1.30
LJ3	L	0.94	1.31
LJ4	L	0.90	1.29
LJ5	L	0.79	1.11
MJ2	M	1.03	1.47
MJ3	M	1.10	1.60
MJ4	M	1.04	1.50
MJ5	M	0.83	1.18
MJ6	M	0.85	1.22
NJ1	N	0.92	1.28
NJ2	N	0.96	1.37
NJ3	N	0.98	1.39
NJ4	N	0.93	1.25
NJ5	N	0.96	1.28
OJ1	O	1.18	1.78
OJ2	O	1.03	1.52
OJ3	O	0.93	1.34
OJ4	O	0.90	1.38
OJ5	O	1.03	1.55
PJ1	P	1.13	1.60
PJ2	P	0.97	1.35
PJ3	P	0.89	1.25
PJ4	P	0.99	1.41
PJ5	P	1.12	1.58
QJ1	Q	0.95	1.29
QJ2	Q	0.93	1.29
QJ3	Q	0.88	1.21
QJ4	Q	0.92	1.22
QJ5	Q	0.88	1.29
RJ1	R	0.98	1.38
RJ2	R	0.92	1.26
RJ3	R	0.95	1.34

RJ5	R	0.90	1.24
RJ6	R	0.90	1.27

Table S4: Sample description of the adults from the field.

Sample	Site	Sex	Standard length (cm)	Total length (cm)	Weight (g)	Fulton's condition index
AL6	A	Female	2.96	4.02	1.3	2.00
AL7	A	Male	4.69	6.76	5.0	1.62
AL8	A	Female	3.29	4.44	1.1	1.14
AL9	A	Female	3.25	4.32	1.0	1.24
AL10	A	Female	3.16	4.18	1.1	1.37
BL6	B	Female	3.16	4.16	1.0	1.39
BL7	B	Male	5.04	6.5	3.6	1.31
BL8	B	Female	3.20	4.20	1.1	1.35
BL9	B	Male	4.41	5.84	3.0	1.51
CL6	C	Female	3.62	4.78	2.1	1.92
CL7	C	Female	4.01	5.25	3.0	2.07
CL8	C	Female	4.13	5.24	2.5	1.74
CL9	C	Female	3.35	4.65	1.8	1.79
CL11	C	Female	4.77	5.92	3.3	1.59
DL6	D	Female	4.18	5.62	2.6	1.46
DL7	D	Female	5.11	6.91	4.8	1.45
DL8	D	Female	3.66	4.89	1.4	1.20
DL9	D	Male	3.34	4.42	1.5	1.74
DL10	D	Female	3.07	4.23	1.2	1.59
EL6	E	Male	4.93	6.52	4.4	1.59
EL7	E	Male	4.04	5.42	2.8	1.76
EL8	E	Female	3.11	4.44	1.5	1.71
EL9	E	Female	3.49	4.61	1.8	1.84
EL11	E	Male	4.38	6.08	3.6	1.60
FL6	F	Male	4.90	6.79	5.0	1.60
FL7	F	Male	4.93	6.37	3.9	1.51
FL8	F	Male	4.77	6.66	4.1	1.39
FL9	F	Female	3.78	5.03	2.4	1.89
FL10	F	Female	3.89	5.30	2.5	1.68
GL6	G	Female	4.21	5.59	2.6	1.49
GL7	G	Male	4.11	5.53	2.9	1.71
GL8	G	Male	4.00	5.32	2.2	1.46
GL9	G	Male	4.41	6.28	3.8	1.53
GL10	G	Female	3.05	4.07	1.5	2.22
HL6	H	Male	4.36	6.00	4.7	2.18
HL7	H	Female	3.36	4.62	1.9	1.93
HL8	H	Male	4.15	5.60	3.1	1.77
HL9	H	Male	4.75	6.35	3.9	1.52

HL10	H	Female	3.07	4.37	1.8	2.16
IL6	I	Male	4.27	5.76	3.0	1.57
IL7	I	Male	3.65	4.94	2.3	1.91
IL8	I	Male	3.21	4.43	1.7	1.96
IL9	I	Female	3.17	4.47	1.7	1.90
IL10	I	Female	3.25	4.40	1.3	1.53
JL6	J	Male	4.20	5.83	3.4	1.72
JL7	J	Male	3.95	5.55	2.7	1.58
JL8	J	Male	4.45	6.09	3.4	1.51
JL9	J	Male	4.15	5.63	2.4	1.34
JL10	J	Female	3.40	4.74	1.8	1.69
KL6	K	Female	2.94	4.18	1.1	1.51
KL7	K	Female	2.70	3.90	1.0	1.69
KL9	K	Female	2.83	3.87	1.0	1.73
KL10	K	Male	3.91	5.64	2.7	1.50
KL11	K	Female	2.92	4.15	1.1	1.54
LL6	L	Female	3.00	3.84	1.0	1.77
LL7	L	Male	3.92	5.31	3.0	2.00
LL8	L	Female	3.30	4.57	1.6	1.68
LL9	L	Female	3.40	4.72	2.3	2.19
LL10	L	Male	4.54	6.30	3.8	1.52
ML6	M	Male	4.71	6.15	5.6	2.41
ML7	M	Male	5.60	7.71	7.5	1.64
ML8	M	Female	3.84	5.08	2.8	2.14
ML9	M	Female	3.80	5.12	2.3	1.71
ML10	M	Female	4.20	5.46	3.1	1.90
NL6	N	Male	4.20	5.78	2.8	1.45
NL7	N	Female	2.83	4.10	1.1	1.60
NL8	N	Male	4.18	5.73	2.4	1.28
NL9	N	Female	3.35	4.5	1.4	1.54
NL10	N	Female	3.72	5.19	2.6	1.86
OL6	O	Female	2.91	3.95	1.6	2.60
OL7	O	Male	4.95	6.75	5.4	1.76
OL8	O	Male	4.25	5.63	3.4	1.91
OL9	O	Female	3.24	4.69	1.9	1.84
OL10	O	Male	4.38	6.24	4.2	1.73
PL6	P	Male	4.37	5.87	3.3	1.63
PL7	P	Female	3.69	4.87	2.4	2.08
PL8	P	Female	3.57	4.71	2.4	2.30
PL9	P	Male	4.12	5.54	3.1	1.82
PL10	P	Female	2.35	4.24	0.9	1.18
QL6	Q	Female	4.21	5.70	2.8	1.51
QL7	Q	Female	3.50	4.8	2.2	1.99
QL8	Q	Male	4.85	6.56	4.9	1.74
QL9	Q	Male	4.59	6.22	3.9	1.62
QL10	Q	Female	3.40	4.75	2.1	1.96

RL6	R	Male	4.16	5.60	2.4	1.37
RL7	R	Male	4.70	6.20	4.2	1.76
RL8	R	Female	3.38	4.58	1.6	1.67
RL9	R	Male	4.37	5.80	3.5	1.79
RL10	R	Female	2.89	4.03	1.2	1.83

Table S5: Sample description of the juveniles from the feeding experiment.

