

Host genotype and macro-chromosomal inversions influence gene expression and microbiome assembly in a complex holobiont

Carles Galà-Camps

`carles.galia@mncn.csic.es`

Museo Nacional de Ciencias Naturales (MNCN-CSIC) <https://orcid.org/0000-0002-0945-0249>

Josep Ramoneda

Center of Advanced Studies of Blanes (CEAB)

Marta Pascual

University of Barcelona <https://orcid.org/0000-0002-6189-0612>

Carlos Carreras

University of Barcelona <https://orcid.org/0000-0002-2478-6445>

Xavier Turon

Centre d'Estudis Avançats de Blanes (CEAB-CSIC)

Celia Schunter

The University of Hong Kong (UHK)

Cinta Peguerols

Universitat Autònoma de Barcelona (UAB)

Article

Keywords: *Styela plicata*, Whole genome sequencing, host transcriptome, microbiome, chromosome inversions, host-symbiont association, non-native species

Posted Date: October 16th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11283491/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest. Please, note that all Supplementary tables (15 table sin total) have been provided in a single excel file. If you encounter any issues when downloading it, please, contact me.

Host genotype and macro-chromosomal inversions influence gene expression and microbiome assembly in a complex holobiont

Carles Galià-Camps^{1,2,3,4,*}, Josep Ramoneda⁴, Marta Pascual^{2,3}, Carlos Carreras^{2,3}, Xavier Turon⁴, Celia Schunter⁵, Cinta Pegueroles⁶

1: Department of Biodiversity and Evolutionary Biology, Museo Nacional de Ciencias Naturales (MNCN-CSIC), Calle José Gutiérrez Abascal 2, 28006, Madrid, Spain

2: Departament de Genètica, Microbiologia i Estadística, Universitat de Barcelona (UB), Avinguda Diagonal 643, 08028 Barcelona, Spain

3: Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain

4: Department of Marine Ecology, Centre d'Estudis Avançats de Blanes (CEAB-CSIC), Accés Cala St. Francesc 14, 17300 Blanes, Girona, Spain

5: Swire Institute of Marine Science, School of Biological Sciences, The University of Hong Kong (UHK), Pokfulam Road, Hong Kong SAR

6: Department of Genetics and Microbiology and Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona (UAB), 08193 Bellaterra, Barcelona, Spain

* corresponding author: cgaliacamps@gmail.com

co-author	e-mail	ORCID
Carles Galià-Camps	carles.galia@mncn.csic.es	0000-0002-0945-0249
Josep Ramoneda	josep.ramoneda@ceab.csic.es	0000-0002-1806-2889
Marta Pascual	martapascual@ub.edu	0000-0002-6189-0612
Carlos Carreras	carreras@ub.edu	0000-0002-2478-6445
Xavier Turon	xturon@ceab.csic.es	0000-0002-9229-5541
Celia Schunter	schunter@hku.hk	0000-0003-3620-2731
Cinta Pegueroles	cinta.pegueroles@uab.cat	0000-0003-0701-9866

Host genotype and macro-chromosomal inversions influence gene expression and microbiome assembly in a complex holobiont

ABSTRACT

Genome architecture and gene regulation can influence phenotypes without requiring new mutations. In multicellular organisms, these processes operate within holobionts, where host fitness depends on interactions with associated microbial communities. However, how structural genomic variation influences gene regulation and microbiome assembly in natural populations remains poorly understood. Here we investigate the non-native ascidian *Styela plicata*, a lower chordate harbouring four chromosomal inversions and a rich microbiome. Using integrative genomic, transcriptomic, and microbiome analyses across populations and inversion karyotypes, we show that inversion karyotypes are associated with extensive genome-wide shifts in gene expression beyond the inversions. While population-specific genetic variation structures overall microbiome diversity, we find specific karyotypes associated with shifts in microbial taxa with putative host-associated functions. We also identify widespread correlations between host overexpressed genes featuring functions enriched in lipid metabolism and immune signalling pathways, and the abundance of specific microbes involved in host–microbe homeostasis. Together, our results demonstrate that holobiont adaptation emerges from the interplay between genome architecture, transcriptional regulation, and microbiome assembly, revealing chromosomal inversions as key modulators of host–microbe evolutionary dynamics.

Keywords: *Styela plicata*, Whole genome sequencing, host transcriptome, microbiome, chromosome inversions, host-symbiont association, non-native species

Abbreviations: WGS=Whole Genome Sequencing, DEG=Differentially Expressed Gene, ASV=Amplicon Sequence Variant, iASV=indicator Amplicon Sequence Variant, Pop=Population, Chr2=Chromosome 2, Chr4=Chromosome 4, Chr 11=Chromosome 11, Chr16=Chromosome 16

MAIN

Adaptation is traditionally understood as a property of host genomes, typically emerging through changes in allele frequencies under natural selection^{1,2}. Yet, evolutionary change arises not only from the genomic variants a species carries, but also on how those variants are structured and affect gene regulation^{3,4}. Changes in genome architecture can thus modify gene expression, potentially leading to adaptive divergence without requiring mutations in genic regions⁵. Multicellular organisms further complicate this picture by existing as holobionts⁶, whose phenotypes emerge from interactions between host biology and associated symbiotic communities^{7,8}. In this scenario, microbiomes can influence host nutrition⁹, immunity¹⁰, development⁹, or stress tolerance¹¹, substantially contributing to the holobiont ecological and adaptive performance^{8,12}. The microbiome can be vertically transferred together with the host genomic information across generations¹³, thus constituting an evolutionary unit on which natural selection can act^{14,15}. Nonetheless, the microbiome can also be horizontally transmitted contributing to fast adaptation in response to environmental change¹², but it remains unclear how host genomic variation interacts with microbial community structure during adaptation in natural populations. Understanding adaptation thus requires consideration of genome content, genome regulation and holobionts interactions. Integrating these perspectives will enhance our understanding of the environmental and ecological factors that drive species' evolutionary responses⁸.

Structural genomic variants, particularly chromosomal inversions, are important contributors to local adaptation by maintaining co-adapted allelic combinations and generating alternative regulatory environments¹⁶⁻¹⁸. By reducing recombination and reorganising chromosomal folding, inversions can promote coordinated shifts in gene expression across large genomic regions, producing distinct transcriptional states potentially associated with ecological divergence^{18,19}. Although inversions are increasingly recognised as key drivers of phenotypic divergence, whether their regulatory consequences extend beyond the host to influence the assembly and function of host-associated microbiomes remains unknown. Microbiome composition varies across host tissues, ontogenetic stages, populations, and environments^{12,20}, with host genetic effects on microbiome structure documented in several systems^{21,22}. Nonetheless, studies trying to link host genome and symbiont community rarely consider genome architecture or gene-regulatory variation. We consequently lack an integrated understanding of how different genomic architectures influence host gene expression and microbiome composition in environmental settings.

Lower chordates such as tunicates offer a powerful opportunity to address how gene expression and microbiome composition are influenced by host genomic architecture in the wild. As early-diverging chordates with relatively simple body plans and intimate microbial associations, they offer insight into the main principles of host–microbe integration^{21,23,24}. Among tunicates, many invasive Ascidiacea species inhabit environmentally heterogeneous marine coastal systems^{25,26}, where local selection can act on standing genomic variation while connectivity among populations can be maintained by human-related repeated introductions^{27,28}. These features make them ideal for examining how genome architecture, gene regulation, and microbiome composition interact during ongoing adaptation. Here, we investigate holobiont-level interactions in the invasive ascidian *Styela plicata*, a species that exhibits four inversions on independent chromosomes¹⁶, using an integrated multi-omic framework that combines host genomics, transcriptomics, and microbiome profiling across heterogeneous populations and alternative chromosomal inversion karyotypes (Fig. 1). Our study explores a complex scenario in which the multilevel interplay of mutation, genome architecture, gene expression, and microbiome community composition is crucial for holobiont system homeostasis.

GENOMIC VARIATION UNDERLIES DIFFERENTIAL EXPRESSION

To investigate how host genomic variation and chromosomal architecture shape holobiont-level responses in natural populations of *S. plicata*, we sampled five adult individuals (>5 cm) per site from three Mediterranean harbours spanning a gradient of anthropogenic impact: the small harbour of Blanes, the medium harbour of ‘Vilanova i la Geltrú’ (abbreviated from now on as Vilanova), and the large harbour of Barcelona (Fig. 1, Supplementary Table S1). Whole-genome sequencing (WGS) was performed on the gill tissue of 15 specimens, and population structure together with inversion karyotype distribution was assessed using 7,188,230 high-quality SNPs (Supplementary Tables S1–S2). Multidimensional scaling (MDS) analyses based exclusively on chromosomes lacking inversions recovered the three localities as genetically distinct clusters (Fig. 1). The great resolution afforded by WGS revealed fine-scale genomic structure among localities separated by only ~45–105 km of coastline (Supplementary Table S1). These results suggest that strong environmental heterogeneity across harbours due to different levels of anthropisation may contribute to the maintenance of locally differentiated genomic backgrounds. In contrast, none of the chromosomes carrying inversions (Chromosome 2 (Chr2), Chromosome 4 (Chr4), Chromosome 11 (Chr11), and Chromosome 16 (Chr16)) exhibited population-specific differentiation (Fig S1). Instead, the same inversion karyotypes were broadly distributed across all localities, with alternative arrangements occurring at unequal

frequencies among localities (Fig. 1; Fig. S1; Supplementary Table S3), consistent with previous observations from a global-scale sampling^{16,29}. The absence of population-specific fixation of inversion karyotypes suggests that chromosomal inversions in *S. plicata* may constitute standing structural variation maintained across populations, providing a reservoir of genomic and regulatory diversity^{16,29}.

