UNIVERSITÀ
DEGLI STUDI
DI TRIESTE

UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXVIII CICLO DEL DOTTORATO DI RICERCA IN

APPLIED DATA SCIENCE AND ARTIFICIAL INTELLIGENCE

ISTITUTO NAZIONALE DI OCEANOGRAFIA E GEOFISICA SPERIMENTALE

DATA-DRIVEN MODELS TO STUDY PHYTOPLANKTON DYNAMICS AND PHYSIOLOGY

Settore scientifico-disciplinare: **MAT/06**

DOTTORANDO / A
ALBERTO GHEDIN

COORDINATORE
PROF. FRANCESCO PAULI

SUPERVISORE DI TESI
PROF. COSIMO SOLIDORO

CO-SUPERVISORE DI TESI
PROF. LUCA MANZONI

CO-SUPERVISORE DI TESI
DOTT. VINKO BANDELJ

ANNO ACCADEMICO 2024/2025

UNIVERSITÀ DEGLI STUDI DI TRIESTE
DIPARTIMENTO DI MATEMATICA, INFORMATICA E GEOSCIENZE

XXXVIII CICLO DEL DOTTORATO DI RICERCA IN
IN
APPLIED DATA SCIENCE AND ARITIFICIAL INTELLIGENCE

DATA-DRIVE MODELS TO STUDY PHYTOPLANKTON DYNAMICS AND PHYSIOLOGY

Settore scientifico-disciplinare: MAT/06

Candidato:
Alberto Ghedin

Relatori:
Prof. Cosimo Solidoro
Prof. Luca Manzoni
Dott. Vinko Bandelj
Coordinatore:
Prof. Francesco Pauli

ANNO ACCADEMICO 2024/2025

Abstract

Phytoplankton are the foundation of marine ecosystems and a major driver of global biogeochemical cycles. Their composition and productivity are shaped by environmental gradients, physical mixing, and nutrient availability, while at the cellular level, their life-cycle transitions are controlled by complex gene regulatory programs. Understanding these processes across ecological and molecular scales is crucial for predicting how marine primary producers respond to environmental change. This thesis addresses these two complementary dimensions: the large-scale spatial and seasonal organization of phytoplankton communities along the Italian coastline, and the regulation of gene expression underlying sexual competence in the pennate diatom *Pseudo-nitzschia multistriata*.

In the first part, We analyzed an extensive dataset comprising over 2,000 phytoplankton samples collected between 2015 and 2017 across the Italian coastline. By integrating taxonomic counts with concurrent environmental variables, We characterized patterns of richness, abundance, and community composition. Species accumulation and indicator-taxon analyses revealed marked biogeographical structuring, primarily driven by nutrient availability and hydrographic connectivity. The northern Adriatic emerged as a high-abudance, diatom-dominated system with clear diatom–dinoflagellate succession, whereas the Ligurian and Tyrrhenian coasts displayed more oligotrophic, low-diversity assemblages. Constrained ordination analyses identified temperature, salinity, and nutrient gradients as the main drivers of community turnover, though together they explained only a modest portion of variance, which suggest additional unresolved influences such as episodic mixing, trace-metal inputs, or local hydrodynamics. These results provide one of the most comprehensive, process-oriented portraits of Italian coastal phytoplankton communities and highlight the strengths and limitations of monitoring data for ecological inference.

The second part explores transcriptional regulation in *Pseudo-nitzschia multistriata*, a model species for studying diatom sexual reproduction. By integrating published and newly processed RNA-seq libraries from both mating types and from cells above and below the sexualization size threshold, We constructed gene co-expression networks using the Weighted Gene Co-expression Network Analysis framework. We identified robust modules of co-regulated genes and linked them to traits of interest. Several modules were consistently associated with mating type and enriched in genes related to metabolism, redox homeostasis, and calcium signaling, while modules related to size were fewer and less stable. The known mating-type determinants exhibited expected expression biases but limited co-clustering, suggesting independent regulatory pathways or condition-specific activation.

Together, the two parts of this thesis connect ecosystem-scale phytoplankton dynamics with the molecular mechanisms governing life cycle regulation in diatoms. At the community level, they reveal how nutrient gradients and hydrography shape coastal biodiversity; at the cellular level, they potentially delineate mating type–specific transcriptional programs coordinating sexualization. By combining large-scale ecological analyses with systems-level transcriptomics, this work provides a multiscale view of phytoplankton, from population assemblages to gene regulatory networks.

Acknowledgments

I wish to thank all people of the National Institute of Oceanography and Applied Geophysics (OGS) for the support and friendship during these three years of study and work. Their presence, help and constant suggestions have been fundamental for the realization of the work here presented. First and foremost, I would like to thank Prof. Cosimo Solidoro for his guidance and for the possibility he gave me to delve into two different but complementary topics in marine biology. I gratefully thank Dott. Vinko Bandelj for his help and guidance on the work related to phytoplankton community ecology and Dott. Maria Immacolata Ferrante for the collaboration and guidance on the gene regulatory chapter.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 On phytoplankton

Marine ecosystems are among the most complex and dynamic environments on Earth [153]. Through a suite of biological and physical processes, they regulate the exchange of gases, most notably carbon dioxide (CO_2), between the ocean and the atmosphere, thereby exerting a major influence on global climate [98, 256]. The ocean acts as a central component of the Earth's climate system, absorbing a substantial fraction of anthropogenic CO_2 emissions and redistributing heat and energy across the planet [317]. These processes are embedded within the broader framework of global biogeochemical cycles: the interconnected pathways through which essential elements such as carbon, nitrogen, phosphorus, and silicon circulate among the biosphere, atmosphere, hydrosphere, and lithosphere [261]. By mediating these elemental fluxes, marine ecosystems sustain biological productivity and help maintain both the chemical balance of the oceans and the stability of Earth's climate [16].

At the root of these processes lies *plankton*: a collective term for organisms that lack sufficient motility to counteract water movements and are thus passively transported by currents. This assemblage spans nearly eight orders of magnitude in size, from viruses ($\sim 0.02\ \mu m$) to large gelatinous animals such as jellyfish (up to $\sim 2\ m$) [273]. A fundamental ecological distinction is made by separating the photosynthetic algae denominated *phytoplankton* from the animal organisms called *zooplankton*. Phytoplankton are predominantly autotrophic, performing photosynthesis convert co2 and inorganic nutirents in organic matter, with the concurrent use of light. In doing so, they represent the largest biological sink for atmospheric CO_2 and contribute roughly half of global net primary production [99]. As primary producers, they form the energetic and biogeochemical foundation of marine and freshwater ecosystems, sustaining nearly all

heterotrophic life. Conversely, zooplankton feed on phytoplankton and act as a key trophic link, transferring energy and organic carbon to higher consumers [33, 153].

Phytoplankton are mostly single-cell organisms encompassing an extraordinary taxonomic and functional diversity, each differing in cellular morphology, pigment composition, life-history strategies, and ecological roles [239]. This diversity is reflected in the roughly 50,000 described species of marine and freshwater microalgae [121], including about 4,000 morphologically distinguishable marine species. Among these, diatoms (1,700 species), dinoflagellates (1,600 species), and coccolithophores (300 species) represent the most ecologically significant groups [279]. Diatoms are siliceous algae enclosed within intricate glass-like frustules, responsible for $\sim$20% of global primary production and major contributors to carbon export to the deep ocean [295, 228]. Dinoflagellates, characterized by their cellulose thecal plates and motility via two flagella, display complex behaviors such as diel vertical migration and bioluminescence. Coccolithophores, unicellular haptophytes that produce calcareous plates (coccoliths), contribute simultaneously to organic carbon fixation and to the formation of biogenic calcium carbonate [239]. Molecular surveys suggest that the true diversity of these groups is likely an order of magnitude higher than morphological estimates indicate [82]. Their remarkable diversity makes the very concept of phytoplankton somewhat ill-defined: many species are not strictly autotrophic but exhibit *mixotrophy*, switching between photosynthetic and heterotrophic modes of nutrition depending on light or nutrient availability [284]. This metabolic flexibility further blurs classical trophic boundaries and underscores the ecological versatility that characterizes planktonic life.

The ecological role of phytoplankton extends far beyond supporting higher trophic levels. The organic matter they produce fuels microbial processes through the *microbial loop*—a network of interactions among bacteria, archaea, viruses, and protists that recycles dissolved organic matter back into the food web [21, 319]. Within this network, phytoplankton function simultaneously as sources and recipients of nutrients and organic substrates, tightly coupling primary production with microbial remineralization and nutrient regeneration [317, 261]. From a biogeochemical perspective, phytoplankton are the primary mediators of global carbon and nutrient fluxes. Their interactions within the microbial loop are intimately connected to the *biological carbon pump*, as both processes represent complementary pathways within the ocean's carbon cycle. While the microbial loop retains carbon and nutrients in the euphotic zone through rapid recycling, the biological pump transfers a fraction of organic matter to deeper layers, where it can be sequestered from the atmosphere on timescales ranging from decades to millennia [310, 93]. When phytoplankton cells die or are grazed upon, some of their carbon is respired in surface waters, while the remainder sinks as detritus or aggregates, contributing to long-term carbon sequestration [317, 261, 51]. Phytoplankton also regulate the cycling of key nutrients such as nitrogen, phosphorus, and silicon, as well as trace

metals like iron and zinc that act as essential cofactors in photosynthetic and respiratory metabolism [317, 261, 51]. Differences among taxa in nutrient uptake strategies, elemental stoichiometry, and storage capacities influence not only species' ecological success but also global nutrient availability and carbon export efficiency [168]. For example, diatoms contribute disproportionately to vertical export, while picocyanobacteria dominate oligotrophic regions where nutrient recycling prevails [106, 330]. Contemporary trait-based and functional-group approaches have helped elucidate how variation in cell size, metabolism, and life-history traits shapes global biogeochemical fluxes and climate feedbacks [107, 313]. Together, these intertwined microbial and biogeochemical processes govern the balance between carbon retention and export in the ocean, ultimately modulating marine productivity and Earth's climate system.

Because of their rapid generation times and high growth potential, phytoplankton respond swiftly to environmental changes, acting as both regulators and sentinels of ecosystem dynamics [87, 280]. Phytoplankton-rich waters often reflect eutrophic conditions resulting from excessive nutrient inputs of terrestrial origin [311, 299] , and shifts in light, nutrient supply, or grazing pressure can rapidly alter community composition and biomass, with cascading effects throughout food webs and biogeochemical cycles [317, 153]. Their tight coupling with environmental conditions also makes them reliable indicators of ecological status, as recognized in marine monitoring frameworks such as the Marine Strategy Framework Directive (MSFD) [89, 203, 96].

The remarkable coexistence of numerous phytoplankton species in seemingly uniform environments poses the classical ecological puzzle known as the *"paradox of the plankton"* [133]. According to the competitive exclusion principle, only a few species should persist when competing for a limited set of resources, yet natural waters often harbor tens to hundreds of coexisting taxa. This paradox has inspired extensive research into the mechanisms maintaining biodiversity in such systems, ranging from temporal environmental variability to nonlinear population dynamics and complex species interactions, but a definitive resolution remains elusive [266, 249]. The complex and nonlinear interactions between phytoplankton, their predators, and the environment have also prompted investigations into the dynamical properties of phytoplankton communities, particularly the potential presence of chaotic behavior [18, 122], which appears to be widespread in natural ecosystems, especially in planktonic systems [245, 246]. Understanding the origin and implications of such chaotic dynamics is central to interpreting the variability observed in field observations and experimental studies [122, 242, 212].

1.2 Mechanisms shaping phytoplankton spatial and temporal variability

Phytoplankton distributions vary enormously across both space and time, forming complex, non-uniform patterns rather than homogeneous fields. Surface chlorophyll organizes in swirling filaments and patches at kilometer scales [317, 164]. Indeed, patchiness is recognized as a ubiquitous feature of plankton in the ocean [298]. This spatial and temporal variability arises from the intricate coupling between biological processes such as growth, mortality, and grazing and physical drivers that acting at mesoscale and submesoscale to deform, concentrate, and disperse plankton fields [231, 175, 269, 244, 192]. Horizontal transport by currents, eddies, and fronts stretches and folds water masses, deforming phytoplankton distributions into filaments and patches [116, 180]. This lateral stirring acts on pre-existing gradients in biomass and nutrients, sharpening contrasts and generating fine-scale spatial heterogeneity [192]. At Fronts (sharp boundaries separating water masses of contrasting temperature, salinity, or density) horizontal convergence enhances both nutrient supply and cell accumulation, while divergence and velocity gradients disperse or dilute existing patches [164, 162]. The scale of this heterogeneity is governed by the interplay between stirring, which intensifies gradients, and diffusion, which tends to smooth them [192]. Turbulence further modulates nutrient fluxes and cell distributions, with different effect based on the scale of the process [85]. At small scales, turbulent diffusion homogenizes concentration differences, while at intermediate scales, vortices and intermittent velocity gradients can transiently concentrate nutrients or cells in short-lived patches rapidly smoothened out by diffusion [52]. The ecological expression of these physical processes depends on biological response times: fast-growing populations can exploit brief nutrient pulses, whereas slower-growing ones experience spatial homogenization before significant growth occurs.

Turbulent and mixing processes strongly influence vertical heterogeneity as well. In stratified waters, phytoplankton often concentrate at specific depths where light and nutrient conditions are balanced, forming a deep chlorophyll maximum (DCM) layer below the surface [75, 102]. In subtropical gyres, for example, a pronounced DCM frequently develops near the base of the euphotic zone, where residual nutrients sustain growth under reduced light exposure, creating vertically confined layer of phytoplankton largely invisible to satellites [99]. Physical mixing processes such as internal waves, shear instabilities, and eddy-driven upwelling can perturb these vertical distributions, periodically lifting deep phytoplankton toward the surface or mixing them downward [52, 196]. The resulting vertical structure depends on the stability of the water column: a well-mixed upper layer produces a more uniform biomass profile, whereas strong

stratification promotes sharp subsurface peaks in chlorophyll and biomass [75, 102, 52]. Such events create transient hotspots of growth that, through advection, evolve into broader spatial contrasts across the seascape [111, 317].

Phytoplankton communities exhibit marked temporal fluctuations across a continuum of timescales, from diel to interannual and decadal variability [144]. In temperate seas, these fluctuations are dominated by a strong seasonal cycle driven by the annual rhythms of light availability, stratification, and nutrient dynamics [317, 261, 153]. During winter, the several processes describe above mix water masses and replenish surface nutrients from deeper layers, but low irradiance, short photoperiods, and intense turbulence constrain phytoplankton growth. As spring advances, increasing solar radiation and the gradual stabilization of the upper water column create conditions favorable for rapid biomass accumulation. This creates an environment where phytoplankton population can exponentially grow, a phenomenon called *bloom* [130]. The collapse of the spring bloom is mainly due to nutrient exhaustion and to the top-down control exerted by concurrent population growth of zooplankton [32]. As the season progresses into summer, thermal stratification isolates nutrients stored in the deeper layers from the resource-depleted surface. The result is an oligotrophic regime which favors smaller phytoplankton and regenerated production [317, 261]. In autumn, surface cooling, stronger winds, and episodic storms deepen the mixed layer and reintroduce nutrients into the euphotic zone, occasionally supporting a secondary bloom before winter conditions return [172]. Beyond seasonal cycles, short-term perturbations can either benefit or hinder phytoplankton dynamics on days to weeks timescales. Storms, wind events and internal waves, for example, can mix the water colum and introduce nutrients into the euphotic zone, thus triggering short-lived blooms [170, 125, 176, 199]. Conversely, storms can also increase turbidity thus reduce the light flux for phytoplankton, which are not able to grow despite a huge supply of nutrients [272]; internal waves or upwelling relaxations that can cause rapid changes in nutrient supply. Overall, these episodic disturbances contribute to high-frequency variability superimposed on the smoother seasonal cycle. Conversely, storm events can enhance water-column mixing and turbidity, thereby reducing light availability and flushing phytoplankton cells out of productive coastal zones, leading to transient declines in biomass and shifts in community structure as species with slower growth or sinking tendencies are selectively removed. [283, 291, 71, 272]. Similarly, internal waves can perturb the vertical distribution of nutrients and light, occasionally transporting phytoplankton into suboptimal depth layers or reducing nutrient availability in the euphotic zone [308, 178]. Upwelling–relaxation cycles can also cause rapid fluctuations in nutrient inputs: during relaxation periods, stratification strengthens and nutrient replenishment weakens, temporarily constraining phytoplankton growth [200]. Overall, these short-term

events illustrate how physical forcing can not only stimulate but also suppress primary production, depending on the balance between light and nutrient perturbations.

1.3 The diatom life cycle

While phytoplankton communities deploy diverse strategies to cope with environmental variability, diatoms are notable for a life cycle that couples size-dependent vegetative growth with periodic sexual reproduction. Grasping this cycle is essential to interpret their ecological success, bloom dynamics, and long-term population renewal. We therefore provide a concise overview of the diatom life cycle before turning, in later chapters, to species-specific sexualization processes such as those of *Pseudonitzschia multistriata*.

Most diatoms share a two-stage life cycle comprising a vegetative and sexual phase. During the vegetative phase, mitotic division supports rapid exploitation of nutrients, leading to their characteristic rapid blooms [97]. Mitosis yields two diploid daughters (bearing two copies of the genome) genetically identical to the parent. The silica cell wall (frustule) comprises two differently sized thecae, the larger epitheca and smaller hypotheca, thus each daughter inherits one parental theca (becoming its epitheca) and must synthesize a new hypotheca [97]. This constraint causes progressive size reduction in lineages: across generations, mean cell size declines and the size distribution broadens. Sex reproduction restores cells to maximal size. Small cells become sexually competent, undergo meiosis to produce haploid gametes, and their conjugation forms a zygote that expands into an auxospore, which can grow to the species-specific initial size due to the absence of the frustule. Details of gamete conjugation differ across diatoms. Centric diatoms (phylogenetically older) are often *homothallic*: a single clone can produce both eggs and sperm. In contrast, most pennate diatoms are *heterothallic*: sexual reproduction requires pairing of opposite mating types (MT+ and MT–, typically referred as male and female, respectively). In both cases, cells must be below a species-specific *sexualization size threshold* (SST) to become sexually competent. The onset of the sexual phase is purely physical: microsurgery experiments showed that cells shrunk below their SST acquire sexual competence [285, 248, 67]. How diatoms sense their own size remains unclear [103, 34] and is a major open question in diatom life cycle regulation. Although this cycle is broadly conserved, environmental triggers vary among taxa: light appears to influence sex in some species of *Chaetoceros* [110], while salinity is implicated in others [115, 201] (see [97] for review). Notably, some model species—*T. pseudonana* and *P. tricornutum*—have not been observed to undergo size reduction or meiosis in laboratory conditions [123, 80] . In *P. multistriata*, successful sexual reproduction in culture initially requires a threshold cell concentration, leading

to the hypothesis that high density facilitates exchange of pheromones between mating types [265]. The presence of chemical signals was previously demonstrated in the freshwater araphid *Pseudostaurosira trainorii* [262] and the marine benthic pennate *S. robusta* [113]. In *S. robusta*, L-proline–derived diproline is released by MT– upon perceiving SIP+ (sexually induced pheromones) from MT+; MT+ then becomes responsive to diproline and glides toward its source [113, 205]. Pheromone exposure induces pronounced transcriptional changes and triggers sexual competence [42]. Analogous pheromone-responsive behaviors and transcriptional changes have been documented in *P. multistriata* [29, 187] and *C. closterium* [147]. In heterothallic pennates, pheromones serve both to locate partners and to coordinate the transition from vegetative growth to meiosis; cell–cell communication is therefore essential for mutual recognition, synchronized cell-cycle arrest, meiotic entry, and gamete exchange.

1.4 Data and methodologies for studying phytoplankton diversity and dynamics

The investigation of phytoplankton communities has historically drawn on multiple methodological traditions, each offering distinct but complementary insights into community composition, dynamics, and ecological function. Over recent decades, the methodological frontier has shifted from purely morphology-based enumeration toward a fusion of optical, molecular, and -omics approaches.[145, 114]. A comprehensive understanding of phytoplankton ecology thus requires integrating these diverse data sources, acknowledging both their complementarity and their limitations.

Microscopy remains the historical cornerstone of phytoplankton research. Conventional light microscopy, often using the Utermöhl sedimentation method [302, 177], enables identification and enumeration of phytoplankton based on morphological traits. This approach has generated decades of long-term datasets that underpin much of our knowledge on seasonal succession, species composition, and size-structured community dynamics [239], and remains crucial for understanding phytoplankton dynamics, ecological interactions, and responses to environmental changes [9]. Despite its robustness, microscopy has well-known constraints: taxonomic resolution is limited for small or morphologically similar species, cryptic diversity remains unresolved, and the process is time-consuming and dependent on expert taxonomists [114, 290]. As with all

count-based approaches, purely numeric measures can misrepresent ecological significance. In particular, when species differ markedly in cell size or carbon density, abundance alone may under- or overemphasize the role of particular taxa. Biomass provides a better proxy for the ecological and biogeochemical significance of phytoplankton communities, as it accounts for cell size and carbon content [124, 198]. Large, less numerous species—such as diatoms or dinoflagellates—may dominate total biomass even when numerically rare, whereas smaller picophytoplankton can dominate abundance but contribute less to total carbon stocks.