Sample	Standard length (cm)	Total length (cm)	Weight (mg)	Fulton's condition index
Fasted 1	0.95	1.22	20.7	1.14
Fasted 2	1.14	1.42	41.2	1.44
Fasted 3	0.98	1.24	20.9	1.10
Fasted 4	1.15	1.47	39.0	1.23
Fasted 5	1.00	1.29	32.9	1.53
Fasted 6	1.00	1.30	33.9	1.54
Fasted 7	1.11	1.43	46.0	1.57
Fed 1	1.25	1.57	65.9	1.70
Fed 2	1.41	1.78	108.8	1.93
Fed 3	1.07	1.40	53.3	1.94
Fed 4	1.37	1.72	108.4	2.13
Fed 5	1.17	1.48	75.1	2.32
Fed 6	1.18	1.46	66.8	2.15
Fed 7	1.18	1.51	63.0	1.83

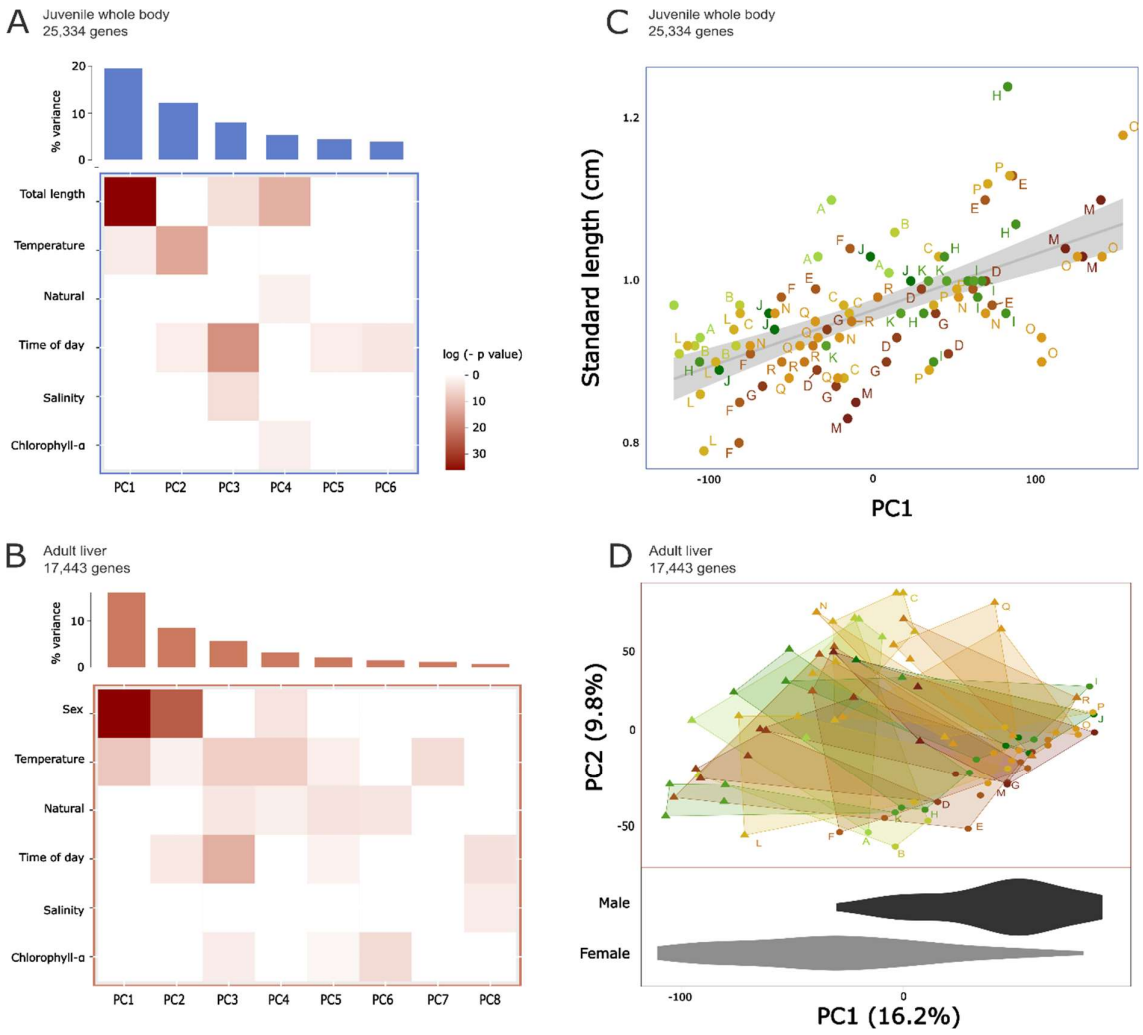
Table S6: Sample description of the adults from the feeding experiment.

Sample	Standard length (cm)	Total length (cm)	Weight (g)	Fulton's condition index	Liver weight (mg)	Hepatosomatic index
Fasted 1	4.13	4.7	1.86	1.79	11	0.59
Fasted 2	3.77	4.43	1.43	1.64	14	0.98
Fasted 3	3.3	3.99	1.15	1.81	8	0.70
Fasted 4	4.99	5.76	3.10	1.62	22	0.71
Fasted 5	4.1	4.93	2.07	1.73	18	0.87
Fasted 6	3.9	4.89	1.84	1.57	15	0.82
Fasted 7	4.46	5.24	2.51	1.74	23	0.92
Fed 1	3.86	4.66	1.86	1.84	28	1.51
Fed 2	4.17	5.02	2.68	2.12	46	1.72
Fed 3	3.87	4.63	1.77	1.78	30	1.70
Fed 4	3.89	4.63	2.12	2.14	25	1.18
Fed 5	4.27	4.97	3.01	2.45	50	1.66
Fed 6	3.49	4.17	1.42	1.96	27	1.90
Fed 7	4.96	5.77	3.85	2.00	47	1.22