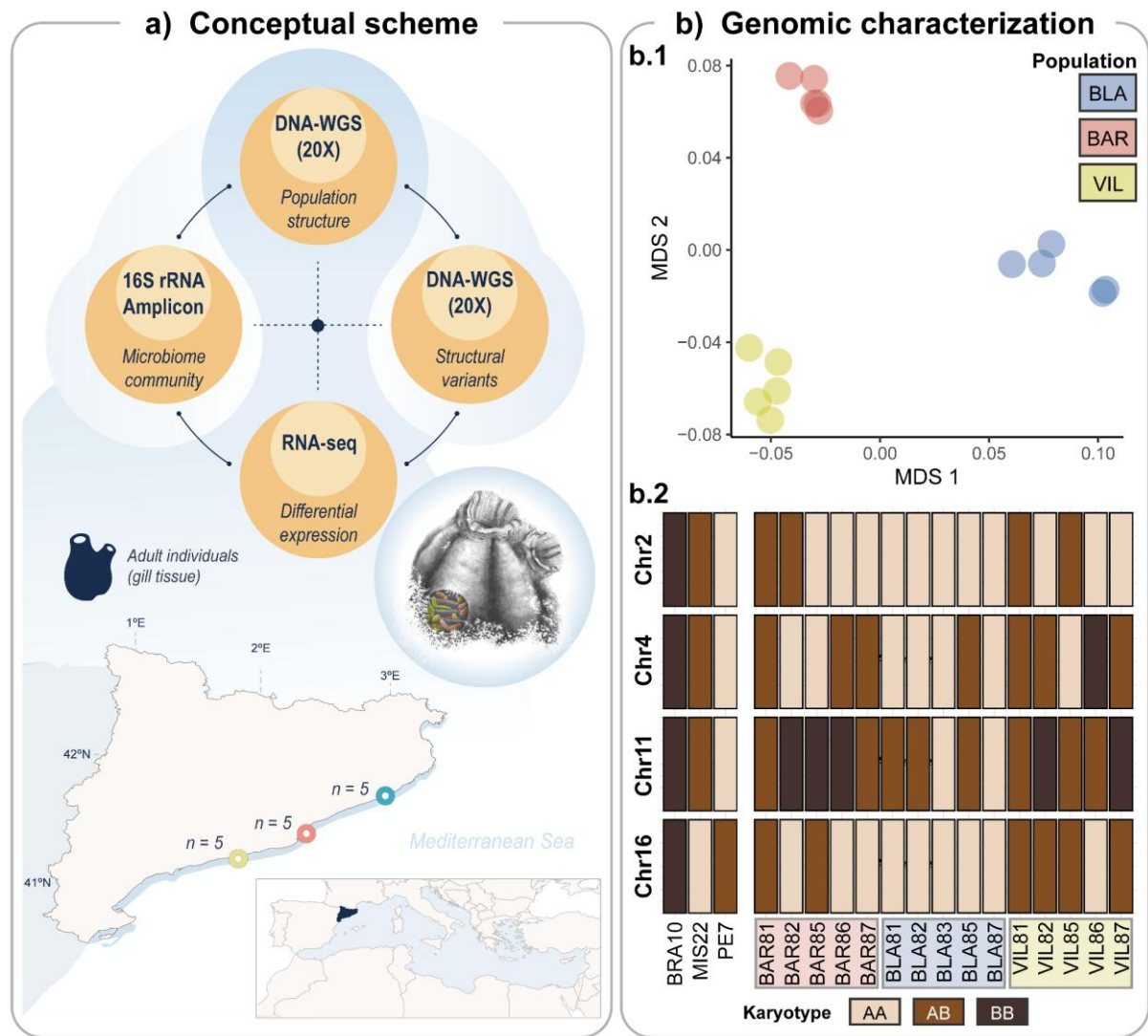


Figure 1 | Conceptual scheme and genomic features of the *Styela plicata* samples. **a)** Conceptual scheme for the study design of this work including the type of data obtained, the analyses performed and the sampling locations. **b)** Genetic structure of the 15 samples based on WGS data, excluding chromosomes with inversions, coloured by population (b.1) and chromosome inversion karyotyping of the samples (b.2). Samples BRA10, MIS22, and PE7¹⁶ were included in the karyotype identification analysis as references. The schematic illustration was designed by Gisela Marín-Capuz. *Styela plicata* hosting microbiome illustration was designed by Jagoba Malumbres Olarte.

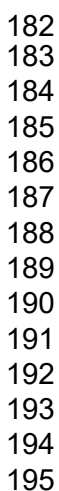
To assess transcriptional variation associated with population identity (Pop), as well as with the different chromosomal inversions, we performed RNA-seq on gill tissue across the same 15 individuals. The ascidian gill, homologous to the vertebrate pharynx³⁰, is a metabolically active

organ that interfaces directly with seawater and constitutes a primary contact zone with environmental microorganisms and potential pathogens ³¹. Because RNA-derived genotypes can differ from genomic genotypes owing to transcription-associated and post-transcriptional sequence variation ^{32–34}, we compared transcriptome-derived and genome-derived genetic structures using MDS for the SNPs from chromosomes without inversions and SNPs for each of the four chromosomes with inversions. Transcriptome-based MDS recapitulated the patterns recovered from whole-genome sequencing, resolving the same three population clusters and identical inversion karyotype distributions across chromosomes (Fig. S2), indicating strong concordance between genomic and transcriptomic datasets. To rule out potential transcriptional artefacts arising from sampling stress or locality-specific effects ³⁵, we performed a principal component analysis (PCA) on all expressed genes. Global transcriptional profiles were highly consistent across samples showing no spurious clustering (Fig. S3).

After restricting the statistical analyses to groups (Pop and each of the karyotypes) represented by more than four individuals to reduce inflation of false positives in low-replicate groups ³⁶, we identified differentially expressed genes (DEGs) among populations and between the most abundant chromosomal inversion karyotypes. To avoid confounding effects, we removed DEGs which were found by two or more analyses (Fig. 2). At the population level, we identified 153 DEGs (Supplementary Table 4). From these, 144 were unique to populations. These 144 unique DEGs showed a differential expression profile in individuals from Blanes relative to those in Vilanova and Barcelona, which exhibited a more similar expression pattern between them (Fig. 2, Fig. S4). This pattern may reflect differences in environmental conditions among harbours, since Blanes is a small harbour with less anthropogenic activity compared with Vilanova and Barcelona, which are more urbanised and industrialised areas and, therefore, associated with higher levels of anthropogenic contamination ^{23,37}. Similar patterns of environmentally associated transcriptional differentiation have been reported across a broad range of taxa in both laboratory and field settings, including sponge holobionts, intertidal crabs, and humans ^{38–40}, among others. When considering chromosomes hosting inversions, Chr11 exhibited a pronounced number of differentially expressed genes between karyotypes, with 125 unique DEGs. In contrast Chr2, Chr4, and Chr16 presented lower numbers of differential signals involving 19, 14 and 9 unique DEGs respectively (Fig. 2, Supplementary Table 4). Differential expression patterns based on unique DEGs grouped individuals according to their karyotype (Fig. 2, Fig. S4), with the exception of a few individuals in Chr2, suggesting that karyotypes are also key drivers of gene expression. Interestingly, DEGs were distributed inside but also

outside focal inversions (Fig. 2), suggesting that transcriptional effects associated with inversions may extend beyond rearranged genomic regions, which might reflect both local and trans-regulatory responses. A hypothesis explaining genome-wide transcriptional signatures associated with inversions is that inversions may influence higher-order chromatin organisation. In this scenario, different inversion karyotypes might affect chromosomal folding and nuclear positioning, potentially altering contacts between genomic regions and reshaping intra and inter-chromosomal interaction landscapes including topologically associating domains (TADs) ⁴¹, as demonstrated with experimental work in model systems ⁴². Together, these findings suggest that structural variants may contribute to genome regulation, although the extent to which such mechanisms drive adaptation in *S. plicata* remains unclear. Future work combining high-resolution imaging with chromatin conformation and accessibility assays will help elucidate whether chromosomal inversions influence chromatin interactions, regulatory accessibility, and gene expression.

To understand which functions were involved in the potentially adaptive processes of *S. plicata* both at the population and inversion levels, we compiled the available GO Terms of DEGs (Supplementary Table S4). Functional annotation was possible for 30% of the genes, and we observed biological processes related to transport and intercellular signalling, lipid biosynthesis, lipid transport, and lipid metabolism, whereas their molecular function categories were related to enzymatic activity, cofactor binding, and transporter functions (Supplementary Table S4).



8

HOST GENOMIC FEATURES INFLUENCE MICROBIOME COMMUNITY

To evaluate whether population structure and chromosomal inversions are associated with variation in microbiome composition, we analysed a previously published 16S rRNA gene amplicon dataset²³ that includes the gill microbiome from the same *S. plicata* individuals used in the present study. The dataset contained 5,377 amplicon sequence variants (ASVs) from gill-associated communities (Supplementary Table S5) belonging to 32 prokaryotic phyla (bacteria and archaea). Alpha-diversity measured as ASV richness significantly differed among populations, with Blanes exhibiting the highest values followed by Barcelona and Vilanova (Fig. S5; Supplementary Tables S6–S7). Regarding ASV richness across inversions, we detected that the reference homozygous karyotype (AA) of Chr2 and Chr16 had significantly higher values than their heterozygous (AB) counterpart. Interestingly, the AA karyotype of Chr2 was the only group across all comparisons which showed a significantly higher Shannon diversity compared to AB (Fig. S5; Supplementary Table S6). We next examined differences in community composition using non-metric multidimensional scaling (nMDS) based on Bray–Curtis dissimilarities. The nMDS ordination showed that prokaryote communities from the different populations had a distinct taxonomic composition (PERMANOVA pseudo-F = 2.907, $R^2 = 0.326$, $p = 0.004$; 10,000 permutations) (Fig. 3; Supplementary Table 8), mainly driven by the populations of Blanes and Barcelona, which correspond to the smallest and largest harbors, respectively. In contrast, the taxonomic compositions of prokaryote communities from hosts with different inversions were overall not significantly different from each other, with the exception of Chr2 (PERMANOVA pseudo-F = 2.603, $R^2 = 0.167$, $p = 0.025$; 10,000 permutations) (Fig. 3; Supplementary Table 8). Overall, these results suggest that while population effects are the dominant driver of microbiome structure, inversions (and more specifically the one in Chr2) contribute to variation in alpha diversity metrics and community composition. Future work using a larger sample representation is advisable to validate our results and to further investigate Chr2 karyotype variation across broader geographic scales to assess whether its association with microbiome diversity is robust and generalisable across its global distribution.