Advancements in imaging technology have enabled automated, high-frequency monitoring of plankton communities. Instruments such as the FlowCAM and Underwater Vision Profiler capture large numbers of particle images and allow semi-automated classification of taxa . These systems have greatly improved temporal and spatial resolution, allowing for near real-time observation of plankton dynamics[217, 227]. Machine-learning algorithms—particularly convolutional neural networks—are increasingly applied to classify plankton images, reducing reliance on manual annotation and expanding analytical throughput [114].

From the earliest DNA-focused studies of plankton [8, 126], genetic resources have expanded dramatically and now complement microscopy-based observations [145]. Metagenomic studies have enabled increasingly detailed exploration of community structure [181, 226], unprecedented assessments of taxonomic richness [206], functional diversity [282], biogeography [173], and clarified the contributions of plankton to carbon export and ocean biogeochemistry [119, 49]. In parallel, metatranscriptomics, metaproteomics, and metabolomics have expanded our capacity to link community composition to function, offering functional snapshots of gene expression and metabolic activity [7]. These integrative omics approaches allow reconstruction of metabolic pathways [289], inform reconstructions of evolutionary history [91, 204], and refine our understanding of plankton metabolism by identifying genes and pathways involved in environmental responses, both in natural assemblages [53, 207] and controlled laboratory experiments [92, 19]. Together, they provide a comprehensive view of phytoplankton diversity, function, and evolution, illuminating their central roles in nutrient cycling, primary production, and global biogeochemical processes [173].

Falling costs for collecting and analyzing omics data have fostered global initiatives, from early large-scale efforts [250] to those led by the Tara Ocean Foundation [229, 135, 257], and have even enabled the emergence of omics-based biogeochemical models [237, 163, 73]. In parallel, the development of model microalgae has further advanced our understanding of phytoplankton biology. The first diatom models—*Thalassiosira pseudonana* [14] and *Phaeodactylum tricornutum* [50]—were later joined by *T. oceanica*

[171], *Fragilariopsis cylindrus* [204], *Seminavis robusta* [66], *Cylindrotheca closterium* [303], *Skeletonema marinoi* [140], and *Pseudo-nitzschia multistriata* [29]. The goal is now to sequence 100 species to better probe their ecological and evolutionary roles in the ocean [1], while large-scale initiatives such as the Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP) have released 650 assembled, functionally annotated transcriptomes spanning 210 eukaryotic genera [146].
Beyond field-based meta-omics approaches, controlled laboratory experiments have increasingly employed RNA sequencing (RNA-seq) to investigate gene expression dynamics in phytoplankton under specific environmental or physiological conditions [41, 40]. While metatranscriptomics captures community-level activity in natural assemblages, laboratory RNA-seq provides a targeted means to dissect cellular responses and molecular regulation within single species or strains, linking genotype to phenotype under experimentally defined stimuli [204]. In addition to physiological acclimation, RNA-seq analyses have become pivotal for exploring life-cycle transitions and cellular differentiation in phytoplankton. In particular, diatoms undergo a complex life cycle alternating between mitotic cell division and a sexual reproduction phase triggered by cell-size reduction and environmental cues [68, 15]. The sexualization phase involves profound transcriptional reprogramming, including the activation of meiosis-related genes, signaling molecules, and sex-specific transcription factors [205]. RNA-seq experiments in model species such as *Pseudonitzschia multistriata* and *Seminavis robusta* have illuminated the molecular basis of this transition, identifying sex-specific gene networks and signaling cascades associated with gametogenesis and mate recognition [103, 66].

1.5 Statistical approaches in marine ecology and genetics

Previous sections elucidate the inherent complexity of marine ecosystems, characterized by high variability across spatial and temporal scales and by the interaction of multiple biotic and abiotic processes. Statistical analyses in marine ecology aim to disentangle this complexity by identifying patterns, quantifying relationships, and testing ecological hypotheses derived from observational or experimental data. The choice of statistical methods depends on both the type of data available (e.g., time series, spatial surveys, community composition matrices, molecular profiles) and on the ecological questions being addressed.
Broadly speaking, relevant questions in marine ecology are:

- Which are the fundamental species in a community?

- What are the main environmental drivers shaping community composition and diversity?
- How do species distributions vary in space and time?
- Are there detectable long-term trends or regime shifts in ecological indicators?
- Can we infer causal mechanisms and predict system responses?

To address such questions, marine ecologists employ a wide spectrum of statistical tools. Descriptive and exploratory analyses (e.g., summary statistics, diversity indices, and correlation analyses) provide initial insights into data structure and variability. This initial process help first of all to identify outliers, redundant information such as highly correlated variables and can give clues of the presence (or absence) of pattern in the data. When the focus lies on testing ecological hypotheses, linear models (LM) are extremely used, given they simplicity, ability to include multiple predictors and their interaction, and most importantly their explainability [331]. Strictly related, analysis of variance (ANOVA) and similar tests (Kruskall-Wallis, Mann-Whitney) are also used for their ability of assessing differences in the distribution of any quantity. Linear models however are hardly applied to raw data because neither the response nor predictors are normally distributed [159]. A possible approach is to apply specific transformation to the data to reduce the non-gaussianity, as an example log-transform abundances greatly helps [159, 47]. Another possible path is instead to use more advanced techniques such as generalized linear models (GLMs) or generalized additive models (GAMs) which are able to capture nonlinear relationships and can accomodate diverse kind of responses (presence/absence data modelled as ones and zeros, and abundance data which are discrete) [318]; also mixed model, allowing a more complex structure of the covariates, are widely used [332]. Though widely used, these methods fall short when dealing with ecological communities, as the response consists of multiple taxa. However, there are multiple methods to address questions. On one hand it is possible to *clusterize* samples according to the observed species using a specific dissimilarity or distance measure. Clusters are extremely helpfull to determine subcommunities inside the entire dataset (by looking which taxa are more relevant inside a cluster) which can be indirectly associated to possible drivers by looking at the properties of the samples belonging to any cluster. On the other, if there is the interest of a more direct association between a communitiy and possible response variables, *ordination methods* are quite the standard approach. Redundancy analysis (RDA), canonical correspondance analysis (CCA) and non-metric multidimensional scaling (nMDS), and several others method, allow to display the community data matrix in a low-dimensional space (generally 2D) and at the same time to find statistically meaningful associations with possible response variables [159, 47, 235, 222]. However, this is just an overview of the most

Software/Package	version
R	4.4.1
car	3.1
caret	7.0
clevr	0.1.2
clusterProfiler	4.14.6
DESeq2	1.46.0
fastqc	0.11.9
ggplot2	4.0.0
indicspecies	1.8.0
iNEXT	3.0.2
MASS	7.3
Rsubread	2.12.3
star	2.7.9
sva	3.54.0
trimmomatic	0.39
vegan	2.7
WGCNA	1.73

TABLE 1.1: List of software and packages used in this thesis alongside their version.

standard techniques used, but many other have been implemented to answer similar or related problems (see [235, 159, 47, 222, 120, 10]).

As briefly explained above, RNA-seq now allows researchers to investigate how organisms respond to controlled environmental perturbations at the transcriptional level. Rather than focusing on community structure or biomass, these approaches quantify changes in gene expression, revealing the molecular mechanisms underlying physiological responses, acclimation, and adaptation. From a statistical perspective, RNA-seq data require specialized analytical pipelines that differ from those used for ecological count data, yet share conceptual similarities in dealing with multivariate, high-dimensional, and often noisy datasets [149]. Differential expression analysis is used to identify genes whose expression levels significantly change under experimental treatments, and they are tipically teated with GLM with negative binomial error distributions [174, 65]. Apart from this, multivariate ordination and network inference methods are increasingly applied to RNA-seq datasets to detect co-regulated gene modules and to explore functional connectivity within cellular systems. Gene co-expression networks and clustering algorithms (e.g., Weighted Gene Co-expression Network Analysis, WGCNA) provide insights into coordinated biological responses, linking transcriptional patterns to physiological states or life-cycle transitions [154].
To keep track of all the different methods and associated packages used in this thesis, we list all of them, with their version, in Tab. 1.1.

1.6 Structure of the thesis

Chap. 2 presents the first set of results and focuses on the ecological and temporal dynamics of phytoplankton communities along the Italian coast. It characterizes patterns of abundance, diversity, and taxonomic composition across space and time, and investigates the environmental drivers influencing community structure. Emphasis is placed on how physical forcing and nutrient regimes shape the variability of coastal phytoplankton assemblages and how these patterns inform our understanding of ecosystem functioning. Chap. 3 shifts the perspective from the community to the molecular scale, exploring the global regulation of gene expression in the diatom *Pseudo-nitzschia multistriata*. Using RNA sequencing and Weighted Gene Co-expression Network Analysis, it reconstructs transcriptional networks underlying mating-type identity and the transition across the sexualization size threshold. The chapter highlights asymmetries between mating types, the organization of co-regulated gene modules, and the integration of known regulators such as the *MRP*/*MRM* genes. Finally, in Chap. 4, we summarize the entire work, highlight its strengths and limitations, and discuss possible directions for future research.

Chapter 2

Characterization of phytoplankton communities along Italian coast

2.1 Introduction

2.1.1 Oceanographic characteristics of the Italian coastline

The Italian peninsula is surrounded by the Adriatic, Ionian, and Tyrrhenian basins, each characterized by distinct hydrographic, morphological, and biogeochemical features. Contrasts in riverine inputs, bathymetry, circulation patterns, and water-mass exchanges drive the spatial and temporal variability of phytoplankton productivity and community structure along the Italian coasts.

Extending along Italy's eastern coastline, the Adriatic Sea exhibits a pronounced north–south depth gradient, ranging from the shallow northern shelf (average depth $\sim 35\ m$) to the deep South Adriatic Pit ($\sim 1,200\ m$) [260]. The Strait of Otranto (maximum depth $\sim$ 780–800 m) partially restricts water exchange with the Ionian Sea, creating distinct sub-basins with contrasting hydrographic regimes: a shallow, river-influenced northern zone; a transitional central sector (average depth $\sim$ 140 m); and a deep, quasi-oceanic southern basin [17].
The Adriatic receives substantial freshwater input, most notably from the Po River (mean annual discharge 1500–1700 m^3/s), which alone accounts for roughly one-third of the total Mediterranean riverine inflow [100, 20]. Additional rivers (e.g., Adige, Brenta, Piave) and lagoonal discharges further enrich the northwestern shelf with nutrients [20, 38, 208]. The basin's dominant cyclonic circulation sustains the Western Adriatic Current (WAC), a persistent coastal flow that transports riverine waters and

plankton southeastward along the Italian coast toward the Strait of Otranto [179]. This southward transport extends the influence of northern nutrient inputs, producing a marked trophic gradient from the eutrophic northwestern Adriatic to the oligotrophic southern and eastern sectors [186, 100]. Associated inshore–offshore gradients favor large phytoplankton (e.g., diatoms) and elevated chlorophyll concentrations near the coast, while offshore waters remain nutrient-poor and clear [43, 277].
During winter, strong cooling and wind-driven mixing frequently result in complete vertical mixing of the northern and central shelves, replenishing surface nutrients and resetting the system for subsequent productivity cycles [208]. Wind-driven upwelling events can further enhance nutrient supply under favorable atmospheric conditions. Northeasterly Bora or along-shore Sirocco winds, for instance, induce offshore Ekman transport that brings deeper, nutrient-rich waters to the surface along portions of the Adriatic coast [219]. These episodic upwellings often trigger transient phytoplankton blooms [100].
Consequently, the northern Adriatic sustains high primary productivity and frequent diatom-dominated blooms, which can occasionally lead to eutrophication and mucilage formation [267, 20, 208, 260]. In winter, turbulent and nutrient-rich conditions favor diatom dominance, whereas in summer, stratification and nutrient depletion promote smaller phytoplankton and subsurface chlorophyll maxima. The interplay of freshwater forcing, vertical mixing, and episodic upwelling makes the Adriatic one of the most productive regions of the Mediterranean [208].

In contrast, the Ionian Sea is among the deepest Mediterranean basins, with depths exceeding $\sim 3,000$–$4,000\ m$ in several areas [150]. This sea lies at a hydrographic crossroads of major Mediterranean water masses: relatively fresh, nutrient-influenced Adriatic water enters from the north, saline and nutrient-poor Levantine Intermediate Water (LIW) flows westward from the Levantine Basin, and Modified Atlantic Water (MAW) enters from the west via the Sicily Channel [36, 37, 150]. Its great depth and strong connectivity with the Eastern Mediterranean contribute to its overall oligotrophic character: the water column is highly stratified and stable for most of the year, with nutrients largely confined to deep layers well below the euphotic zone[76, 189]. Only water-mixing events such as cyclonic gyres or upwelling transport LIW or other nutrient-rich subsurface waters to the upper layers [70].
Along southern Sicily, a persistent upwelling cell driven by the Atlantic–Ionian Stream (AIS) and often reinforced by summer winds produces a ribbon of cold, nutrient-rich water that supports elevated phytoplankton biomass relative to the typically oligotrophic offshore waters [243, 241]. Another key hotspot is the Strait of Messina, where strong

tidal currents and abrupt topography induce intense turbulence and intermittent deep-water upwelling [57]. Phytoplankton productivity in this area can be an order of magnitude higher than in adjacent stratified waters, with subsurface nutrient injection fueling recurrent local blooms and supporting abundant fish populations [61].

West of Italy, the Tyrrhenian Sea is a deep ($\sim 3,600\ m$) and morphologically complex basin, bounded by the Sardinia and Corsica Channels to the west and the Strait of Sicily to the southeast [36]. Its irregular coastline, with embayments such as the Gulfs of Naples and Gaeta, interacts with local bathymetry to steer currents and generate localized upwelling zones [190, 304]. The basin receives surface inflow of MAW via the Sardinia Channel and intermediate inflow of LIW through the Strait of Sicily at depths of $\sim$ 200–400 m [79, 134]. These water masses circulate cyclonically along the Italian margin, generating strong west–east gradients across the basin [54].
The open Tyrrhenian Sea is largely oligotrophic to mesotrophic [183]. Productivity increases locally in dynamic coastal zones—near major river mouths (e.g., Arno, Tiber) and within semi-enclosed embayments such as the Gulfs of Naples and Gaeta [69, 188, 240] The Tiber River, the second largest in Italy, exhibits elevated winter discharge that correlates with increased coastal phytoplankton abundance [305]. Wind forcing also modulates vertical fluxes: depending on wind direction, downwelling or upwelling events may occur along the Tyrrhenian coast [190, 188]. Seasonally, a northwestward-flowing current develops along the Italian coastline during winter and spring, connecting the Tyrrhenian to the Ligurian Sea before weakening in summer and autumn [81].

Overall, the Adriatic—particularly its northwestern region—exhibits mesotrophic to eutrophic conditions, sustained by strong riverine inputs and vertical mixing. The Ionian remains persistently oligotrophic, with productivity limited to localized upwelling zones, while the Tyrrhenian is generally oligotrophic but punctuated by mesotrophic coastal hotspots.

2.1.2 Phytoplankton communities in Italian seas

As previously discussed, phytoplankton dynamics in temperate marine coastal zones result from the complex interplay of environmental drivers and anthropogenic pressures. The major natural factors shaping these communities include solar irradiance,

water temperature, nutrient availability, stratification and mixing dynamics, and physical transport processes (e.g., upwelling, advection, and currents). In parallel, human-induced influences—particularly nutrient enrichment (eutrophication) and coastal development—can markedly alter phytoplankton distribution, abundance, and community structure. Together, these processes determine the spatio-temporal variability of phytoplankton in coastal waters, from seasonal blooms to biogeographic gradients in species composition. Observations across the Adriatic, Ionian, and Tyrrhenian seas reveal distinct community structures, bloom timings, and controlling mechanisms, reflecting the strong influence of physicochemical gradients and anthropogenic impacts in these environments.

Such hydrographic variability promotes taxonomic compositions and turnover patterns typical of temperate coastal zones [275, 296, 239]. Phytoplankton species are often classified along a continuum between *r*- and *K*-strategists, reflecting adaptations to contrasting environmental regimes. *r*-strategists thrive under high-nutrient, turbulent, and light-variable conditions, taking advantage of rapid growth and opportunistic reproduction. In contrast, *K*-strategists are better suited to stable, stratified, and oligotrophic waters, where they compete efficiently for limited resources. Diatoms are typical *r*-strategists, making them well adapted to the late-winter and early-spring periods, when nutrient-rich and well-mixed waters coincide with increasing solar irradiance. Under such conditions, they can form intense blooms that rapidly deplete available nutrients within a few weeks. Dinoflagellates, by contrast, represent *K*-strategists: they exhibit slower growth rates, favor nutrient-limited conditions, and are often negatively affected by strong mixing. Many species are motile and/or mixotrophic, allowing vertical migration and grazing on other phytoplankton [239, 35]. Additionally, numerous bloom-forming dinoflagellates display warm-temperature optima and grow poorly under cold conditions [322, 24]. Consequently, dinoflagellates dominate during stratified, warm-water periods and can persist in nutrient-depleted environments. This ecological contrast results in a well-known diatom–dinoflagellate seasonal succession: diatoms dominate in late winter and early spring, dinoflagellates prevail in late spring and summer, and diatoms reappear in autumn [275, 296, 239]. Their differing life-history strategies thus make each group better adapted to specific environmental conditions and underpin the temporal turnover observed in phytoplankton communities.

In the Adriatic Sea, phytoplankton abundance follows alternating peaks and troughs closely associated with the Po River's flood–drought cycles, underscoring the strong hydrological control on community biomass and seasonality [38]. Assemblages are overwhelmingly dominated by cosmopolitan diatoms, which account for most of the

biomass and abundance, while small phytoflagellates prevail during inter-bloom periods. Despite the overall similarity in species composition across the basin, local abundance patterns vary substantially according to nutrient input, hydrodynamics, and stratification intensity [20, 38].

The northern Adriatic exhibits three main growth phases that recur annually: late winter (February), spring (April), and mid-summer (July) [20]. The most pronounced bloom typically occurs between February and April, driven by increasing light availability and nutrient inputs from river runoff and winter mixing. This bloom is dominated by *Skeletonema marinoi*, followed by smaller peaks in spring and autumn. Other common diatoms thriving under nutrient-rich conditions include multiple species of *Chaetoceros* and *Pseudonitzschia*, and *Lauderia annulata*, whereas coccolithophores, notably *Emiliania huxleyi*, form a minor but persistent component in winter [38, 294]. As nutrients—particularly silicate and phosphorus—become depleted, diatom biomass declines and dinoflagellates (e.g., *Prorocentrum minimum*, *Ceratium furca*, *Scrippsiella trochoidea*) become dominant under stratified, low-nutrient conditions. Secondary blooms often occur in autumn when mixing and rainfall restore nutrient levels, although these tend to be smaller and more species-specific [225].

In the Venice Lagoon and adjacent coastal stations, a extensively studied section of the norther Adriatic, a consistent seasonal succession has been observed: *Chaetoceros spp.* dominate throughout the year; *S. marinoi* peaks in late winter; *Pseudo-nitzschia spp.* in summer–autumn; *E. huxleyi* in autumn–winter; and cryptophytes in colder months [39]. Harmful or toxic taxa also follow seasonal trends: domoic acid-producing *Pseudo-nitzschia* strains regularly constitute approximately $\sim$15% of diatom abundance in the Gulf of Trieste, peaking in spring and autumn [300]. In this area, species from the *P. delicatissima* complex typically peak in spring–summer, whereas those of the *seriata* complex occur mainly in autumn–winter [300]. Although toxicity is generally low, autumn blooms can occasionally comprise up to 80% of total phytoplankton biomass. Along the Romagna–Marche coast, potentially harmful taxa represent about 8% of total phytoplankton, with abundance linked to phosphate enrichment and low salinity—conditions indicative of nutrient stoichiometry imbalances and Po-derived freshwater inputs [225]. Overall, the northern Adriatic remains a high-biomass, diatom-dominated system characterized by marked seasonality and strong interannual variability modulated by river discharge, stratification–mixing cycles, and episodic meteorological events [38].

In the mid-Adriatic, WAC maintains a pronounced coastal–offshore gradient in nutrient concentration, biomass, and cell size structure [208]. Coastal waters, under greater riverine influence, typically experience bloom in winter, whereas offshore sites peak in early summer under more oligotrophic, stratified conditions. Coastal assemblages are

thus more bloom-prone and variable, while offshore communities exhibit higher taxonomic stability and diversity [208]. Therefore, this area shares the basin-scale seasonality of diatom-dominated blooms interrupted by smaller flagellate phases, but with decreasing riverine influence and more frequent summer maxima offshore. Stratification–mixing dynamics and the magnitude and timing of river plumes remain the principal drivers of phenology and interannual variability in this area [208].
Microphytoplankton (diatoms and dinoflagellates) contribute significantly to spring blooms along the Puglia coast, where winter blooms (January–February) are typically followed by bacterial peaks (February–March), indicating strong microbial loop coupling in an otherwise nutrient-poor system [281]. Urbanized and port areas such as Brindisi show locally enhanced diatom biomass, whereas offshore waters remain low in chlorophyll-*a* but display distinct winter maxima [281]. In the Ionian sector near Taranto, community composition is highly variable and structured by phosphorus and nitrogen availability. Larger phytoplankton such as the diatoms *Chaetoceros spp.*, *Cylindrotheca closterium*, and *Pseudo-nitzschia spp.*, together with the dinoflagellates *Alexandrium spp.*, *Ceratium furca*, and *Prorocentrum minimum*, coexist with abundant picophytoplankton, particularly *Synechococcus*, which shows strong seasonal correlations with temperature and nitrogen compounds [56].