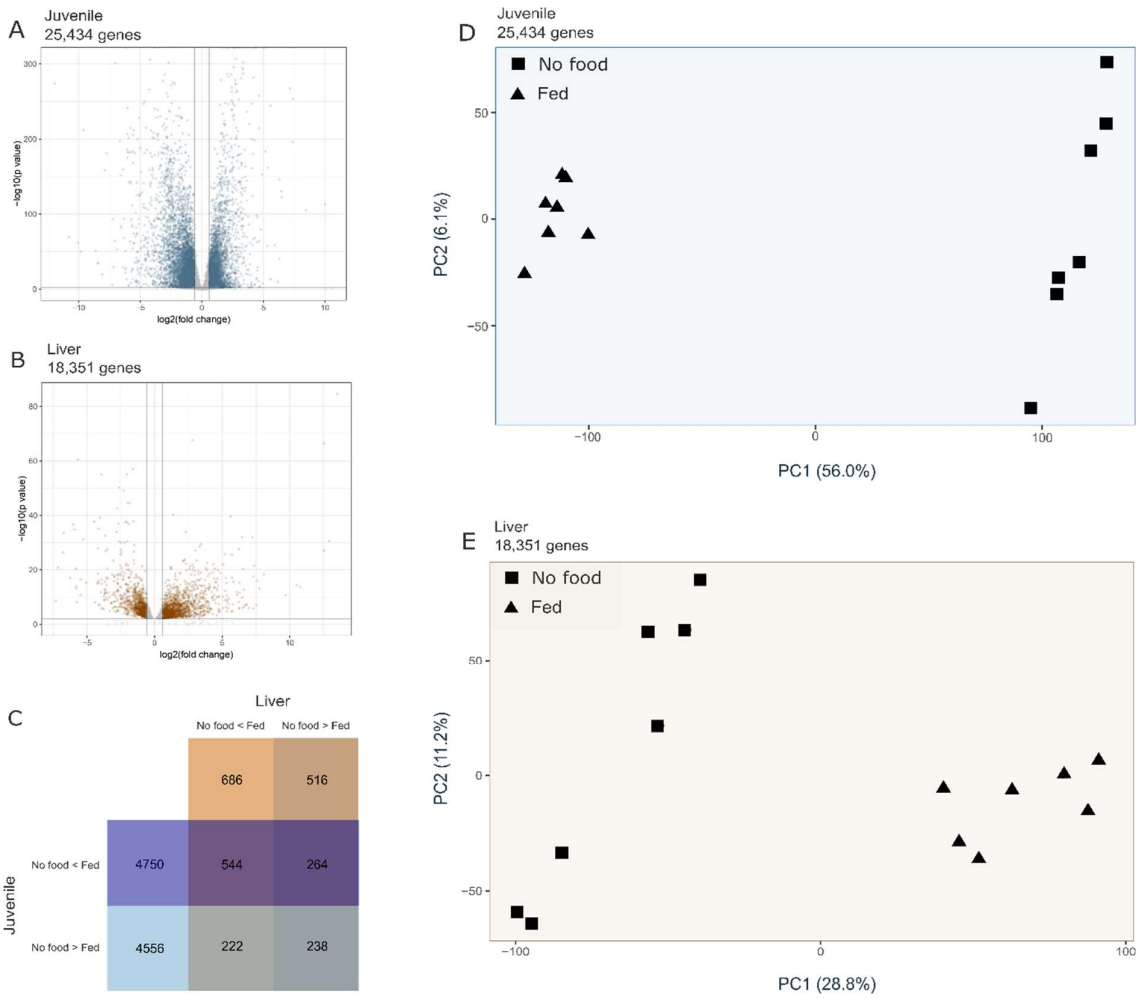
Table S7: Spearman correlation coefficient between the projected coordinate of each sample along the main axis of variance across the top dimensions of Principal Component Analyses for subsets of genes, and the mean expression level of the genes.

Juveniles		
Category	Number of genes	Correlation
Satiety	45	0.58
Glycolysis	63	0.46
Beta-oxidation	15	0.96
TCA cycle	33	0.95
Urea cycle	7	0.48
Fatty acid synthesis	46	0.19
Cholesterol biosynthesis	26	1
Corticoids	25	0.77
Thyroid hormones	27	0.83
Neurotransmission	127	0.97
Phototransduction	45	0.45
Heat shock proteins	36	0.95
Immunity	403	0.96
Reactive oxygen species	20	0.99
Xenobiotic metabolism	219	0.9

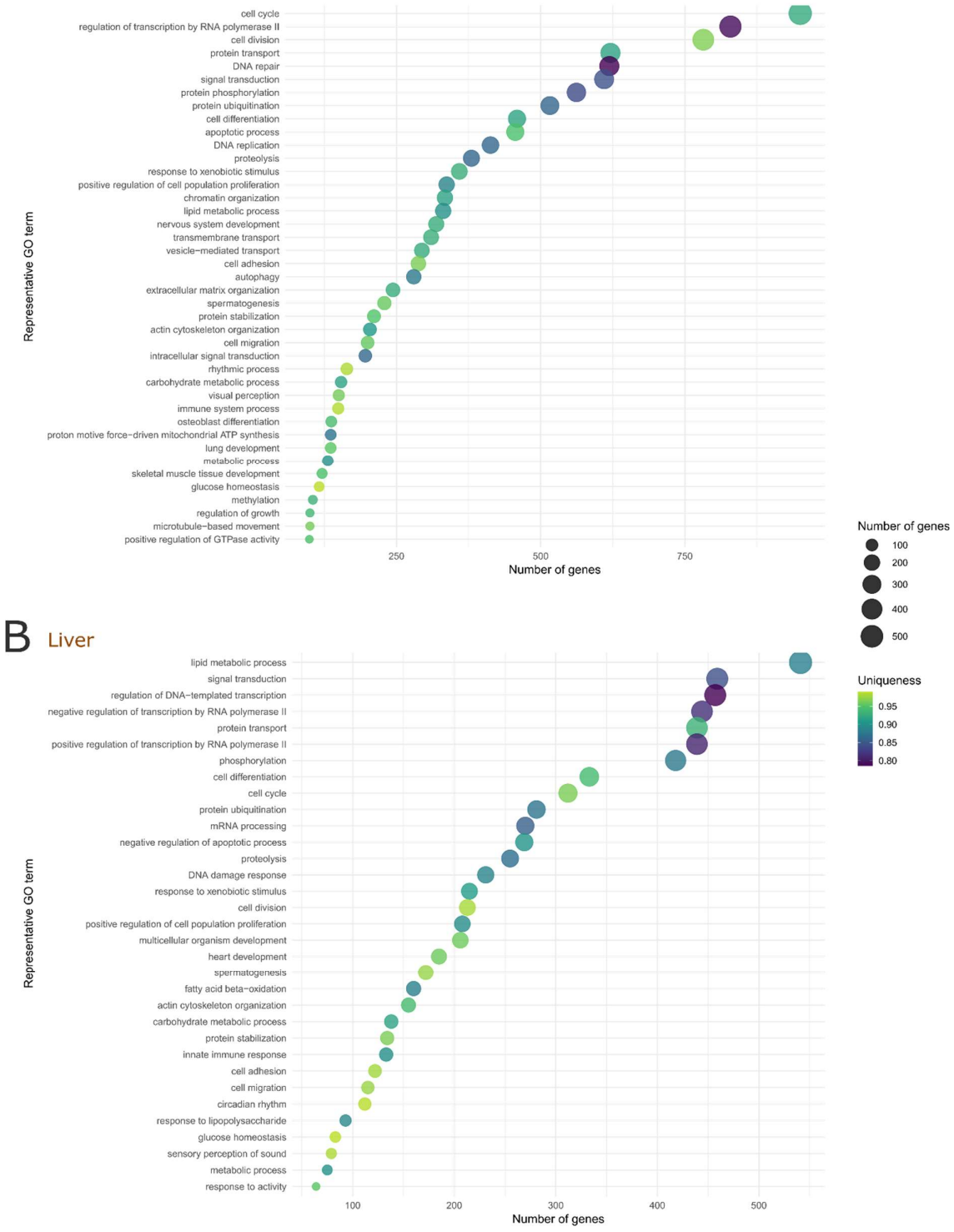
Liver		
Category	Number of genes	Correlation
Satiety	19	0.24
Glycolysis	56	0.27
Beta-oxidation	15	0.92
TCA cycle	31	0.86
Urea cycle	6	0.75
Fatty acid synthesis	39	0.97
Cholesterol biosynthesis	26	1
Corticoids	13	0.26
Thyroid hormones	16	0.92
Heat shock proteins	28	0.93
Immunity	273	0.96
Reactive oxygen species	19	0.84
Xenobiotic metabolism	195	0.96



