Article

The evolution of facultative symbiosis in stony corals

https://doi.org/10.1038/s41586-025-09623-6

Received: 14 February 2025

Accepted: 12 September 2025

Published online: 15 October 2025

Check for updates

Shani Levy[1,6,7]✉, Xavier Grau-Bové[1,7]✉, Iana V. Kim[1], Sebastian R. Najle[1], Ewa Księżopolska[1], Anamaria Elek[1], Laia Montes-Espuña[1,2], Sean A. Montgomery[1], Tali Mass[3] & Arnau Sebé-Pedrós[1,2,4,5]✉

Most stony corals are obligate symbionts that are dependent on nutrients provided by the photosynthetic activity of dinoflagellates residing within specialized cells[1]. Disruption of this symbiotic consortium leads to coral bleaching and, ultimately, mortality[2]. However, a few coral species exhibit facultative symbiosis, allowing them to survive extended periods of bleaching[3,4]. Despite this resilience, the underlying biological mechanisms remain poorly understood. Here we investigate the genomic and cellular basis of facultative symbiosis in *Oculina patagonica*, a thermotolerant Mediterranean coral[5,6]. We sequenced and annotated a chromosome-scale genome of *O. patagonica* and built cell atlases for this species and two obligate symbiotic corals. Comparative genomic analysis revealed karyotypic and syntenic conservation across all scleractinians, with species-specific gene expansions primarily driven by tandem duplications. Single-cell transcriptomic profiling of symbiotic and naturally aposymbiotic wild specimens identified an increase in phagocytic immune cells and a metabolic shift in gastrodermal gene expression from growth-related functions to quiescent, epithelial-like states. Cross-species comparison of host cells uncovered *Oculina*-specific metabolic and signalling adaptations indicative of an opportunistic, dual-feeding strategy that decouples survival from symbiotic state.

Stony corals (Scleractinia) are colonial cnidarians that thrive in tropical and subtropical seas, where they sustain highly biodiverse reef ecosystems[7]. Thermal stress causes coral bleaching, a process in which corals expel the dinoflagellate symbionts that provide corals with nutrients, eventually leading to coral death[2,8,9]. Although bleaching poses a major threat to tropical corals, the temperate species *O. patagonica* regularly recovers from seasonal bleaching[10]. Unlike most other corals, *O. patagonica* establishes facultative symbiosis with dinoflagellates (Symbiodiniaceae)[11,12]. This facultative nature enables *O. patagonica* to survive in both symbiotic and aposymbiotic (without algae) states, which is crucial for its ability to withstand environmental stressors such as thermal fluctuations and increased sedimentation that are prevalent in the Mediterranean Sea[5,13], where *O. patagonica* is a widespread species[5,14–17].

Periodic bleaching of *O. patagonica* occurs as sea-surface temperatures peak during summer months, particularly along the Mediterranean Levantine coasts where temperatures can reach up to 31 °C (ref. 18), and is followed by a recovery phase in autumn as temperatures decrease[5]. This cyclical pattern of bleaching and recovery demonstrates the ability of the coral to regain its symbiotic algae when conditions improve, highlighting a level of resilience that is not commonly seen in many tropical coral species. Further contributing to the resilience of *O. patagonica* is its ability to inhabit diverse light environments, from sunlit environments to shallow water regions with low-light conditions, such as caves and undercut rock formations. In these dimly lit habitats, corals are frequently observed in a naturally bleached or aposymbiotic state, probably owing to the reduced availability of light necessary for sustaining their symbiotic algae[19]. This naturally bleached state is not necessarily indicative of stress, but rather reflects an adaptive strategy that enables *O. patagonica* to thrive across a range of light environments, further underscoring its ecological plasticity.

Here we investigate the genomic and cellular basis of *O. patagonica* facultative symbiosis. We sequenced and assembled to chromosome scale the genome of *O. patagonica*, and compared it to other corals and cnidarians to define evolutionary dynamics of gene content and macrosynteny. We used single-cell RNA sequencing (scRNA-seq) to characterize cell-type composition and gene-expression changes between aposymbiotic and symbiotic *O. patagonica* sampled from the eastern Mediterranean. Additionally, we built scRNA-seq atlases for two tropical obligate symbiotic corals (*Acropora millepora* and *Stylophora pistillata*) and systematically compared them with *O. patagonica*. This comparative genomic and transcriptomic approach aims to clarify the molecular mechanisms underlying coral–algal symbiosis and immunity, and the unique adaptations of these corals to their environment.

Oculina genome organization and evolution

We sequenced and assembled the genome of the scleractinian *O. patagonica*, combining long reads and micro-C chromatin contact maps.

[1]Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology, Barcelona, Spain. [2]Universitat Pompeu Fabra (UPF), Barcelona, Spain. [3]Department of Marine Biology, University of Haifa, Haifa, Israel. [4]ICREA, Barcelona, Spain. [5]Tree of Life Program, Wellcome Sanger Institute, Hinxton, UK. [6]Present address: Department of Biology and Environment, University of Haifa - Oranim, Tivon, Israel. [7]These authors contributed equally: Shani Levy, Xavier Grau-Bové. ✉e-mail: slevy@sci.haifa.ac.il; xavier.graubove@crg.eu; arnau.sebe@crg.eu

The resulting chromosome-scale assembly spans 507 Mb and consists of 14 chromosome-level scaffolds (Fig. 1a,b and Supplementary Fig. 1), in agreement with the number of chromosomes reported in other stony corals[20,21]. We then used strand-specific RNA-sequencing (RNA-seq) data to annotate 39,482 genes. Transposable elements represented 23.6% of the genome, with the major class being LTR retrotransposons (5.4%), followed by helitrons (3.8%) and Mutator DNA transposons (3.2%) (Supplementary Fig. 2a,b, Supplementary Table 1).

A high degree of microsyntenic conservation has been reported across scleractinians[22]. Here we extended these analyses to measure macrosyntenic conservation using ten other available cnidarian chromosome-level assemblies[23,24] (Fig. 1c and Extended Data Fig. 1a–c). We defined 27 cnidarian ancestral linkage groups (ALGs)[25] and reconstructed their evolutionary history along cnidarian phylogeny (Extended Data Fig. 1c). This analysis revealed two fusion-without-mixing events shared by all scleractinians and highlighted a high degree of macrosyntenic conservation in the 14 chromosomes across scleractinian corals (Fig. 1c and Extended Data Fig. 1c), although a higher degree of ALG rearrangement was observed in *A. millepora* (even compared to other *Acropora* species). An exception to the globally conserved genome organization of *O. patagonica* is the existence of a 12 Mb region in chromosome 4 that lacks signal from any cnidarian ALG. This low-conservation region is enriched in transposable elements and in evolutionarily young genes mostly restricted to *O. patagonica* and its close relative *Oculina arbuscula*, as well as conserved genes involved in chromatin remodelling, DNA replication, spermatogenesis and male sexual differentiation (Supplementary Table 2). This region appears insulated in 3D chromatin contacts and genes within it are expressed at very low levels (Fig. 1b and Extended Data Fig. 1d). Given that *O. patagonica* is a gonochoric species, we hypothesize this may represent a sex-determining region, which are often characterized by low recombination rates, high transposable element content, low gene density and rapid divergence[26].

The genome of *O. patagonica* and those of other scleractinians encode high numbers of genes compared with other cnidarians (Fig. 1a,d), which have been hypothesized to be the result of whole-genome duplication events[27] or gene duplications[22,28,29]. We found no evidence of an ancestral whole-genome duplication contributing to this pattern in the scleractinian lineage (Supplementary Fig. 2c), in line with recent studies[21]. However, we identified tandem duplications as a major contributor to the gene family expansions in *O. patagonica*: 2,851 genes belonging to 1,092 gene families were affected (7.2% of the total gene complement; Fig. 1e and Supplementary Table 3). Tandem duplications are, in fact, the dominant mode of paralogy among the families expanded in this and other scleractinian genomes (for example, 80% of the families with paralogues in *O. patagonica* have tandem-duplicated genes). A sizeable fraction of tandem duplication events in *O. patagonica* could be identified in *O. arbuscula* (Fig. 1f). These were enriched in genes encoding proteins that are localized to cell vesicles (endosome, lysosome and endosome–lysosome transport). Conversely, we found little overlap between the tandem-duplicated families of various corals, hinting at independent expansions in each species. For example, only 80 out of 1,092 families that are expanded in *O. patagonica* are shared with *S. pistillata* and *A. millepora*. These ancient gene duplications are enriched in immune-related functions and tyrosine kinase signalling (Fig. 1g).

Overall, the *O. patagonica* genome highlights the high conservation among stony coral genomes, at the level of both karyotype and macrosynteny. Gene content is dominated by tandem duplications and these gene expansions are paralleled by the expansion of transposable elements in these same genomes (Fig. 1d). Together, these processes explain the relatively large genome sizes and high gene counts in scleractinian corals.

Conserved coral cell-type repertoires

To study the cellular basis of *O. patagonica* facultative symbiosis, we sampled 29,723 single-cell transcriptomes from both naturally bleached (aposymbiotic) and non-bleached (symbiotic) adult colonies. We also constructed single-cell atlases from two tropical stony corals: *A. millepora* (28,736 cells) and *S. pistillata* (15,053 cells) (Fig. 2a and Supplementary Fig. 3). In brief, adult colonies from each species were collected, dissociated and fixed with a modified ACME (acetic–methanol) protocol[30,31]. Cells were sorted by fluorescence-activated cell sorting (FACS) to remove doublets, debris and ambient RNA, before encapsulation and transcriptome capture using 10X Genomics 3′-end scRNA-seq technology. We sequenced libraries to an average depth of 35,000 reads per cell to obtain a minimum library sequencing saturation of 80% (Supplementary Fig. 3a). We obtained a median of 1,278, 1,662 and 1,724 unique molecular identifiers (UMIs) per cell in *O. patagonica*, *A. millepora* and *S. pistillata*, respectively. We applied the Metacell algorithm[32] to group cells into transcriptionally coherent clusters (metacells), which constitute our basic unit for downstream analysis (Supplementary Table 4).

On the basis of gene-expression patterns (Supplementary Figs. 4–6 and Supplementary Table 5) and comparison with previously published cnidarian atlases[33–35] (Supplementary Fig. 3d,e), we defined 31, 28 and 25 cell types in *O. patagonica*, *A. millepora* and *S. pistillata*, respectively. We compared and grouped cell types across species using co-expression of orthologous genes (Fig. 2b and Extended Data Fig. 2a). This revealed strong cross-species similarities, further supported by co-expression of key transcription factors driving these cell identities (Extended Data Fig. 2b). For example, all neurons share the expression of an *Asc* homologue, and we identified two broad neuronal groups across species defined by expression of *Islet*, *Gata* and *Ptf1a* and co-expression of *Pou4* and *FoxL2*, as observed in other cnidarians[33]. Moreover, cell-type composition was similar across species: for example, 4–5% of the cells were annotated as skeleton-forming calicoblasts, 32–39% were annotated as gastrodermis, and 15–20% were annotated as epidermis (Fig. 2c). An exception was the higher frequency of gland and digestive filament cells in *O. patagonica* compared with the other species (12% versus 3–5% and 11% versus 1–3%, respectively). Gland cells are involved in the secretion of digestive enzymes, mucus or toxins, whereas digestive filament cells form the supportive tissue of the mesenteries[33].

We used cell-type gene-expression conservation to interrogate functional constraints determining the high microsyntenic conservation observed in *O. patagonica* (Extended Data Fig. 3a,b). Collinear genes that are syntenic in other species show consistently higher expression conservation (Extended Data Fig. 3c). This could be explained by stronger sequence-level conservation in the upstream non-coding (Extended Data Fig. 3d), in line with the idea that *cis*-regulatory constraints are a major determinant of microsyntenic conservation[36]. Cell atlases also enabled us to study the functional fates of the abundant tandem-duplicated gene paralogues in coral genomes (Fig. 1e). In all the species, around 50% of the tandem-duplicated genes show signatures of expression divergence, followed by 25% of cases of redundant expression among tandem duplication genes (Fig. 2d,e). Complete loss of expression of tandem duplication genes appeared to be a rare event (only five genes in *S. pistillata*) consistent with most tandem duplication genes being under purifying selection (ratio of non-synonymous substitutions to synonymous substitutions (K_a/K_s) < 1; Extended Data Fig. 3e). In *O. patagonica*, tandem-duplicated gene families did not appear to be biased towards expression in any particular cell type (Extended Data Fig. 3f).

Our single-cell atlases reveal an extensively conserved cell-type repertoire across stony corals and help us to interrogate the functional constraints that underlie the highly conserved genomic organization in these organisms. The complete dataset can be explored in an interactive database (https://sebelab.crg.eu/multicoral-sc-atlas/).

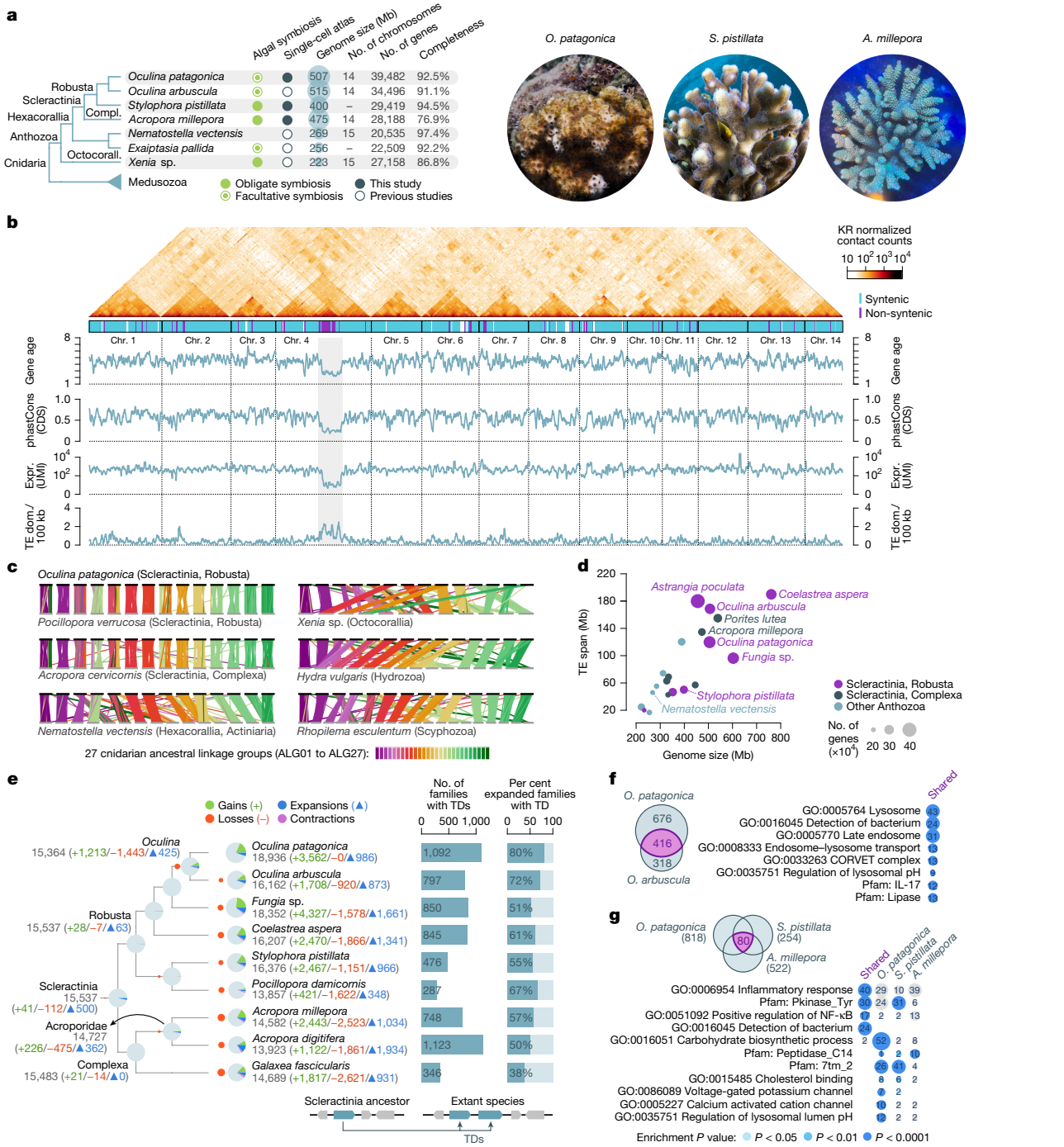


Fig. 1 | The *O. patagonica* genome. a, Left, cladogram depicting phylogenetic relationships among anthozoan cnidarians with available single-cell transcriptomic atlases. For each species, we indicate the type of symbiotic relationship established with dinoflagellate algae, its assembled genome size[20,23,24,29,54] (also represented by dot size), number of genes and BUSCO completeness[20,23,24,29,54]. Right, representative photographs of *O. patagonica*, *S. pistillata* and *A. millepora*. Compl., Complexa; Octocorall., Octocorallia. **b**, Top, micro-C contact map of the 14 chromosomes assembled for the *O. patagonica* genome, measured as Knight-Ruiz (KR) normalized counts. Bottom, median gene age, mean PhastCons conservation in coding regions (CDSs), gene expression (UMI) and transposable element (TE)-related gene density (TE domains (dom.) per 100 kbp) calculated in sliding 1-Mb windows along the chromosomes (20% step). Gene ages are displayed as categorical values (1, *O. patagonica*-specific; 8, Anthozoa). Expr., expression. **c**, Conservation of ALGs between *O. patagonica* (top of each diagram) and selected cnidarian species (bottom). Vertical lines represent synteny between homologous gene pairs, colour-coded by ALG (complete analysis in Extended Data Fig. 1a–c). **d**, Scatter plot of genome size and span of repetitive regions for selected cnidarians. Dot size reflects the number of annotated genes. **e**, Evolutionary reconstruction of gain, loss, expansion and contraction of orthologous groups of genes in Scleractinia. Bar plots to the right of the tree indicate the number of ancestral single-copy orthology groups with tandem-duplicated genes in each extant genome, and the relative frequency of this mode of gene duplication. TD, tandem duplications. **f**, Overlap between the gene families with tandem duplications in *O. patagonica* and *O. arbuscula*, and selected functional terms enriched in the shared set (*P* values result from one-tailed Fisher's exact tests with elim correction for Gene Ontology (GO) categories and false discovery rate (FDR)-adjusted hypergeometric tests for Pfam protein domains). **g**, Overlap between the gene families with tandem duplications in *O. patagonica*, *S. pistillata* and *A. millepora*, and selected functional terms enriched in the species-specific and shared sets (statistical analysis as in **f**).

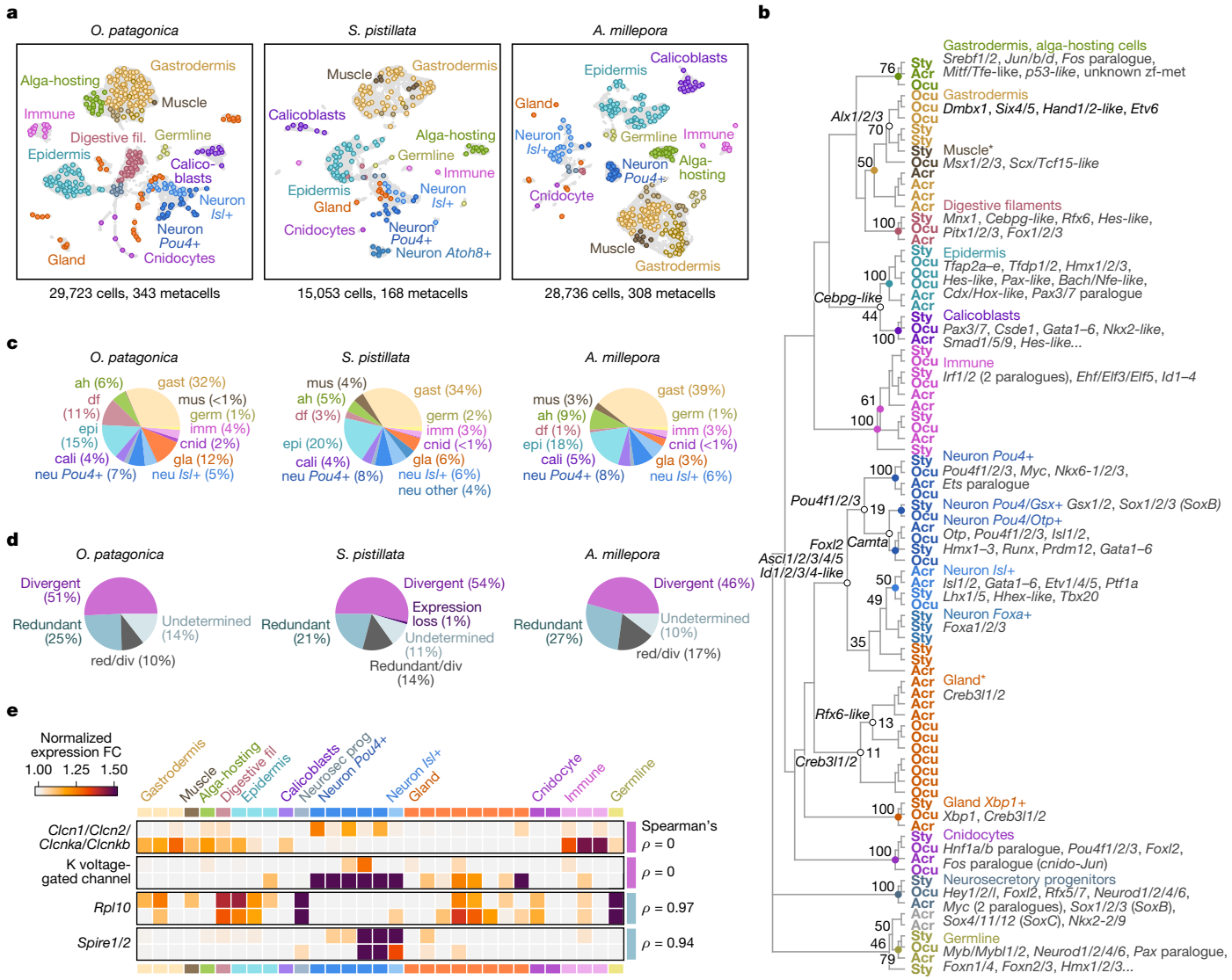


Fig. 2 | Stony coral cell-type diversity and gene-expression programmes. **a**, Two-dimensional uniform manifold approximation and projection (UMAP) projection of the three scleractinian coral single-cell transcriptomic atlases: *O. patagonica*, *S. pistillata* and *A. millepora* (obtained from four specimens, one specimen and two specimens, respectively). Grey dots represent individual cells. The larger coloured dots represent high-granularity cell clusters ('metacells') and are placed at the median centroid coordinates of the cluster. Metacells are colour-coded according to their cell type of origin. Fil., filaments. **b**, Cell-type clustering of the three scleractinian corals obtained using the UPGMA algorithm on binarized gene-expression matrices, with bootstrap supports in selected nodes. In key nodes of the tree, we indicate selected transcription factors shared by the 'descendant' cell types along the tree, obtained using probabilistic ancestral character estimation. Acr, *A. millepora*; Ocu, *O. patagonica*; Sty, *S. pistillata*. **c**, Cell-type frequencies in each coral atlas. ah, alga-hosting; cali, calicoblast; cnid, cnidocyte; df, digestive filament; epi, epidermis; gast, gastrodermis; germ, germline; imm, immune; gla, gland; mus, muscle; neu, neuron. **d**, Frequency of co-expression patterns of the sets of tandem-duplicated genes in each coral species. Pairs of paralogues within each set of tandem duplicates are recorded as having a redundant expression profile within a species if their cell-type-level normalized expression values are correlated (Spearman's $\rho \geq 0.6$) and divergent if they are not correlated ($\rho < 0.4$). red/div, redundant/divergent. **e**, Normalized expression of four sets of tandem-duplicated genes in *O. patagonica*, illustrating two cases of divergence and two of redundancy. neurosec prog, neurosecretory progenitors.