Compared with the Adriatic, the Ionian and Tyrrhenian seas remain relatively understudied, except for specific sites such as the Gulf of Naples and nearby areas [328, 184, 45]. Consequently, information on community structure and temporal variability is still limited.
The northern Ionian (Puglia coast) exhibits a heterogeneous community despite its relatively small spatial extent [56]. Some sites display a winter–early spring bloom dominated by diatoms (*Pseudo-nitzschia spp.*, *Guinardia striata*, *Detonula confervacea*) and dinoflagellates (*Protoperidinium spp.*), whereas adjacent locations lack similar patterns. A secondary bloom in September is typically dominated by *Prorocentrum minimum*, a recurrent HAB species. Similarly, [254] observed in the same region that *Chaetoceros affinis*, *C. curvisetus*, *Leptocylindrus danicus*, *Guinardia delicatula*, and *Thalassionema nitzschioides* prevail in winter–spring, while dinoflagellates (*Prorocentrum micans*, *Scrippsiella trochoidea*, *Tripos furca*, *Protoperidinium granii*) and small flagellates dominate during summer–autumn. On the eastern coast of Sicily, [83] reported that in early autumn, the surface layer supports a dinoflagellate-rich community (*Tripos furca*, *Scrippsiella acuminata*, *Gonyaulax* spp.), whereas the deep chlorophyll maximum (DCM) is dominated by diatoms (*Chaetoceros laciniosus*, *C. lauderi*, *Leptocylindrus danicus* complex, *Pseudo-nitzschia* complex, *Thalassionema bacillare*, *Guinardia striata*) and transient pulses of *Prorocentrum cordatum*.

Limited studies on the Tyrrhenian Sea [191] reveal a typical annual cycle characterized by spring and autumn diatom blooms, followed by a predominance of dinoflagellates during the warmer months. However, most available data concern lagoonal systems rather than open coastal waters, making the verification of large-scale patterns challenging [232, 88]. Finally, studies conducted in the Sardinia region focused on lagoons rather than open coastal waters [232, 88].

2.1.3 Structure of the chapter

The introduction depicts the complex dynamics surrounding the italian coastaline and hints how thw phytoplankton communities are weel characterized only in the Adriatic and in some parts of the Mediterranean. In this chapter we will address this problem by analyzing a national-wide campaign of microscopy-based taxonomic observation of phytoplankton, with the aim of determining the general strcuture of the phytoplankton community, which taxa are more relevant in various sections of the italian coasts and in which period of the year, ans finally we will try to understand to what extent the physico-chemical drivers at disposal are sufficient at describing the overall community. In Sec. 2.2 we describe the dataset, highlighting the limitation it possess; in Sec. 2.3 we desribe all the statistical methods used to answer the questions mentioned above; in Sec. 2.4 we present the results and finally in Sec. 2.5 compare our results with literature and interprest the observed succession and the community-environment relations.

2.2 MSFD data set description

The dataset was collected within the European Marine Strategy Framework Directive (MSFD) program [89, 96]. In Italy, sampling was conducted by the regional *Agenzie Regionali per la Protezione Ambientale* (ARPAs) [137], operating independently under a shared protocol [203]. Sampling occurred between July 2015 and December 2017. The 15 Italian coastal regions—Friuli - Venezia Giulia, Veneto, Emilia-Romagna, Marche, Abruzzo, Molise, Puglia, Basilicata, Calabria, Campania, Lazio, Toscana, Liguria, Sicilia, and Sardegna—are hereafter referred to by the abbreviations in Table 2.1. A total of 162 stations were distributed along 54 transects, each consisting of three stations positioned approximately 3, 6, and 12 nautical miles (NM) from the coast (Fig. 2.1). The number of transects per region scaled with coastline length. Seawater was collected with Niskin bottles; subsamples (250, 500, or 1000 mL) were fixed with acid Lugol's iodine [203]. Phytoplankton cells were counted using inverted light microscopy at 20×–400× magnification [302]. Taxonomic identification was performed by different

operators across regions, potentially introducing variability in accuracy and consistency. Although the protocol recommended bimonthly sampling at the surface and at the deep chlorophyll maximum (DCM, typically 25–30 m), effort was highly heterogeneous across regions. For example, EMR sampled the surface in 29 of 30 months, whereas VEN started only in 2016. BAS did not sample for a total of 17 months. In this time window, not all stations were sampled consistently in the same month (Fig. 2.2). Depth sampling showed similar inconsistencies: some regions sampled more depths than yhe minimum, others only the surface. Data were downloaded from the MSFD portal [136] and merged into a single dataset. The original records included sampling date, station and region identifiers, phytoplankton community composition. Taxon names were not always consistent across or within regions; therefore, all names were matched against the World Register of Marine Species (WoRMS; tool *match taxa*) [320] and corrected accordingly. WoRMS also provided complete taxonomic hierarchies. Samples with missing spatial or temporal metadata that could not be reconstructed were excluded. Concurrently to phytoplankton sampling, measures of physical variables and collection of water samples for the identification of chemical variables in the laboratory, were performed. These variables included: temperature [°C], salinity [psu], Secchi depth [m], electrical conductivity [S/m], nitrate, nitrite, ammonia, total nitrogen, orthophosphate, total phosphorus, chlorophyll-*a*, dissolved oxygen (all [μmol/L]), oxygen saturation [%], and pH. Temperature, salinity, dissolved oxygen, conductivity and chlorophyll-*a* were obtained using a multiparameter CTD probe while the other chemical compounds were collected with a Niskin bottle and subsequently analyzed [203]. Each measurement carried a flag indicating whether the value was below the limit of detection (LOD) or the limit of quantification (LOQ). Values flagged as LOQ indicate the analyte was present but below instrumental sensitivity; LOD indicates presence cannot be confirmed. Thresholds depended on the analyte and instrument. For measurements flagged LOD/LOQ, the corresponding instrument sensitivity was reported. Because sub-threshold values constituted a substantial fraction of the data, we substituted them by drawing a random number from a uniform distribution on $[0, \text{sensitivity}]$. All variables underwent quality control, and values outside acceptable ranges were discarded.

Prior to analysis, we applied the following harmonization steps. (1) Taxa labelled *sp.* were standardized to *spp.* and interpreted as unidentified cells at the genus level, accounting for inconsistency in species-level identification among regions. (2) Only surface samples were retained for the final dataset. (3) One LIG sample with extremely high abundance ($\sim 1 \times 10^6$ cells/L) relative to all other samples in the region was removed. The resulting harmonized dataset comprised 2219 surface samples.

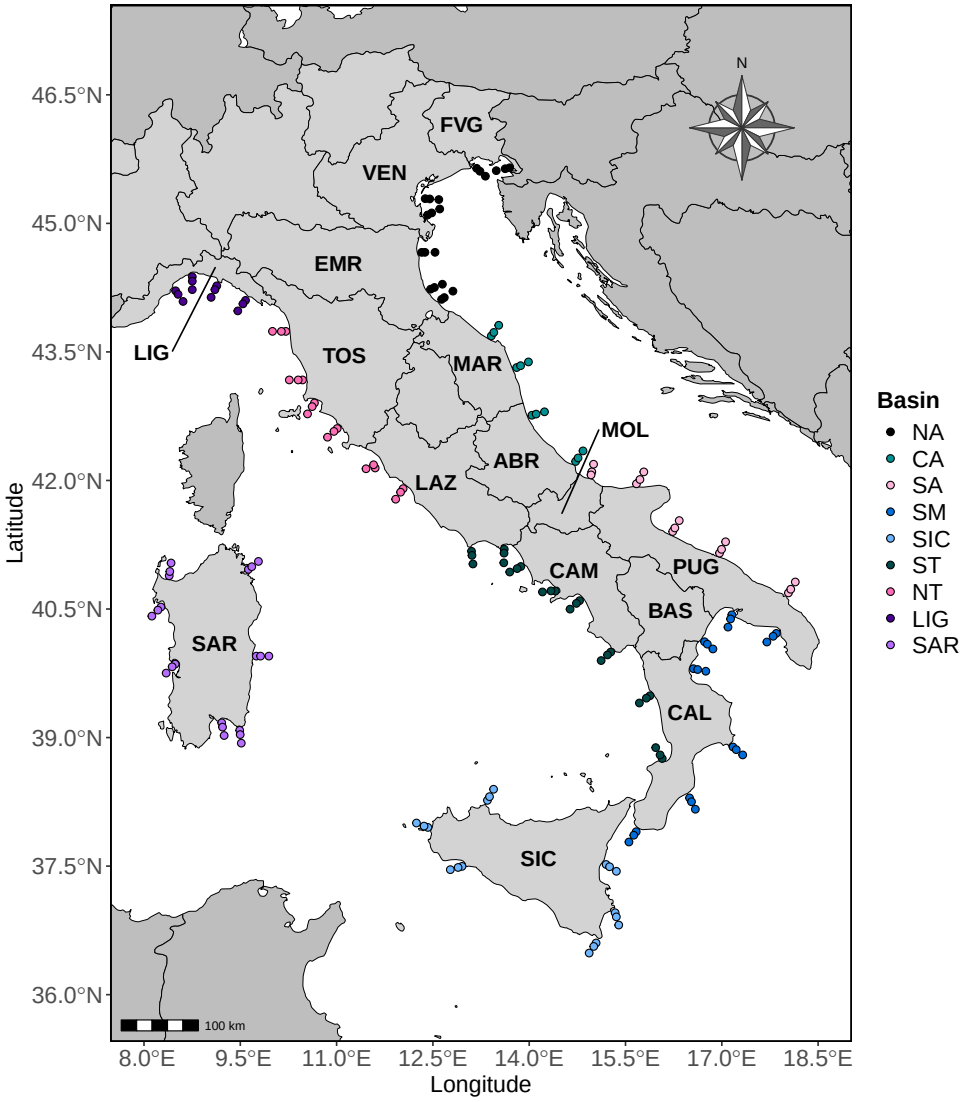


FIGURE 2.1: Map of stations, basin extents, and administrative-region abbreviations. Colours indicate the basin assigned to each station for statistical analyses.

2.3 Statistical analysis

We first examined spatial patterns in total cell abundance along the coast (Fig. 2.5). Unless otherwise specified, differences in abundance were tested using ANOVA on log-transformed data [47]. We also assessed systematic differences in taxonomic resolution by computing, for each region, the average relative contribution of identification levels to total abundance. For simplicity, all identifications above the genus level were pooled into a "Higher category" (Fig. 2.6).

To enable fair comparisons of richness among regions, we constructed species accumulation curves (SACs) using the `R` package `iNEXT` [129, 63]. SACs estimate site richness at a standardized number of individuals or samples, allowing comparisons under unequal sampling effort. `iNEXT` applies a combinatorial approach for rarefaction and

Region	Abbreviation	Station	No. of samples
Friuli-Venezia Giulia	FVG	6	90
Veneto	VEN	6	70
Emilia-Romagna	EMR	9	270
Marche	MAR	6	126
Abruzzo	ABR	6	87
Molise	MOL	3	45
Puglia	PUG	18	270
Basilicata	BAS	3	36
Calabria	CAL	18	95
Campania	CAM	12	180
Lazio	LAZ	12	142
Toscana	TOS	12	144
Liguria	LIG	12	180
Sicilia	SIC	18	216
Sardegna	SAR	21	269

TABLE 2.1: region-related information: name, name abbreviation, number of stations and total number of surface samples

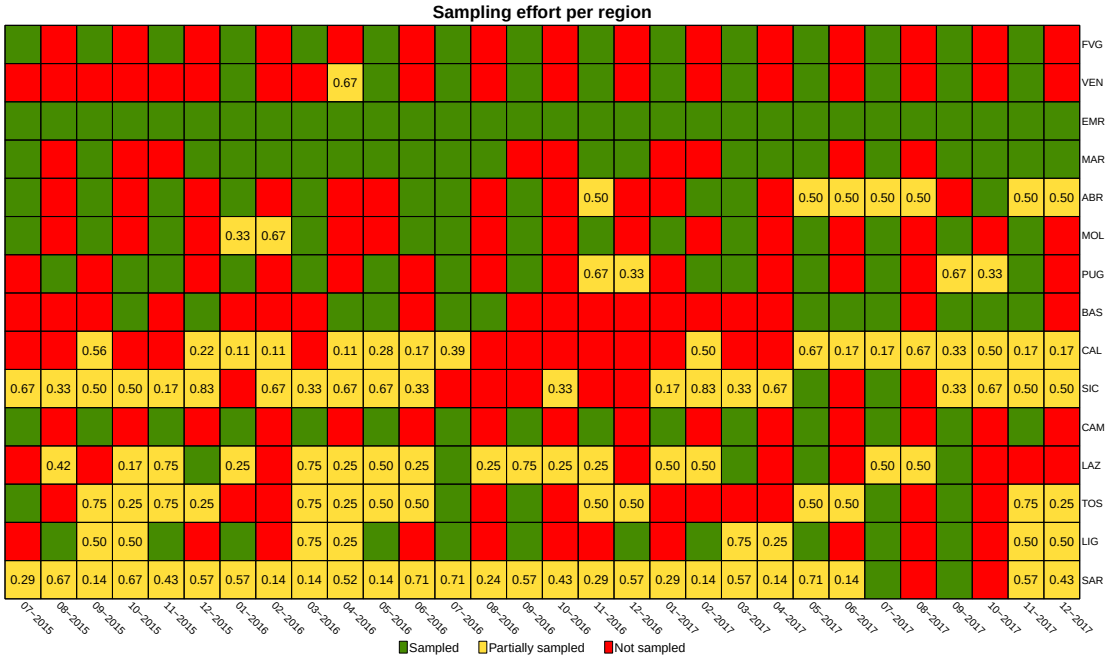


FIGURE 2.2: Temporal distribution of sampling by region. Green squares: all stations sampled in that month; yellow: only a subset of stations sampled (number indicates the fraction); red: no samples collected.

an analytical approach for extrapolation; 95% confidence intervals were obtained by a bootstrap method. We report hypothetical richness at a standardized effort of 50 samples.

To understand the general composition of phytoplankton communities, we calculated: (1) Frequency distributions of the 10 most abundant taxa, classes, genera, and species;

(2) The 10 most species-rich classes and genera. For a broader spatial perspective, we divided Italian waters into nine basins: Northern Adriatic (NA), Central Adriatic (CA), Southern Adriatic (SA), Southern Mediterranean (SM), Sicily (SIC), Southern Tyrrhenian (ST), Northern Tyrrhenian (NT), and Sardinia (SAR) 2.1. These basins were adapted from Copernicus Marine Service sub-regions [258] and modified to avoid splitting coherent areas and to reflect sharp abundance changes (e.g., lower abundances in CA relative to NA and SA).
To identify taxa characterizing basins, we used the Indicator Value (IndVal) index [94, 78], suitable for non-overlapping site groups with frequency and abundance data. For taxon i in group k:

$$\begin{aligned} \text{IndVal}_{ik} &= A_{ik} \cdot B_{ik}, \\ A_{ik} = \frac{N_{ik}}{N_{+k}}, &\quad B_{ik} = \frac{F_{ik}}{F_{+k}}, \end{aligned}$$

where A_{ik} is *specificity*, B_{ik} is *fidelity*; N_{ik} is the mean abundance of taxon i within group k, N_{+k} is the sum of mean abundances of all taxa within group k, F_{ik} is the number of sites in group k where taxon i occurs, and F_{+k} is the total number of sites in group k. The overall IndVal for a taxon is the maximum across groups. A taxon is considered characteristic if its IndVal exceeds a threshold.We used the `R` package `indicspecies` [77] to calculate IndVal and its associated p-value (extracted via a resampling procedure); the threshold for characteristic taxa was 0.25.

Constrained ordination relates a response matrix $\mathbf{Y}$ (sites $\times$ community abundances) to a matrix $\mathbf{X}$ of explanatory variables (sites $\times$ abiotic factors) and projects this relationship onto a low-dimensional space (generally 2) for interpretation [159]. The two most widely used methods in community ecology are Redundancy Anlysis (RDA) and Canonical Correspondance Analysis (CCA). Both essentially perform multiple linear regression followed by a principal component analysis (PCA), with the most relevant difference being that in CCA abundance data are first chi-square transformed [161]. Although RDA does not require transforming the responses, it is customary to apply a Hellinger transformation is commonly used to improve performance [26, 195]. Generally, RDA should be used when species–environment relationships are approximately linear, which is often the case when the observed environmental gradients are short, while CCA when unimodal responses, which is the most likely situation if a sufficiently long gradient is observed [159].
Spatially structured data can exhibit autocorrelation that disturbs the underline ecological signals [292]. It is therefore necessary to take into account the spatial component and ideally to remove its effect when performing a species-environment regression.

Many methods are available, but one of the best suited for ordination analysis is to use Moran's Eigenvector Maps (MeMs), a set of vectors that can be readily used as additional explanatory variables in a model [159, 47]. MEMs are obtained by *metric multidimensional scaling* (MDS) of a predefined *spatial weighting matrix* $\mathbf{W}$. $\mathbf{W}$ is a symmetric matrix encoding spatial connectivity: smaller values indicate stronger correlations and zero entries indicate no correlation. We computed $\mathbf{W}$ as the matrix of shortest-path geographic distances (km) on the minimum spanning tree (MST) of stations and set as distance threshold maximum MST distance (default behaviour of *dbmem* in the R package *adespatial*). The MST was computed using the function *costDistance* from the R package *gdistance* which allows to account for geographical barriers emerging from the Italian peninsula. MDS yields $n-1$ eigenvectors (n being the number of spatial points) with positive or negative eigenvalues; the general procedure, which we also employ, is to use as covariables only eigenvectors with positive eigenvalues.
Because our interest lies in the environmental signal, we used MeMs as *covariables* in a *partial ordination method* [159, 47]. A partial ordination method is used to remove the effect of additional variables (called covariables) from the true response-covariate relations one is interested in. For both RDA and CCA, a partial model is constructed by substituting the matrix $\mathbf{X}$ with the matrix of the residuals of the explanatory variables over the covariables $\mathbf{X}_{\text{res}|\mathbf{W}}$ leaving $\mathbf{Y}$ as it is.
Among the available variables, we selected those we deemed more relevant for the phytoplankton community: temperature (T), salinity, dissolved oxygen (DO), orthophosphate (PO_4), nitrate (NO_3), ammonia (NH_4), pH, silicate (SiO_4), and the molar ratio NO_3/PO_4 (NP). T in this context is a proxy of seasonality, which play an important role in phytoplankton community as previously mentioned; salinity is prognostic of freshwater input; DO is an indirect measure of photosynthesis, as oxygen is a byproduct of this process; the inorganic nutrients are necessary for phytoplankton, and inorganic P-based and N-based compounds are limiting factors for phytoplankton growth [239]. Since RDA/CCA perform best with normally-distributed predictors, we tested normality (Shapiro–Wilk) and applied one-parameter Box–Cox transformations (`boxcox` from `MASS` package) when needed. A categorical factor *Season* was included to capture seasonal effects not fully represented by T.
The best model was chosen according to how much variance was explained by environmental factors and the seasonal component. The variance of each component (procedure called *variation partioning*) can be determined using the function *varpart* from the *vegan* package. The winning model was then subject to two further refinements: 1) variance inflation factor (VIF, function `vif` from R package `car`) was used to detect multicollinearity among predictors and the heuristic rule that variables with and VIF greatly exceeding 10 were discarded. 2) step-wise variable selection was conducted on environmental variables and season in both forward and backward direction to select

the most relevant variables (`ordiR2step`).
We initially fitted eight models changing: (a) ordination method (RDA vs. CCA); (b) number of genera in the response matrix (24 or 63; the sets defined by IndVal thresholds 0.50 and 0.25, respectively); and (c) whether a log transformation was applied to abundances (Hellinger or chi-square transformations were subsequently applied regardless) to reduce the effect of extreme values which might obscure the overall behaviour of the community. Model performance was compared via variation partitioning (`varpart`, `vegan`) among environment, space (MeMs), and season. The selected model then underwent two refinements: (1) multicollinearity screening by variance inflation factors (VIF; `car`), removing variables with VIF $\gg$ 10; (2) bidirectional stepwise selection of environmental variables and Season (`ordiR2step`).

2.4 Results

2.4.1 General overview of observed taxa

We observed 767 taxa in total, most identified to species (544) or genus (195) level. The most frequent taxa were *Gymnodinium spp.*, *Proboscia alata*, *Chaetoceros spp.*, *Dinoflagellata*, and *Cylindrotheca closterium*, occurring in 62%, 58%, 54%, 53%, and 50% of samples, respectively (Fig. 2.3a). The most frequent classes were *Bacillariophyceae* (99%), *Dinophyceae* (98%), and *Cryptophyceae* (59%) (Fig. 2.3b). Common genera included *Chaetoceros* (75%), *Pseudo-nitzschia* (68%), and *Thalassionema* (68%) (Fig. 2.3c); frequent species included *Proboscia alata* (58%), *Cylindrotheca closterium* (50%), and *Thalassionema nitzschioides* (47%) (Fig. 2.3d). In terms of diversity, *Dinophyceae* and *Bacillariophyceae* were the most species-rich classes (251 and 210 species), far exceeding *Coccolithophyceae* (31) and several classes with seven or fewer species (Fig. 2.4a). At the genus level, *Chaetoceros* was the most diverse (54 species), followed by *Protoperidinium* (34) (Fig. 2.4b). Based on occurrence frequencies, taxa were classified as rare (<1%), intermediate (1–10%), and common (>10%) [238], yielding 462 rare, 228 intermediate, and 77 common taxa.