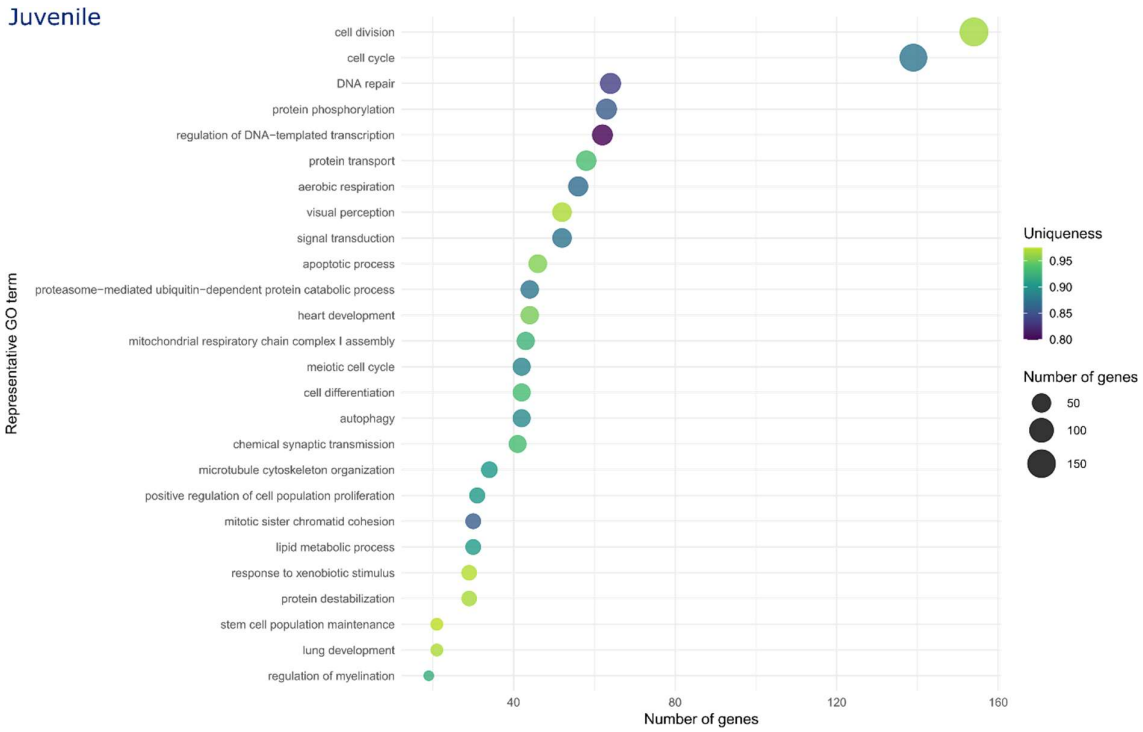
Extended Data Fig. 1: Drivers of the overall data structure in two Principal Component Analyses (PCA), using all genes with counts above 10 in at least two sites, for juvenile and liver samples separately. For juveniles (**a**) and adults (**b**), bar charts of the percentage of contribution to the variance of the top 6 PC dimensions, corresponding to over 50% of the total variance. Tile plots presenting the significance of the linear relationship between various environmental parameters and PC coordinates. The tests were linear mixed effect model with site as a random effect and variables selected from preliminary linear models with lowest BIC value using *regsubset* (*leaps* package in R). **c**, Relationship between PC1 (PCA with all genes with counts above 10 in at least two sites for juvenile samples) and the standard length of the juvenile fish. **d**, Principal Component Analysis of the expression of all genes with normalised counts (from DESEQ2) above an average of 10 in at least two sites, for adult liver samples, *i.e.,* for 17,868 genes. Triangles correspond to female livers and circles to male livers. Sex was assigned based on the tail colour of the fish at sampling (females have transparent caudal fins while males have opaque blue caudal fins). The bottom panel is a violin plot indicating the distribution of the male and female samples along PC1. There is a significant difference in PC1 values depending on sex (Kruskal-Wallis rank sum test, $\chi^2 = 40.4$, df = 1, $p < 10^{-9}$).



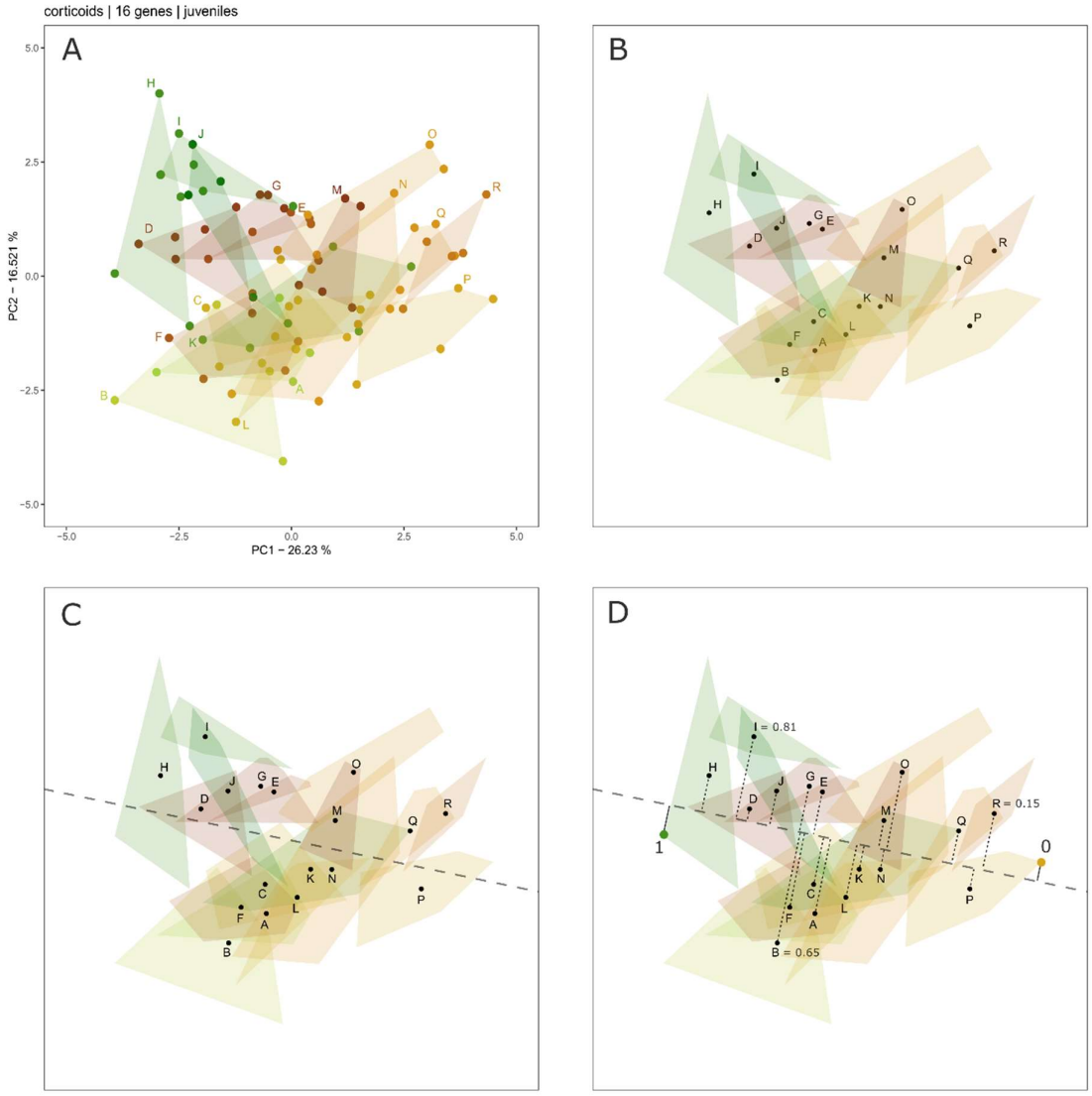
Extended Data Fig. 2: Feeding experiment results for the juvenile and liver samples. a, Volcano plot displaying the number of significantly differently expressed genes in the whole body (absolute Log2 fold change > 0.58, p < 0.01) from DESEQ2 for the juvenile feeding experiment (no food *vs.* fed). **b**, Volcano plot displaying the number of significantly differently expressed genes in the liver (absolute Log2 fold change > 0.58, p < 0.01) from DESEQ2 for the adult feeding experiment (no food *vs.* fed). **c**, Venn Diagram displaying the number of differentially expressed genes (absolute Log2 fold change > 0.58, adjusted p < 0.01) from DESEQ2 in the juvenile and liver samples, with the genes in common shown based on the intersection of the different diagram cells. **d**, Principal Component Analysis of the gene expression level across all genes (with mean expression above 10 counts) for the juvenile feeding experiment samples (7 juvenile fish with no food, 7 fed for 3 days). **e**, Principal Component Analysis of the gene expression level across all genes (with mean expression above 10 counts) for the liver feeding experiment samples (7 adult fish with no food, 7 fed for 3 days).