Cell-type responses to aposymbiosis

We next focused on the analysis of host cells in *O. patagonica*. We confirmed host cell identity in all three species by comparing their transcriptomes with those of FACS-sorted alga-positive coral cells using massively parallel scRNA-seq (MARS-seq)[35]. To this end, we mapped the MARS-seq transcriptomes to each coral genome and a set of reference Symbiodiniaceae genomes from each major clade (*Symbiodinium microadriaticum* (formerly clade A), *Breviolum minutum* (formerly clade B), *Cladocopium goreaui* (formerly clade C) and *Fugacium kawagutii* (formerly clade F))[37], and selected the cells exhibiting signal from both host and at least one algal species. Then, we projected the obtained single-cell transcriptomes into the reference atlas (Methods) to confirm that, in all three species, the majority of alga-positive cells mapped to the gastrodermal alga-hosting cells (more than 72%; Fig. 3a). This experiment also allowed us to identify the dominant Symbiodiniaceae genera in each coral species: *Symbiodinium* and *Breviolum* in *O. patagonica*; *Symbiodinium* in *S. pistillata*, and *Cladocopium* in *A. millepora* (Fig. 3b).

Having confirmed their identity, we observed that *O. patagonica* host cells are less transcriptionally differentiated from gastrodermal cells than the host cells from the obligate symbiotic corals (Fig. 3c). This pattern was also apparent in the single-cell 2D projections (Fig. 2a) and in the gene-expression maps (Supplementary Figs. 4–6), and it could explain why 22% of the cells with detected dinoflagellate in the *O. patagonica* MARS-seq experiments are classified as gastrodermis

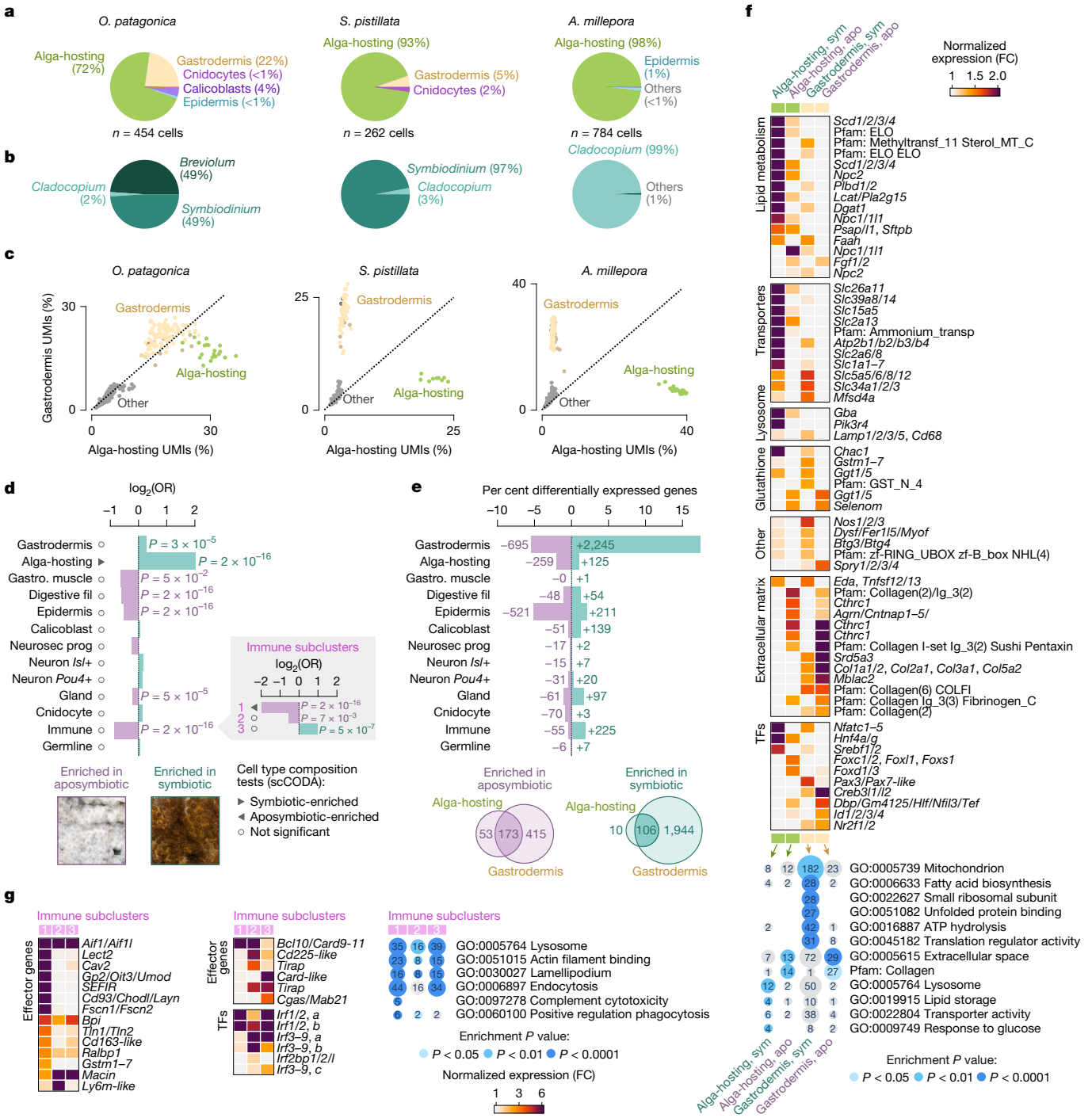


Fig. 3 | Cell-type-specific differences between aposymbiotic and symbiotic *O. patagonica* specimens. a, Cell-type classification of the transcriptomes of individually sorted alga-positive cells in *O. patagonica*, *S. pistillata* and *A. millepora*. Sorted cell transcriptomes were sequenced using MARS-seq and mapped to the reference atlas of each species using anchor genes. **b**, The most abundant Symbiodiniaceae clade in each sorted alga-positive cell, as determined by the number of UMIs obtained when each transcriptome was mapped to representative Symbiodiniaceae species. **c**, Fraction of UMIs from genes assigned to the gastrodermis and alga-hosting gene modules in each coral. **d**, Change in cell-type composition between the symbiotic and aposymbiotic samples of *O. patagonica*, measured as the $\log_2$ odds ratio (OR) of symbiotic/aposymbiotic frequencies for each cell type. Inset, change in composition for three subclusters of immune cells. The statistical significance of the biases was evaluated using two-tailed Fisher's exact tests (*P* values are shown where $P < 0.05$). Triangles and circles next to each cell type indicate consistency with an scCODA analysis of cell-type compositional changes (at probability greater than 95%). Gastro., gastrodermis; OR, odds ratio. **e**, Top, number of differentially expressed genes between the symbiotic and aposymbiotic cells for each cell type in *O. patagonica*. Significance of differential expression was evaluated using a two-tailed Wilcoxon rank-sum test on gene counts, with Bonferroni *P* value correction. Bottom, Venn diagrams indicate the overlap between the genes enriched in symbiotic or aposymbiotic cells for the alga-hosting and gastrodermis clusters. **f**, Top, selected differentially expressed genes between symbiotic (sym) and aposymbiotic (apo) host cells and gastrodermis. Bottom, representative functional terms (dot size represents the number of genes and colour indicates statistical significance in one-tailed Fisher's exact tests with elim correction for Gene Ontologies, and FDR-adjusted hypergeometric tests for Pfam protein domains). TFs, transcription factors. **g**, Selected marker genes for the immune cell subclusters in *O. patagonica* (left and middle) and representative enrichment terms (right). Statistical analysis as in **f**.

according to their gene-expression programme (Fig. 3a). Such plasticity may underlie the facultative nature of *O. patagonica*, enabling it to survive without symbionts by relying on alternative nutritional and metabolic pathways, unlike obligate symbiotic corals that are heavily dependent on their symbionts for survival.

We analysed the differences in cell-type composition and gene expression between symbiotic and aposymbiotic *O. patagonica* specimens by separating the joint single-cell atlas into these two conditions (14,835 and 14,888 cells, respectively). As expected, host cells were strongly enriched in symbiotic specimens (9.5% of the total cells sampled), but we also detected host cells in the aposymbiotic samples (2.5% of the total cells) (Fig. 3d). This probably indicates that many cells expressing the 'host cells' transcriptional programme remain even after the dinoflagellate symbiont is no longer present, and prompted us to examine differences in gene expression between host cell states in aposymbiotic and symbiotic *O. patagonica* (Fig. 3e). These observations, including biases in frequency of alga-hosting cells and the relative lack of differentiation between host and non-host gastrodermal cells, are consistent in another *Oculina* species, *O. arbuscula*[38] (Supplementary Fig. 7).

Host cells in symbiotic animals strongly expressed genes involved in lipid metabolism (including the transcription factor genes *Nfat*, *Hnf4A* and *Srebf1/2*, the latter of which are involved in controlling lipogenesis in mammals), molecular transport, lysosomal activity and glucose metabolism (Fig. 3f and Supplementary Table 6). By contrast, host cells in aposymbiotic animals express extracellular matrix components, such as collagen genes, and matrix regulatory genes, such as *Cthrc1* and *Sparc*. There are also major differences in gene expression between gastrodermal cells (Fig. 3e), with symbiotic gastrodermis expressing many genes involved in mitochondrial activity, protein translation and lipogenesis (Fig. 3f and Supplementary Table 7), suggesting distinct metabolic activities of gastrodermal cells when surrounding host cells with algal symbionts in them. Gastrodermal cells from symbiotic animals also express multiple genes of the glutathione pathway (Fig. 3f), which are likely to be involved in redox protection against oxygen produced by the algae, and specifically express *Nos*, encoding an enzyme that producing the signalling molecule nitric oxide. These results suggest that the presence of algal symbionts affects not only the host cells that harbour them but also the surrounding gastrodermal cells, reflecting a specific response to the symbiotic state.

Another major difference in cell-type composition was the enrichment of a specific type of immune cells in aposymbiotic animals[39,40] (Fig. 3d). This group of immune cells expresses unique combinations of *Irf* transcription factor genes, genes involved in actin cytoskeleton (such as *Talin* and *Ralbp1*), filopodia formation (for example, *Fascin*), phagocytosis (for example, *Caveolin*) and response to pathogens (for example, *SEFIR*, *Gp2* and multiple components of the complement system) (Fig. 3g). Other immune cells that are not enriched in aposymbiotic animals share expression of immune regulators such as *Aif1* and *Irf*s, and different sets of effector genes involved in bacterial (for example, *Tirap*, MACPF genes and genes encoding lipopolysaccharide-binding proteins) and viral (for example, *Cgas*, *Cd225* and *Endod1*) responses. The enriched immune cells in aposymbiotic animals could be amoeboid cells that actively engulf and clear foreign particles from the coral tissues, including remnants of dinoflagellate or host cells, after the dissolution of the symbiotic partnership.

Evolution of the host cell gene programme

To understand the unique features of *O. patagonica* host cells, we reconstructed the shared and novel expressed genes in coral alga-hosting cells across 250 million years of scleractinian evolution (Fig. 4a) and we examined the cell-type expression patterns of functionally relevant genes across multiple coral species (Fig. 4b and Extended Data Figs. 4–7). For all cross-species comparisons, only symbiotic *Oculina* samples were included. All scleractinian alga-hosting cells co-expressed many genes linked to lysosomal function (Fig. 4a,b and Extended Data Fig. 4), as expected given that dinoflagellates reside intracellularly in a modified lysosome that is often called the symbiosome[41]. These include genes encoding lysosomal membrane proteins, such as *Lamp*, diverse transporters for metabolites such as cholesterol (*Npc1* and *Npc2*), vacuolar-type H^+-ATPases (V-ATPases) that maintain acidic pH within the symbiosome (Extended Data Fig. 4), and multiple lysosomal cathepsins, granulins and lipases that are involved in protein degradation and lipid metabolism (Extended Data Fig. 5a). Scleractinian host cells also shared expression of multiple genes involved in the following: (1) redox homeostasis functions, such as the production of bilirubin (*Blvra*), tocopherol biosynthesis and glutathione metabolism (*Ggt7* and *ChaC*) (Extended Data Fig. 7d); (2) nitrogen metabolism, such as glutamine synthetase (*Glul*) (Extended Data Fig. 5d), which has been hypothesized to be involved in regulating nitrogen availability to the algal symbionts[42]; and (3) multiple carbonic anhydrases involved in the conversion of CO_2 to HCO_3^- that is then transported into the symbiosome via Slc26a11 channels[43], which are co-expressed in all scleractinian host cells (Extended Data Fig. 4a,b and 6c). Multiple genes involved in lipid metabolism also showed conserved expression, including genes encoding enzymes involved in synthesis of fatty acids (*Elo*, *Scd1–Scd4*, *Hsd17b1–Hsd17b3* and *Acaca* and *Acacb*), triglycerides (*Dgat1*) and sphingolipids (*Pisd*), lipid transport (*ApoD*), the formation of lipid droplets (*Plin1–Plin5*), and the degradation of diverse lipids (*Plbd1*, *Plbd2* and *Lipa*) (Extended Data Fig. 5). Finally, all three species shared expression in host cells of diverse transmembrane metabolite transporters (summarized in Fig. 4b and Extended Data Fig. 4) and transcription factors that are involved in lipid metabolism (*Srebf1/2* and *Usf1-3*)[44] and lysosomal biogenesis (*Mitf1*)[45] (Extended Data Fig. 7a).

We then focused on the differences observed between the host cell transcriptomes of the facultative symbiotic *O. patagonica* and the obligate symbiotic stony corals. We observed lower expression in *O. patagonica* host cells of some key genes involved in galactose catabolism (especially *Galt* and *Gale*, the latest steps in the Leloir pathway) (Extended Data Fig. 5c), glycogen synthesis (*Ugp2*, *Gys1*, *Gys2* and *Gbe1*) (Extended Data Fig. 5d) and sugar transport (*Slc2a1–Slc2a4* and *Slc23a1*, *Slc23a2* and *Slc23a4*) (Extended Data Fig. 4), suggesting a minor role for glucose metabolism in this species that relies less on its symbionts and can draw nutrients from external sources. By contrast, the gain in expression of genes associated with fatty acid metabolism and lipid storage in the host transcriptomes of *O. patagonica* and *O. arbuscula* (Fig. 4a and Extended Data Fig. 5a,b) suggests a more prominent role for lipid metabolism in facultative symbiosis. In addition, whereas all stony corals expressed key components of the mammalian target of rapamycin (MTOR) signalling pathway[46], such as Lamtor and Rrag proteins, which are involved in the recruitment of mTOR to the lysosomes, we observed an opposite expression pattern of the positive regulator of MTOR *Rptor*[47,48] (high in *O. patagonica* and *O. arbuscula*, and low in *S. pistillata* and *A. millepora*) and the negative regulator of MTOR *Fnip1/2* (ref. 49) (low in *O. patagonica* and *O. arbuscula*, and high in *S. pistillata* and *A. millepora*) (Extended Data Fig. 7b). Given the role of MTORC1 in regulating the cellular anabolic–catabolic balance[50], this could represent a broad physiological trait specific to *Oculina* or, alternatively, it could reflect differences in the metabolic response to opportunistic symbiotic nutrition in *Oculina* host cells. Another remarkable difference was in the expression of light-sensing molecular pathways (Extended Data Fig. 7c). Whereas *A. millepora* and *S. pistillata* host cells express multiple opsins and *S. pistillata* also expressed key enzymes in the retinal pathway (encoded by *Bco1* and *Rpe65*), these genes are not expressed in *O. patagonica*, possibly reflecting differences in habitats (as this coral can be found in dark habitats) or non-obligate symbiotic strategy. Finally, there are multiple genes that are expressed specifically in *O. patagonica* host cells, including *Nos1–Nos3*, which are involved in nitric oxide production[51], several myosins and other genes that are involved

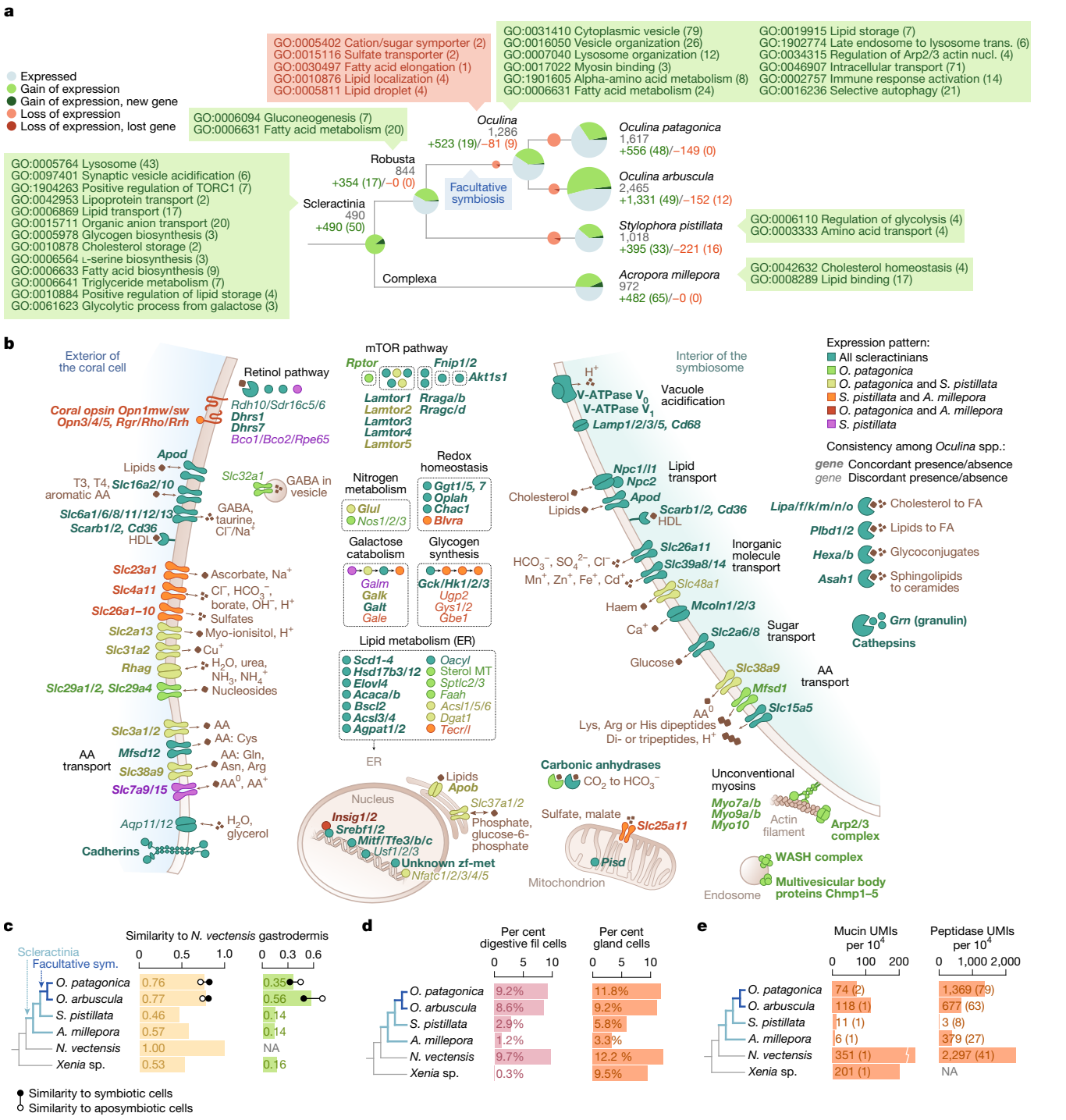


Fig. 4 | Evolution of the alga-hosting cell transcriptome. a, Evolutionary reconstruction of the alga-hosting cell transcriptomes of four scleractinians (*O. patagonica*, *O. arbuscula*, *S. pistillata* and *A. millepora*). We recorded the Anthozoa-level orthology groups specifically expressed in the alga-hosting cells of each coral (fold change > 1.1; Wilcoxon's rank-sum test on counts, FDR < 0.001), and inferred the ancestral presence, gains and losses along the depicted species tree using Dollo parsimony. Pie plots indicate the number of gains and losses in each node. Events of gene expression gain or loss were further classified according to the evolutionary history of the relevant genes (that is, gains of expression of de novo-originated genes or loss of expression accompanied by gene loss). We indicate selected gene and enriched Gene Ontologies gained (in green) or lost (in red) at each node ($P < 0.05$, based on one-sided Fisher's exact tests with elim correction, with the number of genes involved in parentheses). Nucl., nucleation; trans., transport. **b**, Schematic depiction of a stereotypical alga-hosting cell with selected protein components colour-coded according to their expression in each species. AA, amino acid; ER, endoplasmic reticulum; FA, fatty acid; GABA, γ-aminobutyric acid; HDL, high-density lipoprotein. **c**, Global transcriptome similarity between the gastrodermis cells of the sea anemone *N. vectensis* and gastrodermis and alga-hosting cells of four scleractinian corals and the octocorallian *Xenia* sp. Similarity was quantified with SAMap scores. In the case of *Oculina* spp., the similarity scores are also reported for the symbiotic and aposymbiotic samples separately. NA, not applicable. **d**, Frequency of digestive filament cells and gland cells in the transcriptomic atlases of each species. **e**, Fraction of the transcriptome corresponding to mucins and peptidases in the *Mucin*+ and *Peptidase*+ gland cell clusters of each species, measured as the sum of UMIs belonging to these categories per 10^4 UMIs. In **a**,**b**,**d**,**e**, *O. patagonica* and *O. arbuscula* analyses were restricted to cell transcriptomes from the symbiotic samples.

in lysosomal transport[52] (Extended Data Fig. 6b), and the transcription factor gene *Nfat1*, which is involved in regulating immune responses[53].