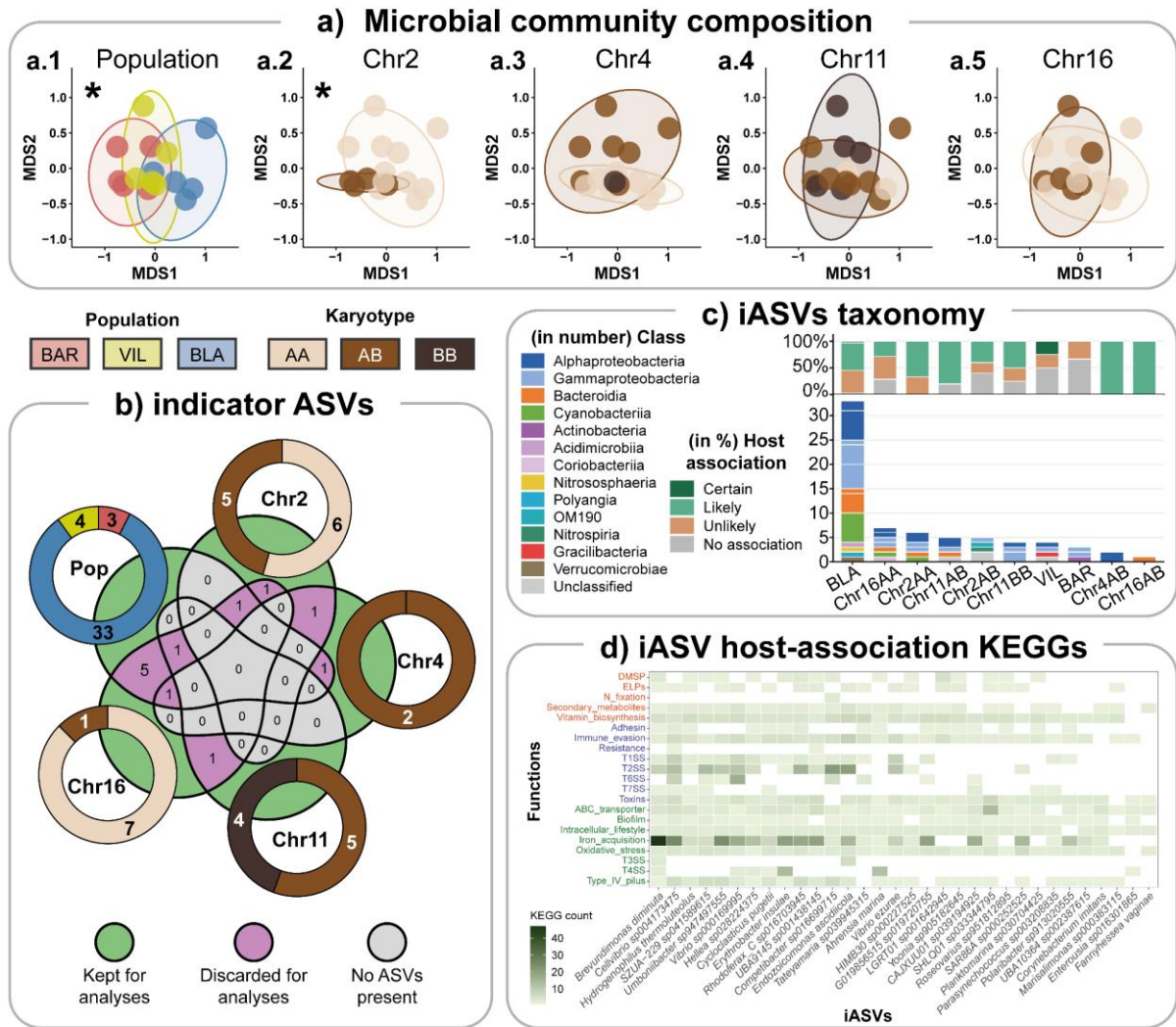


Figure 3 | Microbiome community structure and indicator ASVs (iASVs) in *Styela plicata*. **a)** Non-metric multidimensional scaling (nMDS) ordinations of prokaryotic communities colored by Pop (a.1) and inversion karyotype for Chr2 (a.2), Chr4 (a.3), Chr11 (a.4), and Chr16 (a.5). Asterisks in the top-left corner of the panels indicate significant differences among represented groups based on a PERMANOVA analysis with 10,000 permutations when considering only groups with more than four specimens. **b)** Number of iASVs associated with each population and inversion karyotype. Donuts indicate the number of iASVs uniquely associated with each population or karyotype. Karyotypes with fewer than four specimens were excluded before statistical analyses. **c)** Number and taxonomic assignment of the iASV associated to different groups and potential host-interaction. **d)** KEGG repertoire of iASVs whose genome could be recovered through GTDB. KEGG names in red correspond to symbiont lifestyle functions, KEGG names in blue correspond to parasitic and/or pathogenic functions, and KEGG names in green correspond to host-dependent (symbiont and pathogen) lifestyle functions.

Previous studies in holobiont systems have shown that microbiome composition can be associated with host genotype^{21,22}. Thus, we performed an indicator species analysis to identify which ASVs are most frequently associated with population or chromosomal inversion karyotypes as indicator ASV (iASVs). This approach has previously been applied in *S. plicata* to resolve environmentally associated prokaryotic fractions linked to seasonal adaptation¹². Since low-abundance taxa may contribute disproportionately to host-associated functions⁴³, we kept only those iASVs for which a minimum of 75% of the reads were present in at least 75%

of samples of the group (population or karyotype) (Fig. 3). We identified 70 unique iASVs associated with either populations (40 iASVs) or inversion karyotypes (30 iASVs), spanning 44 families across 11 phyla (Fig. 3, Supplementary Table 9). A large fraction of population-associated iASVs (33/40) were specific to Blanes (Fig. 3), consistent with patterns observed in ASV richness and Shannon analyses (Figure S5). Focusing on chromosomal associations, Chr2 accounted for 11 of the 30 inversion-associated iASVs, with similar representation across AA and AB karyotypes (Fig. 3, Supplementary Table 9). Remarkably, 36 of the 70 iASVs belonged to taxa previously reported as common core members of coral, sponge, or ascidian microbiomes. Examples are those of Francisellaceae⁴⁴, *Nitrosopumilus*⁴⁵, *Erythrobacter*⁴⁶, and *Litorimonas*⁴⁷, among others (Supplementary Table 9). In *S. plicata*, ASV1 and ASV2 were the dominating iASV of adult gill microbiomes²³. While these two ASV mean abundance in our dataset per sample are ~21,000 and ~19,000 respectively, the iASV we obtained in this study for population and karyotype have <200 reads per sample, with the exception of one Chr11-associated ASV3 (unidentified bacteria), which exhibited ~19,152 reads per sample. The role of key low-abundance bacteria in boosting community ecological performance has been widely acknowledged in environmental systems under the term “rare biosphere”⁴⁸. This rare-biosphere can have a disproportionate role in biogeochemical cycles and act as a hidden driver of overall microbiome function⁴⁹. Recently, similar key roles for adaptation of rare microbial taxa have been acknowledged in marine holobiont systems^{12,44}, and this could also be the case for *S. plicata*.

In order to explore the genomic repertoire of these indicator taxa, we obtained representative high-quality genomes from the Genome Taxonomy Database (GTDB; <https://gtdb.ecogenomic.org/>). Genomes were obtained by mapping the 16S rRNA gene amplicons of the iASVs to the genomes in GTDB, with matches at >99% identity being considered generally representative of the genomic makeup of those taxa (an approach that has proven effective in several previous studies)^{50–52}. We obtained the genomes for 32 out of the 70 iASVs (Supplementary Table 10) and identified potential associated functions (Supplementary Table 11). These genomes were then investigated for the presence of functions usually found in host-associated taxa in terms of metabolism (vitamin biosynthesis, Fe acquisition, dimethylsulfoxide catabolism, and production of typical symbiont secondary metabolites), interaction with the host (secretion systems, eukaryotic-like proteins, evasion of the immune system), and adhesion mechanisms (type IV pili, adhesins)^{50–52} (Fig. 3, Supplementary Table 11). This functional gene analysis identified iASV with a rich repertoire

of genes encoding functions commonly associated with host-associated lifestyles (Fig. 3, Figure S6). Among them, we found genomes belonging to two *Vibrio* species (*Vibrio ezurae* and *Vibrio* sp.), *Endozoicomonas ascidiicola*, *Erythrobacter insulae*, and a member of the marine Pseudohangiellaceae family which stood out for the richness of their host-associated gene repertoires (Fig. 3, Figure S6). All these taxa have previously been identified in sponge or tunicate microbiomes^{53,54} underscoring their likely involvement in interactions with the host. Indicator associations included both population-specific and inversion-associated taxa (e.g. *Vibrio* sp. with Chr2; *Pseudohongiellaceae* with Chr11; *Umboniibacter* with Chr16) (Figure S6), suggesting that both population structure and chromosomal variation are associated with shifts in prokaryotic taxa with potentially relevant functions to host fitness. The genomes of the molluscan symbiont *Umboniibacter*⁵⁵ (indicator of Chr16) the free-living bacteria *Competibacter* (indicator of Chr2) and *Rhodoferrax C* (indicator of Pop)⁵⁶, and the cellulose degrading *Cellvibrio* sp.⁵⁷ (indicator of Pop), added to the list of taxa particularly well-equipped for host-symbiont interaction (Fig. 3, Figure S6). Finally, although not presenting a competitive functional toolkit which might be associated with host-symbiont interaction, we also identified typical components of marine sponge and tunicate microbiomes such as other *Erythrobacter* members⁴⁶ and *Litorimonas*⁴⁷.

AN INTERACTIVE TOOLKIT FOR AN ADAPTIVE HOLOBIOME

To explore potential associations between host transcriptional variation and prokaryotic community composition, we independently tested correlations between the abundance of each iASV and the expression level of each DEG across populations and inversion karyotypes using a threshold of $R^2 > 0.5$ to identify strong associations. In the population-level analysis, 38 iASVs (95% of iASVs) showed significant positive correlations with 46 DEGs (32% of DEGs) (Fig. 4; Supplementary Table 12). Such correlation patterns may arise from multiple, non-mutually exclusive processes, including host-mediated regulation of prokaryotic assemblages, prokaryotic modulation of host gene expression, or shared responses of the holobiome as a whole to external environmental drivers. Similar host-microbiome associations have been reported in model systems such as mice and humans, particularly in controlled environments⁵⁸. However, to our knowledge, this represents the first genome-wide evidence of host-microbiome transcriptional associations in a wild non-model invertebrate holobiont, providing a framework for investigating these interactions across natural populations and diverse non-model systems. At the genome scale, genes being differentially expressed which correlate with iASVs were broadly distributed across chromosomes, including both inversion and non-

inversion regions (Fig. 4). Moreover, these DEGs were often associated with multiple prokaryotic taxa (Fig. 4). This suggests that host–microbiome covariation is a polygenic and genome-wide property rather than being restricted to specific chromosomal regions, consistent with complex regulatory architectures observed in other host systems ⁵⁹.

Across inversions, Chr2, Chr4, and Chr11 also exhibited significant positive associations between iASVs and DEGs, although involving fewer correlations compared to population-level patterns (Fig. 4). In Chr2, two iASVs were associated with six DEGs; in Chr4, one iASV was associated with two DEGs; and in Chr11, six iASVs were associated with 22 DEGs (Supplementary Table 12). As previously mentioned, most correlated iASVs corresponded to low-abundance taxa (Supplementary Table 9). These findings suggest that host transcription–microbiome associations across inversion contexts could involve functionally relevant rare-symbionts ⁴³, but also dominant community members such as ASV3 to a lower extent. Note that although Chr2 had a relatively high number of inversion-associated iASVs, only a small subset (2 out of 11) showed correlation with DEGs (Supplementary Table 9, Supplementary Table 12). Importantly, most correlated DEGs were located in other chromosomes (Fig. 4a), suggesting that host–microbiome covariation is not restricted to genes within structural variants. This pattern is consistent with a potential role of genome-wide regulatory networks and higher-order chromatin organisation in shaping host–microbiome interactions ^{5,41,42}.