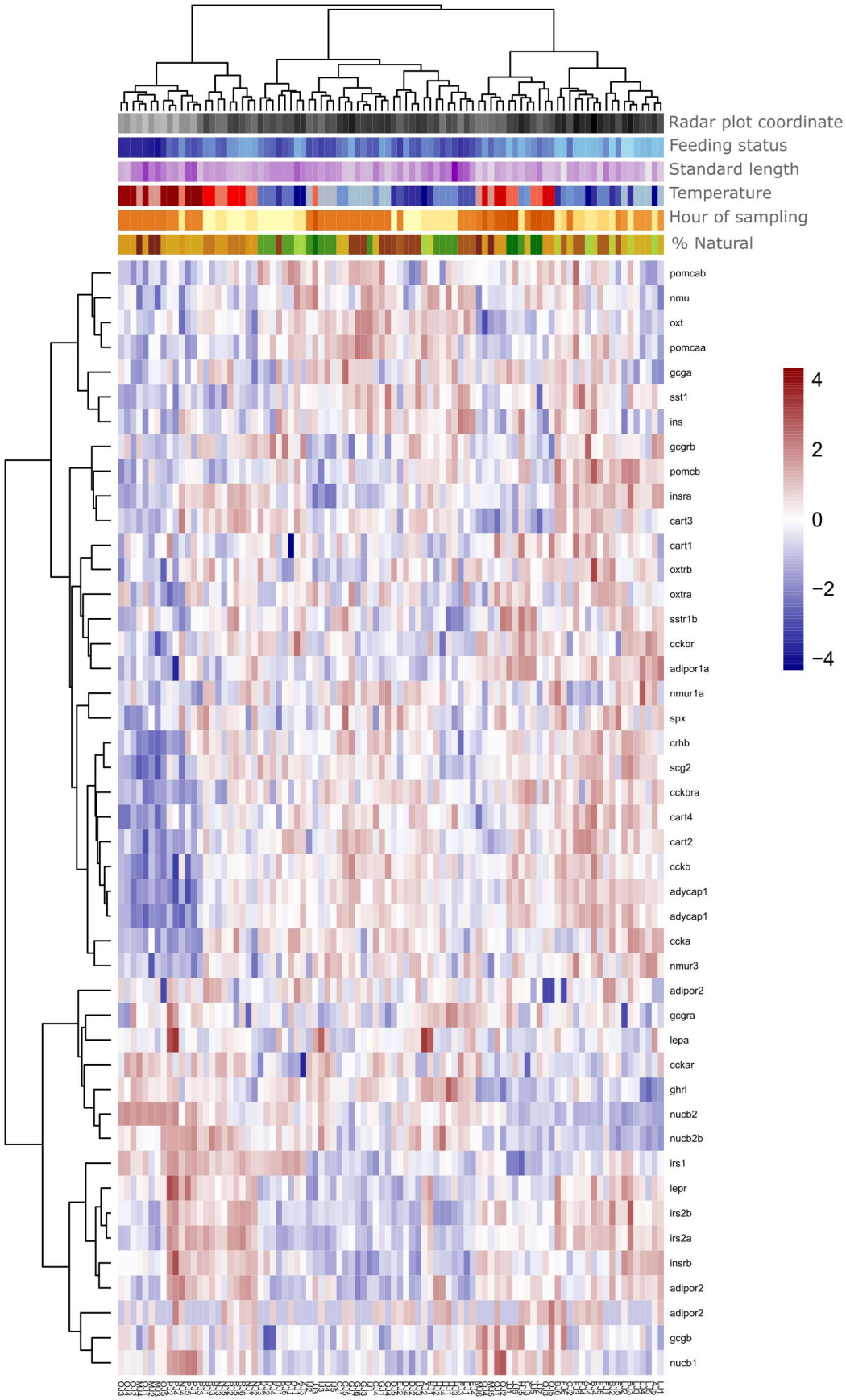
Extended Data Fig. 3: Most common GO terms across the top 10 percent genes driving the loading of dimension 1 on the Principal Component Analysis (PCAs) of the feeding experiment samples for juveniles (a) and adult liver (b). The PCAs are presented on Fig. 3. GO terms were trimmed based on semantic similarity at a threshold of 0.7 using GOSemSim and the most unique terms are presented here.