To better understand the differences between facultative and obligate symbiotic stony corals, we extended our cross-species comparisons to include two outgroup cnidarians: the heterotrophic sea anemone *Nematostella vectensis*[33] and the symbiotic soft octocoral *Xenia* sp.[24]. Transcriptomic analyses of the gastrodermis and host cells in *Oculina* revealed greater similarity to the heterotrophic *N. vectensis* than to the obligate symbiotic corals *S. pistillata* and *A. millepora* (Fig. 4c and Supplementary Fig. 7e). In addition, we found that the proportion of digestive filament cells and gland cells is higher in facultative symbiotic corals compared to obligate symbionts, closely resembling the cellular composition observed in heterotrophic *N. vectensis* (Fig. 4d). Digestive filament ciliated cells are part of the mesenteric structures and facilitate the transport of particles within the gastrovascular cavity. Many gland cell types are also found in the mesenteric edges and contribute to enzymatic digestion and mucus production. In line with this, expression levels of orthologous peptidases and mucins in *Oculina* gland cells were higher relative to those in *S. pistillata* and *A. millepora*, and comparable to expression levels in *N. vectensis* (Fig. 4e and Extended Data Fig. 8).

Collectively, our cross-species, cell-type-resolved transcriptomic analysis revealed metabolic, structural and regulatory functions that are conserved across all scleractinians. Of note, it also uncovered specific adaptations in *Oculina* species that are indicative of an evolutionary shift towards increased heterotrophic capacity in facultative symbiotic corals.

Discussion

Here we uncover the genomic and cellular basis of facultative symbiosis in the thermotolerant stony coral *O. patagonica*. Our study highlights the power of integrating chromosome-scale genomes with single-cell resolution atlases to investigate non-traditional model species across different conditions. This approach provides reciprocal insights—with cell-level gene-expression patterns shedding light on the evolutionary history of coral genomes, and comparative genomics improving our understanding of the diversity and evolution of cell-type expression programmes.

By comparing naturally occurring symbiotic and aposymbiotic *O. patagonica* specimens, we identified key differences in both cell-type composition and gene-expression programmes. Aposymbiotic colonies exhibited a reduction in alga-hosting cells and an increase in a specific immune cell population that showed gene-expression signatures that were compatible with a phagocytic, macrophage-like function. Additionally, gene-expression changes were particularly pronounced in the gastrodermal and host cells of aposymbiotic corals, in which hundreds of genes involved in metabolism, protein production, lysosomal function and symbiont interaction were downregulated, probably reflecting a shift from symbiont-derived nutrition to heterotrophic feeding.

Comparative analysis of alga-hosting cells in *O. patagonica* symbiotic specimens versus two obligate symbiotic corals (*A. millepora* and *S. pistillata*) revealed differences in gene-expression programmes that were indicative of secondary losses and unique adaptations in facultative symbiotic corals. Notably, *Oculina* host cells exhibit limited transcriptional differentiation from gastrodermal cells, in contrast to the more distinct cell populations that we observed in the obligate symbiotic species. Furthermore, *Oculina* host cells express an expanded repertoire of genes related to lipid metabolism and storage, but showed lower expression levels of various genes associated with glucose production and storage, as well as of light-sensing opsins that could couple metabolic states to symbiont photosynthetic activity. We also observed significant differences in genes related to signalling and metabolic regulation, suggesting that facultative stony corals utilize a growth-focused strategy supported by multiple food sources (symbionts and heterotrophic nutrition).

This is further reflected in the higher proportion of digestive filament and mucin- and peptidase-expressing gland cells in *Oculina* species, a ratio that remains largely stable between symbiotic and aposymbiotic specimens and closely mirrors the cellular composition observed in heterotrophic cnidarians such as *N. vectensis*. These features point to the presence of well-developed mesenterial structures adapted for particle capture and extracellular digestion in facultative corals, in contrast to the reduced digestive investment observed in obligate symbiotic species. Together, these findings suggest that heterotrophy is an atavistic trait in facultative symbiotic scleractinians, highlighting a fundamental divergence in nutritional strategy from the ancestral specialization on autotrophic symbiosis shared by most stony corals.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-025-09623-6.

1. Muscatine, L. & Porter, J. W. Reef corals: mutualistic symbioses adapted to nutrient-poor environments. *Bioscience* **27**, 454–460 (1977).
2. Bellwood, D. R., Hughes, T. P., Folke, C. & Nyström, M. Confronting the coral reef crisis. *Nature* **429**, 827–833 (2004).
3. Rivera, H. E. & Davies, S. W. Symbiosis maintenance in the facultative coral, *Oculina arbuscula*, relies on nitrogen cycling, cell cycle modulation, and immunity. *Sci. Rep.* **11**, 21226 (2021).
4. Aichelman, H. E., Zimmerman, R. C. & Barshis, D. J. Adaptive signatures in thermal performance of the temperate coral *Astrangia poculata*. *J. Exp. Biol.* **222**, jeb189225 (2019).
5. Fine, M., Zibrowius, H. & Loya, Y. *Oculina patagonica*: a non-lessepsian scleractinian coral invading the Mediterranean Sea. *Mar. Biol.* **138**, 1195–1203 (2001).
6. Shemesh, T. et al. The effects of elevated temperatures on the reproductive biology of a mediterranean coral, *Oculina patagonica*. *Oceans* **5**, 758–769 (2024).
7. Reaka-Kudla, M. L. in *Biodiversity II: Understanding and Protecting our Biological Resources* (eds Reaka-Kudla, M. L., Wilson, D. E. & Wilson, E. O.) 83–108 (1997).
8. Gates, R. D., Baghdasarian, G. & Muscatine, L. Temperature stress causes host cell detachment in symbiotic cnidarians: implications for coral bleaching. *Biol. Bull.* **182**, 324–332 (1992).
9. Hughes, T. P. et al. Global warming and recurrent mass bleaching of corals. *Nature* **543**, 373–377 (2017).
10. Shenkar, N., Fine, M. & Loya, Y. Size matters: bleaching dynamics of the coral *Oculina patagonica*. *Mar. Ecol. Prog. Ser.* **294**, 181–188 (2005).
11. Leydet, K. P. & Hellberg, M. E. The invasive coral *Oculina patagonica* has not been recently introduced to the Mediterranean from the western Atlantic. *BMC Evol. Biol.* **15**, 79 (2015).
12. Rubio-Portillo, E. et al. Eukarya associated with the stony coral *Oculina patagonica* from the Mediterranean Sea. *Mar. Genomics* **17**, 17–23 (2014).
13. Rubio-Portillo, E., Vázquez-Luis, M., Valle, C., Izquierdo-Muñoz, A. & Ramos-Esplá, A. A. Growth and bleaching of the coral *Oculina patagonica* under different environmental conditions in the western Mediterranean Sea. *Mar. Biol.* **161**, 2333–2343 (2014).
14. Rodolfo-Metalpa, R., Reynaud, S., Allemand, D. & Ferrier-Pagès, C. Temporal and depth responses of two temperate corals, *Cladocora caespitosa* and *Oculina patagonica*, from the North Mediterranean Sea. *Mar. Ecol. Prog. Ser.* **369**, 103–114 (2008).
15. Serrano, E., Ribes, M. & Coma, R. Demographics of the zooxanthellate coral *Oculina patagonica* along the Mediterranean Iberian coast in relation to environmental parameters. *Sci. Total Environ.* **634**, 1580–1592 (2018).
16. Salomidi, M., Katsanevakis, S., Issaris, Y., Tsiamis, K. & Katsiaras, N. Anthropogenic disturbance of coastal habitats promotes the spread of the introduced scleractinian coral *Oculina patagonica* in the Mediterranean Sea. *Biol. Invasions* **15**, 1961–1971 (2013).
17. Terrón-Sigler, A., Casado-Amezúa, P. & Torre, F. E. Abundance and distribution of the rapid expansive coral *Oculina patagonica* in the Northern Alborán Sea (Western Mediterranean). *Mar. Biodivers. Rec.* **8**, e45 (2015).
18. Pastor, F., Valiente, J. A. & Palau, J. L. Sea surface temperature in the Mediterranean: trends and spatial patterns (1982–2016). *Pure Appl. Geophys.* **175**, 4017–4029 (2018).
19. Martinez, S., Bellworthy, J., Ferrier-Pagès, C. & Mass, T. Selection of mesophotic habitats by *Oculina patagonica* in the Eastern Mediterranean Sea following global warming. *Sci. Rep.* **11**, 18134 (2021).
20. Fuller, Z. L. et al. Population genetics of the coroal *Acropora millepora*: toward genomic prediction of bleaching. *Science* **369**, eaba4674 (2020).
21. Stankiewicz, K. H. et al. Genomic comparison of the temperate coral *Astrangia poculata* with tropical corals yields insights into winter quiescence, innate immunity, and sexual reproduction. *G3* **15**, jkaf033 (2025).
22. Ying, H. et al. Comparative genomics reveals the distinct evolutionary trajectories of the robust and complex coral lineages. *Genome Biol.* **19**, 175 (2018).
23. Fletcher, C. & da Conceicoa, L. P. The genome sequence of the starlet sea anemone, *Nematostella vectensis* (Stephenson, 1935). *Wellcome Open Res.* **8**, 79 (2023).

24. Hu, M., Zheng, X., Fan, C.-M. & Zheng, Y. Lineage dynamics of the endosymbiotic cell type in the soft coral *Xenia*. *Nature* **582**, 534–538 (2020).
25. Simakov, O. et al. Deeply conserved synteny and the evolution of metazoan chromosomes. *Sci. Adv.* **8**, eabi5884 (2022).
26. Saunders, P. A. & Muyle, A. Sex chromosome evolution: hallmarks and question marks. *Mol. Biol. Evol.* **41**, msae218 (2024).
27. Mao, Y. & Satoh, N. A likely ancient genome duplication in the speciose reef-building coral genus, *Acropora*. *iScience* **13**, 20–32 (2019).
28. Noel, B. et al. Pervasive tandem duplications and convergent evolution shape coral genomes. *Genome Biol.* **24**, 123 (2023).
29. Voolstra, C. R. et al. Comparative analysis of the genomes of *Stylophora pistillata* and *Acropora digitifera* provides evidence for extensive differences between species of corals. *Sci. Rep.* **7**, 17583 (2017).
30. García-Castro, H. et al. ACME dissociation: a versatile cell fixation–dissociation method for single-cell transcriptomics. *Genome Biol.* **22**, 89 (2021).
31. Najle, S. R. et al. Stepwise emergence of the neuronal gene expression program in early animal evolution. *Cell* **186**, 4676–4693.e29 (2023).
32. Baran, Y. et al. MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions. *Genome Biol.* **20**, 206 (2019).
33. Sebé-Pedrós, A. et al. Cnidarian cell type diversity and regulation revealed by whole-organism single-cell RNA-seq. *Cell* **173**, 1520–1534.e20 (2018).
34. Steger, J. et al. Single-cell transcriptomics identifies conserved regulators of neuroglandular lineages. *Cell Rep.* **40**, 111370 (2022).
35. Levy, S. et al. A stony coral cell atlas illuminates the molecular and cellular basis of coral symbiosis, calcification, and immunity. *Cell* **184**, 2973–2987.e18 (2021).
36. Irimia, M. et al. Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. *Genome Res.* **22**, 2356–67 (2012).
37. LaJeunesse, T. C. et al. Systematic revision of Symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts. *Curr. Biol.* **28**, 2570–2580.e6 (2018).
38. Valadez-Ingersoll, M. et al. Cell type-specific immune regulation under symbiosis in a facultatively symbiotic coral. *ISME J.* **19**, wraf132 (2025).
39. Pinzón, J. H. et al. Whole transcriptome analysis reveals changes in expression of immune-related genes during and after bleaching in a reef-building coral. *R. Soc. Open. Sci.* **2**, 140214 (2015).
40. Wall, C. B. et al. The effects of environmental history and thermal stress on coral physiology and immunity. *Mar. Biol.* **165**, 56 (2018).
41. Barott, K. L., Venn, A. A., Perez, S. O., Tambutté, S. & Tresguerres, M. Coral host cells acidify symbiotic algal microenvironment to promote photosynthesis. *Proc. Natl Acad. Sci. USA* **112**, 607–612 (2015).
42. Matz, M. V. Not-so-mutually beneficial coral symbiosis. *Curr. Biol.* **34**, R798–R801 (2024).
43. Thies, A. B., Quijada-Rodriguez, A. R., Zhouyao, H., Weihrauch, D. & Tresguerres, M. A Rhesus channel in the coral symbiosome membrane suggests a novel mechanism to regulate NH_3 and CO_2 delivery to algal symbionts. *Sci. Adv.* **8**, 303 (2022).
44. Griffin, M. J., Wong, R. H. F., Pandya, N. & Sul, H. S. Direct interaction between USF and SREBP-1c mediates synergistic activation of the fatty-acid synthase promoter. *J. Biol. Chem.* **282**, 5453–5467 (2007).
45. Nardone, C. et al. A central role for regulated protein stability in the control of TFE3 and MITF by nutrients. *Mol. Cell* **83**, 57–73.e9 (2023).
46. Voss, P. A. et al. Host nutrient sensing is mediated by mTOR signaling in cnidarian-dinoflagellate symbiosis. *Curr. Biol.* **33**, 3634–3647.e5 (2023).
47. Hara, K. et al. Raptor, a binding partner of target of rapamycin (TOR), mediates TOR action. *Cell* **110**, 177–189 (2002).
48. Kim, D.-H. et al. mTOR interacts with Raptor to form a nutrient-sensitive complex that signals to the cell growth machinery. *Cell* **110**, 163–175 (2002).
49. Ramírez, J. A. et al. Folliculin interacting protein 1 maintains metabolic homeostasis during B cell development by modulating AMPK, mTORC1, and TFE3. *J. Immunol.* **203**, 2899–2908 (2019).
50. Saxton, R. A. & Sabatini, D. M. mTOR signaling in growth, metabolism, and disease. *Cell* **168**, 960–976 (2017).
51. Alderton, W. K., Cooper, C. E. & Knowles, R. G. Nitric oxide synthases: structure, function and inhibition. *Biochem. J.* **357**, 593–615 (2001).
52. Soni, L. E., Warren, C. M., Bucci, C., Orten, D. J. & Hasson, T. The unconventional myosin-VIIa associates with lysosomes. *Cell Motil. Cytoskeleton* **62**, 13–26 (2005).
53. Fric, J. et al. NFAT control of innate immunity. *Blood* **120**, 1380–1389 (2012).
54. Baumgarten, S. et al. The genome of *Aiptasia*, a sea anemone model for coral symbiosis. *Proc. Natl Acad. Sci. USA* **112**, 11893–11898 (2015).

Methods

Animal sources

One symbiotic colony and one naturally aposymbiotic colony of *O. patagonica* were collected from the wild in July from the Israeli Mediterranean Sea near Michmoret (32.24049° N, 34.51530° E). An *S. pistillata* colony was collected from the Gulf of Eilat/Aqaba, in front of the Interuniversity Institute of Marine Biology (IUI) in Eilat, Israel. The corals were acclimatized and cultured for one month before dissociation in a controlled aquarium system at the Leon H. Charney School of Marine Sciences, University of Haifa. An *A. millepora* colony was obtained from DeJong Marinelife (SPS-MILB-050672). Species identity was confirmed by sequencing 18S rRNA, which showed the highest similarity to the *A. millepora* JS-1 genome (NCBI RefSeq assembly GCF_013753865.1). All corals were maintained in artificial seawater (Red Sea Salt, Red Sea Limited) with a salinity of 39 ppt at 25 °C under a 12 h:12 h light:dark photoperiod with a photosynthetically active radiation level of 50 µmol m^{-2} s^{-1}. Corals were fed twice weekly with planktonic coral food (Reef Snow, Brightwell Aquatics) according to the manufacturer's instructions.

Genome sequencing and assembly

Long-read sequencing. Coral fragments were washed with filtered artificial sea water and incubated in 10 ml of calcium-free ASW containing 5 mM EDTA for 30 min at room temperature in a 6-well plate. Cells were separated from the skeleton by gently scraping with a 10 µl tip and filtered through a 70-µm strainer into 15 ml tubes. The cells were centrifuged at 2,000*g* for 5 min, the supernatant discarded, and the pellet resuspended in 500 µl TNES (Tris, NaCl, EDTA, SDS)-urea lysis buffer. Samples were transferred to 2 ml tubes, treated with 5 µl proteinase K (100 µg ml^{-1}), and incubated at 55 °C for 2.5 h. The lysate was centrifuged at 500*g* for 2 min to remove algal symbionts and debris, with the supernatant transferred to a fresh 2 ml tube. RNase A (10 µl of 1 mg ml^{-1}) was added and incubated at 37 °C for 15 min. Subsequently, an equal volume of phenol was added and mixed by inverting the tube five times, followed by an equal volume of chloroform, also mixed by inversion. The mixture was centrifuged at 12,000*g* for 10 min at room temperature, and the upper aqueous phase was carefully transferred to a fresh tube (400 µl). This chloroform extraction was repeated twice. Sodium acetate (3 M, pH 5.2) was added to the final supernatant to a final concentration of 0.3 M, followed by the addition of 2 volumes of 100% ice-cold ethanol. The sample was incubated at −20 °C for 30 min to precipitate DNA. DNA was pelleted by centrifugation at 12,000*g* for 20 min at 4 °C, washed with 1 ml of 70% ice-cold ethanol, and centrifuged again at 12,000*g* for 10 min at 4 °C. After removing the ethanol, the DNA was dried and resuspended in 40 µl EB buffer (Qiagen). DNA quality was assessed using a TapeStation and NanoDrop (for 260/280 and 260/230 ratios), and DNA concentration was measured with a Qubit.

Obtained high molecular weight genomic DNA was sequenced on Oxford Nanopore using two PromethION flow cells (R9.4.1), generating 2.45 M reads with estimated N50 32.74 kb and 2 M reads with N50 32.45 kb.