The relative contribution of the several identification levels (“Species”, “Genus”, “Higher cat.” and three cases of “Unknown”) varied markedly among regions (Fig. 2.6). For example, ABR identified on average ∼90% of abundance at the species level, whereas SIC and SAR left >50% of abundance as “Unknown”. These differences can be imputed to two main factors: varying capabilities in identifying common taxa or the presence of

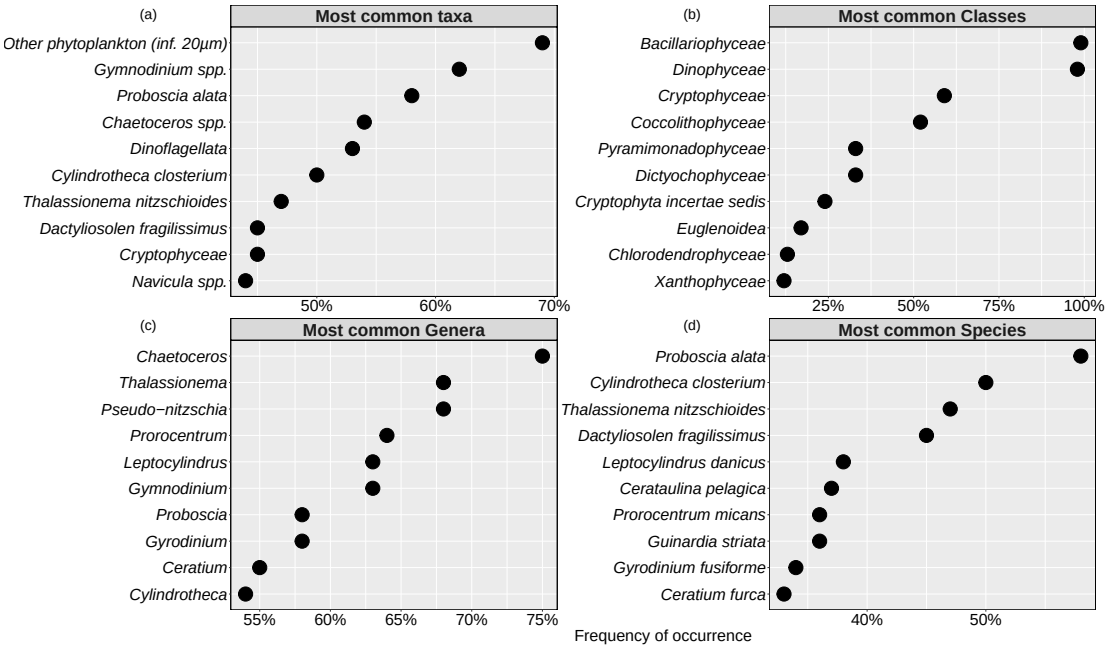


FIGURE 2.3: Top 10 most common taxa (a), classes (b), genera (c), and species (d). On the x-axis the frequency of occurrence is reported.

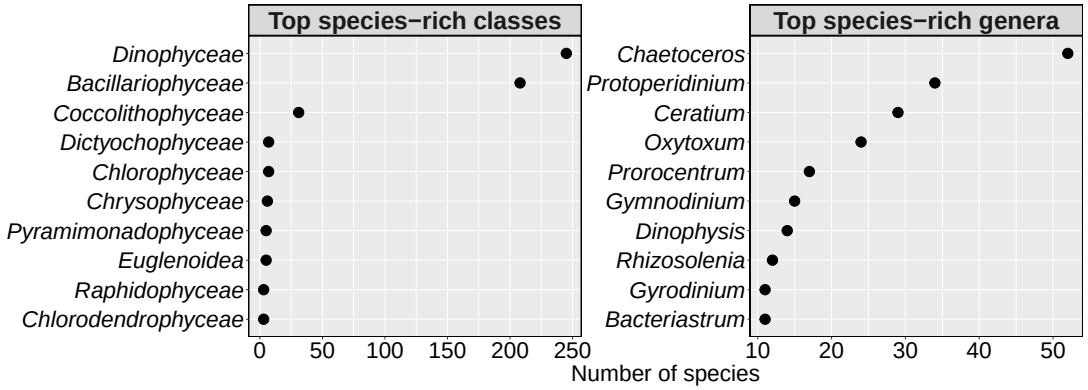


FIGURE 2.4: Top 10 species-rich classes (a) and genera (b). On the x-axis the number of species is reported.

rare species in certain regions, which are harder to recognize.

Despite heterogeneity in identification depth across regions, standardized SACs (50 samples) built from species- or genus-level observations reveal clear richness differences (Fig. 2.7). FVG, EMR, and MAR were estimated to exceed 100 species, whereas VEN was significantly lower; all four were richer than ABR and MOL (both ∼60 species). PUG reached ∼160 species, exceeding neighbouring ABR, MOL, and BAS. CAL and SIC would reach ∼100 species; CAM ∼78 and LAZ ∼92. TOS was the richest region (>200 species), substantially exceeding its Tyrrhenian neighbours (CAM, LAZ, LIG). SAR was the second richest (183 species). Notably, in ABR the very high fraction of

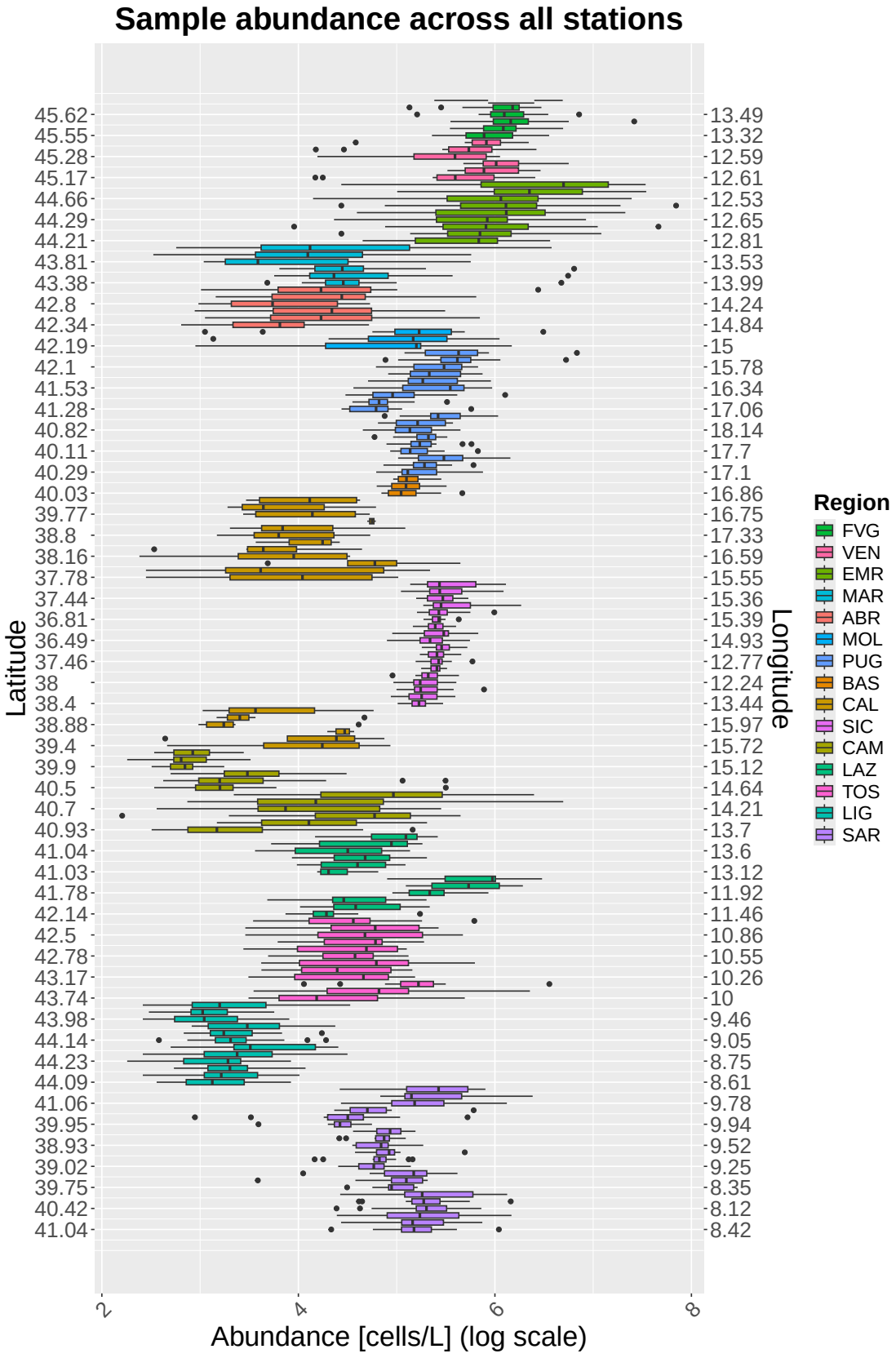


FIGURE 2.5: Spatial distribution of cell abundance (cells/L; log scale) across coastal stations ordered from north-east to north-west, with SAR at the end. Points are grouped by administrative region and plotted by latitude and longitude.

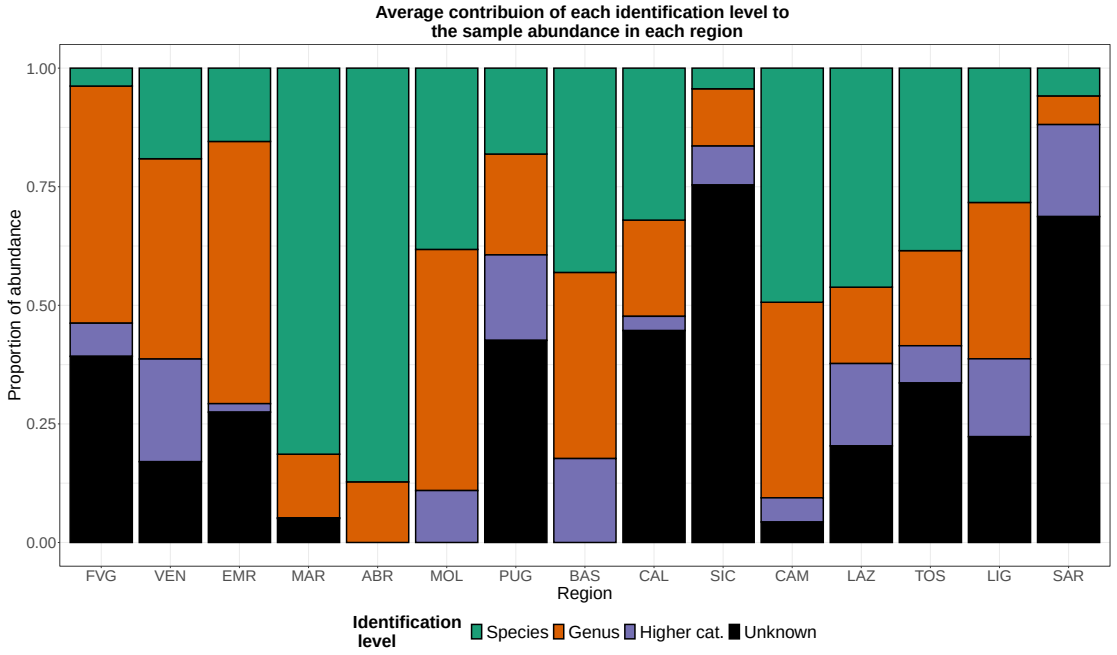


FIGURE 2.6: Average contribution of each identification level to the abundance for each region. The four categories are "Species" (green), "Genus" (orange), "Higher cat." (violet), "Unknown" (black)

species-level identifications does not translate into high richness; its estimate remains well below TOS and SAR, suggesting genuinely lower biodiversity rather than an identification artefact. The same comparison holds for ABR, MOL, and BAS against regions with similar or even lower fractions of species-level identification. TOS appears much richer than LAZ, whereas no firm conclusion can be drawn for LIG given its low percentage of species-level identifications. A plausible contributor to between-region differences is coastline extent: larger areas often encompass greater environmental heterogeneity and thus higher richness, even when sampling effort is standardized. However, its effect is only marginal (See Fig. A.1).

2.4.2 Abundance influenced by local phenomea

Fig. 2.5 shows strong spatial structure in phytoplankton abundance. Regional differences exceeded seasonal variability (ANOVA on log-abundances: $F = 884.5$, $p < 10^{-16}$). In NA (FVG-VEN-EMR), high abundances are consistent with eutrophic conditions driven by multiple river inputs, notably the Po River, whose discharge is advected southward along the western Adriatic by the counter-clockwise circulation [117]. This influence is evident in the sharp southward decrease across EMR stations. Despite the influence of northern inputs, CA exhibits low abundances. To test whether environmental gradients explain this, we ran Mantel tests between region pairs using distance

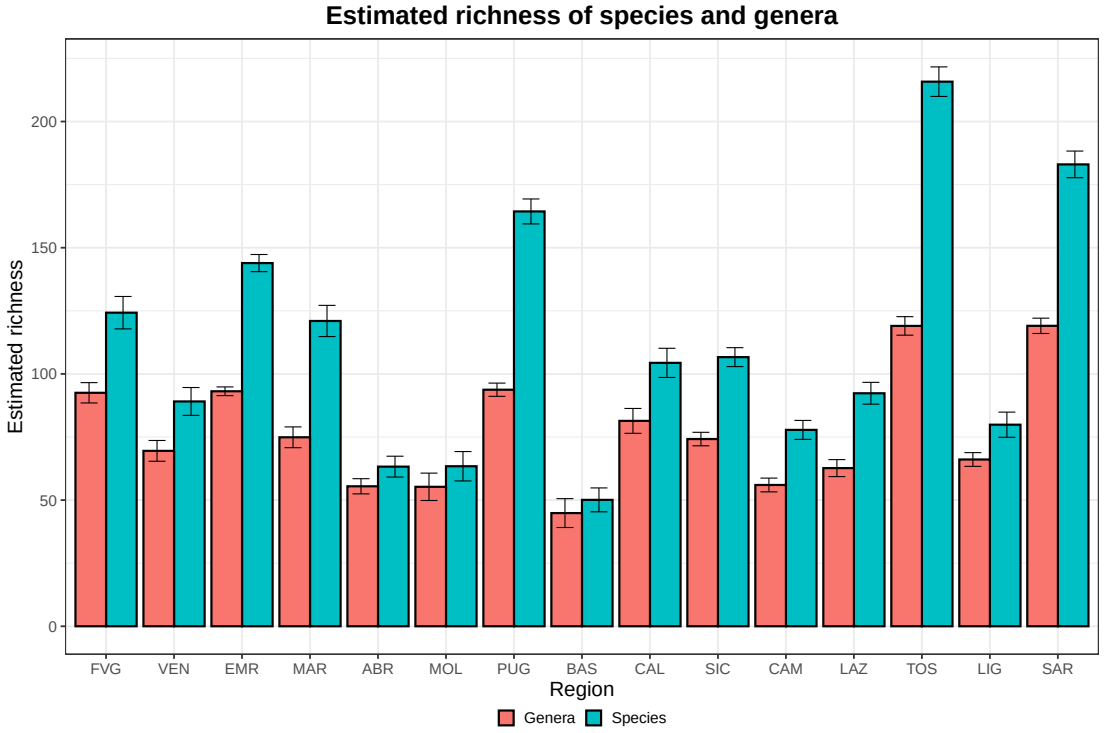


FIGURE 2.7: Estimated species richness and its 95% CI per region at a standardized effort of 50 samples using `iNEXT`

matrices computed on DO, DIN, PO_4, TN, TP, T, pH, and salinity. No significant correlations between environmental similarity and geographic proximity were found for MAR–MOL ($p = 0.24$) and ABR–MOL ($p = 0.48$), suggesting local environmental conditions alone may not account for the patterns. By contrast, SA remains influenced by currents originating from NA [218]. In PUG, nutrient enrichment appears to be driven by the upward movement of LIW, particularly in the Gulf of Taranto [255, 84]. The BAS station in the Ionian Sea is also influenced by local rivers (Bradano, Basento, Agri; see [307]) as indicated by the declining salinity moving offshore. CAL is strongly oligotrophic, with persistently low NO_3, DIN, and PO_4. MAR shows different environmental conditions yet similar abundance levels (Mantel test, $p = 0.63$). Enhanced abundances in SIC might be primarily due to upwelling zones, especially along the south-eastern coast, where cooler, nutrient-rich waters enhance biomass [224, 152, 253]. Additional forcings include local gyres, frontal systems, and the broader effects of Atlantic inflow and Mediterranean surface currents [46, 230, 236, 202]. Although coast-wide studies remain limited [22, 185], available evidence supports high spatial variability and locally elevated productivity [223]. In CAM, the southern transects (Salerno and Cilento, see Fig. A.2) show abundances comparable to or lower than CAL despite higher nutrient concentrations. However, it is not possible to draw conclusions, as most CAL samples from the Tyrrhenian Sea were collected in 2017, a year in which abundances were higher than in 2015 and 2016 (see Fig. 2.5). Indeed, CAL samples from

earlier years show values more consistent with those observed in Salerno and Cilento. Domizio and Napoli, the last two remaining CAM transects, show higher values, likely reflecting proximity to the Volturno and Sarno rivers. A linear model confirms a strong negative relationship between abundance and distance from river mouths ($\beta = -0.076$, $t = -10.04$, $p < 2 \times 10^{-16}$); additionally, the Napoli transect lies within the Gulf of Naples, an area of substantial anthropogenic pressure [45, 293]. In LAZ, phytoplankton abundance peaks at the transect closest to the Tiber mouth, consistent with prior observations of strong riverine nutrient supply [188]. Similarly in TOS, the influence of the Arno is visible, with sharp offshore decreases. LIG appears to be an oligotrophic zone, with nutrient levels comparable to CAL and low overall abundances, due to the absence of major rivers in the region. SAR is also regarded as oligotrophic, although available data remain scarce [31]. The northern SAR sector shows higher abundances than central/southern sectors, without a clear link to nutrient supply, but in the strait of Bonifacio between SAR and Corsica westerly winds create an upwelling zone which sustains high productivity [183]
Overall, complex coastal geometry, currents, river inputs, and upwelling lead to huge fluctuations in phytoplankton abundance that exceed seasonal variability and are not always explained by nutrients alone.

Considering group-level composition across basins and seasons in Fig. 2.8, several patterns emerge. The group that we consider here are *Coccolithophyceae* (COC), *Cryptophyceae* (CRY), *Bacillariophyceae* (DIA), *Dinoflagellata* (DIN), since they are the four most frequent classes, all other classes (ELSE), and unclassified cells split by size when available (UNK, UNK $< 20\mu m$, UNK $> 20\mu m$). In NA, cold seasons (winter, autumn) are characterized by diatom blooms, especially *Skeletonema* (mean: $\sim 3.3 \times 10^6$ cells/L in winter/autumn vs. $\sim 2.8 \times 10^5$ in spring/summer; ANOVA on log-abundances: $F = 8.04$, $p < 0.01$)[1]. The slight summer increase of COC is mainly *Chrysochromulina* (summer $\sim 2.9 \times 10^5$ cells/L vs. other seasons $\sim 1.8 \times 10^5$; $F = 0.72$, $p = 0.39$). In CA, diatoms dominate year-round except in spring, when many unknown cells prevent a determination of the dominant group. The spring decrease in DIA reflects reductions in *Skeletonema* and *Pseudo-nitzschia* (combined spring mean $\sim 4.9 \times 10^3$ vs. $\sim 2.7 \times 10^5$ in other seasons; $F = 6.94$, $p < 0.01$) which have different dynamics. The former is more present in winter and autumn (average: $\sim 7.6 \times 10^5$ cells/L and $\sim 2.5 \times 10^5$ cells/L, respectively), while the latter is more abundant in summer and autumn (average:

[1]For the rest of the paragraph, we will omit to specify all reported F and p values are from ANOVA on log-transformed abundances.