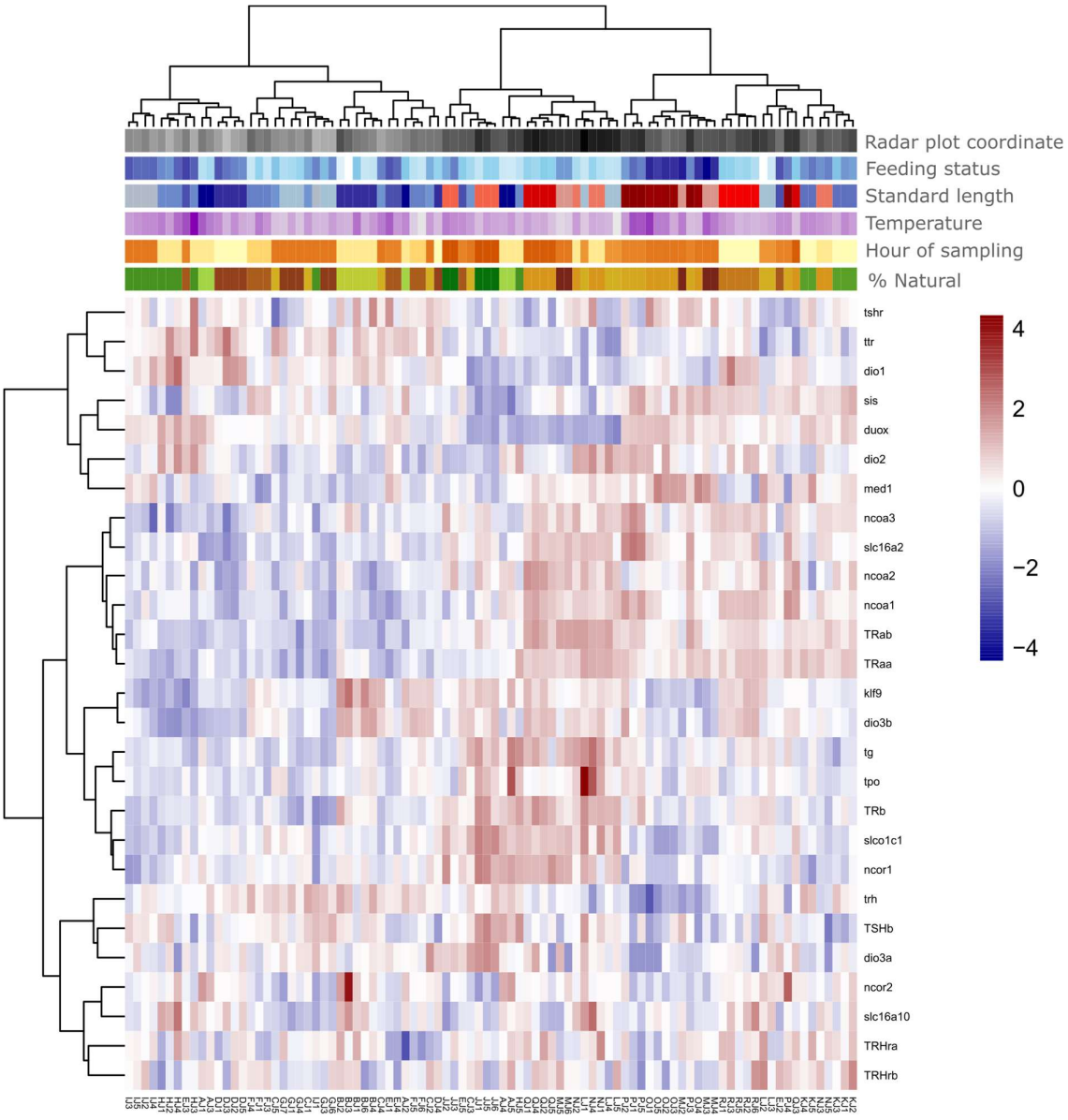
Extended Data Fig. 4: Most common GO terms across the top 10 percent genes driving the loading of dimension 2 on the Principal Component Analysis of the feeding and field samples (Fig. 3b from the main text) in juveniles. GO terms were trimmed based on semantic similarity at a threshold of 0.7 using GOSemSim and the most unique terms are presented here.



Extended Data Fig. 5: Sketch of the method to obtain the radar plots presented in Fig. 4. This is an example where the first two PC dimensions are used. The analysis includes all PC dimensions representing up to 50% of the total variance. **a**, First, the Principal Component Analysis, here based on the normalized expression level of 16 genes linked to corticoids, across 18 sites in Okinawa, with 5 samples per site, is obtained. **b**, The mean PC1/PC2 coordinate of the samples from each of the 18 sites is calculated. **c**, The direction of the main variance in the mean coordinate of the sites is then calculated, here presented as a dashed axis. **d**, The orthogonal projection of all mean site coordinates as well as the projection of each sample onto the main axis of variance is calculated. The two furthermost samples across the main variance axis are set to have values of 0 and 1. Based on this method, the orthogonal projection of all samples and sites was calculated for multiple gene sets. The results are then summarized as a radar plot in the main manuscript.

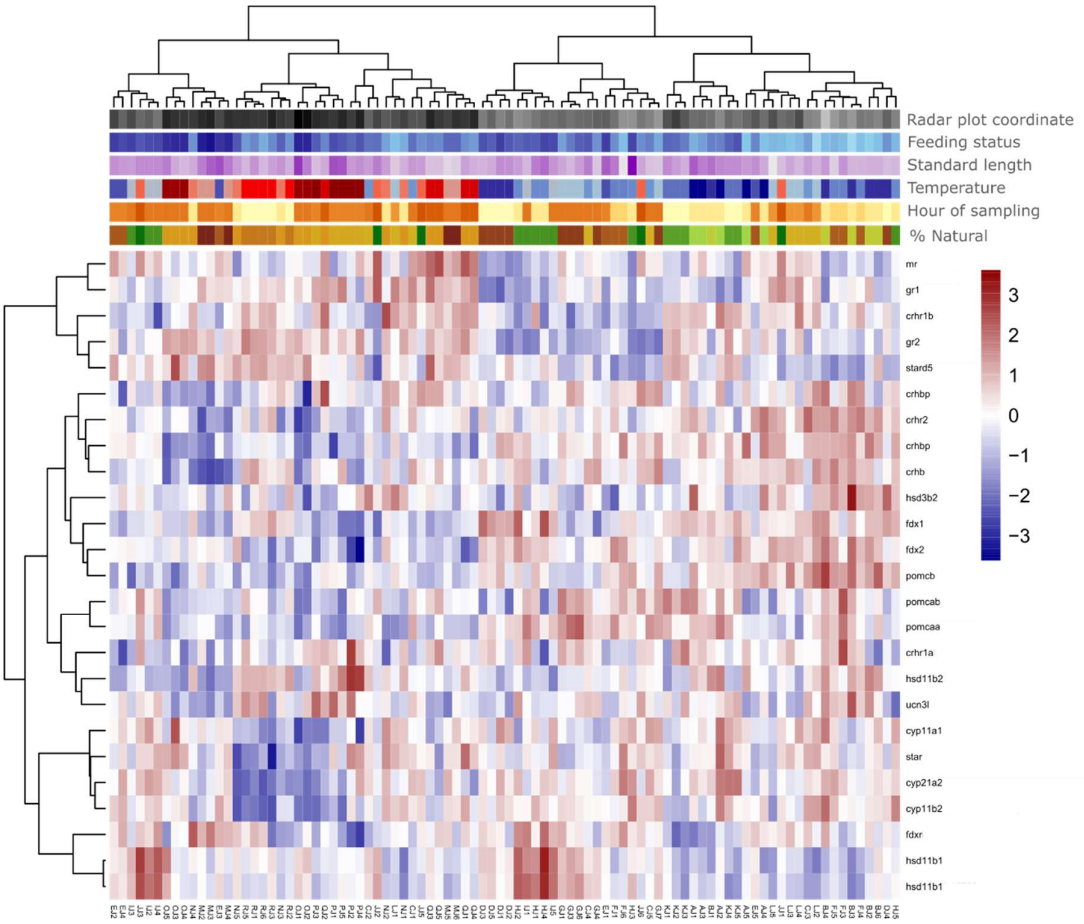

Radar plot coordinate
Feeding status
Standard length
Temperature
Hour of sampling
% Natural
pomcab
nmu
oxt
pomcaa
gcga
sst1
ins
gcgrb
pomcb
insra
cart3
cart1
oxtrb
oxtra
sstr1b
cckbr
adipor1a
nmur1a
spx
crhb
scg2
cckbra
cart4
cart2
cckb
adycap1
adycap1
ccka
nmur3
adipor2
gcgra
lepa
cckar
ghrl
nucb2
nucb2b
irs1
lepr
irs2b
irs2a
insrb
adipor2
adipor2
gcgb
nucb1
4
2
0
−2
−4

Extended Data Fig. 6: Heatmap of the vsd-normalised expression level of 45 appetite-related genes in juvenile field samples. Rows are clustered based on correlation using the Ward D2 method with *pheatmap*. Environmental (temperature, hour of sampling) and intrinsic fish variables (feeding status, standard length) are indicated above the heatmap. Increasingly dark colours indicate higher values for the radar plot coordinates, feeding status, standard length, and hour of sampling. Temperature is indicated from a blue (low temperature) to red (high temperature) scale.