Bulk RNA-seq. Coral fragments were incubated in ACME solution (6.5 ml calcium-free artificial seawater (CaF-ASW), 1 ml glycerol, 1.5 ml methanol, 1 ml glacial acetic acid) for 30 min at room temperature with gentle shaking. The fragments were then transferred to a sterile six-well plate, and the tissue was gently scraped off in a chemical hood with appropriate protective equipment. The resulting cell suspension was filtered through a 70-µm strainer to remove debris and transferred to 15 ml tubes, ensuring no more than 5 ml per tube. Cells were centrifuged at 2,000*g* for 5 min at 4 °C to wash out the ACME solution. The supernatant was discarded, and the cells were washed with PBS in ultrapure water, followed by another centrifugation at 2,000*g* for 5 min at 4 °C. The supernatant was discarded, and the cell pellet was resuspended in 2 ml Trizol. After vortexing, 1 ml of the suspension was transferred to RNase-free 1.5 ml Eppendorf tubes and incubated for 5 min at room temperature. Next, 200 µl of chloroform was added, the tubes were inverted 10 times, and the mixture was allowed to stand for 5 min at room temperature. The samples were then centrifuged at 12,000*g* for 10 min at 4 °C. The upper colourless phase (~500 µl) was carefully transferred to a new RNase-free 1.5 ml Eppendorf tube, mixed with an equal volume of 100% ethanol, and inverted 10 times. RNA extraction was performed using the PureLink RNA Mini Kit (Invitrogen). The mixture was added to the RNA binding column in 600 µl aliquots, centrifuged at 12,000*g* for 30 sec at 4 °C, and the flow-through was discarded. This step was repeated until all the liquid was processed, which bound RNA (and some genomic DNA) to the column. Following this, RNA extraction was performed using the PureLink RNA Mini Kit (Invitrogen), including on-column DNase treatment. Elution was performed with 45 µl of RNase-free water. The concentration and quality of RNA were measured using an Agilent TapeStation system. All samples had an RNA integrity number (RIN) greater than 8.5. Strand-specific mRNA libraries were prepared using Illumina Truseq kits. The resulting libraries were sequenced on an Illumina Novaseq6000 sequencer, obtaining a total of 222M 150 bp paired-end reads. These data were used subsequently for gene prediction.

***O. patagonica* micro-C library preparation and sequencing.** *O. patagonica* fragments were dissociated into single cells by first cleaning the fragments with filtered CaF-ASW. Fragments were then transferred to a 6-well plate containing 10 ml CaF-ASW and 5 mM EDTA. Cells were gently scraped from the skeleton using a 10 µl tip and passed through a 70-µm strainer. Cells were centrifuged at 500*g* for 2 min to remove the algal symbionts; the pellet, which consisted mostly of algae, was discarded, and the cell suspension was transferred to a new tube. Cells were then counted and diluted to a concentration of 10^6 cells per ml. Cells were pelleted again by centrifugation at 2,000*g* for 5 min at room temperature, and the supernatant was discarded. The pellet was resuspended in 1 ml CaF-ASW, and 62 µl of 16% methanol-free formaldehyde (Thermo Scientific, 28906) was added to achieve a final concentration of 1% formaldehyde. Cells were incubated on a rotating wheel at room temperature for 10 min, then the formaldehyde was quenched with glycine (final concentration 128 mM) for 5 min at room temperature, followed by an additional 15 min incubation on ice. Crosslinked cells were pelleted at 4,500*g* for 10 min, and the supernatant was removed. Cells were washed with PBS to remove glycine and pelleted again at 4,500*g* for 10 min. The pellet was resuspended in 990 µl PBS (optimal concentration 1–2 million cells per ml), and 10 µl of freshly prepared 300 mM disuccinimidyl glutarate (DSG) (final concentration 3 mM) was added. Cells were incubated with 3 mM DSG (Thermo Scientific, A35392) in PBS for 40 min at room temperature on a rotating wheel. The DSG was then quenched with glycine (final concentration 400 mM) for 5 min at room temperature. Cells were pelleted at 4,000*g* for 10 min, the supernatant was removed, and the cells were frozen at −80 °C.

Micro-C libraries were prepared as previously described[55,56] with some modifications. Approximately 10^6 cells were resuspended in 500 µl of ice-cold MB1 lysis buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 0.2% NP-40, 1× Protease Inhibitor Cocktail (Roche)) and incubated on ice for 20 min, with gentle resuspension every 5 min. Cells were then centrifuged at 4,500*g* for 5 min at 4 °C, and the supernatant was mostly discarded, leaving 20–40 µl. The pellet containing cells and/or nuclei pellet was washed with 500 µl of MB1 buffer and centrifuged again at 4,500*g* for 5 min at 4 °C. The supernatant was discarded, and the nuclei were resuspended in 100 µl of MB1 buffer. For chromatin digestion, 1 µl of 20 U $µl^{-1}$ Mnase (Takara Bio, 2910a) was added to the sample, and the mixture was incubated for 10 min at 37 °C with shaking at 850 rpm. The reaction was stopped by adding 0.8 µl of 500 mM EGTA, followed by incubation at 65 °C for 10 min. The sample was immediately placed on ice and 500 µl of ice-cold 1× NEB2.1

buffer (10 mM Tris-HCl pH 7.4, 50 mM NaCl, 10 mM $MgCl_2$, 0.1% BSA) was added. The nuclei were centrifuged at 4,500*g* for 5 min at 4 °C, and the supernatant was removed. The nuclei were washed once more with NEB2.1 buffer, followed by centrifugation at 4,500*g* for 5 min at 4 °C. Nuclei were resuspended in 45 µl of an end-chewing reaction mix containing 5 µl 10× NEBuffer 2.1, 1 µl 100 mM ATP, 2.5 µl 100 mM DTT, 34 µl nuclease-free water, and 2.5 µl 10 U μl^{-1} T4 PNK (NEB, M0201). The reaction was incubated for 15 min at 37 °C with shaking at 850 rpm. Subsequently, 5 µl of 5 U μl^{-1} Klenow Fragment (NEB, M0210) was added, and the mixture was incubated for another 15 min at 37 °C with shaking. For biotin fill-in, 25 µl of a fill-in master mix was added to the reaction, making a total volume of 75 µl. The mix included 2.5 µl 10× T4 DNA Ligase Buffer, 11.875 µl nuclease-free water, 5 µl 1 mM Biotin-dATP (Jena Bioscience, NU-835-BIO14-L), 5 µl 1 mM Biotin-dCTP (Jena Bioscience, NU-809-BIOX-L), 0.5 µl 10 mM dTTP + dGTP, and 0.125 µl 20 mg ml^{-1} BSA (200×). The sample was incubated for 45 min at 25 °C with interval mixing. The reaction was then stopped by adding 4.5 µl of 0.5 M EDTA, bringing the final concentration to 30 mM, and the enzymes were inactivated by incubation at 65 °C for 20 min. The chromatin was pelleted by adding 500 µl of MB3 buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$), followed by centrifugation at 10,000*g* for 10 min at 4 °C. The supernatant was removed, and the pellet was washed with an additional 500 µl of MB3 without resuspending, followed by centrifugation at 16,000*g* for 10 min at 4 °C. The chromatin was resuspended in a proximity ligation reaction mix containing 920 µl nuclease-free water, 120 µl 10× T4 DNA ligase buffer, 100 µl 10% Triton X-100, 12 µl 20 mg ml^{-1} BSA, 36 µl PEG4000 (50%, final concentration 1.5%), and 12 µl of 5 U μl^{-1} T4 DNA ligase (Thermo Scientific, EL0012). The reaction was incubated for 3 h at room temperature with gentle rotation at 20 rpm. The chromatin was then pelleted by centrifugation at 16,000*g* for 10 min at 4 °C, and the supernatant was discarded. To remove un-ligated biotin-dNTPs, the chromatin was resuspended in a master mix containing 87 µl nuclease-free water, 10 µl 10× NEBuffer 1, and 1 µl 10 mg ml^{-1} RNase A (Roche, Merck 10109142001). The sample was incubated for 10 min at 37 °C with shaking at 850 rpm, followed by the addition of 2 µl of 100 U μl^{-1} Exonuclease III (NEB, M0206) and incubation for an additional 5 min at 37 °C with shaking. For reverse crosslinking, 12 µl 10% SDS, 6 µl 20 mg ml^{-1} proteinase K (Roche, Merck 3115879001), and 8.5 µl 5 M NaCl were added to the sample. The mixture was incubated overnight at 65 °C with shaking at 1,000 rpm. The samples were then purified using Zymo tubes (DNA Clean & Concentrate-5, Zymo Research, D4014) and eluted in 33 µl Tris–EDTA buffer. The purified DNA was processed for library preparation by biotin pull-down, end-repair, adapter ligation, and final amplification using the NEBNext Ultra II Ligation and End Repair kits (NEB, E7546), following the manufacturer's instructions. Libraries were size-selected using AMPure XP beads and stored at −20 °C until further use. The final libraries, comprising four biological replicates, were sequenced with NextSeq2000 in paired-end format with a read length 101 bp, resulting in a total of 263,962,574 sequenced reads.

Genome assembly and scaffolding. Long reads were assembled into draft scaffolds with the Flye assembler 2.9.3, using the default mode for Nanopore reads (--nano-raw flag for regular reads with 5–20% error rates; the alignment error rate estimated by Flye was 8.41%) and the scaffolding option (--scaffold flag). Then, we used purge_dups 1.2.5 to remove uncollapsed haplotypes from the assembly, as follows: (1) we split the draft assembly into contigs using the split_fa utility in the purge_dups package; (2) we aligned the resulting scaffolds against themselves using minimap2 2.26 with the assembly presets (-x asm5 flag); (3) we calculated coverage cut-offs using the pbcstat (maximum coverage at 1,000×) and calcuts utilities in purge_dups; and (4) we used these coverage cut-offs to identify collapsed and uncollapsed haplotypes on the basis of their coverage distribution with purge_dups (using two rounds of chaining for increased accuracy with the -2 flag), and split the contigs accordingly using get_seqs.

This draft collapsed assembly was then re-scaffolded using proximity-ligation contact information. To generate chromosome-level assembly, micro-C data were mapped on the genome using Juicer[57] (v.1.6) (REF) with the -p assembly option. The genome was then scaffolded using the 3D-DNA pipeline[58] with -r2 -editor-repeat-coverage 2 option. Final manual curation in the Juicebox Assembly Tool[57] resulted in 14 chromosomes. The final genome assembly was polished with the original nanopore reads using Medaka (v.1.5.0). This resulted in a fully collapsed diploid assembly with 14 chromosome-level scaffolds, which coincides with the chromosome count of other scleractinians[20] and anthozoans[23]. Chromosome-level scaffolds were named 1-to-14, on the basis of global syntenic similarity with the *A. millepora* reference genome (see 'Microsynteny analysis').

Finally, we polished the scaffolded assembly with Medaka 1.5.0, as follows: (1) we realigned the raw Nanopore reads to the chromosome using minimap2 (presets as per the Medaka-bundled mini_align utility); (2) used medaka consensus with the sequence model matching the Nanopore flowcell and Guppy versions (PromethION 4.9.1 and Guppy 5.0.7 respectively; -m r941_prom_sup_g507 flag) in order to create consensus sequences for each chromosome; (3) and created a final assembly from the consensus chromosomes with medaka stitch. Completeness and contiguity statistics for the initial and polished versions of the assembly are available in Supplementary Fig. 1, including *k*-mer spectra plots obtained with kat comp spectra-cn tool (with $k = 27$ bp) from the KAT 2.0.8 suite[59].

Gene prediction in *O. patagonica*. We annotated the genes in the chromosome-level assembly of *O. patagonica*, using a combination of tools for de novo and evidence-based gene prediction (BRAKER2[60] 2.1.6, Augustus[61] 3.5.0, StringTie[62] 2.2.1, and GenomeThreader[63] 1.7.1) and optimal gene model selection (Mikado[64] 2.3.4). This procedure is described below.

First, we mapped bulk strand-specific RNA-seq libraries to the reference *Oculina* genome using the read aligner STAR[65] 2.7.10b without multi-mapping reads (flag: --outFilterMultimapNmax 1), only considering uniquely mapping reads for splice junctions (--outSJfilterReads Unique), reporting splice junction-supporting reads and keeping only the reads with junctions that passed filtering (--outFilterType BySJout, --alignSJDBoverhangMin 1, --alignSJoverhangMin 8), and reporting the alignment strand based on intron motifs (--outSAMstrandField intronMotif). The resulting coordinate-sorted BAM file (--outSAMtype BAM SortedByCoordinate) was used to produce an initial set of transcript predictions using StringTie in conservative mode (-t -c 1.5 -f 0.0 flags), from which open reading frames were predicted using TransDecoder 5.7.1. Predicted peptides were collapsed by sequence similarity using CD-HIT 4.8.1 (-c 0.95) and complete genes (those with start and end codons) of non-extreme lengths (>600 and <10,000 amino-acids) were retrieved for later use in the de novo gene-prediction step. Specifically, these were used to train Augustus iteratively within BRAKER2, by aligning them to the reference genome using GenomeThreader (BRAKER2 flag: --prg=gth –trainFromGth) and using the original STAR-produced alignments as further evidence (--bam = <file>).

Second, we used Mikado to select the best gene predictions from each locus, selecting from the output of BRAKER2/Augustus (all exons and coding exons (CDS) were used separately), an unguided Stringtie assembly, and the filtered set of GenomeThreader training peptide alignments to the reference genome. To build the Mikado hints file: (1) all sources of evidence were considered as strand-specific; (2) a score of 1 was associated with the BRAKER2/Augustus predictions and 0 for the others; (3) the CDS from BRAKER2/Augustus were considered as a reference annotation; (4) redundant models were excluded from all samples; and (5) CDS sequences with errors were removed from the model set. To build the Mikado configuration file (mikado configure), we clustered transcripts with a minimum cDNA overlap of 20% (--min-clustering-cdna-overlap 0.2) and any for CDSs

(--min-clustering-cds-overlap 0.0), set the programme to permissive mode with regards to open reading fragment (ORF) splitting policy (--mode permissive). Additional options, suited for animal genomes, were defined in the scoring configuration file. After preparing the transcripts sets with the mikado prepare submodule, we prepared further evidence sources to be considered by Mikado: (1) predicted ORFs for all Mikado models, using TransDecoder; (2) evidence-based splice junction coordinates from STAR (see above), obtained using the junctools convert 1.2.4 module of Portcullis 1.1.2 (ref. 64); and (3) homology models obtained with diamond blastx[66] against the 2022-05 release of UniRef50 (ref. 67), adding the percentage of positive-scoring alignments and the traceback operations fields to the reported output (flag: -f 6 qseqid sseqid pident length mismatch gapopen qstart qend sstart send evalue bitscore ppos btop). These additional sources of evidence were considered by Mikado (mikado serialize submodule, disabling start codon adjustment with –no-start-adjustment). Finally, the best gene model for each locus was selected using the mikado pick module, prioritizing reference models (--reference-update). Transposable elements were annotated using EDTA 2.2 (ref. 68), with the --anno 1 parmeter and providing transcript sequences for masking (--cds flag). Finally, transcript and peptide sequences for each gene model were retrieved using gffread[69] 0.11.7. The completeness of the resulting gene set was evaluated using BUSCO[70] 5.5.0 in protein mode (-m proteins) against the Metazoa database of universal orthologues (-l metazoa_odb10).

Comparative genomic analyses

Orthology analyses and functional gene annotation. We built two orthology databases using different taxon sampling: (1) an Anthozoa-focused one with higher resolution within this taxon (16 species, listed with sources in Supplementary Table 1a), to be used for inter-species comparisons among corals and other cnidarians; and (2) a second one with a wider taxon sampling of metazoans one (24 species, listed in Supplementary Table 1a), used to annotate coral genes to known orthologues in reference species (namely *Mus musculus*). In all cases, only the predicted peptides of the longest isoform per gene was considered for orthology inference.

For both orthology databases, we used a combined approach that relied on proteome-wide analyses with Broccoli[71] 1.2 supplemented by dedicated phylogenies to classify transcription factor families with high accuracy. First, we used Broccoli to identify clusters of orthologous genes (step 3 in Broccoli) and pairs of orthologous genes (step 4), using the maximum-likelihood gene tree inference algorithm between all species in both datasets (setting a *k*-mer length of 10,000 to avoid the filtering of paralogous sequences). In parallel, we computed gene phylogenies to refine the orthology assignments of transcription factors using the Metazoa-wide dataset. In this case, we queried the predicted peptide sequences using the hidden Markov model profiles of DNA-binding regions of transcription factor families (obtained from the Pfam database[72]; Supplementary Table 3) using the hmmsearch utility from the HMMER[73] 3.3.2 toolkit. Each collection of peptide sequences sharing a DNA-binding domain (that is, general transcription factor families) was aligned in an all-to-all manner using diamond[66] v.2.1.8.162 (high-sensitivity mode enabled with the --more-sensitive flag, reporting up to 100 target sequences per query) and clustered into low-granularity homology groups using MCL[74] v.22.282 (ABC clustering mode, using alignment bit-scores as weights, and gene family-specific inflation parameters, as listed in Supplementary Table 3). The resulting homology groups were then aligned using MAFFT[75] 7.475 (E-INS-i algorithm with up to 10,000 refinement iterations); the alignments were pruned using using ClipKIT[76] 1.1.395 (in kpic-gappy mode and using a gap threshold = 0.7); and a gene tree was built from each of the pruned alignments using using IQ-TREE[77] 2.1.0 (running each tree for up to 10,000 iterations until convergence threshold of 0.999 is met for 200 generations; the best-fitting substitution model among LG, WAG and JTT was selected with ModelFinder[78]; statistical supports were obtained from 1,000 iterations of UFBoot[79]). Orthology groups and pairs were then inferred from the final gene trees with Possvm[80] 1.1 (iterative gene tree rooting procedure for up to 10 steps, deriving gene names from *M. musculus* orthologues). Then, for all genes included in transcription factor phylogenies, we replaced their Broccoli-derived orthology information in the Anthozoa and Metazoa datasets with the corresponding Possvm-derived information regarding orthology groups and pairs.

Finally, we annotated all genes in the dataset with the following information: (1) gene names, obtained from *M. musculus* orthologues in the same group or, in the case of transcription factor families, from the phylogenetic information as parsed by Possvm[80]; (2) domain annotations, obtained with Pfamscan and the Pfam database[72] 33.1; (3) Gene Ontology annotations, transferred to each gene from its *M. musculus* orthologues, which were obtained from the November 2022 release of the Mouse Genome Database[81]; (4) KEGG Orthology (KO) categories, also transferred from the *M. musculus* orthologues in our database (the relevant mouse KO annotations were obtained from the Uniprot mappings available in the KEGG Pathways database[82,83] in the 2024-08-19 release); (5) presence of transmembrane domains, using TMHMM[84] 2.0; and (6) presence of signal peptides, using SignalP[85] 5.0b.

Gene family evolution. We inferred the patterns of gain, loss and expansion (duplication) of gene families from the Anthozoa orthology database along its corresponding phylogeny, using a maximum-likelihood phylogenetic birth-and-death model[86] as implemented in Count[87]. First, we trained the model using the presence counts of 2,000 randomly selected domains annotated in the 16 anthozoan species of interest (restricting the sampling to domains present in at least 5% of species). The model was trained in three refinement iterations, starting from a simple model that estimated uniform gain, loss, duplication, and transfer rates across the entire tree, followed by an additional iteration allowing for rate heterogeneity using one Γ distribution for each rate, and a final one allowing two Γ categories per rate. Each training iteration was run for up to 100 rounds of optimization, stopping when the relative change in the model log-likelihood fell by 1% in two consecutive rounds. The resulting model was used to infer the probability of gain, loss, duplication and expansion of each orthology group in the Anthozoa dataset, supplied to Count as a matrix of gene counts in each of the extant genomes. The result was a matrix with the probabilities of gene family presence, gain, loss, and duplication in each of the extant and ancestral nodes of the supplied phylogeny (also available in the model file in Supplementary Table 3).

Gene family age assignment. We used the probabilistic reconstruction of gene family gain/loss (see above) to date each orthogroup in the Anthozoa orthology database. For each orthogroup, Count outputs a vector of probabilities of gain of that family along the nodes of the species tree (ancestral and extant). We used this information to assign ages to each orthogroup on the basis of the most probable node of gain. Orthogroups where the most likely node of gain was unclear (maximum probability <50%) were dated by Dollo parsimony, with Possvm[80].

Microsynteny analysis of conserved collinear gene pairs. We evaluated the conservation of microsynteny among anthozoan species of interest (*O. patagonica*, *O. arbuscula*, *S. pistillata*, *A. millepora*, *N. vectensis* and *Xenia sp.*) by counting the co-occurrence of gene pairs, defined as two genes placed next to each other (that is, collinear) in both a query and target pair of genomes. Gene orientation was recorded but not taken into account to establish syntenic conservation, and we allowed one extra gene to be inserted between two genes in our definition of syntenic gene pair. To match genes between pairs of species, we used the Broccoli-derived orthologue pairs for the Anthozoa orthology database. Genomic range operations were performed in

R using the GenomicRanges 1.54, IRanges[88] 2.36, and rtracklayer[89] 1.62 libraries.

Macrosynteny analysis of ancestral anthozoan linkage groups. We identified linkage groups present in the Cnidaria ancestor and scored their conservation in each of the scleractinian corals of interest (*O. patagonica*, *A. millepora*, and related species). To that end, we used a set of species with chromosome-level genome assemblies: *O. patagonica*, *O. arbuscula*, *Porites lutea*, *A. millepora*, *A. cervicornis*, *A. palmata*, *N. vectensssia* sp., *Rhopilema esculentum* and *Hydra vulgaris*. We used the same approach as Simakov et al.[25]; specifically, we (1) defined homology groups (using local diamond alignments of proteins for all species pairs followed by MCL clustering with an inflation parameter $I = 2.1$); (2) identified unique combinations of at least 10 homology groups present in the same chromosomes for sets of 3 species selected to act as outgroups of the scleractinian corals of interest (namely, the hydrozoan *H. vulgaris*, the scyphozoan *R. esculentum*, and either the octocorallian *Xenia* sp. or the sea anemone *N. vectensis*); and (3) termed these combinations of homology groups cnidarian ALGs, and scored the presence of the constituent homologous genes in the chromosomes of the scleractinian corals of interest (*O. patagonica*, *O. arbuscula*, *Porites lutea*, *A. millepora*, *A. cervicornis* and *A. palmata*) along running windows (with a length of $l = 200$ homologous genes and a step of $s = 50$ genes). Overall, we identified 26 and 27 ALGs for the three-species set consisting of *H. vulgaris*, *R. esculentum* and *N. vectensis* or *Xenia* sp., respectively. We evaluated the degree of ALG segregation (that is, lack of fusion with mixing) in the query scleractinian chromosomes using χ^2 tests of homologous gene counts along non-overlapping windows of $l = 200$ homologous genes per chromosome, as previously described[31].