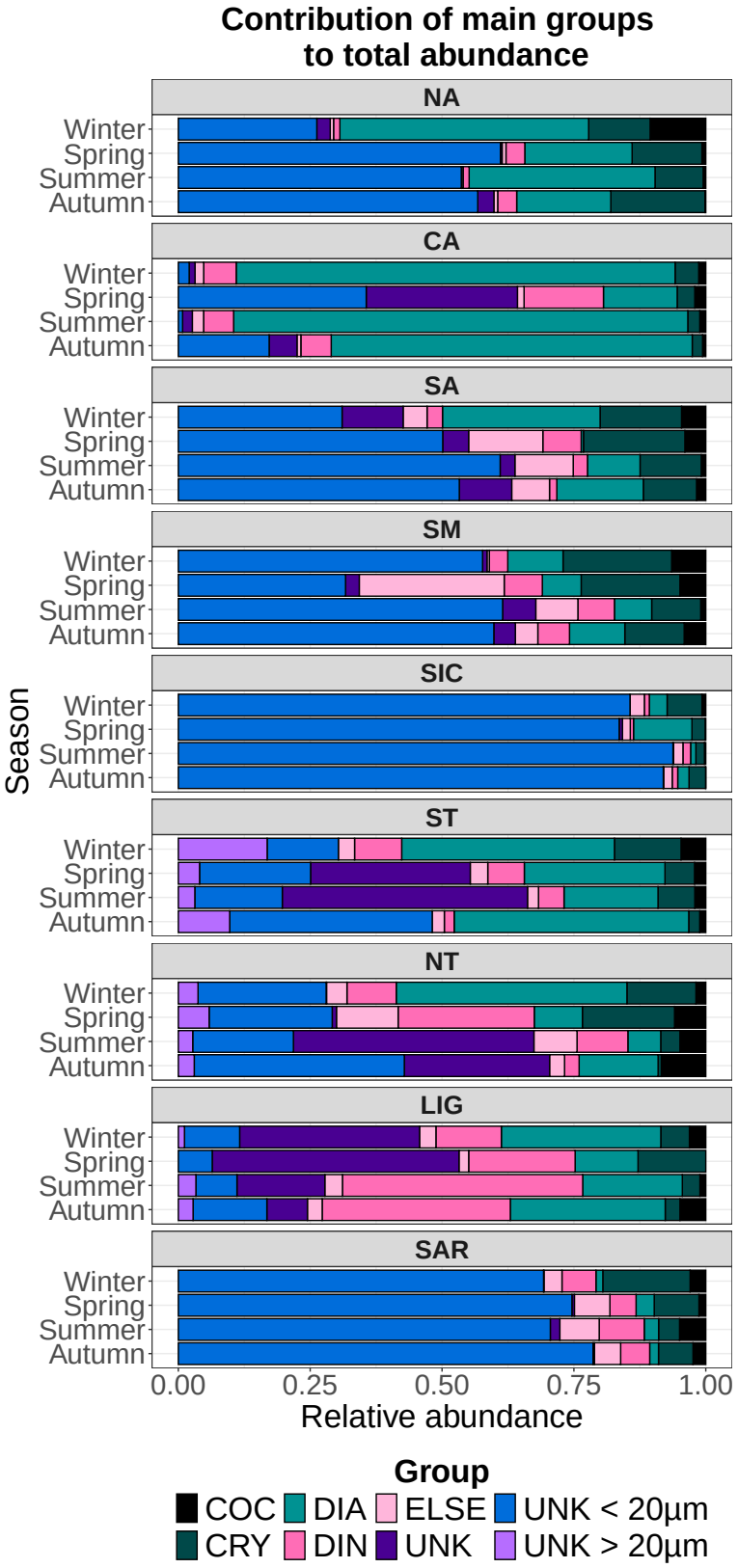


FIGURE 2.8: Average contribution of major phytoplankton groups: *Coccolithophyceae* (COC), *Cryptophyceae* (CRY), *Bacillariophyceae* (DIA), *Dinoflagellata* (DIN), other classes (ELSE), and unclassified cells split by size when available (UNK, UNK $< 20\mu m$, UNK $> 20\mu m$).

$\sim 4.2 \times 10^4$ cells/L and $\sim 2.3 \times 10^4$ cells/L, respectively). In SA, DIA dominates in winter, while CRY and ELSE gain importance outside the cold seasons. *Skeletonema* dominates again in winter (mean $\sim 5.0 \times 10^5$ cells/L vs. ~ 200 in other seasons; $F = 4.19$, $p < 0.05$), and undetermined CRY increase in spring (mean $\sim 5.9 \times 10^4$ vs. $\sim 3.2 \times 10^4$; $F = 7.75$, $p < 0.01$). In SM, patterns are less clear; undetermined CRY are notable in winter (mean $\sim 6.1 \times 10^4$ vs. $\sim 1.7 \times 10^4$; $F = 9.93$, $p < 0.01$), and DIN significantly contributes across seasons. In SIC, DIA modestly increases in spring due to *Skeletonema* (mean $\sim 1.1 \times 10^5$ vs. $\sim 1.9 \times 10^4$; $F = 9.18$, $p < 0.01$), while undetermined CRY rises in winter; however, the zone is difficult to characterize because UNK < 20 μm made up the majority of the abundance. In ST, diatoms consistently dominate throughout the year as a stable group composition. In NT, DIA are reduced in summer due to lower *Chaetoceros*, *Pseudo-nitzschia*, and *Skeletonema* (combined summer mean $\sim 6.4 \times 10^3$ vs. 1.2×10^5; $F = 27.18$, $p < 0.01$); in particular, *Skeletonema* is the most abundant in Spring, while the other two genera are more relevant in the other 3 seasons. CRY increases reflect sporadic *Plagioselmis* blooms (spring mean $\sim 1.5 \times 10^5$ vs. 2.2×10^3; $F = 9.81$, $p < 0.01$). In LIG, DIA dominate in autumn via *Asterionellopsis* (mean $\sim 3.6 \times 10^3$ vs. $\sim 1.1 \times 10^3$; $F = 9.09$, $p < 0.01$), while undetermined DIN are substantially present in summer. For SAR, no strong seasonal or taxonomic signal can be reliably inferred except for a greater abundance of undetermined CRY and *Plagioselmis* in winter/spring (mean 3.2×10^4 vs. 3.3×10^3; $F = 21.70$, $p < 0.01$).

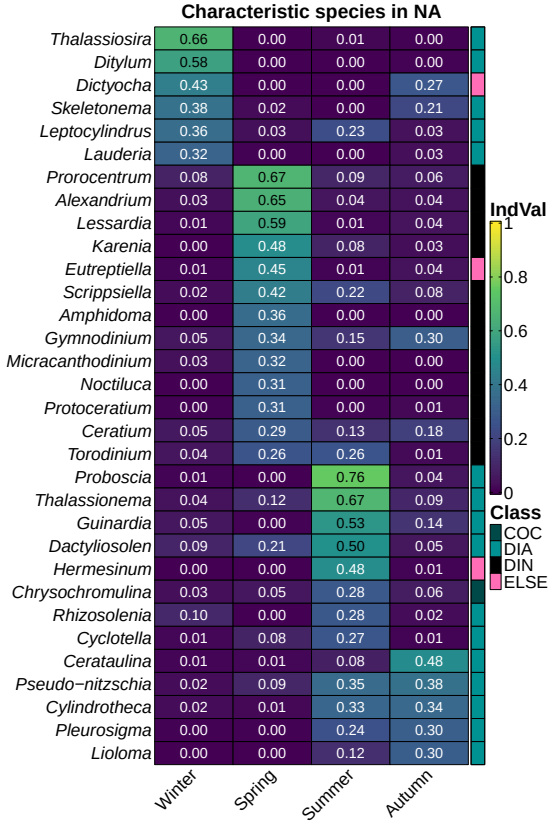
Despite the large fraction of unclassified cells, diatoms appear to be the most abundant group except in SIC and SAR where this cannot be verified. Several diatom genera exhibit seasonal blooms, notably *Skeletonema* and *Chaetoceros* in colder seasons and *Pseudo-nitzschia* in warmer seasons.

2.4.3 Genera characterizing Italian coasts

The IndVal indices (Fig. 2.9) first show that most characteristic genera are diatoms, which is expected given their high frequency and typical abundance dominance. The three Adriatic basins, together with SM, NT, and LIG, display a marked seasonal turnover—i.e., the characterizing genera change from one season to the next. A consistent feature across these basins is that the genera qualifying as characteristic in spring are exclusively dinoflagellates, whereas the other seasons are predominantly characterized by diatoms; only CA is inconclusive in spring because a single genus exceeds the threshold. By contrast, the remaining basins show weaker or irregular seasonal signals. In SIC, almost all characteristic genera occur in spring, with none in summer or autumn; this may reflect the large fraction of unclassified cells in SIC samples outside spring,

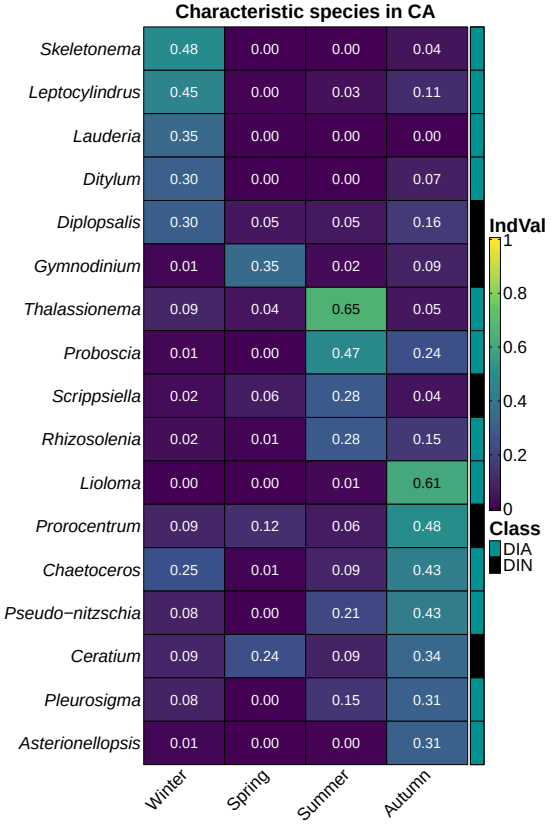
when we previously noted a rise in diatoms in Sec. 2.4.2. ST shows only five characteristic genera, likely because it aggregates stations from CAL and CAM, which have markedly different abundance levels. SAR exhibits a less pronounced turnover, with several genera showing similar IndVal values across seasons (e.g., *Leucocryptos*, *Ceratium*, *Oxytoxum*); also for this basins a more pronounced signal might be obscured by large fraction of unclassified cells.

It is also interesting to look at how many basins each genus characterise in Tab. 2.2. *Thalassionema* is characteristic of most basins (consistent with its being the second most frequently observed genus; Fig. 2.3), and is absent only from ST (a poorly characterized basin) and SAR. Next in cross-basin frequency are *Ceratium* and *Cylindrotheca* (six basins each), followed by *Leptocylindrus*, *Prorocentrum*, and *Pseudo-nitzschia* (five basins), and *Gymnodinium*, *Proboscia*, and *Chaetoceros* (four basins). This prominence is unsurprising given that these genera also rank among the most frequently observed (Fig. 2.3). Seasonality varies by genus and basin. *Thalassionema* characterizes summer in the Adriatic, autumn in SM, spring in SIC, and winter in NT–LIG. This indicates a summer association, while remaining prominent in adjacent seasons depending on basin. *Ceratium* is consistently a spring genus, except in CA where it is autumnal, indicating a preference for mild conditions. *Cylindrotheca* is winter-characteristic in SM, NT, and LIG, spring-characteristic in SAR and SIC, and autumnal in NA, indicating an overall affinity for colder months. *Dictyocha* instead is purely characterizing winter. *Rhizosolenia* shows no single seasonal fingerprint across basins. *Leptocylindrus* is winter-characteristic in the Adriatic but spring-characteristic in SIC, ST, and NT. *Prorocentrum* varies (spring in NA and SAR, autumn in CA, and winter in SA), meaning that its species present along the Italian coasts might have different seasonal cycles. *Pseudo-nitzschia* is predominantly autumnal (NA, CA, SA, SM), but spring in SIC and winter in LIG. *Gymnodinium* is consistently spring-characteristic in the Adriatic and NT. *Proboscia* peaks in summer in the Adriatic but in winter in SIC. *Chaetoceros* appears in different seasons across basins (CA autumn; SA winter; SM summer; SIC spring), consistent with its ubiquity (Fig. 2.3).



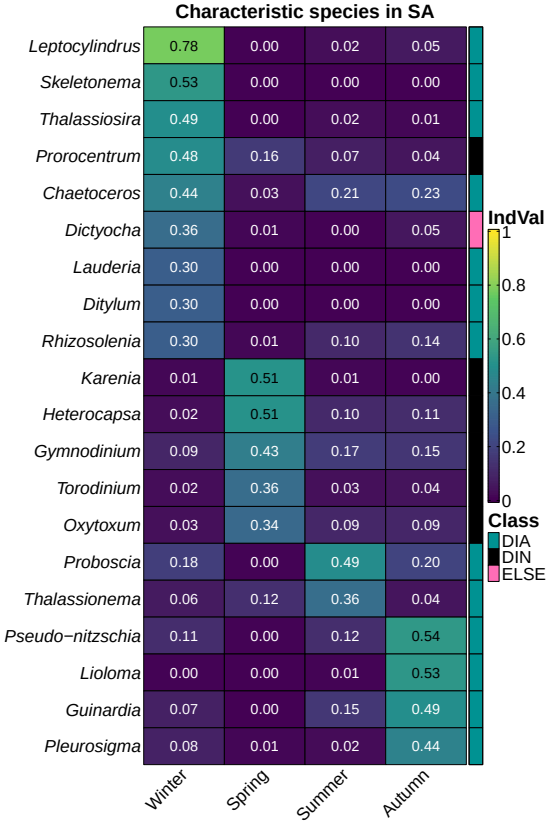
(A) IndVal for genera in NA

FIGURE 2.9: Characteristic genera (statistically significant IndVal > 0.25) across seasons for NA and CA. The rightmost vertical bar indicates which class each genus belongs to: *Bacillariophyceae* (DIAT, light green), *Coccolithophyceae* (COC, dark green) *Dinophyceae* (DIN, black), *Cryptophyceae* (CRY, light pink), or ELSE (dark pink). (Continues.)



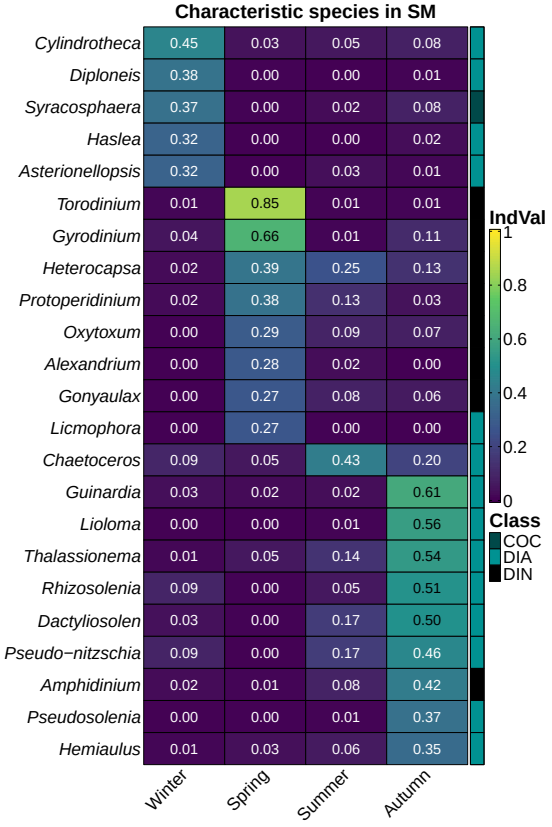
(B) IndVal for genera in CA

FIGURE 2.9: Characteristic genera (statistically significant IndVal > 0.25) across seasons for NA and CA. The rightmost vertical bar indicates which class each genus belongs to: *Bacillariophyceae* (DIAT, light green), *Coccolithophyceae* (COC, dark green) *Dinophyceae* (DIN, black),*Cryptophyceae* (CRY, light pink), or ELSE (dark pink). (Continues.)



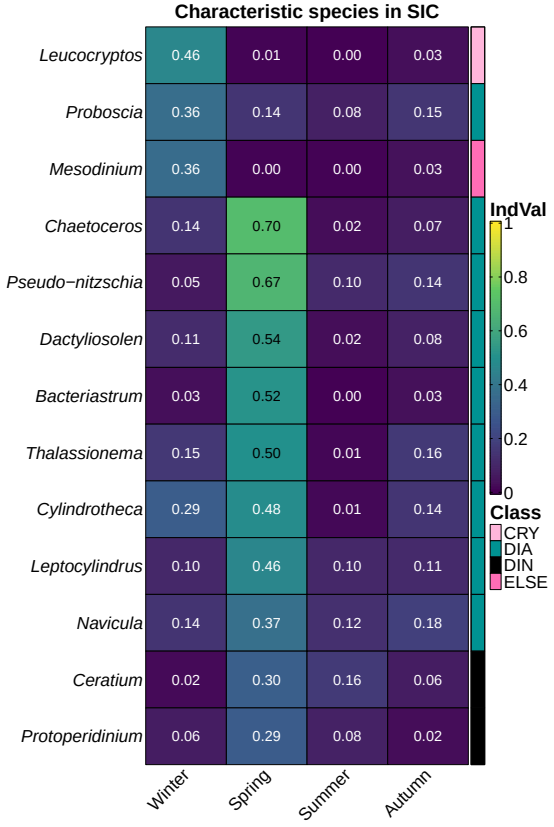
(C) IndVal for genera in SA

FIGURE 2.9: Characteristic genera (statistically significant IndVal > 0.25) across seasons for NA and CA. The rightmost vertical bar indicates which class each genus belongs to: *Bacillariophyceae* (DIAT, light green), *Coccolithophyceae* (COC, dark green) *Dinophyceae* (DIN, black),*Cryptophyceae* (CRY, light pink), or ELSE (dark pink). (Continues.)



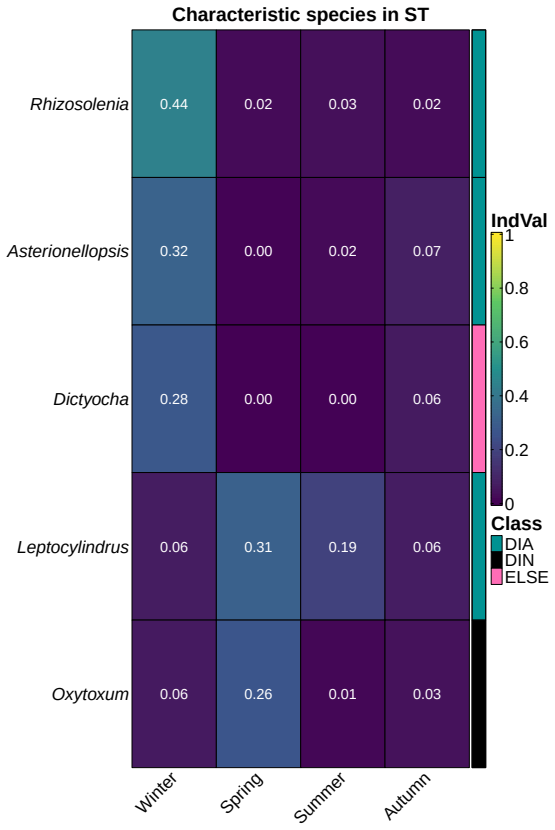
(D) IndVal for genera in SM

FIGURE 2.9: Characteristic genera (statistically significant IndVal $>$ 0.25) across seasons for SM, SIC, ST, NT, LIG, and SAR.TThe rightmost vertical bar indicates which class each genus belongs to: *Bacillariophyceae* (DIAT, light green), *Coccolithophyceae* (COC, dark green) *Dinophyceae* (DIN, black),*Cryptophyceae* (CRY, light pink), or ELSE (dark pink). (Continues.)



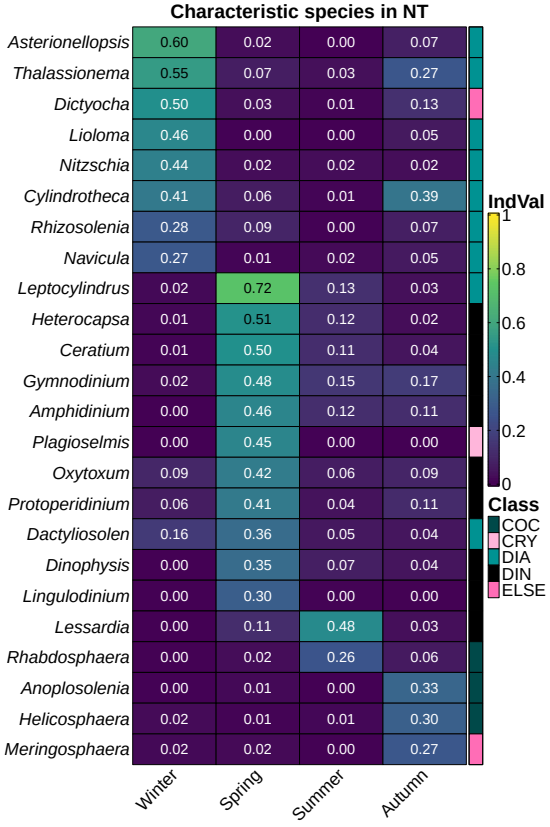
(E) IndVal for genera in SIC

FIGURE 2.9: Characteristic genera (statistically significant IndVal > 0.25) across seasons for SM, SIC, ST, NT, LIG, and SAR. The rightmost vertical bar indicates which class each genus belongs to: *Bacillariophyceae* (DIAT, light green), *Coccolithophyceae* (COC, dark green) *Dinophyceae* (DIN, black),*Cryptophyceae* (CRY, light pink), or ELSE (dark pink). (Continues.)