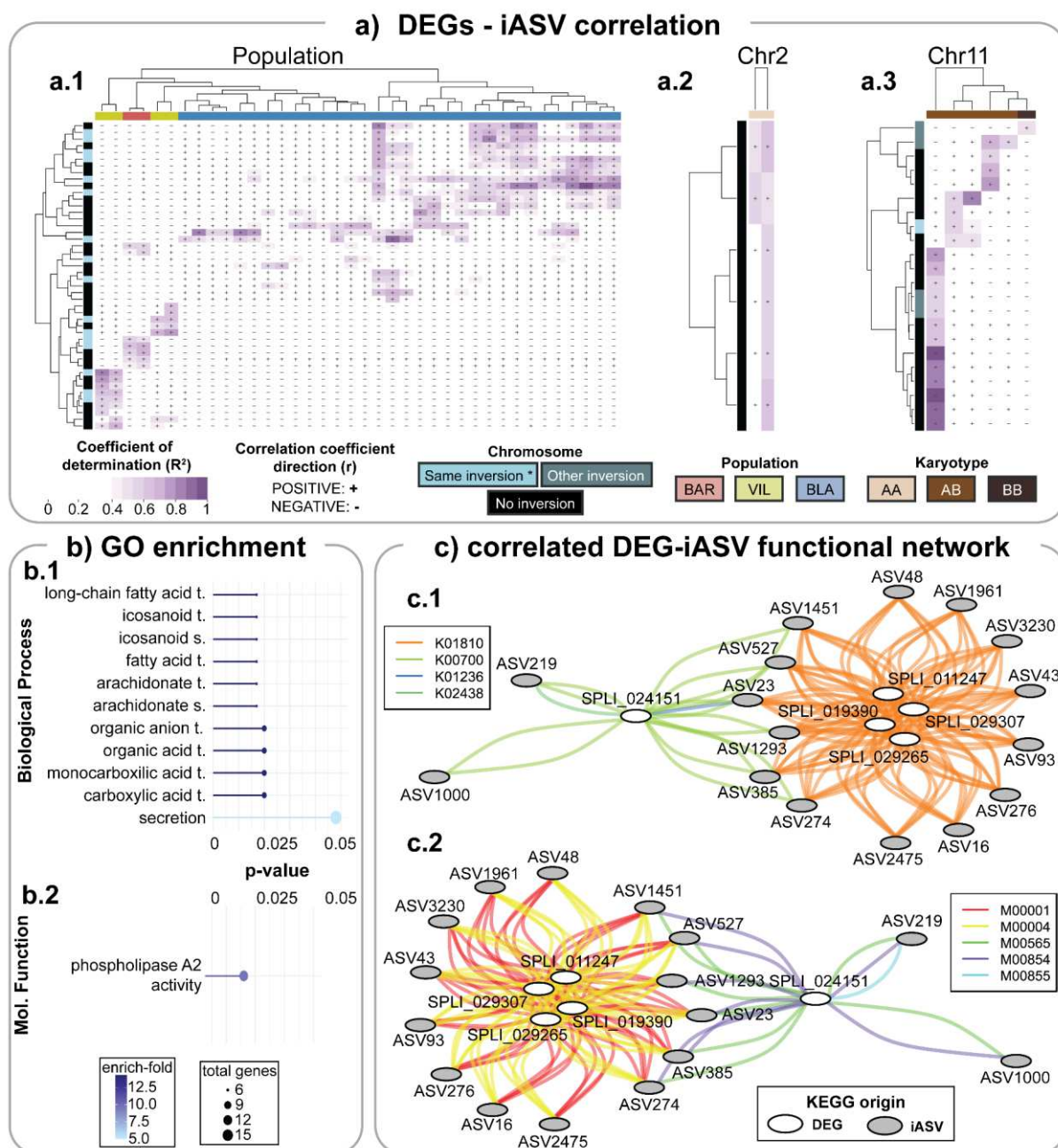


Figure 4 | Host-microbiome functional associations in *Styela plicata*. **a)** Correlation heatmaps linking differentially expressed host genes with indicator ASVs (iASVs) of the different *S. plicata* populations (a.1) and for those indicators of chromosome inversions in Chr2 (a.2) and Chr11 (a.3). Colour intensity denotes the coefficient of determination (R^2), and symbols indicate the direction of the correlation. Gene locations relative to inversions and iASV of population or karyotype are indicated by annotation bars. **b)** Gene Ontology functional enrichment of correlated host genes for Biological Processes (b.1) and Molecular functions (b.2). Note that “t.” stands for transport, and “s.” for secretion. **c)** Functional network of the correlated DEGs and iASV based on KEGG Orthology (c.1) and KEGG Modules (c.2).

In total, we identified 47 iASVs showing significant correlations with DEGs across conditions, spanning 8 phyla and 32 families (Supplementary Table 9). Of these correlated iASV, 58.2% corresponded to taxa previously reported as common core members of ascidian, coral, or sponge microbiomes. Among the correlated taxa, we detected members of Francisellaceae, a

group that includes opportunistic pathogens previously associated with coral disease states ^{60–62}. We also found correlations with *Nitrosopumilus*, a well-documented archaeal symbiont in marine tunicates ^{63–65}. Other correlated groups generally found in coral, sponge, or tunicate microbiomes were the families Flavobacteriaceae (1 ASV; Chr11), Rhodobacteraceae (6 ASVs; Pop, Chr2, and Chr11), Saprospiraceae (3 ASVs; Pop), and Vibrionaceae (2 ASVs; Pop and Chr2) (Supplementary Table 9). We obtained representative genomic information from 19 genomes (Supplementary Table 10) and evaluated within them the presence of functions usually found in host-associated taxa in terms of metabolism, interaction with the host, and adhesion mechanisms (Supplementary Table S11). Five genomes showed a higher representation of these functional categories, including *Vibrio ezurae*, an unidentified *Vibrio*, *Endozoicomonas ascidiicola*, *Erythrobacter insulae*, and a member of the Pseudohangiellaceae family (Figure S7). These taxa have all previously been identified in sponge or tunicate microbiomes ⁵³, underscoring their likely involvement in interactions with marine invertebrate hosts.

We next investigated the functional profiles of DEGs associated with iASV variation. Functional enrichment analysis revealed a predominance of pathways related to lipid metabolism, including processes involving arachidonate and eicosanoid transport and secretion (Fig. 4; Supplementary Table 13). These lipid-derived signalling molecules are known to play important roles in immune modulation and host–microbiome interactions in the intestinal microbiome of vertebrates, as their inhibition leads to imbalances and illnesses of the host ⁶⁶. Similar functionality for these molecules have been reported in diverse invertebrate taxa, including corals, where lipid-mediated signalling has been linked to disease resistance ⁶⁷. In insects, these lipid-derived pathways partially contribute to immune regulation of the organism ⁶⁸. In this context, the presence of lipid-related signalling and metabolic pathways in *S. plicata* suggests that components of host–microbiome regulatory machinery may be broadly conserved across metazoans. Future comparative analyses across additional early-diverging chordates and vertebrates are advised to further support our findings and to clarify when and how these functions became integrated into host–microbiome physiological interactions.

To assess potential functional complementarity between host and bacterial genomes, we examined the distribution of KEGG metabolic modules represented within the DEGs and iASVs. Five host DEGs were linked with metabolic pathways that were either missing a single KEGG entry or fully completed by 16 specific iASVs (Fig. 4; Supplementary Table 14), including five carbohydrate-related metabolic modules: Glycolysis (M00001), Pentose phosphate pathway (M00004), Trehalose biosynthesis (M00565), Glycogen biosynthesis

(M00854), and Glycogen degradation (M00855). Similar patterns of distributed metabolic capacity between host and symbionts have been reported in other holobiont systems, including sponges, where incomplete host metabolic pathways are complemented by prokaryotic functions⁶⁹.

Most DEGs associated with these pathways were shared across multiple prokaryotic taxa, consistent with a potentially redundant and distributed functional landscape across the microbiome. The only exception was the Chr11 DEG SPLI_24151, which was associated with KEGG orthologs K01236 (maltooligosyltrehalose trehalohydrolase) and K02438 (glycogen debranching enzyme) and linked to two specific ASVs (*Corynebacterium imitans* and *Hellea* sp.), suggesting a more restricted set of host–prokaryote associations for this gene (Fig 4, Supplementary Table 10, Supplementary Table 14).

EVOLUTIONARY SIGNALS OF HOST-SYMBIONT INTERACTIONS

To assess the genetic differentiation signatures of iASV-correlated DEGs, and non-correlated DEGs compared with the background genome, we calculated the genome-wide allele fixation index (F_{ST}) in 10,000 bp windows using a 2500 bp sliding window approach, grouping samples by population and by karyotype for Chr2, Chr4, Chr11, and Chr16 (minimum $n = 4$ per group). At the population level, DEGs were broadly distributed across the genome (Fig. 5). However, iASV-correlated DEGs tended to cluster in specific chromosomes (notably Chr11, Chr13, Chr14, and Chr15), suggesting non-random genomic distribution of loci involved in host–microbiome-associated transcriptional variation. Non-correlated DEGs had significantly higher F_{ST} values compared to both iASV-correlated DEGs and the genomic background (Fig. 5, Supplementary Table 15), indicating greater genetic differentiation at these loci and suggesting that mutations involved in genetic differentiation contribute to observed expression differences^{70,71}. In contrast, iASV-correlated DEGs displayed F_{ST} values comparable to genome-wide background levels (Fig. 5, Supplementary Table 15), indicating limited genetic differentiation at these loci. One possible interpretation is that genes associated with host–microbiome correlations at the population level are highly conserved. Taken together, these results suggest that host–microbiome covariation is largely decoupled from strong population-level sequence differentiation at correlated loci, and may instead reflect coordinated changes in gene regulation and prokaryotic community composition across environments, rather than divergence driven by fixed genetic innovation.

For chromosomal inversions, F_{ST} analyses comparing Chr4 and Chr16 karyotypes revealed a similar scenario, with a limited number of DEGs showing higher differentiation than the genomic background (Fig. 5, Supplementary Table 15). However, these estimates are based on 16 DEGs (4 of them correlated with iASVs), and statistical comparisons (t-test and KS-test) may be limited by reduced power, warranting cautious interpretation. In contrast, Chr2 and Chr11 exhibited a larger number of DEGs and iASV–DEG associations distributed across the genome. In Chr2, non-correlated DEGs showed higher differentiation than the genomic background but lower than iASV-correlated DEGs. A similar pattern was observed for Chr11, although involving a larger set of interacting DEGs and prokaryotic taxa (Fig. 5). In both cases, DEGs displayed larger F_{ST} values than the genomic background, suggesting that mutations linked to different karyotypes of the chromosome inversions may influence their expression levels but also the host-associated bacterial community^{21,22,70,71}. However, the relationship between sequence divergence, gene expression, and prokaryotic associations is not uniform across chromosomes, and may reflect heterogeneous genomic architectures rather than a single shared mechanism. Future work combining karyotype comparisons with chromatin architecture profiling will be necessary to disentangle the relative contributions of sequence variation and regulatory organisation to host–microbiome interactions.