Whole-genome alignment and conservation analysis. We calculated conservation scores for selected scleractinian genomes (*O. patagonica*, *A. millepora* and *S. pistillata*) using whole-genome alignments and PHAST (Phylogenetic Analysis with Space/Time) models. First, we aligned nine genomes to each other using Cactus[90] 2.6.4, following a progressive approach guided by the species trees of scleractinians, namely: ((((((*O. patagonica*, *O. arbuscula*), *Fungia spp.*), *Goniastrea aspera*), (*Pocillopora verrucosa*, *S. pistillata*)), ((*A. millepora* (*Acropora palmata*, *Acropora cervicornis*)), *Porites lutea*)). Second, we used the hal2maf utility from the HAL[91] toolkit 2.2 to create MAF alignments of the chromosome of the reference species of interest. Then, we used the rphast 1.6.1 implementation of the Phast[92] toolkit to identify conserved regions in each reference genome, as follows: (1) we used phyloFit[93] to create an initial null model of neutral change on the basis of the four-fold degenerate codon positions from coding regions, using a general reversible nucleotide transition matrix and the predefined species tree (the model was trained using only the longest chromosome in each reference genome); (2) we used PhastCons to optimize this model using the expectation-maximization procedure, re-estimating the transition probabilities and tree parameters at each iteration; and finally (3) we calculated the PhastCons and phyloP scores for individual bases in the reference genome, as well as running windows of $l = 200$ bp and step $s = 100$ bp. Genomic range operations were performed in R using the GenomicRanges, IRanges and rtracklayer libraries.

Whole-genome duplication analysis. We tested the existence of ancestral whole-genome duplications in *O. patagonica* using ksrates[94] 1.1.4. This tool compares the Ks distribution of paralogues from a focal species and orthologues from outgroup species along a predefined phylogeny, adjusting the split times in the tree according to branch-specific evolutionary rates. For these analyses, we used *O. patagonica* as the focal species in three different runs, and the the same anthozoan dataset and phylogenetic tree described above; using transcript coding sequences and transcript genomic coordinates in the GTF format instead of predicted peptides. ksrates calculates Ks values using the wgd[95] package, and relies on BLASTP to identify one-to-one orthologous sequences using the reciprocal best hit criterion, and MCL to identify clusters of paralogues, and MUSCLE to construct multiple sequence alignments[96]. We ran ksrates using both the whole set of paralogues (paranome = yes option in the configuration file) and only collinear gene pairs (collinearity = yes), which relies on i-ADHoRe[97] to identify sets of collinear genes. We set the maximum number of outgroup species for the rate optimization procedure to four (max_number_outgroups = 4) and used the mean among all outgroups as the criterion to define the adjusted rates (consensus_mode_for_multiple_outgroups = mean among outgroups). Only paralogous families with less than 200 copies were considered (max_gene_family_size = 200).

Single-cell transcriptomics

Specimen dissociation and cell fixation. To dissociate and fix coral cells for 10X scRNA-seq, we used a modified version of the ACME maceration protocol[30]. Coral fragments, 2–3 cm in length, were washed with filtered (0.22 µm) CaF-ASW (10 mM Tris-HCl pH 8, 2.1427 mM $NaHCO_3$, 10.7309 mM KCl, 426.0123 mM NaCl, 7.0403 mM Na_2SO_4) and transferred to a 50 ml tube containing 10 ml of ACME maceration solution without BSA, ensuring that the solution fully covered the fragments. The ACME solution was prepared as follows: 6 ml CaF-ASW, 1 ml glacial acetic acid, 1 ml glycerol, 1.5 ml methanol, and 0.5 ml EDTA (a 13:2:2:3:1 ratio). Samples were incubated at room temperature for 30 min with periodic pipetting. The cell suspension was then filtered through a 70-µm strainer into a new 50 ml tube, kept on ice, and aliquoted into 1.5 ml portions in 2 ml tubes. Aliquots were centrifuged at 1,000*g* for 10 min at 4 °C. The supernatant was discarded, and the pellet was resuspended in 1 ml of RB1 buffer (prepared by mixing 1 ml 10× PBS, 3.3 ml 2.4 M sorbitol, 5.7 ml nuclease-free water, and 20 µl RNAsin). Cells were washed again with RB1 buffer under the same centrifugation conditions. Cells were counted using DAPI staining (9 µl of cell suspension mixed with 1 µl DAPI, 1 mg ml^{-1}). Cell concentration was calculated by multiplying the average cell count from 4 sets of 16 squares by 10,000. The sample volume was adjusted to obtain aliquots of 100 µl containing 400,000 cells each.

ClickTag barcoding for 10X scRNA-seq. Fixed cells were barcoded using a modified version of ClickTags[31,98]. To optimize the labelling reaction in ACME fixative, the amine-reactive cross-linker TCO-NHS used by Gehring et al.[98] was replaced with TCO-PEG4-TFP (Click Chemistry Tools), which offers improved stability against hydrolysis in aqueous media. Barcoding DNA oligonucleotides (ClickTags) with a 5′-amino modifier (Integrated DNA Technologies) were activated by derivatization with methyltetrazine-NHS ester as originally described[98]. For cell tagging, we used combinations of three different MTZ-derivatized oligonucleotides per sample. Cell samples were pre-incubated by adding 4.5 µl of 1 mM TFP-TCO and incubating for 5 min at room temperature, protected from light. Premixed MTZ-activated tags (12 µl total) were then added, followed by thorough mixing. Samples were incubated for 30 min at room temperature on a rotatory platform, protected from light. The reaction was quenched by adding 13 µl of 100 mM Tris-HCl (final concentration 10 mM) and 0.65 µl of 10 mM MTZ-DBCO. Samples were incubated for an additional 5 min at room temperature. Each pool was mixed with 2 volumes of RB1, inverted three times, and centrifuged at 1,000*g* for 10 min at 4 °C. The pellet was washed with 1 ml RB1 and centrifuged again under the same conditions. Finally, cells were resuspended in 900 µl RB1 and 100 µl DMSO. Samples were stored at −80 °C until sorting for scRNA-seq.

Cell sorting and scRNA-seq. Single-cell transcriptomes were obtained using the Chromium Single Cell 3′ Gene Expression kit v.3.1 (10X Genomics). Frozen samples were thawed on ice, and cells were collected by centrifugation at 2,000*g* for 5 min at 4 °C. After a wash

with 2 ml of Resuspension Buffer 2 (RB2; 1× PBS, 0.5% BSA, 40 U ml^{-1} RNasin Ribonuclease Inhibitor), cells were pelleted again and resuspended in 1 ml of RB2. Cells were stained with 1:330 DRAQ5 (Thermo 62251) for nuclei staining, and 1:400 Concanavalin-A CF 405 m (biotium 29074), for membrane staining. For single-cell isolation, 36,000 cells were sorted into a well of a 96-well plate using a FACSAria II SORP cell sorter, following 10X Genomics' guidelines. Non-cellular particles were excluded by selecting only DRAQ5-positive cells, and doublets and multiplets were removed using forward scatter width (FSC-W) versus forward scatter height (FSC-H). To specifically sort coral host cells containing the algal symbiont, we employed a targeted strategy by selecting cells positive for DRAQ5, Concanavalin-A 405, and Cy7, the latter indicating the autofluorescence of the algal symbiont. From this population, an additional 4,000 cells were sorted, except in the bleached (apo-symbiont) *O. patagonica* samples, from which only 2,000 host cells could be sorted. In total, 40,000 cells were sorted for each coral sample. Cells were immediately encapsulated after sorting, and barcoded cDNA and sequencing libraries were prepared according to 10X Genomics' protocols. For ClickTag library preparation, ClickTag cDNA was separated from cellular cDNA after the cDNA amplification step, using differential size-selection purification with AMPure XP beads (Beckman Coulter). ClickTag sequencing libraries were prepared as previously described[98]. The size distribution and concentration of the final libraries were assessed using a TapeStation (Agilent) and Qubit (Invitrogen). Libraries were sequenced on an Illumina NextSeq 500 sequencer using high-output 75-cycle V2 kits (Illumina).

Cell sorting and MARS-seq. Cells were dissociated and fixed as described above and saved in −80 °C until sorting for MARS-seq. Frozen samples were thawed on ice, and cells were collected by centrifugation at 2,000*g* for 5 min at 4 °C. After a wash with 2 ml of Resuspension Buffer 2 (RB2; 1× PBS, 0.5% BSA, 40 U ml^{-1} RNasin Ribonuclease Inhibitor), cells were pelleted again and resuspended in 3 ml of RB2. Cells were stained with 1:330 DRAQ5 (Thermo 62251) for nuclei staining, and 1:400 Concanavalin-A CF 405 m (biotium 29074), for membrane staining. Cells were sorted using a FACSAria II SORP cell sorter and distributed into 384-wells capture plates (all coming from the same plate production batch) containing 2ul of lysis solution: 0.2% Triton X-100 and RNase inhibitors plus barcoded poly(T) reverse transcription primers for scRNA-seq. For the MARS-seq, only coral host cells containing the algal symbiont were sorted. To sort these specific cells, we employed a targeted strategy in which non-cellular particles were excluded by selecting only DRAQ5-positive cells, and doublets/multiplets were removed using forward scatter width (FSC-W) versus forward scatter height (FSC-H). From the singlets we sorted only cells positive for DRAQ5, Concanavalin-A 405, and Cy7, the latter indicating the autofluorescence of the algal symbiont. Sorted plates were immediately spun down at 800*g*, to ensure cell immersion into the lysis solution, kept in dry ice and then frozen at −80 °C until further processing. Single-cell libraries were prepared using MARS-seq[99]. For each coral species, all single-cell libraries were prepared in parallel: 5 libraries for *O. patagonica* (5× 384-well plates), 4 for *S. pistillata*, and 7 for *A. millepora*. First, using a Bravo automated liquid handling platform (Agilent), mRNA was converted into cDNA with an oligonucleotide containing both the UMIs and cell barcodes. PEG8000 (0.15%) was added to the reverse transcription reaction to increase efficiency of cDNA capture. Unused oligonucleotides were removed by exonuclease I treatment. cDNAs were pooled (each pool representing the original 384-wells of a MARS-seq plate) and linearly amplified using T7 in vitro transcription and the resulting RNA was fragmented and ligated to an oligonucleotide containing the pool barcode and Illumina sequences, using T4 ssDNA:RNA ligase. Finally, RNA was reverse transcribed into DNA and PCR amplified. The size distribution and concentration of the resulting libraries were calculated using a Tapestation (Agilent) and Qubit (Invitrogen). scRNA-seq libraries were pooled at equimolar concentration and sequenced to saturation (R6 reads per UMI, in most cases > 10 reads per UMI) on an Illumina NextSeq 500 sequencer and using high-output 75 cycles v.2.5 kits (Illumina). We obtained 550 million reads in total, resulting in a median of 90,000 uniquely mapped reads per cell.

Mapping of single-cell transcriptomic libraries. All seven 10X cDNA libraries were mapped to their respective reference genome (four to *O. patagonica*, one to *S. pistillata*, and two to *A. millepora*) using CellRanger 7.2.0 in order to obtain counts of unique molecular identifiers (UMI) per gene and cell. To that end, we used whole gene regions to count UMIs in each genome (--transcriptome flag), which were extended to include proximal downstream regions (defined using the same scRNA-seq data; see below).

The extension of 3′ regions is meant to compensate for the low-quality annotation of untranslated regions in non-model species[35,100]. In the case of *S. pistillata*, we reused the extended gene annotation reported in by Levy et al.[35]. Specifically, we did the following: (1) mapped the second reads of the 10X datasets to each reference genome using *STAR* 2.7.10b (tolerating 3 mismatching positions per read and 5 multi-mapping positions); (2) called peaks separately on each strand using the callpeaks module of MACS2 (ref. 101) 2.2.7.1, adjusting the effective genome size to the ungapped length of each assembly, retaining at most 20 duplicates from different libraries (--keep-dup 20), retaining peaks with a false discovery rate $q \leq 0.01$ (-q 0.01 flag), excluding peaks shorter than 30 bp (--min-length 30), and disabling the modelling of peak extension for chromatin immunoprecipitation with sequencing (ChIP–seq) libraries (--nomodel flag), and the resulting peaks were extended by 20 bp (--extsize 20); and, finally, (3) assigned the resulting peaks to nearby genes on the same strand if they were located up to 5 kb downstream of a gene, or less if another gene was closer to that peak.

To map and quantify MARS-seq data, we used as reference the same gene models for each coral genome as for the 10X libraries, as well as annotations for four candidate symbiont algae, corresponding to four genera with available reference genomes: *S. microadriaticum* (Symbiodiniaceae clade A), *B. minutum* (clade B), *C. goreaui* (clade C), and *F. kawagutii* (clade F). In brief, we first mapped reads onto the corresponding genome using STAR[65] (with parameters: --outFilterMultimapNmax 1 --outFilterMismatchNmax 8 --alignIntronMax 6500). Mapped reads were further processed and filtered as previously described[99]. UMI filtering include two components, one eliminating spurious UMIs resulting from synthesis and sequencing errors, and the other eliminating artifacts involving unlikely in vitro transcription product distributions that are likely a consequence of second strand synthesis or in vitro transcription errors. The minim FDR *q*-value required for filtering in this study was 0.02.

Filtering of single-cell transcriptomes by removal of doublet and empty droplets. First, we used Clicktag-multiplexed libraries to experimentally identify doublet cells in each of the seven 10X single-cell cDNA sequencing experiments. As previously described[102], we sequenced matching Clicktag libraries from the same cells and used Clicktag barcode counts to identify doublet cells in the overloaded cDNA experiments, by assigning combinations of *n* unique barcodes (two or three, depending on the sample) to each sample. Specifically, we: (1) mapped the Clicktag-derived read set to a mock transcriptome consisting of the all barcode sequences with all possible site-wise substitutions, using the kallisto 0.46.2 framework—that is, mapping (bus module) followed by correction, sorting and counting (correct, sort, and count modules); (2) we retained cells with at least 20 total counts in the top-scoring barcode combination; (3) we normalized the barcode counts per cell by dividing them by the total number of counts of each barcode across the dataset, and compared the ratio of normalized counts of each set of same-sample barcodes to the second most abundant set, retaining cells with a first-to-second ratio >2; (4) recorded the first and *n*th most

abundant barcodes in each cell and retained those for which both of them were associated with the same initial sample. Cells with sufficient counts, concordant top barcodes, and high first-to-second normalized count ratios were classified as singlets. Cells with sufficient counts but discordant top barcodes were flagged as doublets and removed from downstream analyses, as were unclassifiable cells.

Second, we used the list of known doublet cells identified using the Clicktag experiment for each 10X sample to flag further possible doublets using the scDblFinder function in the scDblFinder R library[103] (provided with the knownDoublets argument). All of these were also removed from downstream analyses.

Third, we used the distributions of UMIs and unique genes identified per cell in each of the seven 10X samples to distinguish bona fide cells from empty droplets. Specifically, we discarded cells with low number of UMIs upon manual identification of the low- and high-UMI per cell peaks in each library (Supplementary Fig. 3a), and we also discarded cells with a low number of unique genes, defined as a z-score < −3.09 (that is, $P < 0.001$, or $P < 0.01$ in *Oculina*) of the distribution of log-scaled counts of observed genes per cell.

Cell clustering and batch integration of single-cell transcriptomes. We merged the filtered UMI matrices from 10X experiments for each species separately (four for *O. patagonica*, one in *S. pistillata*, and two in *A. millepora*), and stored as counts in a Seurat[104] 5.0.3 object. Then, we applied the sctransform 0.4.1 algorithm (Seurat::SCTransform function) to obtain normalized and variance-stabilized counts (hereafter SCT counts) for each sample separately ('assays', in Seurat parlance), scaling residuals to unit variance (do.scale = TRUE) and zero mean (do.center = TRUE). Using SCT normalized counts improves the accuracy in the selection of variable genes to use in dimensionality reduction and cluster identification[105,106], as well as the removal of batch effects.

We identified cell clusters using two parallel approaches, which are described in detail below: (1) high-granularity metacells[32]; and (2) deliberately over-split Leiden clusters[107] in Seurat that can be curated and manually merged if needed. Both levels of clustering were highly concordant in all species, which offered the possibility of annotating both of them simultaneously in a hierarchical manner based (see below).

For both clustering approaches—metacell and Leiden—we started from batch-integrated principal components analysis (PCA) dimensionality reductions for each species, obtained using the Harmony algorithm[108] as implemented in Seurat (except for *S. pistillata*, for which a single batch was used in the main analyses and the co-integrated dataset was only used for cell cluster annotation purposes; see below). Specifically, we used the SCT normalized counts to run a PCA dimensionality reduction (Seurat::RunPCA function, set to calculate 100 principal components), and integrated the cell-level coordinates across batches using Harmony (Seurat::IntegrateLayers function with method = HarmonyIntegration, normalization.method = "SCT", and k.weight = 50). To decide the appropriate number of principal components to use for downstream clustering procedures, we plotted the fraction of s.d. captured by each principal component in the integrated PCA reduction and identified the 'elbow' in the graph using the first derivative criterion as implemented in the findPC R package[109]. This resulted in 19–32 principal components being selected in each species (Supplementary Fig. 3b).

Then, we proceeded with the two clustering procedures. First, to identify high-granularity metacells, we used the balanced co-clustering algorithm as implemented in the MetaCell[32] and tgstat R packages, as follows: (1) calculated pairwise Pearson's correlation coefficients among the variable expression components selected above (tgstat::tgs_cor function); (2) used these values to select the K nearest neighbours of each cell (tgstat::tgs_knn with K = 200); (3) build a directed graph of cell–cell similarities (tgstat::tgs_graph) of up to 100 neighbours per cell (knn = 100, with edge filtering parameters set to k_expand = 10 and k_beta = 3); (4) create a cell–cell co-clustering graph by randomly resampling 75% of markers in 100 iterations (tgstat::tgs_graph_cover_resample function with p_resamp = 0.75 and n_resamp = 100, and setting the following parameters: min_cluster_size = 10, cooling factor = 1.05, and burn_in = 10); (5) filter this graph to retain the top $K = 100$ neighbours of each cell with an alpha_relaxation factor = 2 in order to obtain a balanced reciprocal co-clustering graph; and (6) obtain the metacell clusters from this pruned cell–cell similarity graph (tgstat::tgs_graph_cover with cooling = 1.05, burn_in = 10). This resulted in 168–343 batch-integrated metacells identified in each species (Supplementary Fig. 3c and Supplementary Table 4).

Second, we applied a Leiden clustering procedure as implemented in Seurat, as follows: (1) identification of the closest neighbours of each cell (Seurat::FindNeighbors function on the batch-integrated PCA dimensionality reduction, using the top N informative principal components defined for each species as the dims parameter); (2) identification of high-granularity (over-split) Leiden clusters from the cell–cell nearest neighbours graph (Seurat::FindClusters function with algorithm = 4 [Leiden], method = "igraph", and resolution = 4). This resulted in 64–67 batch-integrated clusters identified in each species, all of which consisted of at least one metacell (Supplementary Fig. 3c).

In parallel, we calculated cell-level UMAP dimensionality reductions for each species using the batch-integrated PCA coordinates and the same number of top principal components selected for clustering analyses (Seurat::RunUMAP function). We then used the centroid UMAP coordinates (specifically, the median) of all the cells in each metacell cluster to place each metacell in the UMAP space.

We measured gene-expression values using the regularized geometric mean of gene counts for each cell group (metacells, Leiden clusters or cell types) and dividing this value by the median across all clusters, as described in the metacell R package. We refer to these normalized gene-expression values as fold changes across this Article.

The same clustering procedure was also applied to a focused analysis of the gastrodermal (including all cells annotated as gastrodermis, alga-hosting or muscle cells in each of the species) and neurosecretory cell clusters (including all neurons, gland and neurosecretory progenitor cells). Similarly, we repeated the same procedure to reanalyse the previously published single-cell transcriptomic datasets of *Xenia* sp.[24]. and *N. vectensis*[33].

Finally, we evaluated the batch effects in the integrated atlases for each species using cluster-specific sample compositional bias testing[110], as well as silhouette scores[111]. Specifically, we evaluated the batch effects within each cell type by resampling $n = 100$ local cell neighbourhoods (of size equal to the mean number of cells per metacell in each cell type), and testing compositional bias in each resampling using χ^2 tests. Then, for each cell type, we report the FDR-adjusted mean P value of the χ^2 tests, and the distribution of silhouette scores across all resamplings (Supplementary Fig. 3e–g). This approach is equivalent to the kBET procedure if it were run separately per cell type using a k_0 parameter equal to the mean metacell size[110], except that the expected batch fractions within each resampling correspond to those of the global dataset (rather than being cell-type specific).

Annotation of cell types in the coral single-cell transcriptomes. To curate and annotate the cell clusters to known cell types, we used various sources of information: (1) in the case of *S. pistillata*, co-integration of the new 10X dataset with the old MARS-seq atlas[35]; and (2) in the case of *O. patagonica* and *A. millepora*, cell-type-level comparisons with *S. pistillata* itself[35].