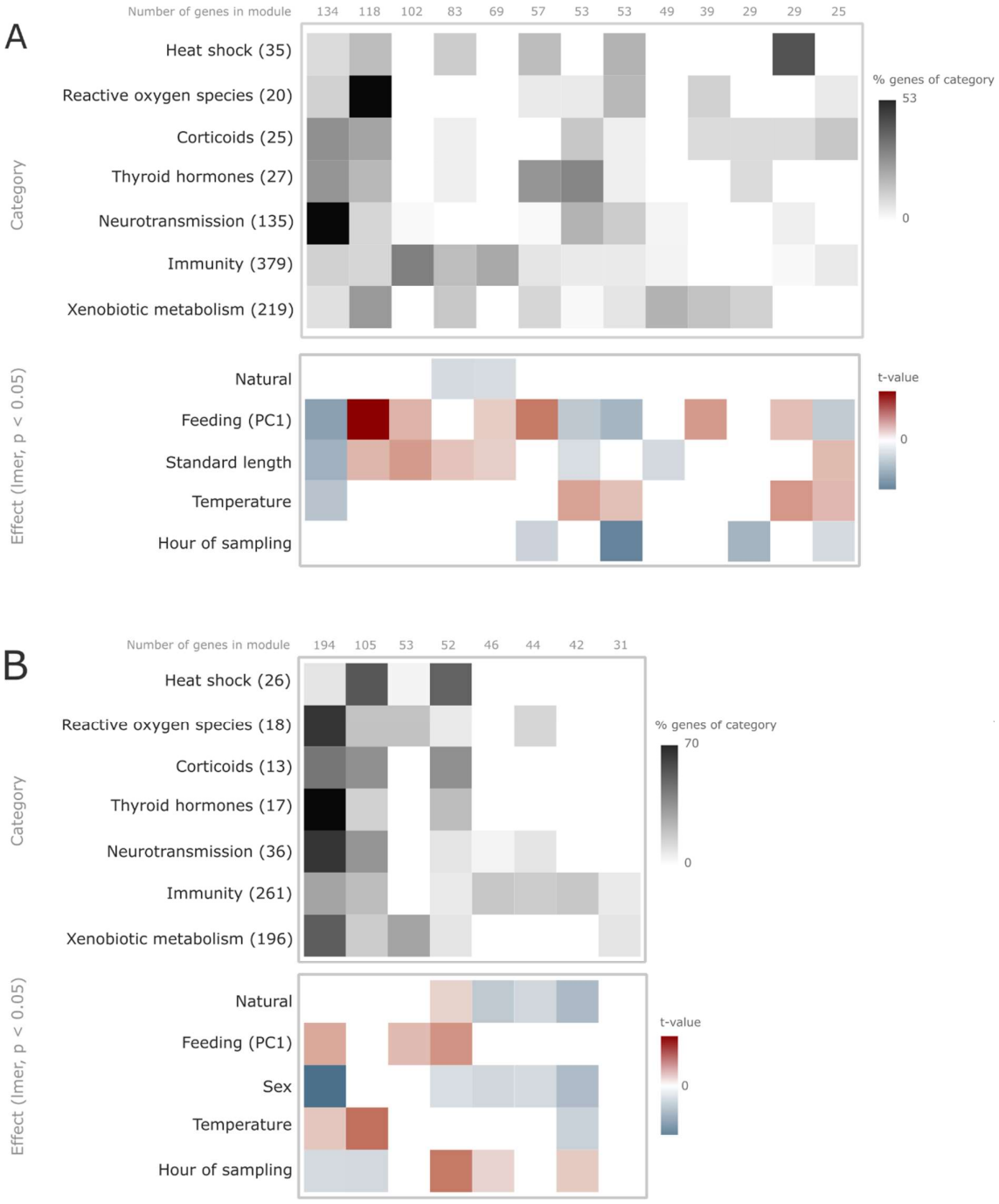


Extended Data Fig. 7: Heatmap of the vsd-normalised expression level of 27 thyroid-related genes in juvenile field samples. Rows are clustered based on correlation using the Ward D2 method with *pheatmap*. Environmental (temperature, hour of sampling) and intrinsic fish variables (feeding status, standard length) are indicated above the heatmap. Increasingly dark colours indicate higher values for the radar plot coordinates, feeding status, standard

length, and hour of sampling. Temperature is indicated from a blue (low temperature) to red (high temperature) scale.