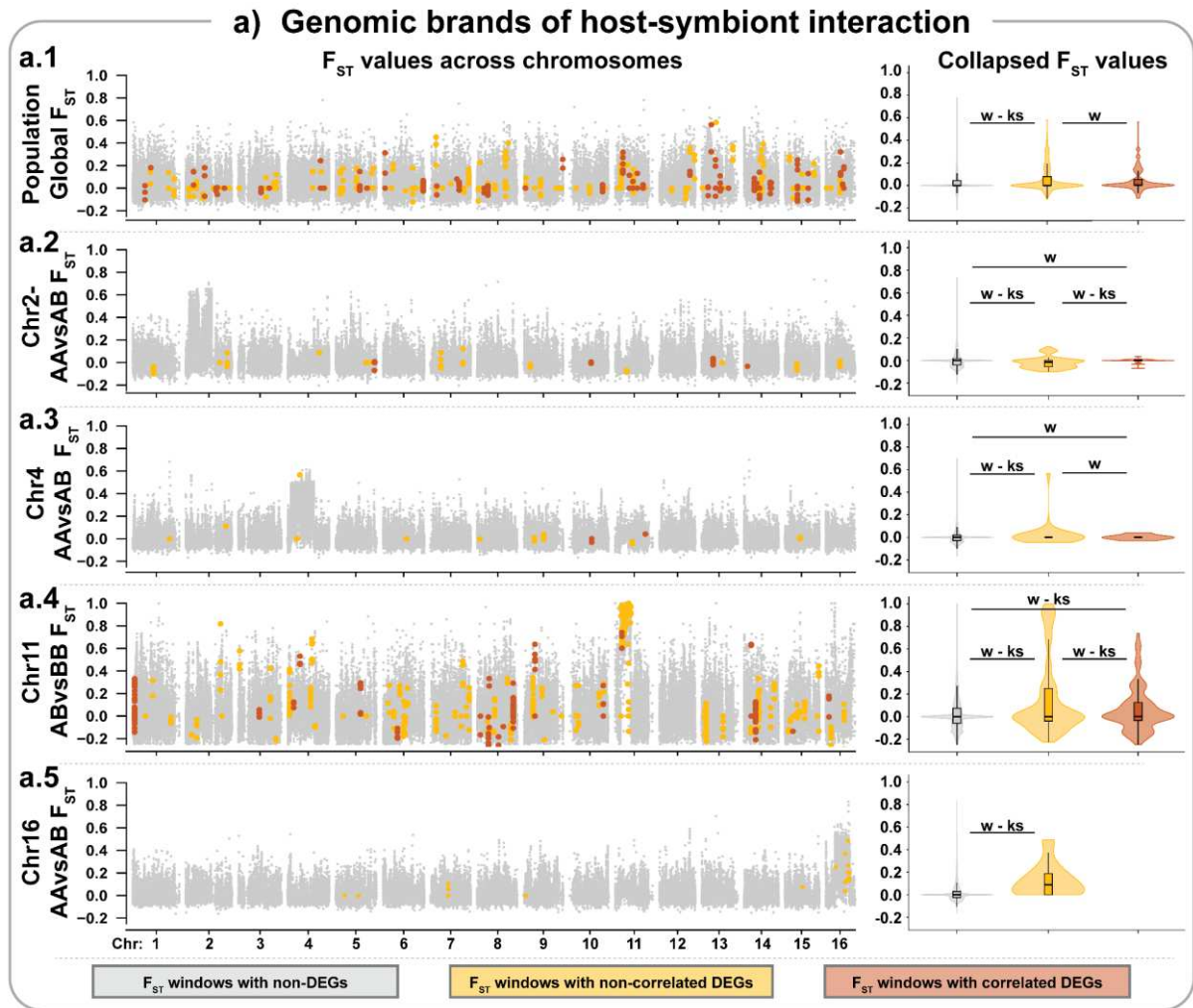


Figure 5 | Genome-wide F_{ST} values across chromosomes and their association with gene expression and indicator ASV (iASV) presence. The left panels show genome-wide F_{ST} values plotted across chromosomes for different comparisons: Pop (a.1), Chr2 (a.2), Chr4 (a.3), Chr11 (a.4), and Chr16 (a.5). Each dot represents a 10,000 bp window. Right panels depict boxplots of collapsed F_{ST} values for each group, with statistical significance ($p < 0.05$) indicated by Mann–Whitney U tests (w), Kolmogorov–Smirnov tests (ks), or both, as annotated. Boxplots within violins represent medians and interquartile ranges.

CONCLUSIONS

Together, our results provide a proof of concept for dissecting the genomic architecture of host–microbiome interactions in natural holobionts. By integrating genome structure, host transcription and microbiome composition, we uncover coordinated variation between host regulatory states and microbial communities across populations and chromosomal backgrounds. Although these associations do not establish causality, their distribution across the genome, including regions both within and outside chromosomal inversions, suggests that host–microbiome interactions are unlikely to be governed by isolated loci alone, but may instead emerge from distributed regulatory networks connecting multiple components of the host genome. The functional complementarity observed between host and microbial metabolic

repertoires further highlights how these interactions may contribute to the functional integration of the holobiont. More broadly, identifying genomic regions associated with microbiome composition provides a route to identifying candidate host genes, regulatory pathways and microbial partners involved in microbiome assembly and host–microbe communication, offering a framework for testing the genetic basis of microbial recruitment and maintenance. Such a framework could help elucidate how host genomes shape microbiome flexibility across environments, and how these interactions contribute to environmental adaptation, host–microbe coevolution, and the ecological capacity of holobionts to respond to environmental change. By bringing genome architecture, gene regulation and microbial ecology into a common evolutionary framework, our approach opens a path towards understanding adaptation not simply as a property of host genomes, but as an emergent feature of interacting host and microbial genomes, with broader relevance to the evolution and ecological dynamics of metazoan holobionts.

MATERIAL AND METHODS

Sample collection and nucleic acids extraction

Five adult individuals from each of the Spanish harbours of Blanes (41.674, 2.799), Barcelona (41.375, 2.182), and Vilanova i la Geltrú (41.215, 1.727) were collected in April 2022. Blanes is a small harbour (total linear quay length ~1500 m), Vilanova’s enclosure is medium sized (total linear quay length ~3000 m) and characterized by an elevated fishing activity, and the harbour of Barcelona is the largest one (total linear quay length > 20,000 m) with high anthropogenic impacts. In the field, each specimen was divided in two halves immediately after collection; one was preserved in 96% ethanol (with several changes afterwards) and kept at -20°C, while the other half was snap frozen in the field with liquid nitrogen and maintained at -80°C until extraction. For DNA-based approaches we used the half specimens in ethanol, from which we excised a gill fold using pincers. The DNA of each tissue fragment was extracted using the Gentra® Puregene® Tissue Kit (QIAGEN), following the manufacturer’s recommendations. For RNA-based approaches, we used the other half of the specimens, kept at -80°C. For each specimen, we waited until the surface of the individual was thawed and a gill fold was excised using pincers. The RNA of each tissue fragment was extracted using the protocol described in Ghangal et al., 2009⁷².

DNA and RNA extractions were sent to Novogene (UK) for library construction and sequencing. For DNA, a whole genome sequencing (WGS) approach was used, while for RNA,

cDNA transcriptome libraries were generated. Both for WGS and for RNAseq, libraries were prepared with an insert size of 350 bp. DNA and RNA libraries were sequenced following a 150 bp paired-end strategy on an Illumina NovaSeq 6000 platform (San Diego, CA), targeting 5 Gb output for each DNA sample aiming for an approximate coverage of 12X per individual¹⁶, and 10 Gb for each RNA sample.

Host DNA Sequence analyses

After sequencing, we removed adaptors and low-quality genomic data with Trimmomatic v.0.39⁷³, using the parameters LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15, MINLEN:36 (Table S1). The data quality was verified with FastQC v.0.11.9⁷⁴. *Styela plicata* has 16 chromosomes, with four large chromosomal inversions (~19.5 Mb inverted region in chromosome 2, ~17 Mb in chromosome 4, ~10.5 Mb in chromosome 11, and ~14 Mb in chromosome 16)¹⁶. To identify the inversions' karyotype in each of our samples, we used as reference three previously published samples (ENA Accession codes: BRA10=SAMEA114503380, MIS22=SAMEA114503395, PE7=SAMEA114503401)¹⁶ having all possible karyotypic combinations for each of the four inversions (homozygous for the reference ordination, heterozygous, homozygous for the alternative ordination) We mapped the clean reads of the 18 samples against the reference genome of *Styela plicata* (Accession code: GCA_963675895.1)¹⁶ with the software bwa v.0.7.12⁷⁵ using the BWA-MEM algorithm⁷⁶. We used the command “mpileup” from the software bcftools v.1.13-60⁷⁷ to call all individuals' SNPs. We used the software vcftools v.0.1.17⁷⁸ to keep only those SNPs which were biallelic (--min-alleles 2 --max-alleles 2), had a minimum depth of 5 reads (--minDP 5), a mean maximum depth of 24 (--max-meanDP 24), a minimum allele frequency of 3% (--maf 0.03), and were present in all samples (--max-missing 1). The mean maximum depth threshold was calculated as 1.5 times the interquartile range of depth across all SNPs. We generated two datasets, one including all 18 samples for chromosome inversion assignment, and another one to infer regional population structure excluding the non-Mediterranean samples facilitating chromosomal inversion assignment (BRA10, MIS22, and PE7). For the file including all samples we generated four different datasets, each of them including a single chromosome hosting an inversion using the function “--chr” from vcftools (Table S2). For the other file, we removed non-Mediterranean individuals using the function “--remove-indv”, and we kept only those chromosomes without inversions using the function “--chr”.

To define each individual chromosome inversion karyotype, we conducted a Discriminant Analysis of Principal Components (DAPC) on each inversion dataset with the R package adegenet v.2.1.10 using the command “dapc”⁷⁹. We defined three potential groups as priors (prior=3), and we only retained one axis for both discriminant and principal components analyses (n.da=1, npca=1). The posterior probability of each individual being assigned to a given genomic group was represented with the R package ggplot2 v.3.4.4⁸⁰. Additionally, we transformed all datasets to plink using the function “--plink” from vcftools, generated Multi-Dimensional Scalings (MDS) of the genomic data using the function “--mds-plot” from plink v.1.9.⁸¹, and visually represented the results with ggplot2.

RNA Sequence analyses

We obtained ~40.92 million raw transcriptome reads per sample (Table S1), which were reduced to ~40.41 million reads after filtering low-quality data with Trimmomatic using the same parameters as for genomic data. We mapped the reads of each individual to the reference genome with the software hisat2 v.2.2.1, intended for transcriptome data⁸². We retrieved transcriptomic genotypes for all samples with bcftools, filtered them with vcftools, and carried out MDS analyses to evaluate differences between the transcriptomic and the genomic information.

We obtained an RNA read count table for each sample using the function “featureCounts” implemented in the R package Rsubread v.2.8.2⁸³. We normalized the counts table with the function “data.Normalization” from clusterSim v.0.51-3⁸⁴, and conducted a PCA to check any differences at the population level on the overall transcriptome using the function “prcomp” implemented in base R. We plotted the ordination with ggplot2. We analyzed differential gene expression with the R package DESeq v.1.42.0⁸⁵. We evaluated differential expression between populations (Pop), and between the most abundant chromosomal karyotypes for each of the 4 chromosomes with inversions (Chr2, Chr4, Chr11, and Chr16). To improve statistical robustness, we restricted the analyses to karyotypes with more than four specimens to ensure stable dispersion estimates and reduce inflation of false positives in low-replicate groups³⁶. Thus, we analysed the presence of differentially expressed genes (DEGs) across Pop (Blanes vs. Vilanova vs. Barcelona), Chr2 (AA vs. AB), Chr4 (AA vs. AB), Chr11 (AB vs. BB) and Chr16 (AA vs. AB). To keep only robust DEGs, we retained those with a q-value below 0.05 and at least two reads per sample. We represented DEGs across comparisons through Venn diagrams using the R package ggVennDiagram v.1.5.2⁸⁶. For further analyses we removed

those genes to be identified as differentially expressed in two or more comparisons (Fig. 2) to avoid confounding effects. We conducted PCAs with the normalized transcript abundance matrices of the exclusive DEGs in each comparison, and represented them with ggplot2. We also represented transcription abundance values as heatmaps with the function “pheatmap” from the same R package v1.0.12 ⁸⁷.

We retrieved the nucleotide sequences from all DEGs exclusive from a single comparison using the function “getfasta” from the software BEDTools v.2.27.1 ⁸⁸, and functionally characterized them using eggNOG-mapper v.2 ⁸⁹ for Gene Ontology (GO) Terms and Kyoto Encyclopedia of Genes and Genomes Ontology (KO) Terms, and INTERPROSCAN ⁹⁰ for Protein families (Pfam). We visually represented the GO-Terms of DEGs across all analyses using the R package treemap v.2.4-4 ⁹¹. We classified gene families as identified by Pfam based on their most likely involvement in immune system processes from the literature.