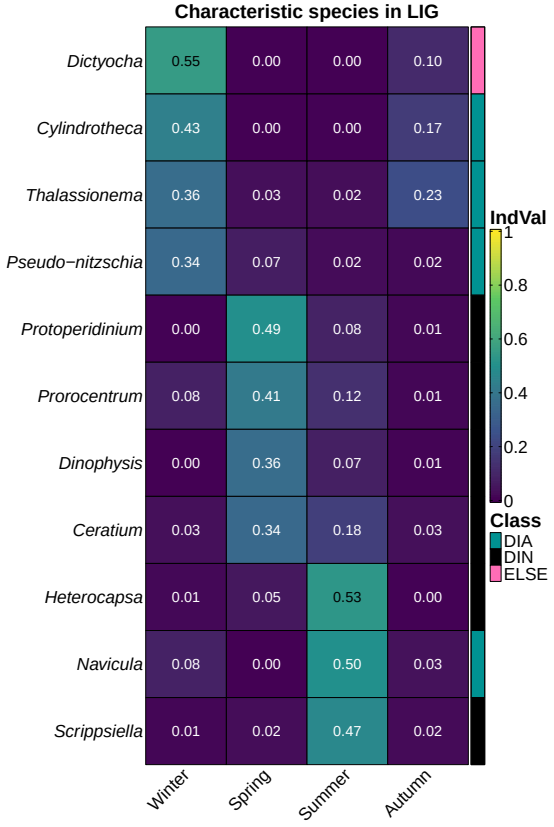
(F) IndVal for genera in ST

FIGURE 2.9: Characteristic genera (statistically significant IndVal > 0.25) across seasons for SM, SIC, ST, NT, LIG, and SAR. The rightmost vertical bar indicates which class each genus belongs to: *Bacillariophyceae* (DIAT, light green), *Coccolithophyceae* (COC, dark green) *Dinophyceae* (DIN, black),*Cryptophyceae* (CRY, light pink), or ELSE (dark pink). (Continues.)



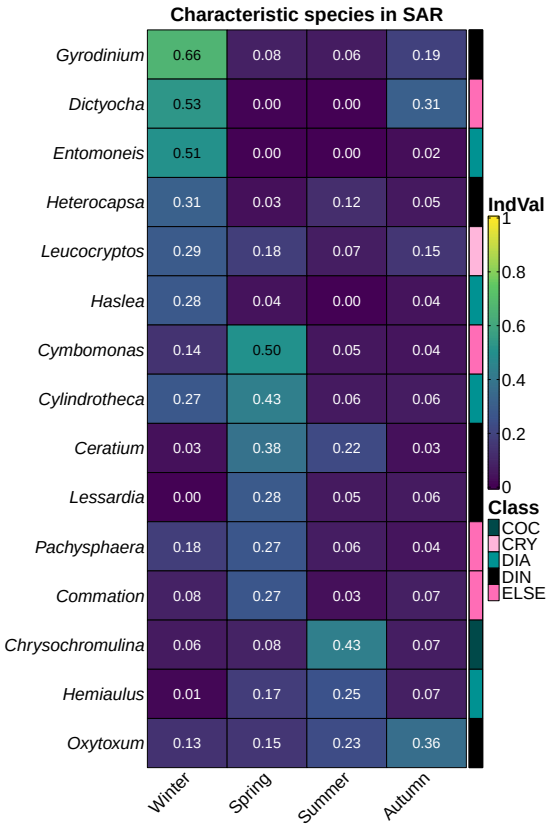
(G) IndVal for genera in NT

FIGURE 2.9: Characteristic genera (statistically significant IndVal > 0.25) across seasons for SM, SIC, ST, NT, LIG, and SAR. The rightmost vertical bar indicates which class each genus belongs to: *Bacillariophyceae* (DIAT, light green), *Coccolithophyceae* (COC, dark green) *Dinophyceae* (DIN, black),*Cryptophyceae* (CRY, light pink), or ELSE (dark pink). (Continues.)



(H) IndVal for genera in LIG

FIGURE 2.9: Characteristic genera (statistically significant IndVal > 0.25) across seasons for SM, SIC, ST, NT, LIG, and SAR. The rightmost vertical bar indicates which class each genus belongs to: *Bacillariophyceae* (DIAT, light green), *Coccolithophyceae* (COC, dark green) *Dinophyceae* (DIN, black),*Cryptophyceae* (CRY, light pink), or ELSE (dark pink).. (Continues.)



(I) IndVal for genera in SAR

FIGURE 2.9: Characteristic genera (statistically significant IndVal > 0.25) across seasons for SM, SIC, ST, NT, LIG, and SAR. The rightmost vertical bar indicates which class each genus belongs to: *Bacillariophyceae* (DIAT, light green), *Coccolithophyceae* (COC, dark green) *Dinophyceae* (DIN, black),*Cryptophyceae* (CRY, light pink), or ELSE (dark pink).

2.4.4 Dominant species in Italy across seasons

To aid interpretation of characteristic genera, we plotted for each basin and season either the 10 characteristic genera with highest mean abundance or those cumulatively contributing up to 95% of community abundance (Fig. 2.10). In Winter (Fig. 2.10a), the whole Adriatic is dominated by *Skeletonema* and minimally by *Thalassiosira* in NA, and by *Leptocylindrus* and *Chaetoceros* in SA. SM is more diverse as it includes *Chaetoceros*, *Pseudo-nitzschia*, *Asterionellopsis* and *Cylindrotheca* and other genera with marginal contributions. Same happens in SIC, though the most relevant genera are *Chaetoceros*, *Leucocryptos*, *Pseudo-nitzschia* and *Leptocylindrus*. ST is poorly characterized and so only two genera appear with *Asterionellopsis* being the most abundant. *Asterionellopsis* dominates also in the NT, though other genera have a non-negligible contributions: *Thalassionema*, *Cylindrotheca*, *Leptocylindrus* and *Dactyliosolen*. In LIG we have *Pseudo-nitzschia* as dominant genera, followed by *Cylindrotheca*, *Thalassionema* and *Dictyocha*. Finally, in

Genus	Abbr.	Frequency
Thalassionema	Tham	7
Ceratium	Cera	6
Cylindrotheca	Cyli	6
Dictyocha	Dict	6
Leptocylindrus	Lept	6
Pseudo-nitzschia	Pseu	6
Rhizosolenia	Rhiz	6
Heterocapsa	Hete	5
Lioloma	Liol	5
Oxytoxum	Oxyt	5
Asterionellopsis	Aste	4
Chaetoceros	Chae	4
Dactyliosolen	Dact	4
Gymnodinium	Gymn	4
Proboscia	Prob	4
Prorocentrum	Pror	4
Protoperidinium	Prot	4
Ditylum	Dity	3
Guinardia	Guin	3
Lauderia	Laud	3
Lessardia	Less	3
Navicula	Navi	3
Pleurosigma	Pleu	3
Scrippsiella	Scri	3
Skeletonema	Skel	3
Torodinium	Toro	3
Alexandrium	Alex	2
Amphidinium	Amph	2
Chrysochromulina	Chry	2
Dinophysis	Dino	2
Gyrodinium	Gyro	2
Haslea	Hasl	2

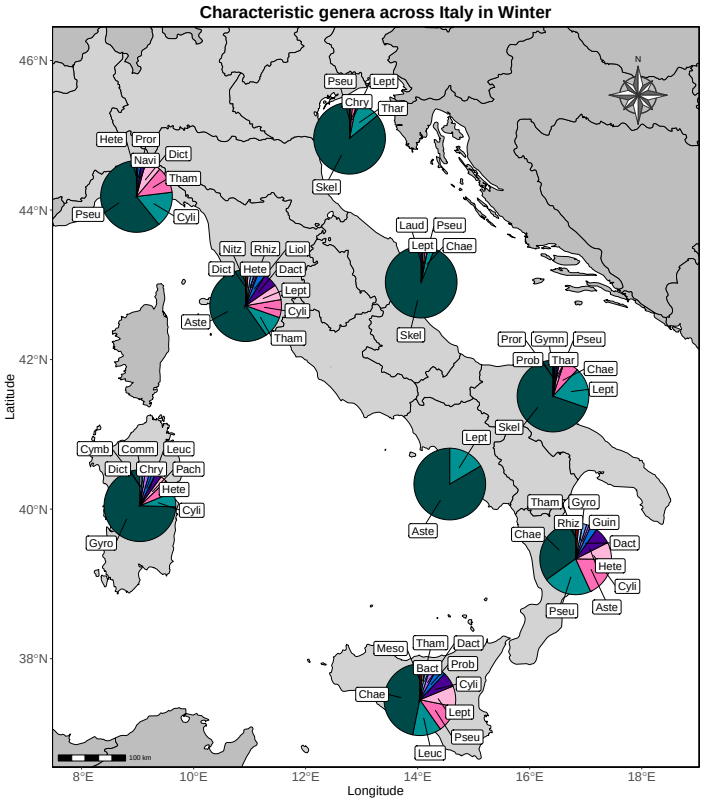
Genus	Abbr.	Frequency
Hemiaulus	Hemi	2
Karenia	Kare	2
Leucocryptos	Leuc	2
Thalassiosira	Thar	2
Amphidoma	Amph	1
Anoplosolenia	Anop	1
Bacteriastrum	Bact	1
Cerataulina	Cera	1
Commation	Comm	1
Cyclotella	Cycl	1
Cymbomonas	Cymb	1
Diploneis	Dipl	1
Diplopsalis	Dipl	1
Entomoneis	Ento	1
Eutreptiella	Eutr	1
Gonyaulax	Gony	1
Helicosphaera	Heli	1
Hermesinum	Herm	1
Licmophora	Licm	1
Lingulodinium	Ling	1
Meringosphaera	Meri	1
Mesodinium	Meso	1
Micracanthodinium	Micr	1
Nitzschia	Nitz	1
Noctiluca	Noct	1
Pachysphaera	Pach	1
Plagioselmis	Plag	1
Protoceratium	Prot	1
Pseudosolenia	Pseu	1
Rhabdosphaera	Rhab	1
Syracosphaera	Syra	1

TABLE 2.2: Characteristic genera, abbreviation of their names, and number of basins they are characteristic of (Frequency).

SAR *Gyrodinium* is the dominant genera with *Cylindrotheca* and *Heterocapsa* as two other main contributors.

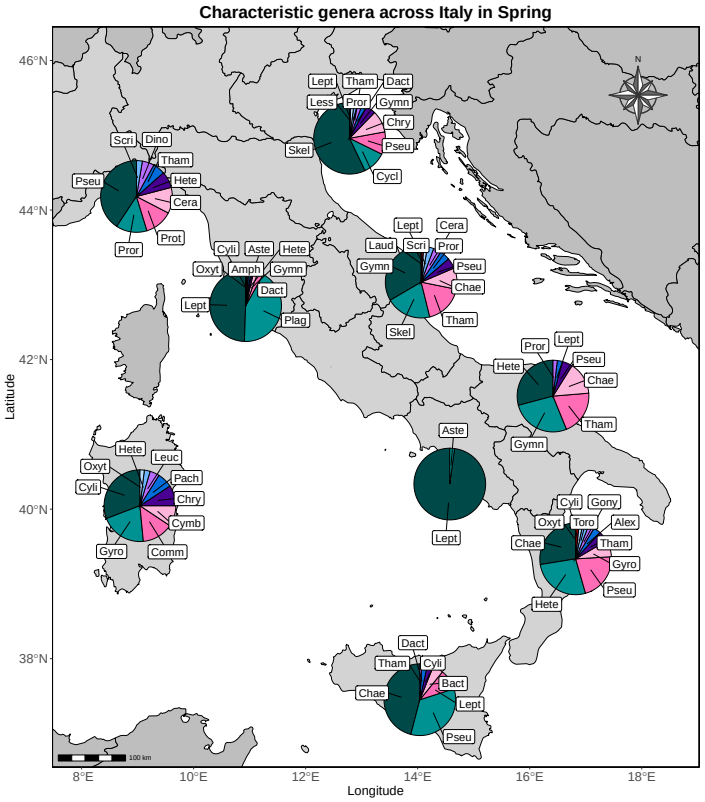
In spring (Fig. 2.10b), we observe changes in community compositions. NA remains *Skeletonema*-dominated but with increased *Cyclotella*, *Pseudo-nitzschia*, *Chrysochromulina*. In CA, abudances are more evenly distributed between *Gymnodinium*, *Skeletonema*, *Thalassiosira*, *Chaetoceros* and *Pseudo-nitzschia*. Compared to winter, SA community shifts to *Heterocapsa*, *Gymnodinium*, *Thalassionema*, *Chaetoceros*, *Pseudo-nitzschia*. SM remains dominated by *Chaetoceros* and *Pseudo-nitzschia*, with *Heterocapsa* important. In ST we

have a complete change as *Leptocylindrus* becomes the only relevant genera to the detriment of *Asterionellopsis*. Also NT shows a different community which is now mainly dominated by *Leptocylindrus* and *Plagioselmis*. LIG retains *Pseudo-nitzschia* but with notable contribution from *Prorocentrum*, *Protoperidinium*, *Ceratium*, and *Heterocapsa*. SAR becomes more even: *Gyrodinium* and *Cylindrotheca* have comparable abundances and are followed by *Commation*, *Cymbomonas*, *Chrysochromulina* and *Pachysphaera*.
In summer (Fig. 2.10c), *Skeletonema* nearly disappears from the Adriatic, replaced by *Pseudo-nitzschia*, *Thalassiosema*, and *Chaetoceros* (plus *Chrysochromulina* and *Cyclotella* in NA and *Gymnodinium* and *Proboscia* in SA). The southern zone remains dominated by *Pseudo-nitzschia* and *Chaetoceros*, with the addition of *Heterocapsa* in SM and *Leptocylindrus* in SIC. The Tyrrhenian is dominated almost entirely by *Leptocylindrus*(with *Asterionellopsis* still relevant in ST and a richer mix in NT). In LIG, characteristic species of summer change: now *Heterocapsa* dominates, it is followed by *Pseudo-nitzschia*, which was relevant also in the previous season, and by *Navicula* and *Scrippsiella* which instead are characteristic of this season. In SAR we observe *Chrysochromulina* as the new dominant genus together with *Heterocapsa* which gained more importance, while *Gyrodinium* retains its relative contribution and *Cylindrotheca* losses importance.
In autumn (Fig. 2.10d), *Skeletonema* reappears as dominant in NA and CA, while *Pseudo-nitzschia* and *Chaetoceros* remain important across the entire Adriatic. SM qualitatively retains its community composition, with reduced relative contribution of *Chaetoceros* and increased contribution of *Dactyliosolen*. Similarly, SIC remains dominated by *Pseudo-nitzschia*, *Chaetoceros*, *Leptocylindrus*. In the Tyrrhenian, *Leptocylindrus* and *Asterionellopsis* equally contribute, and again *Cylindrotheca*, *Gymnodinium* and *Thalassionema* are also present in NT. LIG community shifts to a dominance of *Cylindrotheca*, *Thalassionema*, *Pseudo-nitzschia*. In SAR *Gyrodinium* and *Chrysochromulina* are still the most relevant, with the former being now the most abundant.
From these four panels it is possible to extract general spatio-temporal patterns in the phytoplankton comminity along Italian coast. Especially evident is the succession of genera going from colder to warmer seasons. *Skeletonema* almost entirely dominates the Adriatic during cold months and is gradually replaced by genera more acclimated to warmer seasons such as *Gymnodinium*, *Thalassionema* and *Heterocapsa*; *Pseudo-nitzschia* blooms in summer and persists into autumn; *Chaetoceros* is important year-round (especially spring–summer). In the southern zone, *Chaetoceros* and *Pseudo-nitzschia* occur year-round with fluctuating contributions that do not exhibit a consistent seasonal pattern. In the Tyrrhenian, there is a pronounced succession between *Asterionellopsis* during cold-water periods and *Leptocylindrus* during warm-water periods. In SAR, *Gymnodinium* is present throughout the year but peaks during colder months. A repeatable seasonal succession is observed: *Gyrodinium* dominates in winter, *Cylindrotheca* and *Commation* in spring, *Chrysochromulina* in summer, and *Gyrodinium* again in autum



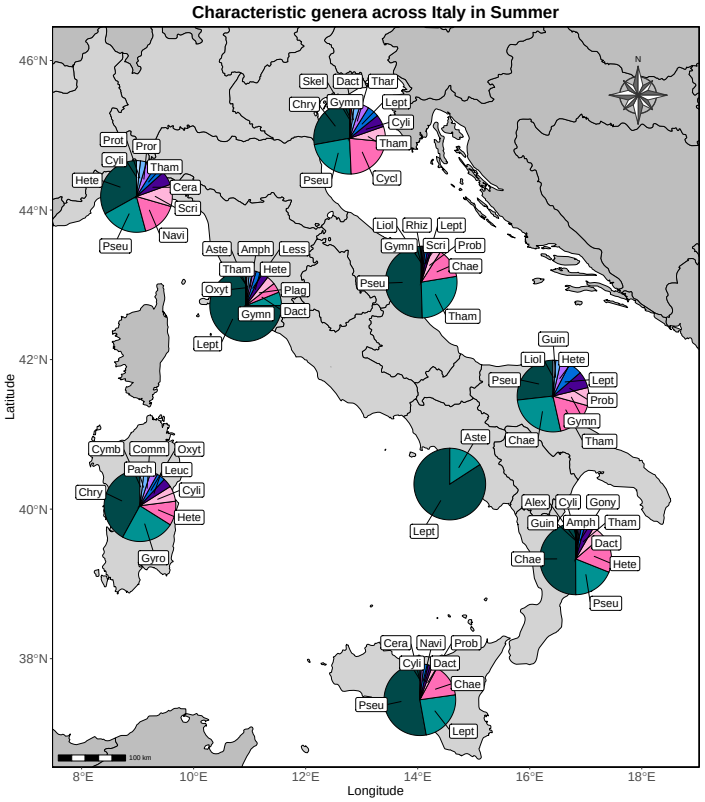
(A) Characteristic genera across Italy in Winter

FIGURE 2.10: Contribution of the most abundant characteristic genera in each basin in Winter (a), Spring (b), Summer (c) and Autumn (d). Colours are only used for visual aid. Abbreviation can be consulted in Tab. 2.2. (Continues on the following pages)



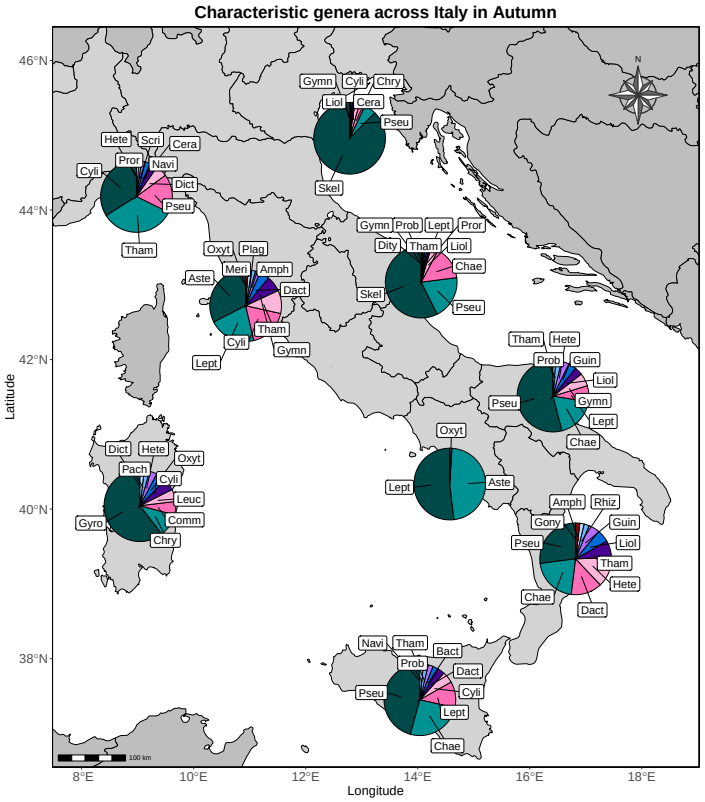
(B) Characteristic genera across Italy in Spring

FIGURE 2.10: Contribution of the most abundant characteristic genera in each basin in Winter (a), Spring (b), Summer (c) and Autumn (d). Colours are only used for visual aid. Abbreviation can be consulted in Tab. 2.2. (Continues on the following pages)



(C) Characteristic genera across Italy in Summer

FIGURE 2.10: Contribution of the most abundant characteristic genera in each basin in Winter (a), Spring (b), Summer (c) and Autumn (d). Colours are only used for visual aid. Abbreviation can be consulted in Tab. 2.2. (Continues on the following pages)



(D) Characteristic genera across Italy in Autumn

FIGURE 2.10: Contribution of the most abundant characteristic genera in each basin in Winter (a), Spring (b), Summer (c) and Autumn (d). Colours are only used for visual aid. Abbreviation can be consulted in Tab. 2.2.

2.4.5 Multivariate regression of communities over environmental variables

Variation partitioning for all models outlined in Sec. 2.3 is reported in Table 2.3. All models explain but a small fraction of the total variance ranging from $\approx$ 22% in the RDA models with 24 genera to the $\sim$ 25% of the CCA models with 24 genera. The robustness of the results suggests either highly nonlinear or non-unimodal species–environment relationships not captured by RDA and CCA, or the influence of unmeasured drivers such as iron supply [131, 132] or stratification [312]). Space (MeMs) explained the largest fraction of variance in all models excluding CCA with 24 genera and no log-transformation. This is consistent with strong basin-scale differences discussed above. Environmental variables by themselves explain $\approx$ 3%, slightly exceeding the seasonal component ($\approx$ 2%). Strong interactions between environment and space (MeMs + Env) and environment and seasonality (Env + Season) are notable. This is expected as the extreme heterogeneity of the Italian coast influence the distribution of nutrients (e.g., Po river) and variables as temperature have a strong seasonal signal.