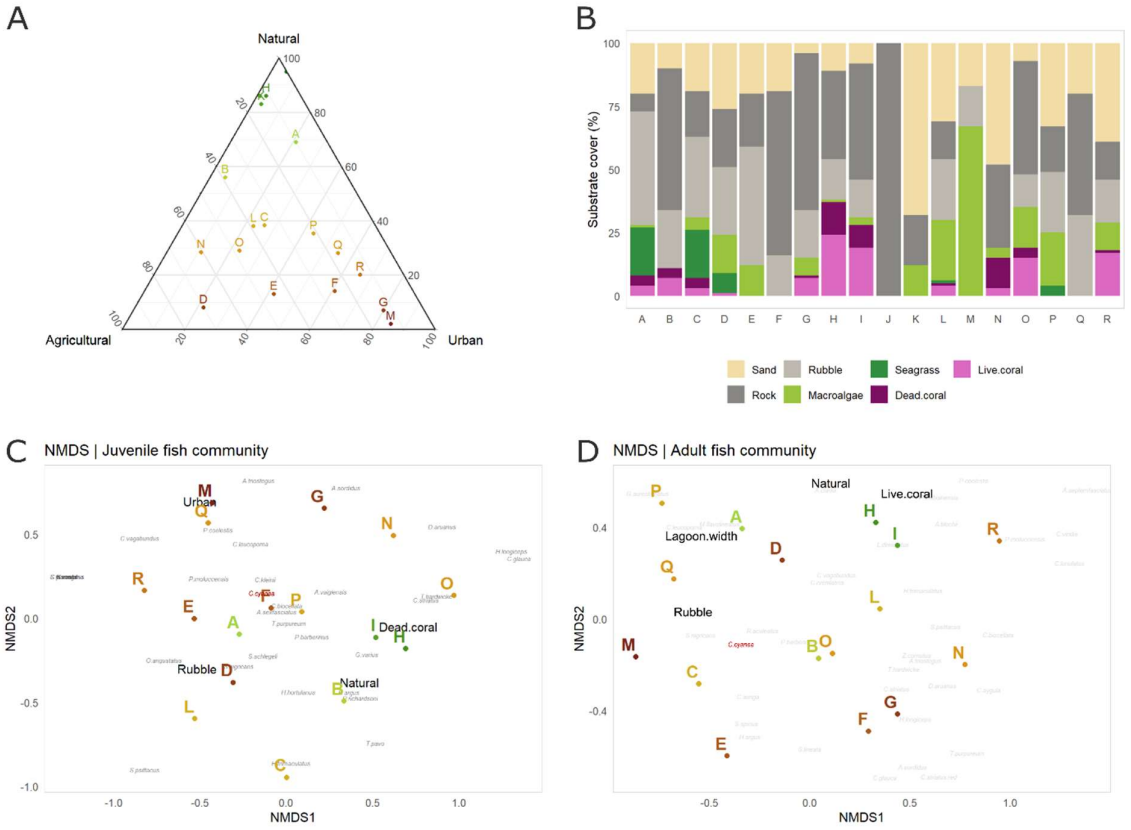


Extended Data Fig. 8: Heatmap of the vsd-normalised expression level of 25 corticoid-related genes in juvenile field samples. Rows are clustered based on correlation using the Ward D2 method with *pheatmap*. Environmental (temperature, hour of sampling) and intrinsic fish variables (feeding status, standard length) are indicated above the heatmap. Increasingly dark colours indicate higher values for the radar plot coordinates, feeding status, standard length, and hour of sampling. Temperature is indicated from a blue (low temperature) to red (high temperature) scale.

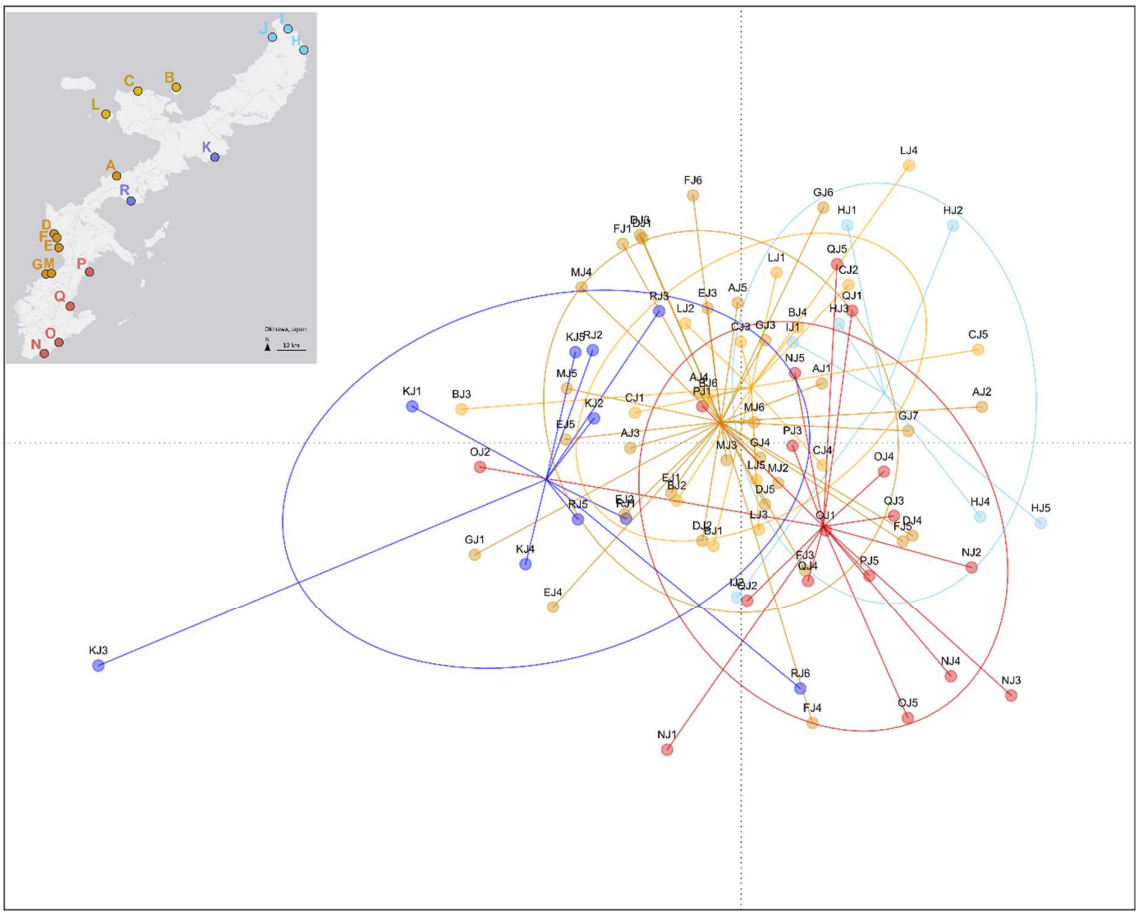


Extended Data Fig. 9: Results from a module analysis for the juvenile whole-body and adult liver gene expression. Analysis performed using WGCNA (Weighted Correlation Network Analysis) with a minimum module size of 20 genes for juveniles and 13 genes for adult liver samples (corresponding to the lowest number of genes within a category) and a soft-thresholding power of 9 for juveniles and 6 for livers. 13 modules of genes with similar expression patterns across the samples were retrieved for the juveniles and 8 for the livers. The following categories of genes were used, with 974 genes retrieved in total: heat shock proteins, reactive oxygen species, corticoids, thyroid hormones, neurotransmission, immunity, and xenobiotic metabolism. Environmental and intrinsic fish variables (percentage of natural zones

within a kilometre of the site, feeding status as derived from Fig. 3 of the main text, standard length for juveniles, sex for livers, temperature, and hour of sampling) were used in a linear-mixed effect model (with site as a random effect and a selection of fixed effects based on model dredging) to determine the relationship between the gene expression patterns in each module and environmental or intrinsic fish drivers. The resulting t-values are presented. Higher t-values (in red) indicate positive relationships between the gene expression level within the module and the given variable, and inversely (in blue). Top: Percentage of genes of each category retrieved within a given module. Bottom: t-value of a linear mixed-effect model retrieving significant relationships between environmental or intrinsic fish characteristics and gene expression patterns within the modules.



Extended Data Fig. 10: Environmental characterisation of the sites. a, Ternary plot presenting the percentage of natural, agricultural, and urban zones within 1km of the sites (excluding water). **b**, Substrate cover on three 25m transects performed parallel to shore near the sampling (maximum depth: 2m). **c**, NMDS analysis of the juvenile fish community structure and contributing environmental variables based on *vegan* function envfit ($p < 0.1$). **d**, NMDS analysis of the adult fish community structure and contributing environmental variables based on *vegan* function envfit ($p < 0.1$). No fish community structure was recorded on sites J and K due to logistical constraints.



Extended Data Fig. 11: First two discriminant functions of a Discriminant Analysis of Principal Components (DAPC) of *C. cyanea* juveniles. Colours indicate 5 zones: northwest, north, east, south, west. The analysis is based on filtered SNPs (mind = 0.25, geno = 0.25) following variant calling using the Genome Analysis Toolkit (GATK) pipeline.