Microbiome analyses

The microbiome obtained from the gills of the same individuals was extracted from the amplicon sequence variant (ASV) abundance table in Galià-Camps et al. ²³. In short, that table was obtained by removing paired-end reads chimeras with UCHIME ⁹² using as a reference the SILVA database ⁹³, and a denoising step using DADA2 ⁹⁴ in QIIME2 software ⁹⁵. Singleton sequences were discarded and the remaining ASVs were taxonomically assigned using the classify-sklearn plugin of QIIME2 against the SILVA database.

Based on the ASV abundance matrix, we calculated the Shannon diversity index for each sample with the function “diversity” from the package vegan v.2.5-6 ⁹⁶. To calculate the species richness, we rarefied all samples’ counts to the sample with the lowest number of reads using the function “rarefy” from vegan. Shannon index and species richness values were graphically visualized as boxplots with ggplot2. We inferred normality of both Shannon and Richness values using a Shapiro-Wilk test with the function “shapiro.test” implemented in base R and subsequently, rank-transformed them using the function “rank” of base R. We generated five independent General Linear Models (GLM) for both Shannon and Richness using the function “glm” of R stats: one testing for differences among populations, and four additional ones testing for differences among each chromosome’s inversion karyotypes. For significant GLMs with more than three groups being compared, we conducted pairwise comparisons using the function “contrast” of the R package emmeans v.1.8.7 ⁹⁷.

Next, we transformed the ASV abundance table to relative frequencies and calculated Bray-Curtis distances among samples using the function “vegdist” from vegan. We obtained a non-metric MDS with the function “metaMDS” from vegan, and graphically represented the results using ggplot2. To test significant differences among populations, and among each chromosomal inversion karyotype, we carried out PERMANOVA tests with 10,000 permutations using the function “adonis2” from vegan.

We inferred the presence of indicator ASVs (iASVs) of populations and chromosomal karyotypes. To do so, we used the “multipatt” function of the R package indicpecies v.1.7.14.⁹⁸, with 999 permutations to test statistical significance. We kept only those iASVs for which at least 75% of the reads were present in at least 75% of samples of the group (population or karyotype). We identified iASV across analyses, represented them through Venn diagrams using the R package ggVennDiagram, and removed shared iASVs for subsequent analyses.

We represented the taxonomic assignment of iASVs at the class level, as lower taxonomic ranks were unavailable for most iASVs. For each taxon, we classified the taxa into four categories: certain interaction, likely interaction, unlikely interaction, or no interaction based on literature searches for host association. Taxa were classified as showing a certain interaction when the iASV was a well-described symbiont or parasite of marine ascidians, corals, or sponges. Likely interactions were assigned to marine taxa that are frequently reported as core members of ascidian, coral, and/or sponge microbiomes. Unlikely interactions corresponded to taxa of marine origin but which are rarely reported in host-associated microbiomes, such as pelagic groups of *Synechococcus sp.*. Taxa classified as showing no interaction were generally associated with terrestrial environments (see Table S9 for full details).

To obtain representative genomes of the iASV, we aligned the 16S rRNA amplicon sequences of the iASVs to genomes in the Genome Taxonomy Database GTDB r226⁹⁹ using vsearch v.2.21.1¹⁰⁰. **We only retained matches with >99% sequence identity and genomes with completeness $\geq$ 90% and <5% contamination based on CheckM v.1.1.6¹⁰¹.** We then matched the genome accessions to those in globDB¹⁰² to obtain annotations of the genes involved in interactions with the host. We used precomputed KEGG ortholog (KO) annotations from the KOfam database¹⁰³ to obtain gene information from both metabolic pathways and specific functional modules. We classified genes based on their most likely involvement in pathogenesis, symbiosis, or both from the literature^{50–52,104–107}.

596 *Host-microbiome interactions*

597 Based on the DEGs and iASVs analyses of the five datasets evaluating differences between
598 Pop, Chr2, Chr4, Chr11, and Chr16, we generated count tables of the differentially expressed
599 genes and abundance tables of the iASVs for each comparison. The matrices only included
600 unique DEGs and iASVs. We normalized the matrices using the “data.Normalization” function,
601 and calculated pairwise Pearson’s correlations (r) and coefficients of determination (R^2)
602 between DEGs transcription levels and iASV abundance for each specimen. We established a
603 threshold of $R^2 > 0.5$ to identify correlations of substantial magnitude, and we visually
604 represented them with the R package pheatmap.

605 We then evaluated the genomic toolkit based on Gene Ontology (GO) terms of the DEGs of the
606 host involved in host-iASV interaction. To do so, we used topGO v.2.54¹⁰⁸ to evaluate
607 enrichment on the genes being correlated with iASV. We compared using t-tests a list of
608 correlated genes and all the genes within *S. plicata* using the function “runTest”, the algorithm
609 “classic”. We repeated so for “biological processes” (BP) and “molecular function” (MF),
610 independently. We retained all enriched GO-terms with a significant p-value below 0.05 and at
611 least three genes with that GO-term. To further address the function of correlated DEGs, we
612 retrieved the Pfam and KO terms previously recovered with INTERPROSCAN and eggNOG-
613 mapper, respectively, and manually evaluated their functions. Similarly, for the correlated
614 iASV toolkit, we represented their potential to create associations with the host and their KO
615 annotations as above mentioned for the iASVs.

616 Finally, to evaluate whether the correlated DEGs and the iASV functions interact with each
617 other, we created a list of KO associated to the DEGs of the host, and a second one with all the
618 KO annotations recovered from the genomes of correlated iASVs. We used the Reconstruct
619 option of KEGG Mapper v.5.2¹⁰⁹, implemented in the KEGG webpage
620 (<https://www.genome.jp/kegg/mapper/>), to evaluate whether any of the almost complete
621 metabolic modules recovered (only one KEGG missing) were composed by both the host DEGs
622 and the symbiont iASVs. We transformed the resulting table into an edge map and graphically
623 represented them as network structures in R.

624 **F_{ST} analysis and modulation of gene expression and the microbiome**

625 To evaluate whether differences of gene expression are associated with genetic divergence, we
626 calculated F_{ST} values in 10,000 bp windows with a 2500 bp sliding step across the genome

among populations for chromosomes without inversions and between the most abundant karyotypes for each chromosome with inversions. We represented the F_{ST} window values as Manhattan plots with the R package CMplot v.4.5.1¹¹⁰, identifying windows with non-DEGs, windows hosting non-correlated DEGs, and windows with correlated DEGs. We represented the F_{ST} distribution of these three genomic fractions using violin plots in ggplot2. We finally carried out Mann–Whitney U tests and KS tests to evaluate significant differences in the means and distributions of F_{ST} values among windows with non-DEGs, non-correlated DEGs, and correlated DEGs for the population and each of the four chromosomes with inversions.

AUTHOR'S CONTRIBUTION

Carles Galià-Camps: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Josep Ramoneda:** Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Marta Pascual:** Conceptualization, Methodology, Validation, Resources, Writing - Review & Editing, Visualization, Supervision, Project administration, Funding acquisition. **Carlos Carreras:** Conceptualization, Resources, Writing - Review & Editing, Supervision, Project administration, Funding acquisition. **Xavier Turon:** Conceptualization, Resources, Writing - Review & Editing, Visualization, Supervision, Project administration, Funding acquisition. **Celia Schunter:** Conceptualization, Methodology, Writing - Review & Editing, Supervision. **Cinta Pegueroles:** Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing, Supervision.

DATA AVAILABILITY STATEMENT

All new data generated in the present study has been deposited to the European Nucleotide Archive (ENA) under Bioproject XXXX (Uploaded upon acceptance). All scripts have been deposited in the corresponding author GitHub webpage (<https://github.com/CGaliaCamps>) to enable the study's replicability.

FUNDING

This research was funded by the projects BlueDNA (PID2023-146307OB), funded by MCIN/AEI/10.13039/501100011033 and ERDF/EU from the Spanish Government and MARGECH (PID2020-118550RB), funded by the Spanish Ministry of Science and Innovation MICIU/AEI/10.13039/501100011033]. The authors CGC, MP, CP, and CC are members of the

research group SGR2021-01271, and XT is member of the research group SGR2021-00405, both funded by the Generalitat de Catalunya (AGAUR). CGC held a predoctoral contract [PRE-2018-085227-MCIN/AEI/10.13039/501100011033] by the Spanish Ministry of Science, Innovation and Universities, and by ERDF “A way of making Europe” at the beginning of the study, and currently holds a Juan de la Cierva fellowship (JDC 2024-055057-I) supported by the Spanish Ministry of Science, Innovation, and Universities. JR was funded by the Spanish Ministry for Science, Innovation and Universities/Agencia Estatal de Investigación (MicroDroughtPredict PID2023-151209NA-I00, SALT-MEMO PID2025-168843NB-I00, and Ramón y Cajal Fellowship RYC2024-049086-I), and CSIC (Proyectos Intramurales Especiales, PREDEX 20263AT004).

ACKNOWLEDGEMENTS

We thank the harbor authorities of Blanes, and Vilanova, as well to the Reial Club Nàutic de Barcelona for providing access to their facilities for sampling. We also want to thank Jagoba Malumbre Olarte for illustrating *Styela plicata* along with its microbiome and DNA used in Figure 1a. Also, to Gisela Marin Capuz to compile all the objectives and create the visual scheme represented as Figure 1a.

REFERENCES

- Günther, T. & Coop, G. Robust identification of local adaptation from allele frequencies. *Genetics* **195**, 205–220 (2013).
- Wu, X. *et al.* Rapid adaptation and extinction in synchronized outdoor evolution experiments of. *Science* **391**, eadz0777 (2026).
- King, D. G. Tandem repeat polymorphisms shape local adaptation. *Trends Ecol Evol* **40**, 331–334 (2025).
- Steward, R. A., de Jong, M. A., Oostra, V. & Wheat, C. W. Alternative splicing in seasonal plasticity and the potential for adaptation to environmental change. *Nat Commun* **13**, 755 (2022).
- Li, Y. *et al.* Chromatin Topology Reconfiguration Orchestrates Thermotolerant Male Fertility via GhAL5 in Cotton. *Adv Sci (Weinh)* **13**, e07766 (2026).
- Bordenstein, S. R. & Theis, K. R. Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes. *PLoS Biol* **13**, e1002226 (2015).
- Rosenberg, E. & Zilber-Rosenberg, I. Symbiosis and development: the hologenome concept. *Birth Defects Res. C Embryo Today* **93**, 56–66 (2011).