Although the best raw performance was obtained by CCA with 24 genera and log-transformed abundances, we selected the corresponding RDA because its triplot was easier to interpret at the cost of only 4% less explained variance and CCA triplot and axes have the same interpretation as RDA. After VIF screening and stepwise selection, retained predictors were T, salinity, DO, PO_4, NO_3, pH, SiO_4, and Season. In Fig. 2.11 we have displayed the RDA triplot with covariates (top panel) and genera (bottom panel). Though the first two axes (RDA1 and RDA2) explain only $\approx 8\%$ of the total variability we can see how samples clearly separate in 4 groups based on the season. Temperature and salinity are highly correlated and point toward summer samples, which is consistent with the increase of evaporation rate and subsequently of salinity in summer due to the increasing heat [270]. Nutrients, DO and pH instead points toward winter samples and are mutually positively correlated and negatively correlated with T and salinity. More nutrients are expected in winter as the increased water mixing brings to the surface nutrients stored in deeper layers [317] while an increase in DO can be the result of increased productivity. Environmental loadings in Fig. 2.12 display the correlation of each environmental variable with the two RDA axes. T and salinity positively correlated with RDA1 (with T nearly collinear), while nutrients are negatively correlated (with PO_4 nearly collinear but opposite in sign). These relations remain valid for RDA2, the differences being that salinity is now more collinear with it compared to T and the correlations of all nutrients are slightly weaker. RDA1 divides warm and salty water masses generally observed in summer from their opposite fresher and colder water masses typical of winter; the negative correlation with nutrients might also suggest that RDA1 follows an oligotrophic gradient. RDA2 instead divides autumn and spring water masses; autumn water masses are saltier as a result of the strong stratification in summer while spring waters are mixed and richer in nutrients. The positioning of genera arrows indicates seasonal succession consistent with the IndVal analysis: winter genera (*Thalassiosira*, *Skeletonema*, *Ditylum*) align with NO_3 and SiO_4; cold-water genera (*Dictyocha*, *Chaetoceros*, *Asterionellopsis*) align more with PO_4; autumn indicators (*Lioloma*, *Leptocylindrus*, *Guinardia*) align with higher salinity; summer genera (*Proboscia*, *Torodinium*) point toward warmer, nutrient-poor waters; and dinoflagellates (*Karenia*, *Heterocapsa*, *Prorocentrum*, *Alexandrium*) point toward spring waters.

2.5 Discussion and conclusion

Despite heterogeneous sampling among regions, several robust, process–based patterns emerge. Total phytoplankton abundance shows a strong basin-scale structure

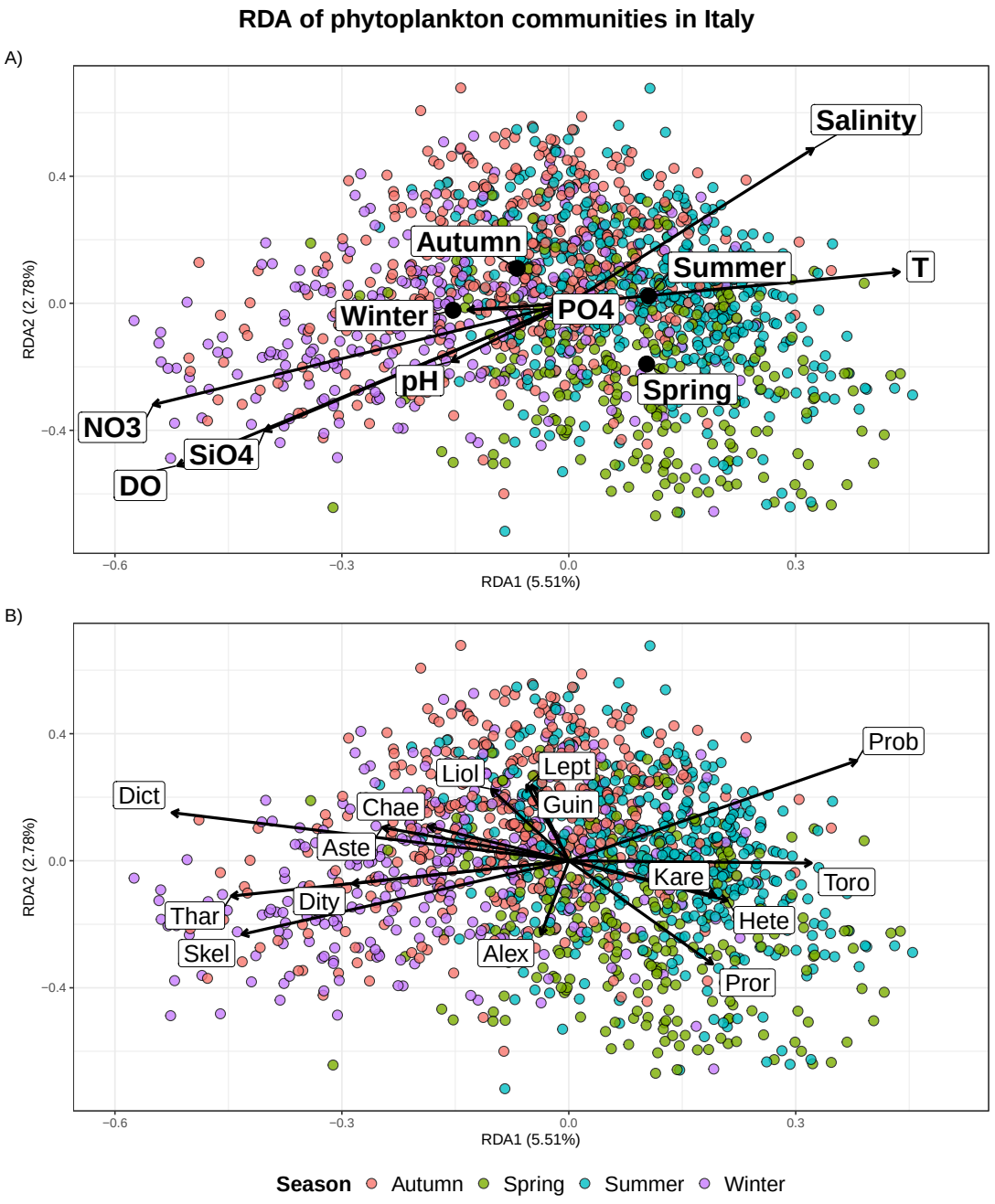


FIGURE 2.11: RDA triplots using 24 genera with log-transformed abundances after VIF screening and stepwise selection. Points are samples, coloured by season. For readability, explanatory variables and genera are shown in separate panels (A and B). Only 15 of 24 genera are displayed (see Table 2.2 for names).

	model	RDA				CCA			
	No. of genera	24		63		24		63	
	log	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
Term (%)	MeMs	9.84	11.49	11.26	13.01	11.4	5.52	12.1	8.54
	Env	2.74	3.39	2.24	3.05	3.72	8.37	2.84	6.08
	Season	2.16	1.83	1.81	1.61	2.37	1.5	1.95	1.41
	MeMs + Env	3.58	3.07	3.62	2.97	4.03	1.7	3.87	1.69
	Envs + Season	3.05	2.11	2.31	1.75	4.25	6.54	2.39	3.53
	MeMs + Season	0.05	0.06	0.06	0	0.13	0.1	0	
	Mems + Env + Season	0	0	0	0	0	1.49	0	1.64
	Residuals	78.69	78.36	78.73	77.84	74.27	74.75	76.82	77.17

TABLE 2.3: Variation partitioning for all ordination models. Rows report the variance explained by each component and their interactions; the remainder is residual.

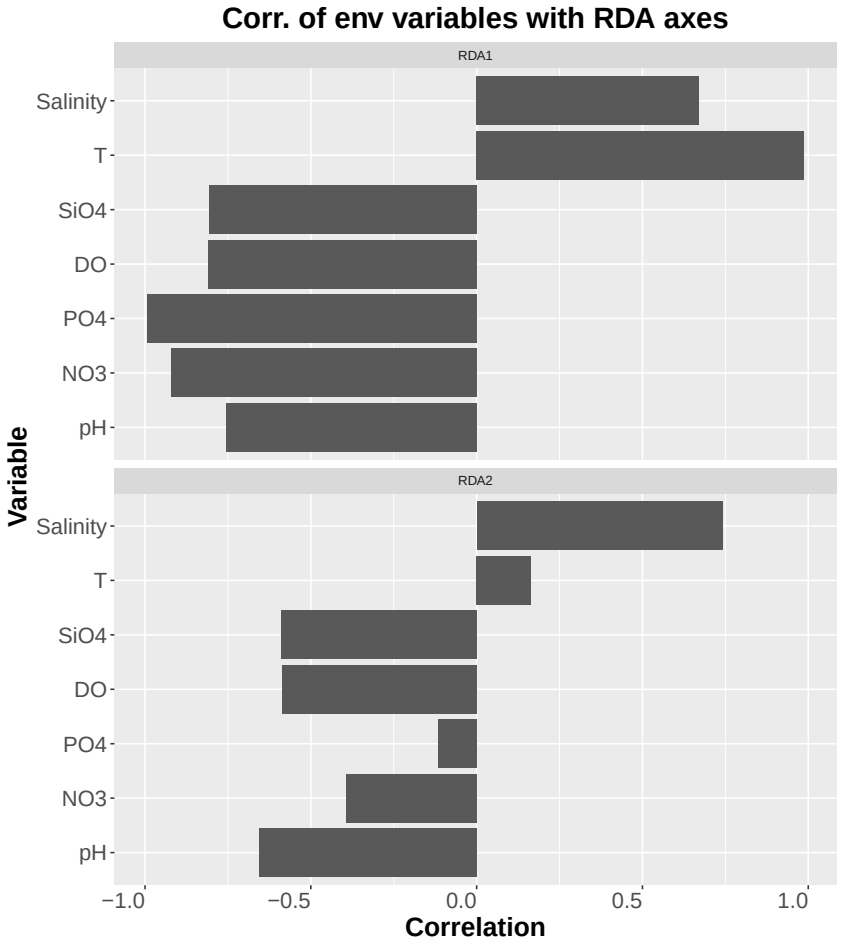


FIGURE 2.12: Correlation of step-wise-selected environmental variables with the first two RDA axes.

that exceeds seasonal variability. This is consistent with hydrographic forcing: Po-driven eutrophication in the northern Adriatic, local upwelling (e.g., along parts of SIC), and mesoscale features (fronts, gyres) that locally enhance nutrient supply; in contrast, oligotrophic sectors (CAL, LIG, much of SAR) exhibit lower abundances. The negative abundance–distance relationships near major river mouths (Volturno, Sarno, Tiber, Arno) support the role of nearshore nutrient injections. That said, simple chemistry–abundance correspondence does not always hold (e.g., EMR vs. MAR), and station layout likely contributes: some transects deliberately face large rivers (Po, Tiber, Arno), whereas others (e.g., in CAL) are far from persistent sources, potentially inflating among-region contrasts. Standardized species accumulation curves show large differences in potential richness. TOS and SAR are the richest regions, whereas ABR, MOL, and BAS are comparatively poor. Crucially, ABR's high fraction of species-level identifications does not inflate richness, arguing against a taxonomic artefact and supporting genuinely lower diversity. VEN estimation yields fewer species and genera than FVG and EMR despite the high eutrophic status of the north Adriatic and the positioning of the station near the Venice lagoon and the Adige river; this might be related to issues in the identification. Coastline extent and associated environmental heterogeneity likely contribute to these contrasts [274] but cannot fully explain them (e.g., TOS vs. PUG). However, these results should be interpreted with caution. Although species accumulation curves can reduce biases arising from unequal sampling effort, they cannot correct for differences in the taxonomic expertise of the analysts who processed the samples. With the available data, it is impossible to determine whether the exceptionally high richness observed in TOS, greater than in surrounding regions, reflects a genuine ecological pattern or a methodological bias.

Constrained ordination indicates a clear seasonal organization. Although the first two RDA axes explain a modest fraction of total variance (∼8%), samples separate cleanly by season: T and salinity align with summer, while nutrients, DO, and pH align with winter. Genera scores map onto this gradient: winter diatoms (*Thalassiosira*, *Skeletonema*, *Ditylum*) with NO_3 and SiO_4; autumn-associated (*Lioloma*, *Leptocylindrus*, *Guinardia*) with higher salinity; summer taxa (*Proboscia*, *Torodinium*) toward warm, nutrient-poor waters; and spring dinoflagellates toward intermediate, recharging conditions. In variation partitioning, space (MEMs) explains the largest share (∼10–13%), with environment and season each explaining only a few percent but with appreciable shared fractions (Env+MeMs; Env+Season), consistent with strong hydrographic control and the seasonal co-variation of physics and chemistry. The lack of the explanatory power of the environmental component has several and possible causes. First, the absence of fundamental drivers such as iron [131, 132, 210] and other metals [127, 301], of any measure of stratification, mixed-layer depth, irradiance, wind stress [312], and episodic meteorology (e.g., rainfall) that reorganizes communities [118, 4]; in general, coastal areas

are characterized by site-specific phenomena and interannual variability that limit generalization [72, 58]. Secondly, issues related to the measurement and sampling strategy. Phytoplankton respond to nanomolar nutrient dynamics at daily timescales [210, 13] and show stochasticity [58], while the campaign consisted of monthly samples and the instruments employed had detection level on the order of tenths of a micromole (nutrients in the marine environment are difficult to measure [259] and this is confirmed by the significant percentage of measurements below LOD/LOQ). Together, these factors obscure part of the signal. In appendix A we briefly discuss the effect of nutrients on the pure abundances, regardless of the community composition, and also in this case understanding their effect is not straightforward.

We previously mentioned that in temperate costal ecosystem it is expected a diatom-dinoflagellate-diatom succession throughout the year, entailing a turnover between *r*-strategist and *K*-strategist species, and across the Adriatic (NA–CA–SA) this succession is particularly clear. In winter, the community (Fig. 2.10a) is dominated by diatoms, especially *Skeletonema*, and to a lesser extent *Thalassiosira*, *Chaetoceros*, and *Leptocylindrus*. Other genera present include *Ditylum*, *Dictyocha*, *Lauderia*, *Prorocentrum*, and *Rhizosolenia*—all typical of the season [239, 309]. In spring, dinoflagellates become prominent, notably *Gymnodinium*, *Heterocapsa*, and *Prorocentrum*. In summer, the assemblage shifts toward diatoms such as *Pseudo-nitzschia*, *Proboscia*, and *Thalassionema*. By autumn, diatoms again dominate, with a resurgence of *Skeletonema* in the northern and central Adriatic. The seasonal distribution of these and the other IndVal-selected taxa is consistent with prior studies in the region [59, 306, 254, 38]. In Southern Mediterranean (SM), *Chaetoceros* and *Pseudo-nitzschia* are persistent features. *Chaetoceros* in particular blooms in summer but occurs year-round, in line with reports of species-specific seasonal blooms and bimodal abundance patterns [254, 306, 328]; most of the genera selected by IndVal have also been previously reported [56]. In sicily (SIC), composition is comparatively uniform through the year, with *Chaetoceros* more prevalent in winter–spring and *Pseudo-nitzschia* in summer–autumn. The second-half assemblage mirrors patterns observed in the other basins. These observations are consistent with previous work [60], although the area remains under-studied. As noted, the Southern Tyrrhenian (ST) is poorly characterized in our dataset, with only *Asterionellopsis* and *Leptocylindrus* emerging, probably as a consequence of large abundance differences between CAL and CAM. Direct comparison with the extensive literature for the Gulf of Naples is therefore not straightforward [328, 184, 45]. In Northern Tyrrhenian (NT), characteristic genera indicate a shift from diatoms to dinoflagellates over the year; the most abundant taxa are still mainly diatoms (*Asterionellopsis*, *Cylindrotheca*, *Thalassionema*), together with the cryptophyte *Plagioselmis*. In LIG, dinoflagellates (*Protoperidinium*, *Prorocentrum*, *Ceratium*) prevail in spring–summer, whereas diatoms (*Pseudo-nitzschia*, *Cylindrotheca*, *Thalassionema*) dominate in winter–autumn. By

contrast, Sardinia (SAR) does not display a clear diatom–dinoflagellate succession: *Gyrodinium* is present throughout the year and dominates in winter–autumn, while warmer months include contributors from multiple classes—a pattern unique among basins. The results for this reagion are difficult to verify because many available studies focus on lagoons rather than open coastal waters [232, 88]. The succession diatoms-dinoflagellates-diatoms seem to be slightly anticipated with respect to expectation. This is due to how months are grouped into seasons: diatoms bloom between February and April and these are split between winter and spring; analogously in June, which belong to spring, we can have dinoflagellates bloom [20]. This however does not change the interpretation of the results.
A possible explanation for the pronounced genera turnover is basin trophic status. In eutrophic settings such as much of the Adriatic-where shallow bathymetry, vigorous winter mixing, and sustained river inputs generate recurrent nutrient pulses—the resulting conditions may favour rapid uptake by a few highly efficient r-strategists (e.g., *Skeletonema*. As stratification develops and nutrients decline, dominance can shift toward K-strategists; the transition may be relatively abrupt compared with oligotrophic sectors, where external pulses are weaker and variability is governed more by high-frequency fluctuations than by strong seasonal forcing [278]. In the latter settings, communities may remain compositionally more stable throughout the year, and detectable turnover is correspondingly weaker. Consistent with this view, strong turnover is also observed in SM, which may still be influenced by exchanges with the Adriatic and is a semi-enclosed gulf receiving nutrients from several rivers; likewise, in NT, recurrent inputs from the Tiber (LAZ) and Arno (TOS) may play a similar role [194]. Comparable links between nutrient pulsing and turnover have been reported for freshwater phytoplankton and zooplankton [326, 112]. SIC and SAR exhibit weaker or irregular seasonal signals, potentially for two non-exclusive reasons. First, the high fraction of unclassified cells likely reduces the power to detect succession. Second, both basins are more exposed to open-ocean Mediterranean circulation (e.g., propagation of Modified Atlantic Water and Levantine Intermediate Water), which could impose mesoscale variability not phase-locked to the seasonal cycle.

In this work we presented one of the first assessments of Italian coastal phytoplankton communities. The Marine Strategy Framework Directive (MSFD) dataset enabled us to examine sectors of the coastline that remain poorly studied (e.g., much of the western coast, Sardinia) and to compare temporal dynamics across the peninsula. This proves the value of publicly-funded monitoring programs both as tools for environmental status assessment and as resources for addressing ecological questions. Abundance is

strongly structured in space, reflecting multiple forcings—most prominently river inputs. Community richness, at both species and genus levels, also appears spatially structured. Community composition is broadly consistent with expectations for temperate coastal ecosystems: small, fast-growing diatoms dominate across much of the year, whereas highly stratified, nutrient-depleted conditions in spring/early summer favour slower-growing dinoflagellates. Basin-specific assemblages differ, but several genera recur across regions (e.g., *Thalassionema*, *Pseudo-nitzschia*, *Chaetoceros*). The diatom–dinoflagellate succession is weak or absent in some sectors, plausibly reflecting lower degrees of eutrophication and more persistent oligotrophy. A more in-depth, quantitative estimation of the environmental drivers was constrained by the limited explanatory power of the ordination models. This likely reflects missing drivers (e.g., iron and other trace metals), a sampling design coarse compared to the rapid phytoplankton dynamics, heterogeneous data availability. In general, phytoplankton communities are shaped by temperature (strongly seasonal and linked to stratification), salinity, and macronutrient inputs (N and P). However, several features hampered finer-scale analyses: the bimonthly sampling objective was not met consistently by all regional agencies, precluding resolution of dynamics finer than seasonal; samples at two depths (surface and DCM) were not collected consistently, limiting vertical comparisons; critically, there were large differences in taxonomic resolution among regions—even at class level—making it difficult to distinguish true biogeographic contrasts from identification artefacts. Intercalibration among taxonomists and harmonized QA/QC would substantially improve the scientific value of future campaigns.

Chapter 3

Exploring global gene expression regulation in the diatom Pseudo-nitzschia multistriata

Dr Maria Immacolata Ferrante contributed to supervising the work presented in this Chapter and in Appendix B and C.

3.1 Introduction

As mentioned in Sec. 1.3, diatoms possess a two-stage life-cycle comprising a vegetative phase and sexual phase and one of the major controls of this switch is the cell size. In this chapter we will focus on *Pseudo-nitzschia multistriata* and will explore its gene expression regulation at these trwo different phases.
Pseudo-nitzschia multistriata, one of the 62 species of a cosmopolitan genus of diatoms, has been observed in variuos coastal regions of the world and is known to be one of the 26 species to produce domoic acid [30], though never at dangerous levels [11, 220]. It belongs to the most evolutionarily recent diatom lineage, the motile raphid pennates. Its elongated cells bear a raphe, a longitudinal slit in the silica wall through which protein complexes are secreted, enabling a slight vertical gliding [148, 97]. There are many interesting aspects regarding diatoms and *Pseudo-nitzschia*, but here we will focus on sexual reproduction phase, described in more detail in Sec. 3.1.1.