First, we co-integrated the new 10X and old MARS-seq atlases of *S. pistillata* to validate and annotate cell-type clusters in the former. We used the batch integration and cell clustering procedure outline above to obtain joint metacell and Leiden clusters. The resulting clusters corresponding to the major cell types described in *S. pistillata* (gastrodermis, alga-hosting cells, muscle, epidermis, calicoblasts, cnidocytes, neurons, gland and immune cells) were composed of cells from both datasets in approximately balanced proportions (Supplementary

Fig. 3d,e). There were two minor exceptions. First, the old 'mitotic host cells' and 'unknown' clusters did not have matching cells in the new dataset. Both clusters corresponded to previously undescribed cell states[35] and were absent in the other two scleractinian atlases (see below). Second, the cnidocyte cell cluster was much more abundant in the old than in the new dataset. Given that the new proportions appeared more similar to those of the other two scleractinian corals too (Fig. 2c), this was also deemed a minor issue.

Second, we compared cell-type-level gene-expression similarity measured as the Pearson correlation coefficients of normalized fold changes between the new *O. patagonica* and *A. millepora* datasets and the previously published *S. pistillata* atlas[35], using one-to-one orthologous gene pairs. This analysis supported the annotation of the relevant clusters as the major scleractinian cell types (gastrodermis, alga-hosting cells, muscle, epidermis, calicoblasts, cnidocytes, neurons, gland and immune cells).

Finally, for all three species, we manually curated the cluster annotations of each major cell type to accommodate heterogeneity within the major cell types. Essentially, we annotated individual Leiden clusters and metacells into general groupings representing cell types or sub-cell types in an ad hoc manner on the basis of examination of the expression profiles of differentially expressed genes among clusters (Supplementary Figs. 4–6).

Within the neurosecretory types, we distinguished neuron from gland metacells by combining cross-species similarity analyses (see above) with a measure of the number of cell-type-specific ion channels, secreted proteins and G-protein-coupled receptors; reasoning that ion channels would predominate in bona fide neurons, as previously observed[35]. And, for neurons specifically, we used the best transcription factor markers to annotate each cluster, namely as *Pou4*+ neurons (matching the previously described *Pou4*+ or *Pou4/Foxl2*+ neurons in other cnidarians[33–35,112]) or *Etv/Isl*+ neurons (as with *Gata*+ or *Isl/Gata*+ neurons[33,35]), among others.

Cell-type compositional analysis between symbiotic and aposymbiotic samples. We evaluated the change in cell-type proportions among the symbiotic and aposymbiotic samples of *O. patagonica* with two procedures: (1) Fisher's exact tests using the global cell counts from symbiotic and aposymbiotic samples as the expected proportions, and adjusting *P* values using the FDR procedure; and (2) using the scCODA Bayesian model[113] as implemented in the pertpy 0.10.0 Python library[114] to test for compositional biases in a sample-aware manner, using the calicoblast cell cluster as a reference (chosen because Fisher's exact tests did not identify significant changes in cell-type proportions in this cell type).

Differential gene-expression analysis between symbiotic and aposymbiotic samples. To identify genes differentially enriched in the symbiotic or aposymbiotic samples of our *O. patagonica* single-cell transcriptomic atlas, we used the FindMarkers function in Seurat. Specifically, we performed Wilcoxon rank-sum tests for each individual cell type, comparing transcript counts between the cells coming from symbiotic and aposymbiotic samples (only for genes detected in at least 1% of the cells of that cell type). *P* values were adjusted using the Bonferroni correction procedure.

Gene module inference and annotation. We used gene-expression fold changes at the metacell level to obtain gene modules in each coral dataset, using WGCNA[115] 1.72-5. Specifically, we: (1) selected variable genes with a fold change ≥ 1.25 in at least one metacell; (2) built a gene co-expression matrix by calculating the Pearson correlation coefficient of each gene, using the average hierarchical clustering algorithm and a soft power parameter = 5 (determined independently for each species using the WGCNA pickSoftThreshold function); (3) used a hierarchical clustering approach to define gene modules using the cutreeHybrid function in the dynamicTreeCut R library[116] with a split parameter = 2 and ignoring clusters with few genes (minClusterSize = 10, cutHeight = 0.99); (4) we assigned each gene to one or more gene module (overlapping memberships) in an inclusive approach, applying a correlation threshold (≥0.5); and (5) calculated the module eigenvectors of each of the resulting modules in each metacell, using the moduleEigengenes function in WGCNA.

Cross-species cell-type transcriptome comparisons. We constructed a multi-species clustering of cell types from scleractinian corals (two sets of species: *O. patagonica*, *S. pistillata*, and *A. millepora*; and *O. patagonica* and *O. arbuscula*), using the UPGMA algorithm from the phangorn[117] 2.9.0 R library. Specifically, we built Log-Det distance matrices obtained from the binarized expression of shared orthologues across cell types of each species (setting the value to 1 if a gene was expressed in that cell type with a cell-type-level fold change ≥2.0; set to 0 otherwise). The most similar (in terms of expression) pairs of orthologous genes between species were selected using the iterative comparison of co-expression (ICC) procedure[31,118], retaining only pairs shared across all species in each comparison. Node supports were obtained from 1,000 iterations of Felsenstein's bootstrap[119] implemented in phangorn.

The ICC approach we used to select the most similar (in terms of expression) pairs of orthologous genes between each two species in the dataset. ICC is able to include in cross-species comparisons additional pairs of homologous genes that are not strictly one-to-one between two species, based on the principle that one of the paralogues may exhibit more similar expression patterns than the other and that such pairs are more informative for gene-expression conservation analyses[118]. We have exposed the implementation of ICC to single-cell transcriptomic data in detail before[31]. In brief, for a pair of species *a* and *b*, we retrieved gene-expression matrices consisting of metacell-level expression of matched one-to-one orthologues (defined using the procedure outlined above), representing using n genes across m_a and m_b conditions. Then, we calculate the Pearson correlation matrix of these two expression matrices, resulting in a $n \times n$ matrix, where the diagonal represents a vector of correlation values between orthologous gene pairs. Third, the n-length correlation vector (range: −1 to +1) is used to obtain a vector of weights (by setting negative values to 0) that quantify the expression conservation of each pair of orthologous genes ($\mathbf{EC_0}$). Next, the $\mathbf{EC_0}$ vector is used to recalculate the $n \times n$ matrix using weighted Pearson correlation (instead of unweighted as in the initial iteration) to produce a new weight vector ($\mathbf{EC_1}$), and this step is repeated for up to i iterations until the final weight vector ($\mathbf{EC}_i$) is approximately equal to the previous one (convergence is achieved using the following criterion: $\sum(\mathbf{EC}_i - \mathrm{EC}_{i-1})^2 < 0.05$). At this point, EC_i represents the expression conservation scores between each pair of one-to-one orthologous gene pairs. Any set of genes with one-to-many or many-to-many paralogy relationships between species *a* and *b* are then sequentially added to this reference matrix of one-to-one orthologues to identify the most conserved pair (in terms of expression similarity) by recalculating EC values across the whole matrix (one-to-one orthologues and the 'test' set of paralogous gene pairs), and selecting the pair with the highest EC_i value. Crucially, in the ICC procedure, the EC scores are defined on the basis of the similarity between gene-gene correlation matrices rather than the similarity of expression between specific cell types. Therefore, it does not require the a priori definition of matching cell clusters between species.

In parallel, we used the SAM[120] 1.0.2 and SAMap[121] 1.0.16 libraries in Python 3.10 to measure transcriptional similarities between cell types in the same sets of species (three scleractinians: *O. patagonica*, *S. pistillata*, and *A. millepora*; two *Oculina* species: *O. patagonica* and *O. arbuscula*; and six anthozoans: *O. patagonica*, *O. arbuscula*, *S. pistillata*, *A. millepora*, *N. vectensis* and *Xenia* sp.). We built a database of pairwise local protein alignments with *blastp* 2.5.0 and used cell-level UMI counts

of each gene to calculate the SAMap mapping scores for each pair of cell types, using the 90% of cells with highest cross-species alignment scores to calculate pairwise similarities between cell types (through the n_top flag in the samap.get_mapping_scores function). Highly variable genes in each dataset were identified with the highly_variable_genes function in Scanpy[122] 1.9.3.

Processing of single-cell transcriptomes for other anthozoan species. We have used single-cell transcriptomic atlases for *Xenia* sp.[24], *Nematostella vectensis*[33], the previously published MARS-seq-based dataset of *S. pistillata*[35], and *O. arbuscula*[38]. In the case of *Xenia* sp., *N. vectensis* and *S. pistillata*, we reused the previously published UMI count matrices and cell-type annotations. In the case of *O. arbuscula*, we produced a de novo gene annotation with the same procedure described above for *O. patagonica*, using the we used the genome assembly and bulk transcriptome datasets provided by the Darwin Tree of Life project (NCBI BioProject accession: PRJEB82731). The single-cell transcriptomic libraries of *O. arbuscula*, corresponding to a symbiotic and an aposymbiotic-induced specimen, were obtained from NCBI (BioProject accession: PRJNA1122932) and mapped to its reference genome using the same procedure as described for our 10X transcriptomic libraries, including dataset integration with Harmony, two-level clustering (metacell and cell-type clusters), cell-type annotation and cell-type composition and differential gene-expression comparisons between the symbiotic and aposymbiotic samples.

Functional gene enrichment analyses. We performed functional enrichments tests for GO terms using the topGO[123] 1.0 library in R. Specifically, we computed the enrichments using counts of genes belonging to each relevant category (enriched markers in a cell type, gene module, and so on) relative to all annotated and expressed genes, using Fisher's exact test and the elim algorithm for GO graph weighting.

Functional enrichment tests of Pfam domain annotations were performed using hypergeometric tests in the R stats library (R Core Team 2024), comparing the frequencies of presence of Pfam domains in each module to the same frequencies in the whole gene set (using unique domains per gene).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw and processed high-throughput sequencing data are available in GEO repository under accession number GSE289546. The *O. patagonica* genome assembly and annotation are available at NCBI under BioProject PRJNA1270949.

Code availability

Scripts to reproduce the data processing and downstream analysis are available at GitHub (https://github.com/sebepedroslab/oculina-coral-sc-atlas/). In addition, datasets can be explored in the interactive web application (https://sebelab.crg.eu/multicoral-sc-atlas/).

55. Hsieh, T.-H. S. et al. Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Mol. Cell* **78**, 539–553.e8 (2020).
56. Krietenstein, N. et al. Ultrastructural details of mammalian chromosome architecture. *Mol. Cell* **78**, 554–565.e7 (2020).
57. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
58. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
59. Mapleson, D., Garcia Accinelli, G., Kettleborough, G., Wright, J. & Clavijo, B. J. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics* **33**, 574–576 (2017).
60. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics Bioinformatics* **3**, lqaa108 (2021).
61. Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435 (2006).
62. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
63. Gremme, G., Brendel, V., Sparks, M. E. & Kurtz, S. Engineering a software tool for gene structure prediction in higher organisms. *Inf. Softw. Technol.* **47**, 965–978 (2005).
64. Venturini, L., Caim, S., Kaithakottil, G. G., Mapleson, D. L. & Swarbreck, D. Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *Gigascience* **7**, giy093 (2018).
65. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
66. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
67. Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B. & Wu, C. H. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**, 926–932 (2015).
68. Ou, S. et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 275 (2019).
69. Pertea, G. & Pertea, M. GFF utilities: GffRead and GffCompare. *F1000Res* **9**, 304 (2020).
70. Waterhouse, R. M. et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018).
71. Derelle, R., Philippe, H. & Colbourne, J. K. Broccoli: combining phylogenetic and network analyses for orthology assignment. *Mol. Biol. Evol.* **37**, 3389–3396 (2020).
72. Punta, M. et al. The Pfam protein families database. *Nucleic Acids Res.* **40**, D290–301 (2012).
73. Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A. & Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **41**, e121 (2013).
74. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002).
75. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
76. Steenwyk, J. L., Buida, T. J., Li, Y., Shen, X.-X. & Rokas, A. ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol.* **18**, e3001007 (2020).
77. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
78. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
79. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
80. Grau-Bové, X. & Sebé-Pedrós, A. Orthology clusters from gene trees with Possvm. *Mol. Biol. Evol.* **38**, 5204–5208 (2021).
81. Blake, J. A. et al. Mouse Genome Database (MGD): knowledgebase for mouse–human comparative biology. *Nucleic Acids Res.* **49**, D981–D987 (2021).
82. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
83. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592 (2023).
84. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
85. Almagro Armenteros, J. J. et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**, 420–423 (2019).
86. Csűrös, M. & Miklós, I. in *Research in Computational Molecular Biology* (eds. Apostolico, A. et al.) 206–220 (Springer, 2006).
87. Csurös, M. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* **26**, 1910–1912 (2010).
88. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
89. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25**, 1841–1842 (2009).
90. Armstrong, J. et al. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020).
91. Hickey, G., Paten, B., Earl, D., Zerbino, D. & Haussler, D. HAL: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics* **29**, 1341–1342 (2013).
92. Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinformatics* **12**, 41–51 (2011).
93. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
94. Sensalari, C., Maere, S. & Lohaus, R. ksrates: positioning whole-genome duplications relative to speciation events in KS distributions. *Bioinformatics* **38**, 530–532 (2022).
95. Zwaenepoel, A. & Van de Peer, Y. Inference of ancient whole-genome duplications and the evolution of gene duplication and loss rates. *Mol. Biol. Evol.* **36**, 1384–1404 (2019).
96. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).
97. Proost, S. et al. i-ADHoRe 3.0—fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Res.* **40**, e11 (2012).
98. Gehring, J., Hwee Park, J., Chen, S., Thomson, M. & Pachter, L. Highly multiplexed single-cell RNA-seq by DNA oligonucleotide tagging of cellular proteins. *Nat. Biotechnol.* **38**, 35–38 (2020).

99. Keren-Shaul, H. et al. MARS-seq2.0: an experimental and analytical pipeline for indexed sorting combined with single-cell RNA sequencing. *Nat. Protoc.* **14**, 1841–1862 (2019).
100. Zolotarov, G., Grau-Bové, X. & Sebé-Pedrós, A. GeneExt: a gene model extension tool for enhanced single-cell RNA-seq analysis. Preprint at *bioRxiv* https://doi.org/10.1101/2023.12.05.570120 (2023).
101. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
102. Chari, T. et al. Whole-animal multiplexed single-cell RNA-seq reveals transcriptional shifts across *Clytia medusa* cell types. *Sci. Adv.* **7**, eabh1683 (2021).
103. Germain, P.-L., Lun, A., Garcia Meixide, C., Macnair, W. & Robinson, M. D. Doublet identification in single-cell sequencing data using scDblFinder. *F1000Res* **10**, 979 (2022).
104. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
105. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
106. Choudhary, S. & Satija, R. Comparison and evaluation of statistical error models for scRNA-seq. *Genome Biol.* **23**, 27 (2022).
107. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233–12 (2019).
108. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
109. Zhuang, H., Wang, H. & Ji, Z. findPC: an R package to automatically select the number of principal components in single-cell analysis. *Bioinformatics* **38**, 2949–2951 (2022).
110. Büttner, M., Miao, Z., Wolf, F. A., Teichmann, S. A. & Theis, F. J. A test metric for assessing single-cell RNA-seq batch correction. *Nat. Methods* **16**, 43–49 (2019).
111. Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).
112. Tournière, O. et al. Nv*POU4*/*Brain3* functions as a terminal selector gene in the nervous system of the cnidarian *Nematostella vectensis*. *Cell Rep.* **30**, 4473–4489.e5 (2020).
113. Büttner, M., Ostner, J., Müller, C. L., Theis, F. J. & Schubert, B. scCODA is a Bayesian model for compositional single-cell data analysis. *Nat. Commun.* **12**, 6876 (2021).
114. Heumos, L. et al. Pertpy: an end-to-end framework for perturbation analysis. Preprint at *bioRxiv* https://doi.org/10.1101/2024.08.04.606516 (2024).
115. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
116. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720 (2008).
117. Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).
118. Tirosh, I. & Barkai, N. Comparative analysis indicates regulatory neofunctionalization of yeast duplicates. *Genome Biol.* **8**, R50 (2007).
119. Felsenstein, J. Phylogenies and the comparative method. *Am. Nat.* **125**, 1–15 (1985).
120. Tarashansky, A. J., Xue, Y., Li, P., Quake, S. R. & Wang, B. Self-assembling manifolds in single-cell RNA sequencing data. *eLife* **8**, e48994 (2019).
121. Tarashansky, A. J. et al. Mapping single-cell atlases throughout Metazoa unravels cell type evolution. *eLife* **10**, e66747 (2021).
122. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15 (2018).
123. Alexa, A., Rahnenführer, J. & Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**, 1600–1607 (2006).

Acknowledgements The authors thank T. Shemesh and L. Uziahu for their assistance in collecting *O. patagonica* colonies; and H. Nativ for providing the photograph of *O. patagonica* and *S. pistillata* colonies shown in Fig. 1a. Research in the A.S-.P. group has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 851647), the BBVA Foundation (Ayudas Fundacion BBVA a Proyectos de Investigacion Cientifica 2021 - Proyecto CoralCellSeq), and the Spanish Ministry of Science and Innovation (PID2021-124757NB-I00 funded by MCIN/AEI/10.13039/501100011033/ FEDER, UE). This project has received funding from the European Union's H2020 research and innovation programme under Marie Sklodowska Curie grant agreements No 101065294 (S.L.) and 101031767 (X.G.-B.). The publication is part of the grant PRE2019-087793SO (A.E.) and PRE2022-105558 (L.M.-E.), funded by MCIN/AEI/10.13039/501100011033 and by the FSE+. S.A.M. is supported by the EMBO postdoctoral fellowship ALTF 066-2022. We acknowledge support of the Spanish Ministry of Science and Innovation through the Centro de Excelencia Severo Ochoa (CEX2020-001049-S, MCIN/AEI /10.13039/501100011033) and the Generalitat de Catalunya through the CERCA programme. We are grateful to the CRG Core Technologies Programme for their support and assistance in this work.

Author contributions S.L. and A.S.-P. designed the project. I.V.K. and X.G.-B. sequenced, assembled and annotated the *Oculina* genome. S.L., S.R.N., L.M.-E., E.K. and A.S.-P. performed single-cell experiments. X.G.-B., I.V.K. and S.A.M. analysed genomic data. X.G.-B. and A.S.-P. analysed single-cell transcriptomics data. A.E. built the interactive web application. T.M. provided biological samples. X.G.-B. created visualizations. S.L., X.G.-B. and A.S.-P. wrote the manuscript with contributions from all authors.

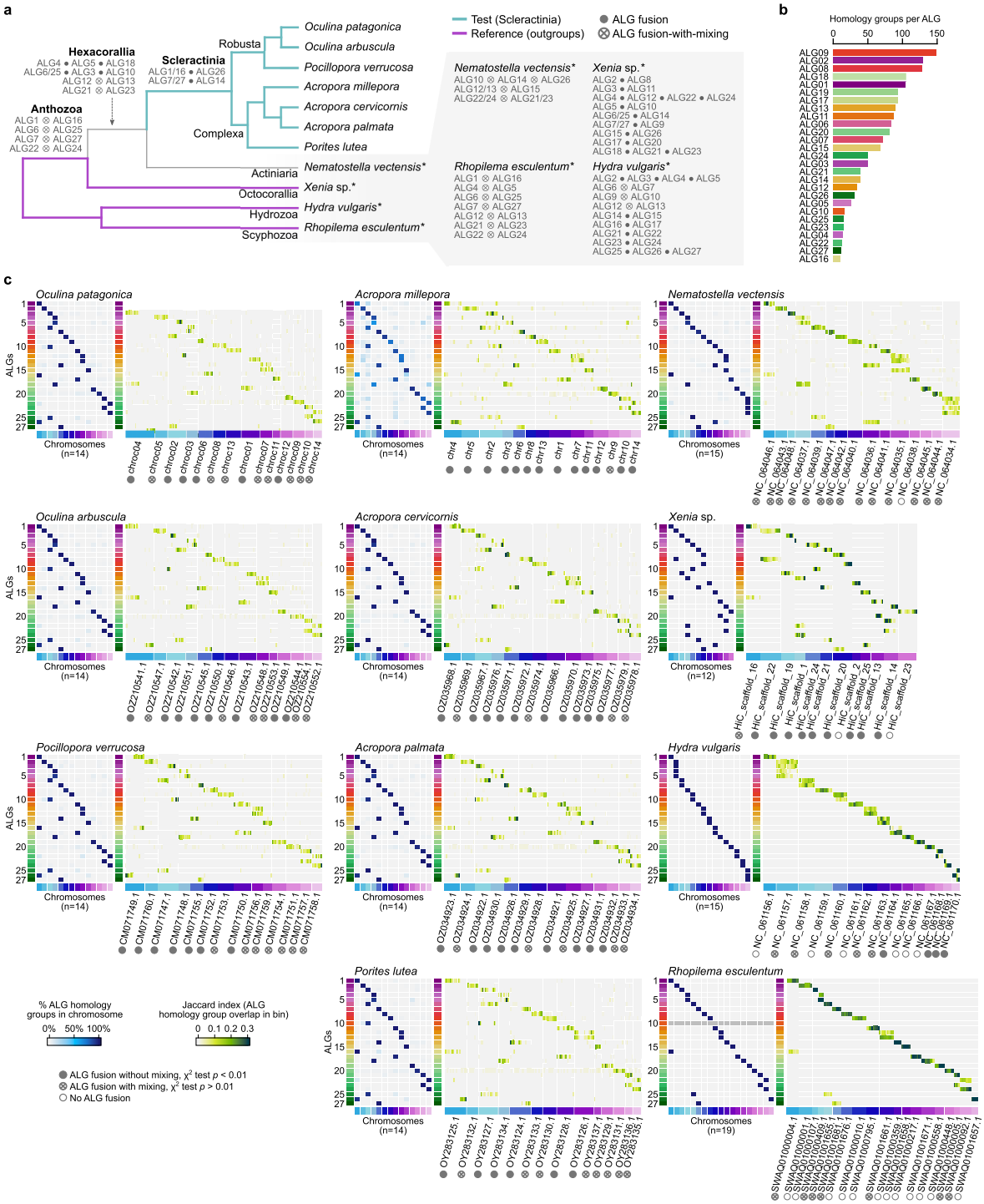
Competing interests The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-025-09623-6.