8. Petersen, C. *et al.* Host and microbiome jointly contribute to environmental adaptation. *ISME J* **17**, 1953–1965 (2023).
9. Sommer, F. & Bäckhed, F. The gut microbiota--masters of host development and physiology. *Nat Rev Microbiol* **11**, 227–238 (2013).
10. Kunst, C. *et al.* The Influence of Gut Microbiota on Oxidative Stress and the Immune System. *Biomedicines* **11**, (2023).
11. Rea, K., Dinan, T. G. & Cryan, J. F. The microbiome: A key regulator of stress and neuroinflammation. *Neurobiol Stress* **4**, 23–33 (2016).
12. Galià-Camps, C. *et al.* Navigating spatio-temporal microbiome dynamics: Environmental factors and trace elements shape the symbiont community of an invasive marine species. *Mar. Pollut. Bull.* **203**, 116477 (2024).
13. Rosenberg, E. Transmission of hologenomes between generations: Mothers matter most. in *Microbiomes* 161–194 (Springer International Publishing, Cham, 2021).
14. Rosenberg, E. & Zilber-Rosenberg, I. The hologenome concept of evolution after 10 years. *Microbiome* **6**, 78 (2018).
15. Roughgarden, J., Gilbert, S. F., Rosenberg, E., Zilber-Rosenberg, I. & Lloyd, E. A. Holobionts as units of selection and a model of their population dynamics and evolution. *Biol. Theory* **13**, 44–65 (2018).
16. Galià-Camps, C. *et al.* De novo genome assembly, inversion detection, and worldwide adaptation on the invasive species *Styela plicata*. *Sci Rep* **15**, 40328 (2025).
17. Schaal, S. M., Haller, B. C. & Lotterhos, K. E. Inversion invasions: when the genetic basis of local adaptation is concentrated within inversions in the face of gene flow. *Philos Trans R Soc Lond B Biol Sci* **377**, 20210200 (2022).
18. Berdan, E. L. *et al.* How chromosomal inversions reorient the evolutionary process. *J. Evol. Biol.* **36**, 1761–1782 (2023).
19. Durmaz Mitchell, E., Kerdaffrec, E., Schmidt, P., Flatt, T. & Kittelmann, S. A Balanced Inversion Polymorphism Exhibits a Dominance Reversal at the Gene Expression Level that Depends on Developmental Context. *Mol Biol Evol* **42**, (2025).
20. Turon, M. *et al.* Microbiome changes through the ontogeny of the marine sponge *Crambe crambe*. *Environ Microbiome* **19**, 15 (2024).

21. Casso, M., Turon, M., Marco, N., Pascual, M. & Turon, X. The Microbiome of the Worldwide Invasive Ascidian *Didemnum vexillum*. *Front. Mar. Sci.* **7**, (2020).
22. Díez-Vives, C. *et al.* On the way to specificity - Microbiome reflects sponge genetic cluster primarily in highly structured populations. *Mol. Ecol.* **29**, 4412–4427 (2020).
23. Galià-Camps, C., Baños, E., Pascual, M., Carreras, C. & Turon, X. Multidimensional variability of the microbiome of an invasive ascidian species. *iScience* **26**, 107812 (2023).
24. Goddard-Dwyer, M., López-Legentil, S. & Erwin, P. M. Microbiome Variability across the Native and Invasive Ranges of the Ascidian *Clavelina oblonga*. *Appl. Environ. Microbiol.* **87**, (2021).
25. Nydam, M. L., Stefaniak, L. M., Lambert, G., Counts, B. & López-Legentil, S. Dynamics of ascidian-invaded communities over time. *Biol. Invasions* **24**, 3489–3507 (2023).
26. Hutchings, B., Stiles, E., Erwin, P. M. & López-Legentil, S. Hurricane events facilitate the dominance of introduced invertebrate species in harbors. *Biol. Invasions* **25**, 2495–2506 (2024).
27. Chen, Y., Lin, Y. & Zhan, A. Multi-omics insights into environmental adaptation and invasion potential of the ascidian *Molgula manhattensis*. *Oikos* <https://doi.org/10.1002/oik.12111> (2026) doi:10.1002/oik.12111.
28. Lin, Y. *et al.* Genetic signatures of natural selection in a model invasive ascidian. *Sci Rep* **7**, 44080 (2017).
29. Galià-Camps, C., Enguídanos, A., Turon, X., Pascual, M. & Carreras, C. The past, the recent, and the ongoing evolutionary processes of the worldwide invasive ascidian *Styela plicata*. *Mol. Ecol.* e17502 (2024).
30. Rychel, A. L., Smith, S. E., Shimamoto, H. T. & Swalla, B. J. Evolution and development of the chordates: collagen and pharyngeal cartilage. *Mol Biol Evol* **23**, 541–549 (2006).
31. Franchi, N. & Ballarin, L. Immunity in Protochordates: The Tunicate Perspective. *Front Immunol* **8**, 674 (2017).
32. Gordon, A. J. E., Satory, D., Halliday, J. A. & Herman, C. Lost in transcription: transient errors in information transfer. *Curr Opin Microbiol* **24**, 80–87 (2015).
33. Zawisza-Álvarez, M. *et al.* Exploring functional conservation in silico: a new machine learning approach to RNA-editing. *Brief Bioinform* **25**, (2024).
34. Hamar, R. & Varga, M. The role of post-transcriptional modifications during development. *Biol Futur* **74**, 45–59 (2023).
35. Son, K., Yu, S., Shin, W., Han, K. & Kang, K. A Simple Guideline to Assess the Characteristics of RNA-Seq Data. *Biomed Res Int* **2018**, 2906292 (2018).

36. Halasz, G. *et al.* Optimized murine sample sizes for RNA sequencing studies revealed from large scale comparative analysis. *Nat Commun* **16**, 10173 (2025).
37. Zarcero, J. *et al.* Spatio-temporal metabarcoding surveys in ports reveal homogenised communities of non-indigenous species with high genetic diversity and connectivity. *Sci Rep* **16**, (2026).
38. Wu, H., Eckhardt, C. M. & Baccarelli, A. A. Molecular mechanisms of environmental exposures and human disease. *Nat Rev Genet* **24**, 332–344 (2023).
39. Vad, J., Duran Suja, L., Summers, S., Henry, T. B. & Roberts, J. M. Marine Sponges in a Snowstorm - Extreme Sensitivity of a Sponge Holobiont to Marine Oil Snow and Chemically Dispersed Oil Pollution. *Front Microbiol* **13**, 909853 (2022).
40. Baratti, M. *et al.* Differential gene expression and chemical patterns of an intertidal crab inhabiting a polluted port and an adjacent marine protected area. *Sci Total Environ* **822**, 153463 (2022).
41. Ghavi-Helm, Y. Functional Consequences of Chromosomal Rearrangements on Gene Expression: Not So Deleterious After All? *J Mol Biol* **432**, 665–675 (2020).
42. Wang, Y. *et al.* Chromosome territory reorganization through artificial chromosome fusion is compatible with cell fate determination and mouse development. *Cell Discov* **9**, 11 (2023).
43. Han, G. & Vaishnava, S. Microbial underdogs: exploring the significance of low-abundance commensals in host-microbe interactions. *Exp Mol Med* **55**, 2498–2507 (2023).
44. Carrier, T. J., Melzner, F., Jung, S. & Hentschel, U. Temperature-induced shift in the rare microbiota of the sponge *Haliclona*. *bioRxiv* (2026) doi:10.64898/2026.04.22.720130.
45. Tian, R.-M. *et al.* The deep-sea glass sponge *Lophophysema eversa* harbours potential symbionts responsible for the nutrient conversions of carbon, nitrogen and sulfur. *Environ Microbiol* **18**, 2481–2494 (2016).
46. Engelberts, J. P. *et al.* Characterization of a sponge microbiome using an integrative genome-centric approach. *ISME J* **14**, 1100–1110 (2020).
47. Schellenberg, J. *et al.* Proposal of *Litorimonas haliclona* sp. nov., isolated from a marine sponge of the genus *Haliclona*. *Int J Syst Evol Microbiol* **68**, 835–843 (2018).
48. Sogin, M. L. *et al.* Microbial diversity in the deep sea and the underexplored ‘rare biosphere’. *Proc Natl Acad Sci U S A* **103**, 12115–12120 (2006).
49. Jousset, A. *et al.* Where less may be more: how the rare biosphere pulls ecosystems strings. *ISME J* **11**, 853–862 (2017).

50. Alex, A. & Antunes, A. Genus-wide comparison of *Pseudovibrio* bacterial genomes reveal diverse adaptations to different marine invertebrate hosts. *PLoS One* **13**, e0194368 (2018).
51. Kwan, J. C. *et al.* Host control of symbiont natural product chemistry in cryptic populations of the tunicate *Lissoclinum patella*. *PLoS One* **9**, e95850 (2014).
52. Li, J. *et al.* Cultured Bacteria Provide Insight into the Functional Potential of the Coral-Associated Microbiome. *mSystems* **7**, e0032722 (2022).
53. da Silva, D. M. G., Costa, R. & Keller-Costa, T. A genomic view of the bacterial family Endozoicomonadaceae in marine symbioses. *Commun Biol* **8**, 1418 (2025).
54. Schreiber, L. *et al.* Endozoicomonas Are Specific, Facultative Symbionts of Sea Squirts. *Front Microbiol* **7**, 1042 (2016).
55. Romanenko, L. A., Tanaka, N. & Frolova, G. M. *Umboniibacter marinipuniceus* gen. nov., sp. nov., a marine gammaproteobacterium isolated from the mollusc *Umbonium costatum* from the Sea of Japan. *Int J Syst Evol Microbiol* **60**, 603–609 (2010).
56. McIlroy, S. J. *et al.* ‘*Candidatus Competibacter*’-lineage genomes retrieved from metagenomes reveal functional metabolic diversity. *ISME J* **8**, 613–624 (2014).
57. Attia, M. A. *et al.* In vitro and in vivo characterization of three glycoside hydrolase family 5 members reveals potent xyloglucan backbone-cleaving functions. *Biotechnol Biofuels* **11**, 45 (2018).
58. Nichols, R. G. & Davenport, E. R. The relationship between the gut microbiome and host gene expression: a review. *Hum Genet* **140**, 747–760 (2021).
59. Li, N. *et al.* Large-scale multi-omics unveils host-microbiome interactions driving root development and nitrogen acquisition. *Nat Plants* **12**, 319–336 (2026).
60. Rosales, S. M. *et al.* Microbiome differences in disease-resistant vs. susceptible *Acropora* corals subjected to disease challenge assays. *Sci Rep* **9**, 18279 (2019).
61. Gignoux-Wolfsohn, S. A., Aronson, F. M. & Vollmer, S. V. Complex interactions between potentially pathogenic, opportunistic, and resident bacteria emerge during infection on a reef-building coral. *FEMS Microbiol Ecol* **93**, (2017).
62. Selwyn, J. D., Despard, B. A., Vollmer, M. V., Trytten, E. C. & Vollmer, S. V. Identification of putative coral pathogens in endangered Caribbean staghorn coral using machine learning. *Environ Microbiol* **26**, e16700 (2024).

63. Erwin, P. M., Pineda, M. C., Webster, N., Turon, X. & López-Legentil, S. Down under the tunic: bacterial biodiversity hotspots and widespread ammonia-oxidizing archaea in coral reef ascidians. *ISME J.* **8**, 575–588 (2014).
64. Martínez-García, M. *et al.* Ammonia-oxidizing Crenarchaeota and nitrification inside the tissue of a colonial ascidian. *Environ Microbiol* **10**, 2991–3001 (2008).
65. Evans, J. S., Erwin, P. M., Shenkar, N. & López-Legentil, S. Introduced ascidians harbor highly diverse and host-specific symbiotic microbial assemblages. *Sci Rep* **7**, 11033 (2017).
66. Rowley, A. F., Knight, J., Lloyd-Evans, P., Holland, J. W. & Vickers, P. J. Eicosanoids and their role in immune modulation in fish—a brief overview. *Fish Shellfish Immunol.* **5**, 549–567 (1995).
67. MacKnight, N. J. *et al.* Disease resistance in coral is mediated by distinct adaptive and plastic gene expression profiles. *Sci Adv* **8**, eabo6153 (2022).
68. Kim, Y., Ahmed, S., Stanley, D. & An, C. Eicosanoid-mediated immunity in insects. *Dev Comp Immunol* **83**, 130–143 (2018).
69. Rusanova, A. *et al.* Taxonomically different symbiotic communities of sympatric Arctic sponge species show functional similarity with specialization at species level. *mSystems* **10**, e0114725 (2025).
70. Snyman, M. & Xu, S. The effects of mutations on gene expression and alternative splicing. *Proc Biol Sci* **290**, 20230565 (2023).
71. Gerstung, M. *et al.* Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. *Nat Commun* **6**, 5901 (2015).
72. Ghangal, R., Raghuvanshi, S. & Chand Sharma, P. Isolation of good quality RNA from a medicinal plant seabuckthorn, rich in secondary metabolites. *Plant Physiol Biochem* **47**, 1113–1115 (2009).
73. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
74. Andrews, S. FastQC: a quality control tool for high throughput sequence data. Preprint at <https://github.com/s-andrews/FastQC> (2010).
75. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
76. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. <https://doi.org/10.48550/ARXIV.1303.3997> (2013) doi:10.48550/ARXIV.1303.3997.
77. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).

78. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
79. Jombart, T. & Ahmed, I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071 (2011).
80. Wickham, H., Chang, W. & Wickham, M. H. Package ‘ggplot2’. *Create elegant data visualisations using the grammar of graphics. Version 2*, 1–189 (2016).
81. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
82. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
83. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
84. Walesiak, M. & Dudek, A. ClusterSim: Searching for optimal clustering procedure for a data set. *CRAN: Contributed Packages* The R Foundation <https://doi.org/10.32614/cran.package.clustersim> (2007).
85. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
86. Gao, C.-H., Yu, G. & Cai, P. ggVennDiagram: An Intuitive, Easy-to-Use, and Highly Customizable R Package to Generate Venn Diagram. *Front. Genet.* **12**, 706907 (2021).
87. Kolde, R. pheatmap: Pretty Heatmaps. *CRAN: Contributed Packages* The R Foundation <https://doi.org/10.32614/cran.package.pheatmap> (2010).
88. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
89. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol* **38**, 5825–5829 (2021).
90. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
91. Tennekens, M. treemap: Treemap Visualization. *CRAN: Contributed Packages* The R Foundation <https://doi.org/10.32614/cran.package.treemap> (2010).
92. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200 (2011).

93. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–6 (2013).
94. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
95. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
96. Oksanen, J. *et al.* Package ‘vegan’. *Community ecology package, version 2*, 1–295 (2013).
97. Lenth, R., Singmann, H., Love, J., Buerkner, P. & Herve, M. emmeans: estimated marginal means. R package version 1.4. 4. *Am. Stat.* (2020).
98. De Cáceres, M., Sol, D., Lapiedra, O. & Legendre, P. A framework for estimating niche metrics using the resemblance between qualitative resources. *Oikos* **120**, 1341–1350 (2011).
99. Parks, D. H. *et al.* GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* **50**, D785–D794 (2022).
100. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
101. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043–1055 (2015).
102. Speth, D. R. *et al.* GlobDB: a comprehensive species-dereplicated microbial genome resource. *Bioinform Adv* **5**, vbaf280 (2025).
103. Aramaki, T. *et al.* KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).
104. Fiore, C. L., Jarett, J. K., Steinert, G. & Lesser, M. P. Trait-based comparison of coral and sponge microbiomes. *Sci. Rep.* **10**, 2340 (2020).
105. Waterworth, S. C., Parker-Nance, S., Kwan, J. C. & Dorrington, R. A. Comparative genomics provides insight into the function of broad-host range sponge symbionts. *MBio* **12**, e0157721 (2021).
106. Matos, A. *et al.* Shotgun proteomics of ascidians tunic gives new insights on host-microbe interactions by revealing diverse Antimicrobial Peptides. *Mar. Drugs* **18**, E362 (2020).

895 107. Gavrilidou, A. *et al.* Candidatus Nemesobacterales is a sponge-specific clade of the candidate
896 phylum Desulfobacterota adapted to a symbiotic lifestyle. *ISME J.* **17**, 1808–1818 (2023).
897 108. Adrian Alexa, J. R. *topGO*. (Bioconductor, 2017). doi:10.18129/B9.BIOC.TOPGO.
898 109. Kanehisa, M., Sato, Y. & Kawashima, M. KEGG mapping tools for uncovering hidden features
899 in biological data. *Protein Sci.* **31**, 47–53 (2022).
900 110. LiLin-Yin. CMplot: Circle Manhattan Plot. *CRAN: Contributed Packages* The R Foundation
901 <https://doi.org/10.32614/cran.package.cmplot> (2015).

Supplementary Figures Captions

Figure S1 | a) Identification of inversions using MDS analyses for Chr2 (a.1), Chr4 (a.2), Chr11 (a.3), and Chr16 (a.4). Each cluster in the first axis corresponds to a karyotype arrangement. The samples are colour coded by locality and the reference samples with known karyotype are coded in greyscale.

Figure S2 | a) Genetic differentiation based on SNP genotyping from RNAseq data for the 15 *S. plicata* samples. Multidimensional scaling (MDS) of individuals with SNPs from chromosomes without inversions (a.1), Chr2 (a.2), Chr4 (a.3), Chr11 (a.4), and Chr16 (a.5).

Figure S3 | a) Transcriptome expression for *S. plicata* samples using all RNAseq data.

Figure S4 | a) Differential expression patterns of *S. plicata* represented as vectorised PCAs for the Population model (a.1), the Chr2 model (a.2), the Chr4 model (a.3), the Chr11 model (a.4), and the Chr16 model (a.5). Note that each arrow corresponds to the weight of a differentially expressed gene on the PCA.

Figure S5 | a) Alpha diversity across populations and karyotypes. Boxplots show ASV richness (left column) and Shannon's diversity index (right column) for samples grouped by Population (a.1) and karyotype for Chr2 (a.2), Chr4 (a.3), Chr11 (a.4), and Chr16 (a.5). Boxes indicate interquartile ranges with medians. Asterisks in the top-left corner of each panel indicate significant differences among represented groups.

Figure S6 | a) Host-associated functions of iASV bacteria. a.1) KEGG repertoire of iASVs whose genome could be recovered through GTDB. KEGG names in red correspond to symbiont lifestyle functions, KEGG names in blue correspond to parasitic and or pathogenic functions, and KEGG names in green correspond to host-dependent (symbiont and pathogen) lifestyle functions. **a.2)** Principal components analysis (PCA) based on the presence of KEGGs involved in microbial interactions with the host, representing the functional gene composition of iASVs with a representative GTDB genome.

Figure S7 | a) Host-associated functions of the iASV correlated with DEGs. a.1) KEGG repertoire of iASVs whose genome could be recovered through GTDB. KEGG names in red correspond to symbiont lifestyle functions, KEGG names in blue correspond to parasitic and or pathogenic functions, and KEGG names in green correspond to host-dependent (symbiont and pathogen) lifestyle functions. **a.2)** Principal components analysis (PCA) based on the presence of KEGGs involved in microbial interactions with the host, representing the functional gene composition of representative genomes of correlated iASVs.

930

Supplementary Table Captions

931 **Table S1 | Sample metadata and sequencing statistics for genomic and transcriptomic datasets.** Rows
932 correspond to individual *Styela plicata* specimens, with columns reporting population, sample identifiers, inversion
933 karyotypes and sequencing information for the DNA and RNA datasets.

934 **Table S2 | Summary of SNP datasets used for population structure and chromosomal inversion analyses.**
935 Rows correspond to the different SNP datasets analysed, with columns summarizing the sequencing source,
936 number of samples and SNPs retained, and the population or chromosomal inversion comparison for which each
937 dataset was used.

938 **Table S3 | Chromosomal inversion karyotypes across samples.** Rows correspond to individual specimens, with
939 columns reporting the inferred karyotype for each of the four chromosomal inversions (Chr2, Chr4, Chr11 and
940 Chr16), together with the population assignment where applicable.

941 **Table S4 | Differentially expressed genes and functional annotations across populations and chromosomal
942 inversion karyotypes.** Rows correspond to differentially expressed genes, with columns reporting the comparison
943 in which each gene was identified, expression statistics and functional annotations, including Gene Ontology and
944 other available annotations.

945 **Table S5 | ASV abundance and taxonomic composition across samples.** Rows correspond to amplicon
946 sequence variants (ASVs), with columns reporting their read abundance across individual samples and their
947 taxonomic assignments.

948 **Table S6 | Alpha-diversity indices and comparison diagnostics across populations and chromosomal
949 inversion karyotypes.** Rows correspond to the fitted models for richness and Shannon diversity, with columns
950 reporting model statistics and diagnostic tests, including normality and homoscedasticity assessments.

951 **Table S7 | Post-hoc pairwise comparisons of alpha diversity among populations.** Rows correspond to pairwise
952 population comparisons, with separate results for richness and Shannon diversity and columns reporting the
953 estimated differences and associated statistical tests.

954 **Table S8 | PERMANOVA tests of population and chromosomal inversion effects on microbiome community
955 composition.** Rows correspond to the population and chromosomal inversion models, with columns reporting
956 degrees of freedom, sums of squares, pseudo-F statistics, explained variation and permutation-based significance
957 values.

958 **Table S9 | Indicator ASVs, taxonomic assignments and putative host associations across populations and
959 chromosomal inversion karyotypes.** Rows correspond to indicator ASVs identified for population or inversion
960 groups, with columns reporting indicator statistics, abundance/correlation information, taxonomic assignments and
961 the associated host genomic or transcriptional comparison.

962 **Table S10 | Genome-resolved characterization and taxonomic assignment of indicator ASVs.** Rows
963 correspond to indicator ASVs linked to reconstructed microbial genomes, with columns reporting sequence
964 identity and alignment statistics, taxonomic assignments from GTDB and NCBI, genome completeness and
965 contamination, and rRNA gene annotations.

966 **Table S11 | KEGG functional annotations and putative host-associated functions of indicator ASV genomes.**
967 Rows correspond to functional annotations identified in the genome-resolved indicator ASVs, with columns
968 reporting KEGG Orthology identifiers, functional categories and the corresponding ASV/genome assignments.

969 **Table S12 | Correlations between indicator ASVs and differentially expressed host genes across populations
970 and chromosomal inversion karyotypes.** Rows correspond to ASV–gene pairs, with columns reporting
971 comparison, indicator ASV, host DEG, and coefficient of determination (R^2).

972 **Table S13 | Gene Ontology enrichment of host genes correlated with indicator ASVs.** Rows correspond to
973 enriched Gene Ontology terms, with columns reporting the GO identifier, functional description, ontology
974 category, number of annotated genes, number of correlated genes and enrichment statistics.

975 **Table S14 | KEGG metabolic modules linking indicator ASVs and correlated host genes.** Rows correspond
976 to KEGG metabolic modules, with columns reporting module identifiers and descriptions, the KEGG Orthology
977 components contributing to each module, and the indicator ASVs and host genes associated with the module.

978 **Table S15 | Distributional comparisons of F_{ST} values among non-DEGs, non-correlated DEGs and iASV-**
979 **correlated DEGs across populations and chromosomal inversion karyotypes comparisons.** Rows correspond
980 to population and chromosomal inversion comparisons, with columns reporting the F_{ST} distributions for each
981 gene category and the statistical comparisons among categories.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [4MultiomicsSuppFigures.docx](#)