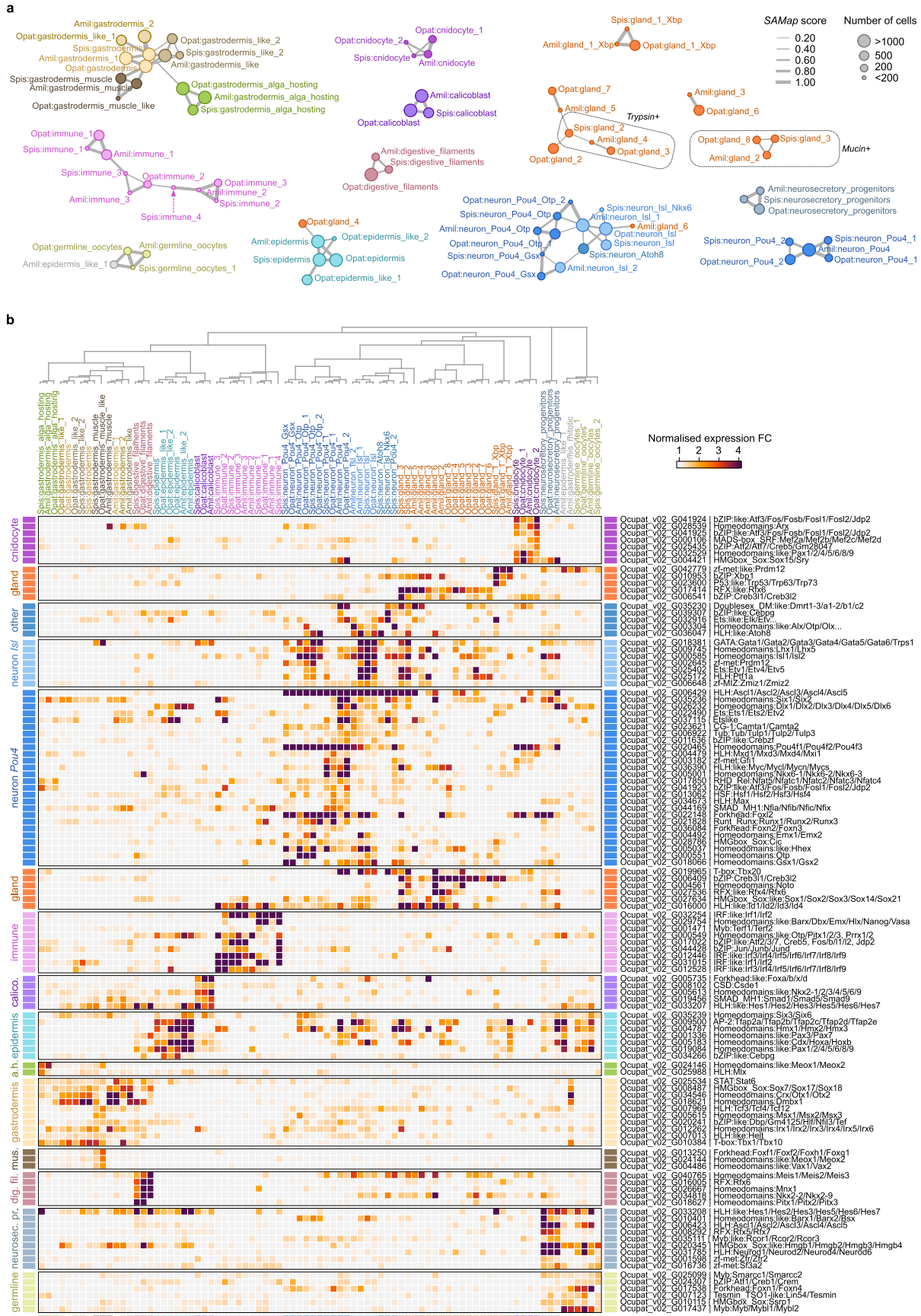
Correspondence and requests for materials should be addressed to Shani Levy, Xavier Grau-Bové or Arnau Sebé-Pedrós.

Peer review information *Nature* thanks José Martín-Durán and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



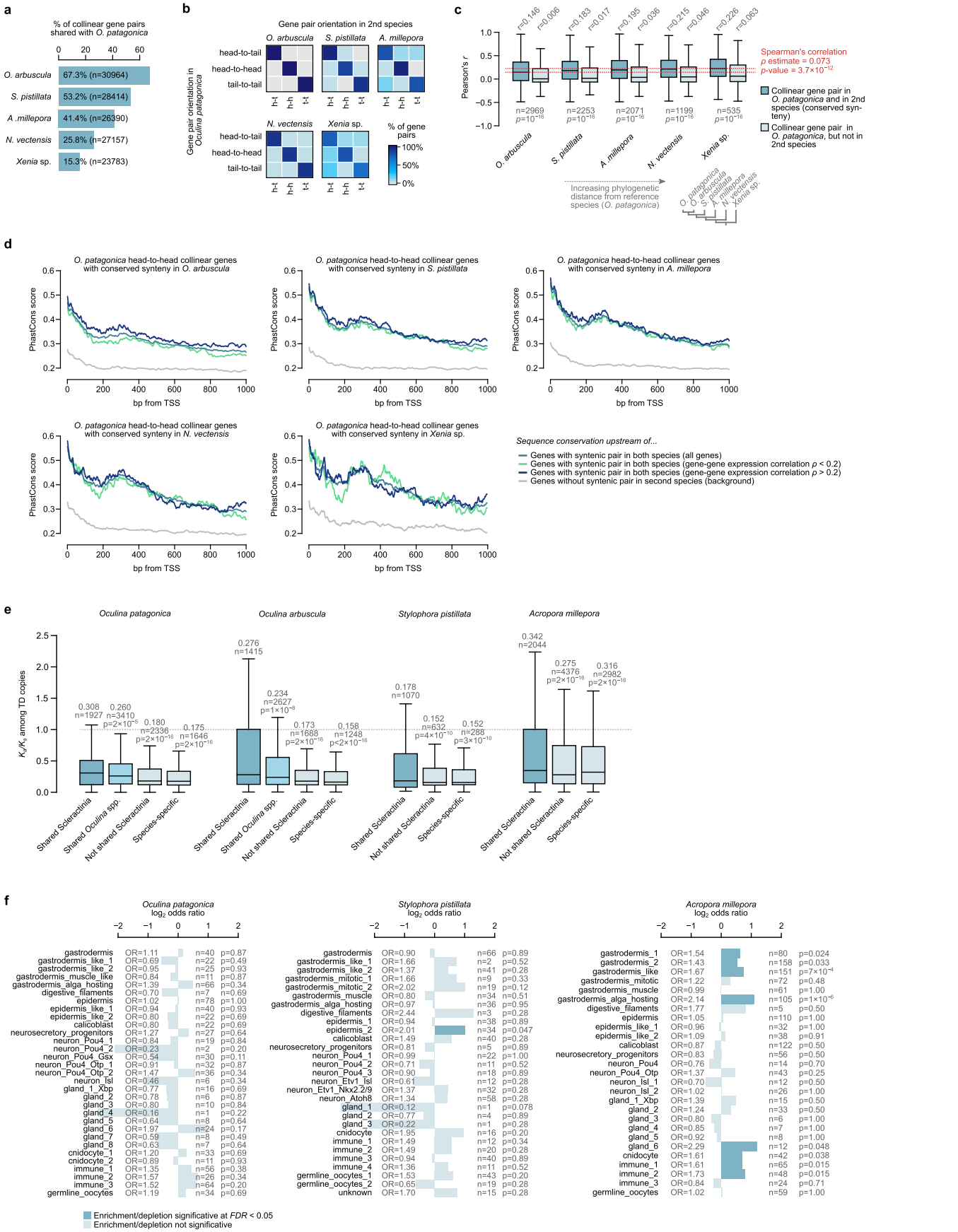
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Macrosynteny analysis. a, Cladogram depicting phylogenetic relationships among the cnidarians with available chromosome/near-chromosome-scale genomes used in the inference of ancestral linkage groups (ALG) for macrosyntenic conservation analysis. The three outgroup species used to infer the combinations of gene homology groups present in the cnidarian ancestor are highlighted in purple (i.e. the octocorallian *Xenia* sp., the hydrozoan *Hydra vulgaris* and the scyphozoan *Rhopilema esculentum*). At each node, we summarise the ALG fusion events, including fusions-with-mixing (denoted with ⊗) and non-mixed fusions (i.e. centric insertions or Robertsonian translocations, denoted with ●). **b**, Number of genes belonging to each of the 27 inferred cnidarian ALGs with more than 10 genes. **c**, For each species in panel **a**, we reported the fraction of genes from each ALG present a given chromosome (heatmaps to the left), and the contribution of each ALG in bins along each chromosome (right-side heatmaps; measured as Jaccard overlap index between genes from a given ALG relative to all genes classified as part to any ALG). For each chromosome, we report whether the observed fusion events correspond to unmixed fusions (denoted with ●, testing for ALG mixedness using two-tailed χ^2 tests in non-overlapping windows) or mixed fusions (denoted with ⊗). **d**, Fraction of syntenic genes originating from each ALG along chromosome 4 of *O. patagonica*, *O. arbuscula* and *P. verrucosa*. The region in the centre of *O. patagonica* chromosome 4 (28 to 44 Mb) is highlighted in grey and is shared only with *O. arbuscula*. **e**, Top 20 Pfam domains enriched in the *Oculina*-specific region of *O. patagonica* and *O. arbuscula* (barplots represent one-sided *p*-values from hypergeometric enrichment tests truncated at $p = 10^{-6}$, counts denoted as *n*). **f**, Synteny conservation between the fourteen chromosomes in *O. patagonica* (*x* axes) and *O. arbuscula* (*y* axes). Each dot represents the placement of an orthologous gene along a given chromosome in each species (only genes from homologous chromosomes are shown). The synteny plot to the right provides a closer look at the syntenic arrangement of chromosome 4, with the *Oculina*-specific region highlighted in purple. All chromosomes exhibit high collinear synteny, with inversions identified in chromosomes 3 and 9.



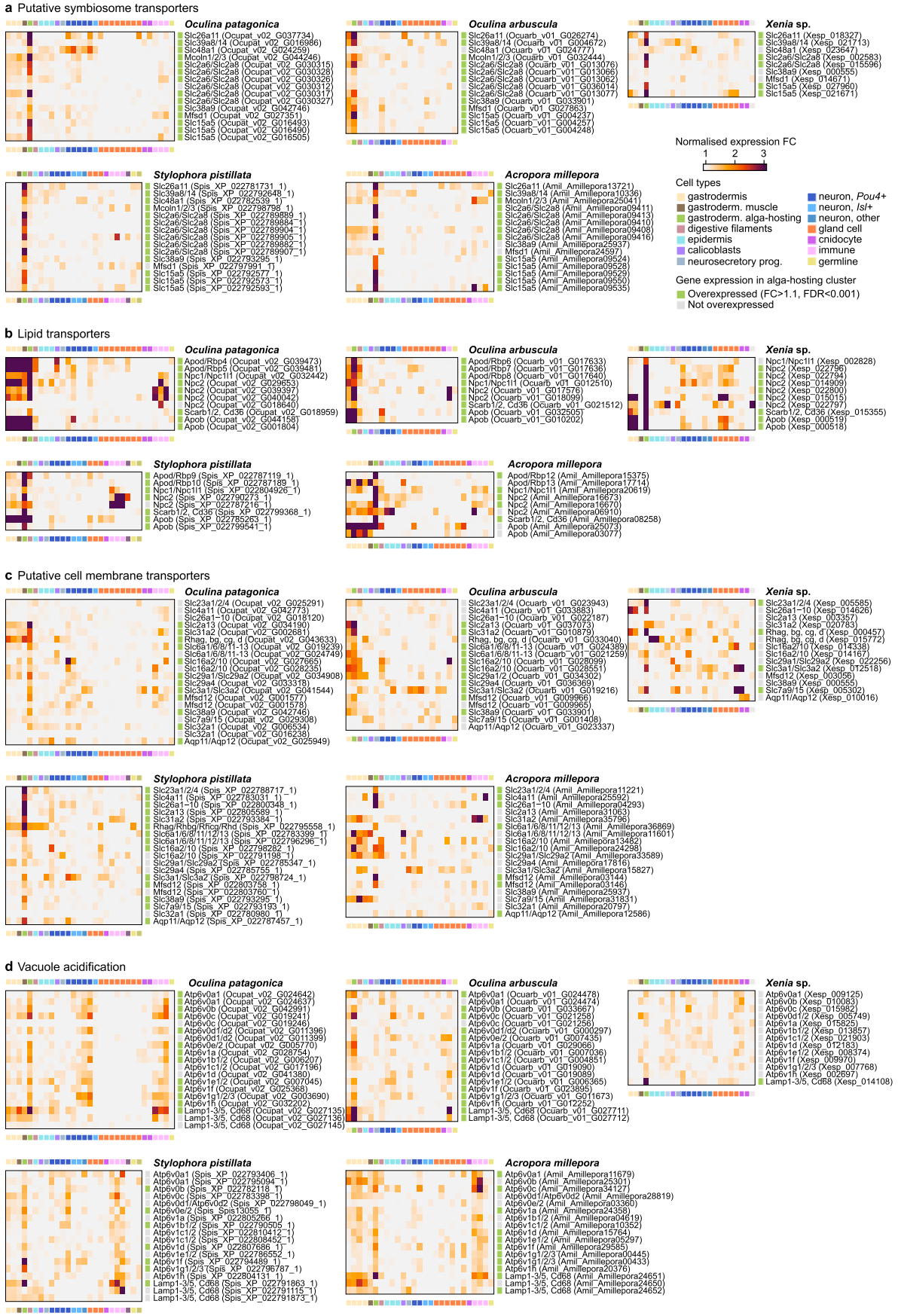
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Cross-species cell type comparisons and transcription factor code. a, Force-directed network of cell type similarity across three scleractinian species (*O. patagonica*, *S. pistillata* and *A. millepora*), using the weighted Fruchterman-Reingold algorithm on each individual connected component. Nodes represent cell type clusters (node is proportional to the number of cells in the cluster) and edge widths represent pairwise similarities across species as *SAMap* similarity scores (threshold at score ≥ 0.25). **b**, Top, cell type clustering of the three scleractinian corals obtained using the *UPGMA* algorithm on binarised gene expression matrices, with bootstrap supports. Bottom, normalised expression fold change values of transcription factors shared by multiple cell types across species. Gene IDs from *O. patagonica* are indicated for reference, but expression values in other corals correspond to their best-matching *ICC*-derived orthologs.



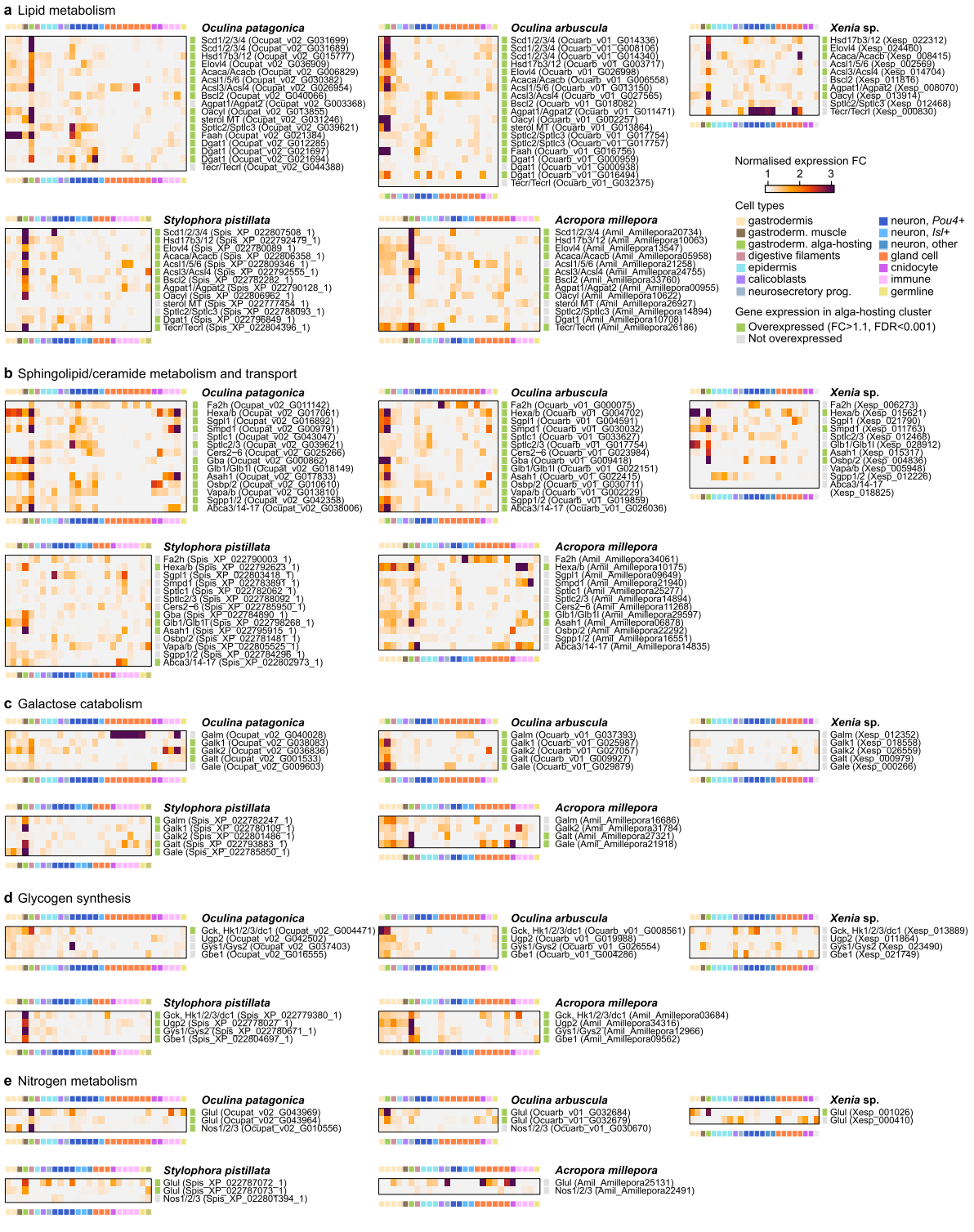
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Microsynteny analysis and tandem duplicated genes. **a**, Fraction of genes pairs in *O. patagonica* with orthologs in collinear arrangement in other cnidarians (*O. arbuscula*, *S. pistillata*, *A. millepora*, *N. vectensis* and *Xenia* sp.). **b**, Conservation of relative gene arrangement (head-to-head [h-h], head-to-tail [h-t], tail-to-tail [t-t]) between *O. patagonica* and other cnidarians. **c**, Distribution of the correlation of gene expression (Pearson's *r*) between pairs of collinear genes of *O. patagonica* with conserved synteny and non-conserved syntenic in other cnidarians (median *r* values reported on top of each distribution, *p*-values from two-sided Kolmogorov-Smirnov tests). The dotted red lines highlight the positive association between gene-gene correlation values and synteny conservation at longer phylogenetic distances from the point of view of *O. patagonica* (ranking the other species from the closest, i.e. *O. arbuscula*, to the most distant one, *Xenia* sp.; and tested using Spearman's ρ for rank correlation). **d**, Conservation of DNA sequence (measured as *PhastCons* scores) upstream of syntenic head-to-head genes between *O. patagonica* and the other cnidarians. The conservation score is reported for all genes with syntenic conservation across species and the subset of genes with correlated expression with their head-to-head pair (Spearman's correlation $\rho > 0.2$), and the subset of genes with low correlation ($\rho < 0.2$). The conservation upstream of a background set of non-syntenic genes is reported for comparison. **e**, Distribution of K_a/K_s values for pairwise alignments of tandem-duplicated genes in each of the three scleractinian corals. For each species, we report the distribution among the set of tandem duplication events shared with the other two corals, the distribution for non-shared families, and the species-specific families (medians reported). The statistical significance of the difference between the latter two distributions and the shared one was assessed using Kolmogorov-Smirnov two-sided tests. **f**, Relative enrichment of genes involved in tandem duplication events among the genes specific to each cell type of *O. patagonica*, *S. pistillata* and *A. millepora* (Wilcoxon rank-sum test on cell type counts, fold-change > 1.1 and *FDR* < 0.001, for genes detected in 1% of the cells). Enrichments were assessed with two-sided Fisher's exact test, for which we report the odds ratio and *FDR*-adjusted *p*-values (if $p < 0.05$). For each cell type, we also report the number of cell tandem-duplicated genes specifically expressed in the cluster (*n*). In panels c and e, box bounds and middle lines in each boxplot represent the first, third and second quartile (i.e. median) respectively; top whiskers represent the value of the third quartile plus 1.5 times the inter-quartile range (or the maximum, whichever is lower); and *id*. for bottom whiskers and values below the first quartile.



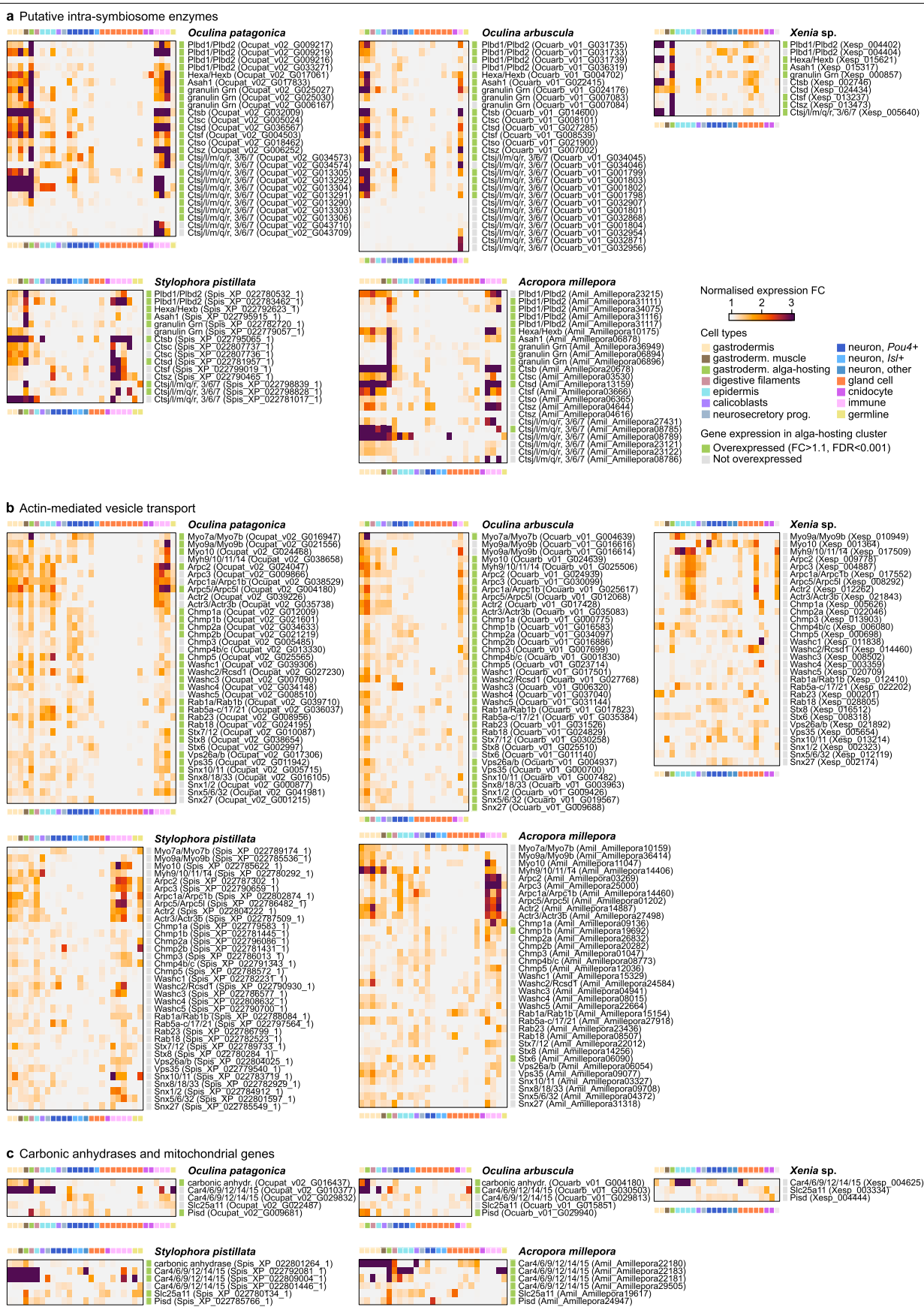
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Expression profile of alga-hosting marker genes across species, focusing on membrane transporters. Expression of selected genes involved in transport functions in the symbiosome (panel **a**), lipid transport (**b**), cell membrane transport (**c**), and vacuole acidification (**d**). Expression is shown as normalised fold changes at the cell type level, for five coral species (*O. patagonica*, *O. arbuscula*, *S. pistillata*, *A. millepora* and *Xenia* sp.). Each gene is colour-coded according to whether it is significantly over-expressed in the alga-hosting cell cluster of each species (Wilcoxon rank-sum test on cell type counts, fold-change > 1.1 and *FDR* < 0.001).



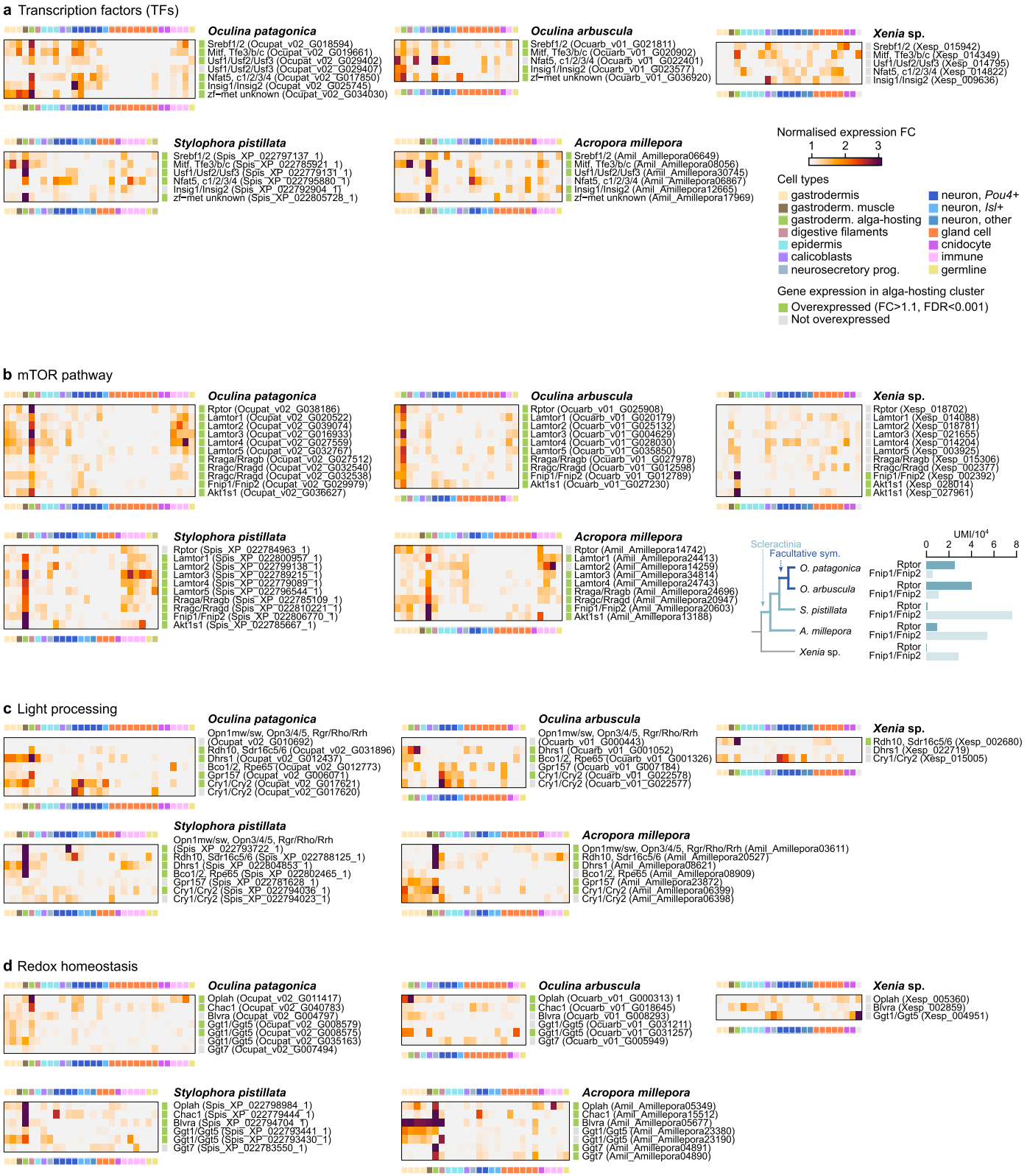
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Expression profile of alga-hosting marker genes across species, focusing on metabolic functions. Expression of selected genes involved in lipid metabolism (panel **a**), sphingolipid metabolism (**b**), galactose catabolism (**c**), glycogen synthesis (**d**), and nitrogen metabolism (**e**). Expression is shown as normalised fold changes at the cell type level, for five coral species (*O. patagonica*, *O. arbuscula*, *S. pistillata*, *A. millepora* and *Xenia* sp.). Each gene is colour-coded according to whether it is significantly over-expressed in the alga-hosting cell cluster of each species (Wilcoxon rank-sum test on cell type counts, fold-change > 1.1 and *FDR* < 0.001). **f**, Fraction of the transcriptome devoted to various metabolic functions in the alga-hosting cells of each of the five corals: lipid and sphingolipid metabolism, galactose catabolism, and glycogen synthesis (measured as UMI per 10^4). **g**, Expression levels as UMI per 10^4 of key genes in the galactose catabolism and glycogen synthesis pathways in the alga-hosting cells of each of the five corals.

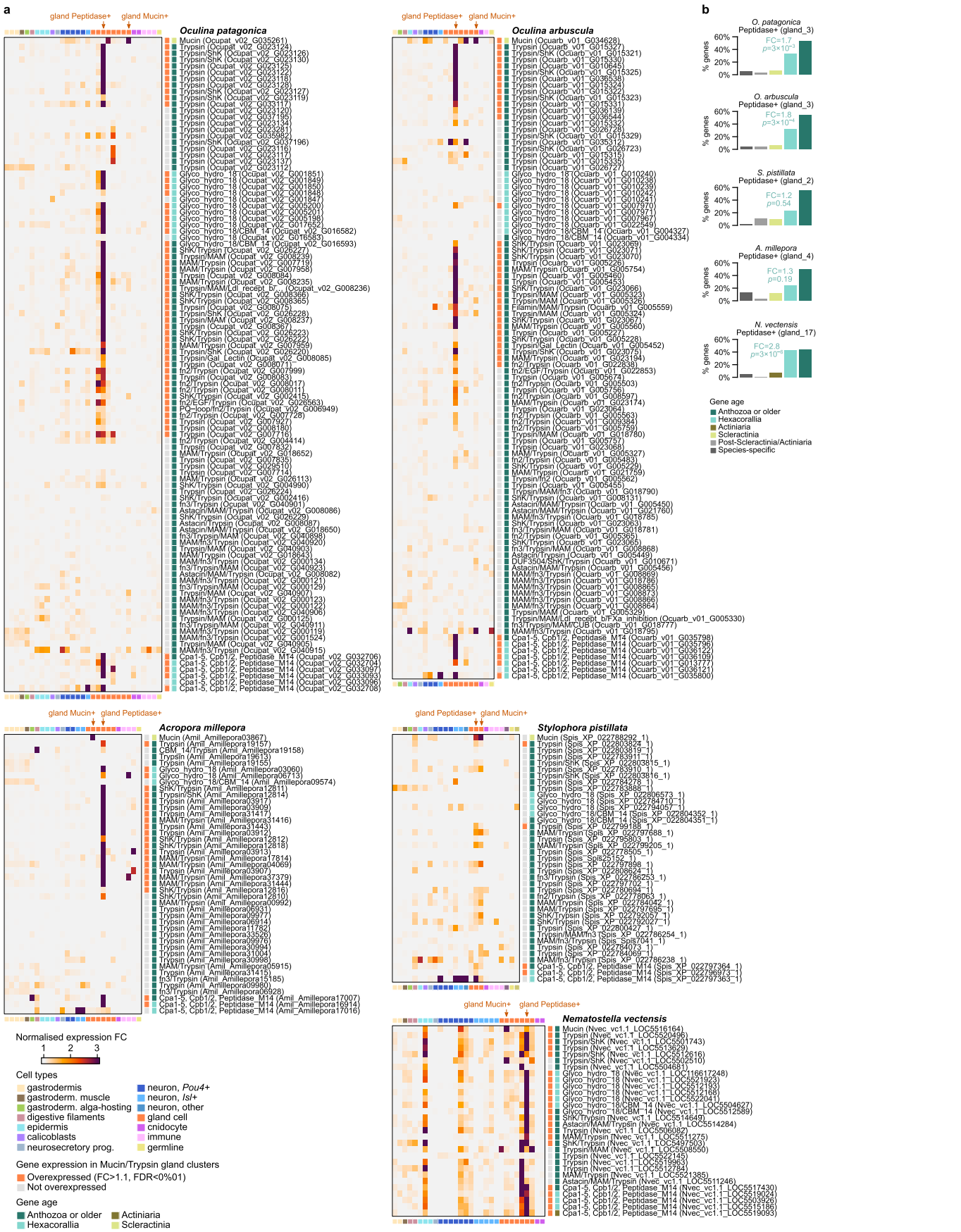


Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Expression profile of alga-hosting marker genes across species, focusing on vesicle activity and transport. Expression of selected genes involved in intra-symbiosome enzymatic processes (panel **a**), actin-mediated vesicle transport (**b**), and other processes (**c**). Expression is shown as normalised fold changes at the cell type level, for five coral species (*O. patagonica*, *O. arbuscula*, *S. pistillata*, *A. millepora* and *Xenia* sp.). Each gene is colour-coded according to whether it is significantly over-expressed in the alga-hosting cell cluster of each species (Wilcoxon rank-sum test on cell type counts, fold-change > 1.1 and *FDR* < 0.001).



Extended Data Fig. 7 | Expression profile of alga-hosting marker genes across species, focusing on regulatory functions. Expression of selected transcription factor genes (panel **a**), mTOR pathway regulators (**b**), light processing proteins (**c**), and redox homeostasis (**d**). Expression is shown as normalised fold changes at the cell type level, for five coral species (*O. patagonica*, *O. arbuscula*, *S. pistillata*, *A. millepora* and *Xenia* sp.). Each gene is colour-coded according to whether it is significantly over-expressed in the alga-hosting cell cluster of each species (Wilcoxon rank-sum test on cell type counts, fold-change > 1.1 and *FDR* < 0.001). In panel b, we report the expression levels (UMI per 10^4) of two mTOR pathway regulators (*Rptor* and *Fnip1/2*) in the alga-hosting cells of each of the five corals.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Expression profile of gland cell marker genes across species. a, Expression of mucin genes and selected peptidases (with Trypsin, Peptidase_M14 and Glyco_hydro_18 Pfam domains). Expression is shown as normalised fold changes at the cell type level, for four coral species (*O. patagonica*, *O. arbuscula*, *S. pistillata*, *A. millepora*) and a heterotrophic sea anemone (*N. vectensis*). Each gene is colour-coded according to whether it is significantly over-expressed in either the Mucin+ or the Peptidase+ gland cell cluster of each species (Wilcoxon rank-sum test on cell type counts, fold-change > 1.1 and *FDR* < 0.001); and according to its age (node of origin in the anthozoan phylogeny). **b**, Gene age distribution of the genes expressed in the Peptidase+ gland cell cluster of each species (for genes with normalised fold-change values > 1.5). Statistical over-representation of genes in each age category was assessed using two-tailed χ^2 tests relative to the rest of the genome.

nature portfolio

Corresponding author(s): Arnau Sebe-Pedros

Last updated by author(s): Aug 26, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

Sofwtare used to collect/preprocess data in this study (package/version/source):
bcl2fastq 2.20 Illumina
Guppy 5.0.17 Oxford Nanopore
Megalodon 2.5 Oxford Nanopore
CellRanger 7.2.0 10x genomics

Data analysis

Scripts to reproduce the data processing and downstream analysis are available in GitHub: https://github.com/sebepedroslab/oculina-coral-sc-atlas/.

Open source software/packages used in for data analysis in this study (package/version/source):
data.table 1.13.0 cran.r-project.org
metacell 0.3.41 https://github.com/tanaylab/metacell
preprocessCore 1.44.0 cran.r-project.org
circlize 0.4.10 cran.r-project.org
ComplexHeatmap 2.5.3 cran.r-project.org
SparseM 1.77 cran.r-project.org
topGO 2.34.0 Bioconductor
GO.db 3.7.0 Bioconductor
S4Vectors 0.20.1 cran.r-project.org

BRAKER 2.1.6 https://github.com/Gaius-Augustus/BRAKER
Stringtie 2.2.1 https://ccb.jhu.edu/software/stringtie/
Augustus 3.5.0 https://github.com/Gaius-Augustus/Augustus
Mikado 2.3.4 https://mikado.readthedocs.io/
ggplot2 3.3.2 cran.r-project.org
phytools 0.7-47 Bioconductor
treeio 1.6.2 Bioconductor
tidytree 0.3.3 Bioconductor
rstatix 0.6.0 cran.r-project.org
ape 5.0 Bioconductor
zoo 1.8-8 cran.r-project.org
stringr 1.4.0 Bioconductor
mcl 22.282 bioconda
pfam_scan 1.6 bioconda
hmmer 3.3 bioconda
diamond 2.1.8.162 bioconda
easel 0.45 bioconda
ete3 3.1.1 pypi
iqtree 2.1.0 bioconda
mafft 7.471 bioconda
clipkit 0.1.2 bioconda
GenomicRanges 1.54.1 Bioconductor
GenomicFeatures 1.54.3 Bioconductor
bwa 0.7.17 https://github.com/lh3/bwa
pairtools 0.2.2 https://github.com/open2c/pairtools
cooler 0.8.11 https://github.com/open2c/cooler
cooltools 0.5.1 https://github.com/open2c/cooltools
coolpuppy 1.1.0 https://github.com/open2c/coolpuppy
coolbox 0.3.8 https://github.com/GangCaoLab/CoolBox
pybbi 0.4.0 https://github.com/nvictus/pybbi
SIP_HiC 1.6.1 https://github.com/PouletAxel/SIP

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All sequencing data is deposited in GEO under accession number GSE289546. The O. patagonica genome assembly and annotation are available in NCBI, under BioProject PRJNA1270949. All generated datasets can be explored in an interactive database https://sebelab.crg.eu/multicoral-sc-atlas/.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No formal sample size was defined, since this study did not test predefined hypotheses across experimental groups. Rather, sequencing depth and the number of libraries were chosen to provide adequate power to capture cell type diversity and generate a dataset sufficient to support the study's key conclusions.
Data exclusions	No data were excluded from the analysis.
Replication	For Oculina patagonica single-cell atlas profiling, two different specimens were collected for each condition (aposymbiotic and symbiotic).
Randomization	Randomization was not applicable in this study because the experimental design did not involve treatment groups, interventions, or subjective outcome assessments where investigator knowledge could introduce bias.
Blinding	Blinding was not applicable in this study because the experimental design did not involve treatment groups, interventions, or subjective outcome assessments where investigator knowledge could introduce bias. A single-cell atlas study is primarily descriptive and exploratory, aiming to comprehensively map cell types, states, and molecular features within a given tissue or organism.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Acropora digitifera
Wild animals	Oculina patagonica. Adult symbiotic and naturally aposymbiotic colonies of Oculina patagonica were collected from the wild in July from the Israeli Mediterranean Sea near Michmoret (32.24049°N, 34.51530°E). The corals were acclimated and cultured for a month before dissociation in a controlled aquarium system at the Leon H. Charney School of Marine Sciences, University of Haifa. They were maintained in artificial seawater (Red Sea Salt, Red Sea Ltd.) with a salinity of 39 ppt at 25°C under a 12-hour light/dark photoperiod with a PAR level of 50 µmol m^{-2} s^{-1}. Corals were fed twice weekly with planktonic coral food (Reef Snow, Brightwell Aquatics) according to the manufacturer's instructions. Stylophora pistillata. Stylophora pistillata colonies were collected from the Gulf of Eilat/Aqaba, in front of the Interuniversity Institute of Marine Biology (IUI) in Eilat, Israel (29.501775°N, 34.917846°E)
Reporting on sex	NA
Field-collected samples	NA
Ethics oversight	The study did not require an ethical approval

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Frozen samples were thawed on ice, and cells collected by centrifugation at 2000 xg for 5 min at 4 °C. Cells were washed once with 2 ml of Resuspension Buffer 2 (RB2; 1X PBS, 0.5% BSA, 80 U/ml Ribolock), pelleted again and finally suspended in 1 ml of RB2. Cell nuclei were stained with 1:300 DRAQ5 (Thermo #62251). Either 8,000 or 40,000 cells (for multiplexed experiments) were sorted into a well of a 96-well plate using a FACSAria II SORP cell sorter following the recommendations by 10X Genomics. Non-cellular particles were discriminated by selecting only DRAQ5-positive cells and doublets/multiplets were excluded using forward scatter width (FCS-W) versus forward scatter height (FCS-H).
Instrument	BD FACSAria II Flow Cytometer
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
Cell population abundance	All single cells were sorted in order to sampling cell type diversity with minimal biases.
Gating strategy	We applied 2 successive gates (detailed in Methods): 1. DRAQ5+ particles, to distriminate between cells and non-cellular particles. 2. Single-cells: double/multiplet exclusion was performed using FSC-W versus FSC-H We did not apply any size gating, as we expect high heterogeneity in cell size/granularity in a whole-organism.

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.