3.1.1 The sexualization phase of P. multistriata and its genetic fingerprint

A central breakthrough in the understanding of the diatom life-cycle has been the discovery of *MRP3* (Mating-type Related Plus 3) as a mating-type–determining gene [252]. *MRP3* is expressed exclusively in MT+ strains and exhibits monoallelic expression. A short, repeated DNA sequence upstream of *MRP3* is consistently associated with MT+, acting as a genetic "tag". Artificial overexpression of *MRP3* in MT– causes sex reversal: transformed cells behave as MT+. This switch is accompanied by a coherent flip in other mating-type-biased genes: MT+ biased *MRP1* and *MRP2* are upregulated, whereas MT– biased *MRM1* and *MRM2* are downregulated. These results establish *MRP3* as a *master regulator* at the apex of a simple but effective regulatory hierarchy in *P. multistriata*, providing the first clear genetic program for heterothallism in diatoms. Within this hierarchy, *MRP1* and *MRP2* encode proteins predicted to be secreted or receptor-like (consistent with roles in signaling and pheromone exchange), *MRM1* contains a heat-shock factor–like DNA-binding domain (a candidate transcriptional/chromatin-level regulator of the MT– state), and *MRM2* encodes a leucine-rich repeat (LRR) receptor-like protein strongly expressed during pairing and gamete fusion [252]. Comparative transcriptomics and phylogenomics indicate that *MRM2*-like proteins are widely conserved across pennate diatoms and are likely to be involed in partner recognition [34]. This suggests a receptor-based recognition apparatus is ancestral in pennate diatoms, whereas *MRP3* appears to be a clade-specific innovation restricted to the *Pseudo-nitzschia/Fragilariopsis* lineage [252, 34]. The ids for each of these five genes (which we will refer to in the following) are listed in Tab. 3.1
Multi-omics studies show that sexualization begins with rapid, asymmetric responses in MT+ and MT–. Early sexual stages (gametes/zygotes) appear $\approx$ 24 h, auxospores by $\approx$ $48-72$ h, and initial cells by $\approx$ $60-72$ h [29, 13]; the SST is approximately 55 μm. Within hours from partner perception, populations undergo a G1 phase (first phase leading to mitosis) arrest, despite nutrient-replete conditions. Only a minority ($\approx$ 20%) proceeds to meiosis and gametogenesis, while most cells remain quiescent for several days and later resume growth [29, 13]. This arrest likely reflects pheromone-driven synchronization of cell cycle before gamete release. MT– tends to initiate meiosis slightly earlier than MT+, consistent with sequential exchange of chemical signals. Distinct pheromones, produced by different mechanisms, mediate this dialogue and trigger mating type–specific responses. Together, these interactions establish a positive-feedback loop in which MT– enhances stimulation of MT+, thereby accelerating progression toward sexualization. The cues are asymmetric: MT– secretes proteinaceous signals, whereas MT+ releases non-protein cues [187]. MT– also shows more differentially expressed genes than MT+, supporting functional asymmetry between mating

types [29]. This asymmetry underpins a gene-response hierarchy that clarifies their distinct roles.

It was observed that numerous genes across different metabolic pathways change behaviour during sex [29, 13, 187]. At the population scale, sexualization entails prolonged growth slowdown and metabolic arrest: photosynthesis is broadly suppressed, with down-regulation of light-harvesting complexes (LHCs) and photosystem subunits, while photoprotection modules such as *LHCX* are activated to safeguard the cell during this delicate phase [13]. Downregulation of transporters of inorganic nutrients needed for growth (silicate, nitrate/nitrite, ammonium, formate transporters), as well as diatom-specific cyclins associated with nutrient status and the cell cycle, indicate a deliberate reallocation of resources [29]. Canonical meiotic toolkit genes and cohesin-complex genes are activated, consistent with engagement of meiotic recombination [29, 13]. G protein–coupled receptors are upregulated, echoing their roles in mating-cue perception in yeast [29]. In MT– there is upregulation of *Cathepsin D*, whose function is similar to those of proteins associated with pheromone responses [28]. A substantial fraction of differentially expressed genes are diatom-, Bacillariales-, *Pseudo-nitzschia*-specific or orphans, consistent with life-cycle uniqueness and species-specific solutions to mate recognition and fusion. These analyses enabled compilation of a sexually induced gene (SIG) set required at discrete stages of sexualization (see [29, 13, 187] for full lists). A similar study aimed at finding genetic fingerprints of sexualization was conducted for *C. closterium* [34]. No *MRP3* homolog has been identified, and transcriptomic surveys have not revealed a comparable top-level switch, reinforcing that *MRP3* is characteristic of the *Pseudo-nitzschia* clade. By contrast, *MRM2* appears widely conserved across pennate diatoms, suggesting a shared origin of a receptor-based recognition system and a key role for *MRM2* in mating type identity [34]. Gene expression analyses reveal clusters of mating type-biased and SIP-responsive genes, many of them are evolutionarily young and taxon-specific. Notably, a block of three MT+ specific, SIP-induced genes occurs in close genomic proximity, suggesting coordinated regulation, although functions remain largely unknown. Some genes encode putative secreted proteins and receptor-like candidates reminiscent of the *MRM2*-type recognition system, but without evidence for a unifying regulatory node. The apparent idiosyncrasy of the indetified clusters implies that size sensing and mating type specification may lack a universal master regulator across pennate diatoms.

Both *P. multistriata* and *C. closterium* exhibit pronounced mating type asymmetries—most plausibly rooted in pheromone emission and perception—underscoring the value of analyzing MT-specific programs. The relatively focused set of differentially expressed genes suggests that transcriptomic differentiation between size classes and mating types is concentrated in traits tied to size sensing and sexual dimorphism, while core metabolic architecture remains largely conserved.

Gene name	gene id
MRP1	24820
MRP2	122240
MRP3	20770
MRM1 (model 1)	41130
MRM1 (model 2)	85380
MRM2	6960

TABLE 3.1: Gene names and their associated gene ids for the mating-type related genes discovered in [252].

3.1.2 RNA-seq data

Sequencing technologies determine the order of nucleotides in DNA or RNA, yielding a digital representation of genetic information. Over the past two decades, high-throughput approaches such as next-generation sequencing (NGS) and, more recently, third-generation sequencing, have transformed biology by enabling rapid, cost-effective decoding of genomes and transcriptomes [271]. When applied to RNA, these technologies form the basis of RNA sequencing (RNA-seq), a method that allows researchers to detect and quantify the expression of diverse RNA populations, including messenger RNA (mRNA) and non-coding RNAs. The study of RNA expression is often more informative than genome sequence analysis when the goal is to understand an organism's functional state: because mRNA is translated by ribosomes into proteins, RNA-seq reveals which metabolic pathways are active under specific conditions, and transcript abundances are used as a proxy for protein synthesis, thereby indicating their biological relevance at that time.
The generation of RNA-seq data begins with the extraction of total RNA or messenger RNA (mRNA) from the organism, followed by conversion into complementary DNA (cDNA) libraries, which are then sequenced using high-throughput sequencing platforms such as Illumina, Oxford Nanopore, or PacBio. Short-read sequencing technologies (e.g., Illumina) typically produce hundreds of millions of short reads (50–250 base pairs), whereas long-read technologies (e.g., PacBio SMRT, Nanopore) generate reads that can span entire transcripts but at lower throughput and higher error rates, especially at the end of the sequence [86]. During library construction, adapters are ligated to cDNA fragments and must later be removed computationally [149]. The primary output of an RNA-seq run is a collection of nucleotide strings (raw reads). To quantify expression, these reads are processed to produce raw counts, i.e., per-gene (or per-feature) read summaries suitable for downstream analysis.

The pre-processing steps necessary and the specific tools used are:

- Library quality assessment with `FastQC` [2] to evaluate per-base quality, GC content, adapter contamination, and sequence length distributions. This step is necessary to to determine the length of the adapter sequence (at the head) and the start of the low-accuracy portion of the sequence.
- Trimming: Low-quality bases, adapter sequences, and poor-quality tails were removed using `Trimmomatic` [44]
- Sequence alignment: raw reads do not incorporate information about the gene they come from. Reads were mapped to the reference genome with `STAR` [90].
- Estimation of expression: Gene-level counts were obtained using the function `featureCounts` of the R package `Rsubread` [166, 165].

3.1.3 WGCNA

High-throughput technologies such as RNA-seq have shifted transcriptomic analysis from single-gene investigations to the study of networks of co-regulated expression [329, 221]. In a gene co-expression network (GCN), genes are represented as nodes connected by weighted edges that reflect expression similarity across samples [268]. The central premise is that co-expressed genes often participate in the same pathways or complexes and are controlled by common regulatory influences [329, 214]. Studying expression in a network context reveals modules—coherent groups of genes with coordinated behavior—thereby exposing emergent properties such as transcriptional programs, global responses, or system-level organization that are not evident from gene-by-gene analyses [221, 193]. A key advantage of co-expression analysis is its ability to generate functional hypotheses. Modules can be characterized by enrichment analysis [287], and uncharacterized genes consistently co-expressed with functionally annotated ones can often be assigned putative roles. This approach has been especially valuable in eukaryotic systems, where many genes lack annotation. For example, [286] demonstrated that conserved modules across species could be used to predict gene function, experimentally validating novel roles in cell proliferation. These findings establish co-expression networks as a systems-level framework for interpreting transcriptomes [154].

Co-expression clustering also reduces data complexity. Thousands of gene profiles can be distilled into a small number of module-level summaries, simplifying interpretation and mitigating multiple-testing issues [109, 193]. Modules often contain hub genes—highly connected members that serve as key regulators or essential pathway components—making them attractive candidates for further study [109]. This not only simplifies interpretation but also mitigates the multiple-testing problem that plagues

genome-wide studies, since one evaluates a limited number of modules instead of every gene individually [109, 193]. Indeed, co-expression network analysis has repeatedly yielded novel biological insights by highlighting system-level organization, such as revealing network-wide changes in developmental transitions, stress responses, or disease states that are not evident from single-gene analyses [193].
Another strength of co-expression networks is their ability to suggest system-level regulatory architecture. If a set of genes forms a tight co-expression module, it often implies the presence of common upstream regulators (such as a transcription factor or epigenetic modification) coordinating that module. By integrating network data with other information (e.g., gene annotation), one can pinpoint candidate master regulators for a module's activity [109]. In cancer transcriptome studies, for example, network analyses have shown that prognostic genes tend to cluster in specific co-expression modules rather than acting as isolated hubs, reflecting underlying regulatory programs driving tumor phenotypes [323].
In short, co-expression networks provide a holistic view of gene regulation, helping to connect genes into pathways and pathways into larger regulatory circuits. This approach is particularly advantageous in complex eukaryotic systems where emergent behaviors (like differentiation or stress responses) arise from the collective action of many genes.

Several methods have been developed to construct gene co-expression networks [263]. In this study, we employed Weighted Gene Co-expression Network Analysis (WGCNA) [154], a widely adopted framework in systems biology. WGCNA has become a standard approach for identifying modules of highly correlated genes and relating them to biological traits, and it has been extensively applied across diverse research domains [108, 234, 286, 64, 315].
WGCNA has also been applied in marine plankton research, demonstrating its versatility across different omic data types and biological contexts [119, 6, 327, 169, 228]. For instance, [119] used metagenomic data from the Tara ocean global expedition to construct networks encompassing eukaryotic, prokaryotic, and viral lineages, linking them to carbon export processes and identifying the organisms most strongly associated with this function. Similarly, [228] applied WGCNA to metabarcoding data from the same expedition to identify modules associated with nutrient availability and phytoplankton community composition, enabling inference of sub-communities responsive to specific nutrient regimes. Other studies have focused on the model diatom *Phaeodactylum tricornutum*. In [6], WGCNA was applied to RNA-seq libraries to explore functional, subcellular, and evolutionary relationships among co-expressed genes. This analysis revealed modules underlying anabolic, photosynthetic, and respiratory metabolism, as

well as possible candidate for transcription factors involved in the regulation of chloroplast and mitochondrial pathways, and genes coordinating the interactions between these two organelles. [327] examined transcriptional responses of *P. tricornutum* under varying light conditions, identifying modules related to fatty acid biosynthesis, photosynthesis, and carbon fixation, along with potential transcription factors regulating these pathways. Finally, [169] identified a gene highly correlated with cell density whose involvement in calcium ion regulation was subsequently confirmed through a gene-knockout experiment.
The approach's popularity stems from its combination of statistical rigor and biological insight: WGCNA provides a comprehensive pipeline to construct, detect, and interpret gene co-expression networks, making it a go-to method in genomics studies where system-level interpretation is needed.

Starting from normalized count data, we build a weighted, fully connected network where each gene is a node and edge weights are proportional to co-expression. Formally, the adjacency matrix is defined as:

$$a_{ij} = \begin{cases} corr(g_i; g_j)^{\beta} & \text{if} \quad corr(g_i; g_j) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

where $corr(g_i; g_j)$ is the sample-wise correlation between two genes and β is a positive integer that has the effect of lowering the low-level correlation and increasing the high-level ones. The choices of which kind of correlation to use (WGCNA implements spearman, Pearson and bi-weight mid-correlation [325]) and which value β to choose are guided by the *scale-free* topology criterion for the network, which requires the degree distribution $P(k)$ to follow approximately

$$\log P(k) \sim \gamma k \tag{3.2}$$

where k is the network degree and γ a constant between 2 and 3. The meaning of this relation is that the number of nodes with degree k decreases exponentially as k increases: the majority of nodes possess few and weak connections, while a few nodes have a lot of extremely strong connections. At the time of publication of WGCNA, there was evidence that most biological networks had a scale-free topology [138, 27], however, this has been recently questioned [276]. Still, this parameter retains its functionality of emphasizing robust correlations. The heuristic is to select the value of β for which the R^2 obtained by fitting Eq. 3.2 reaches a plateau.
As previously mentioned, co-expression networks can be partitioned into modules,

that is clusters of highly co-expressed genes. In general, clustering requires both a pairwise similarity measure, which quantifies how reasonable it is to group two genes together, and a clustering algorithm. One could use $1 - a_{ij}$ as similarity, but spurious correlations and weak links may generate erroneous modules. A more robust alternative is the *topological overlap matrix* (TOM) [128, 325], in which the similarity of two nodes increases with the number of neighbours they share ($a_{ij} \neq 0$). Thus, similarity depends not only on direct correlation but also on the broader network topology. There is empirical evidence that TOM yields biologically meaningful modules [324, 213, 55]. Clustering is now ubiquitous in data analysis, with numerous algorithms optimized for different problem settings [321, 264]. In this work, we adopted the default WGCNA clustering method `DynamicTreeCut` in its "hybrid" form, which has consistently provided meaningful partitions in diverse applications [154]. The hybrid procedure combines hierarchical clustering with partition-around-medoids [47, 159]. For a detailed description, see [155]; here it is sufficient to note that DynamicTreeCut also defines a special grey module, which contains genes not assigned to any cluster—i.e., genes with expression patterns that markedly differ from all others.

Once the modules have been identified, there is the necessity to i) identify those most strongly associated with traits of interest (here, mating type and the state of sex competence); ii) extract within those clusters the most relevant genes, because they are optimal candidates to be master regulators or transcriptional factors. WGCNA introduces several concepts and proposes a pipeline for this purpose. The most important is the *module eigengene* (E_i), defined as the first principal component of the module's expression matrix. This eigengene represents the weighted average expression profile of the module and can be correlated with sample traits to identify biologically relevant modules. At the gene level, *gene significance* (GS) quantifies the correlation between a gene and a trait, while *module memebership* (MM) measures how strongly a gene correlates with its module eigengene. Genes with both high GS and MM are typically central to trait-associated modules and represent strong candidates for further study [128, 325]. A detailed description of these quantities can be found in Tab 3.2. Additional steps recommended in the WGCNA pipeline include: (i) hierarchical clustering of samples (we use Ward linkage) to detect outliers, and (ii) merging of modules with highly correlated eigengenes (here, $corr(E_i, E_j) \geq 0.75$). The latter produces higher-order *meta-modules*.

3.2 Material and methods

The full list of RNA-seq libraries used in this work is reported in Tab. 3.3; some originate from published studies, others are unpublished. For a subset of libraries it was

WGCNA concept	Mathematical formulation	Description
Eigengene	E_i: first principal component of a module	A vector of length equal to the number of samples. Each component represents the weighted average expression of all genes in the module. Correlation between an eigengene and a trait can suggest its genes are related to that trait.
Gene significance	$GS_{jk} = corr(g_j; T_k)$	Samples-wise correlation of gene j with a trait k. Other definitions exist [128], but we report only the one used here.
Module membership	$MM_{ij} = corr(E_i; g_j)$	Correlation between a gene and its module eigengene. High MM indicates that the gene is representative of the overall cluster behavior.

TABLE 3.2: Key WGCNA concepts used to identify trait-associated modules and prioritize biologically relevant genes. g_j denotes the j-th gene and T_k the k-th trait.

necessary to re-ran preprocessing steps needed for downstream analysis[1], and the set of non-default parameters used for these softwares are listed in Tab. C.1. These libraries come from different experiments, each of which with its different purpose. From [13] and [29], we took only the parental mono-culture controls. From [252], we used only the wild-types cultures. The sX (X = 1...14) series targeted expression differences above/-below the SST. PTX/noPTX compared phosphorus sufficiency (control) versus deprivation across time points; PmT4_441 and PmT4_50 analogously contrasted nitrogen repletion versus deprivation. The last column of Tab. 3.3 indicates which samples were re-preprocessed for this study. In Tab. 3.4 shows the distribution of samples between the two MTs and the sexualization state, which is associated to the cell size being below or above the SST.

3.2.1 Removing unwanted variance and batch effects

Before constructing networks, it is essential to remove technical artifacts and biases [3]. Because WGCNA relies on correlations, genes with consistently low expression can lead to spurious connections; thus, we excluded genes with zero expression in more than 25 samples. This threshold effectively reduced the excessive zero peak in the raw count distribution. Each sample also differs in library size (total read count), meaning that apparent differences in gene expression may simply reflect sequencing depth

[1] Any series of statistical and computational methods to identify a biological signal

Sample id	MT	Size	Ref.	Pre-processing	Abbreviation
lv130_T2_A	M	Small	[13]	✓	Ann
lv130_T2_B	M	Small			
lv130_T2_C	M	Small			
lv193_T2_A	F	Small			
lv193_T2_B	F	Small			
lv193_T2_C	F	Small			
B938_C_T1	M	Small	[29]	✓	Basu
B857_C_T1	F	Small			
B856_C_T1	M	Small			
B939_C_T1	F	Small			
B938_C_T2	M	Small			
B857_C_T2	F	Small			
B856_C_T2	M	Small			
B939_C_T2	F	Small			
HCUH	M	Small			
HCUN	F	Big			
HCUO	M	Small			
HATT	F	Big			
CIIO1	M	Small	[252]	✓	Russo
CIIO2	M	Small			
CIIP2	F	Small			

TABLE 3.3: RNA-seq samples used for the study of *P. multistriata*, indicating mating type (MT), cell size (small: below the SST; big: above the SST), pre-processing status (✓: preprocessing required; ✗: raw counts available), and label with which experiments will be referred as along the thesis (Abbreviation) (continues on the following table).

[149]. Furthermore, gene-wise expression exhibits a strong mean–variance relationship (generally, a linear relationship on a log-log scale), with highly expressed genes showing greater fluctuations that can mask variation in lower-abundance genes. Both issues were addressed using *variance-stabilizing transformation* (`vst`) from the R package `DESeq2` [12, 174], which makes data approximately homoskedastic by removing the mean-variance trend estimated with a negative binomial generalized linear model (NBGLM), and normalize with respect to library size. A final major artifact is the presence of *batch effects*, defined as "sub-groups of measurements that behave differently across conditions and are unrelated to biological variables" [157]. In this dataset, batch effects corresponded primarily to experiment of origin. We corrected these using `ComBat` from the R package `sva` [158], which employs an empirical Bayes framework to adjust for batch-specific effects [141].

Sample id	MT	Size	Ref.	To pre-process	Abbreviation
s14	F	Small	Unpublished	✗	Unpub
s1	F	Small			
s2	F	Small			
s3	F	Small			
s4	F	Big			
s5	F	Big			
s6	F	Big			
s7	M	Big			
s8	M	Big			
s9	M	Big			
s10	M	Big			
s11	M	Big			
s12	M	Big			
s13	M	Small			
PT4_1	F	Small	Unpublished	✗	PO
PT4_2	F	Small			
PT4_3	F	Small			
PT6_1	F	Small			
PT6_2	F	Small			
PT6_3	F	Small			
noPT4_1	F	Small			
noPT4_2	F	Small			
noPT4_3	F	Small			
PmT4_441_1	F	Small	Unpublished	✗	Nit
PmT4_441_2	F	Small			
PmT4_441_3	F	Small			
PmT4_50_1	F	Small			
PmT4_50_2	F	Small			
PmT4_50_3	F	Small			

TABLE 3.3: RNA-seq samples used for the study of *P. multistriata*, indicating mating type (MT), cell size (Small: below the SST; Big: above the SST), and preprocessing status (✓: pre-processing required; ✗: raw counts already available), and label with which experiments will be referred as along the thesis (Abbreviation).

		Size	
		Under SST	Above SST
MT	M	14	6
	F	32	5

TABLE 3.4: Distribution of samples between mating types (MT) and sex competence state, which is determined by the size (SST).

3.2.2 Additional statistical analysis for cluster association to traits

WGCNA primarily relies on correlations, which might not be the appropriate inference method when dealing with dichotomous quantities as mating type and sexualization state. Additionally, we are in a case of severe unbalanced classes and we lack complete knowledge of the nutrients availability. which can have a confounding effect. For this reason, we performed additional tests to detect meta-modules that are really associated to our traits.

We faced strong imbalance because only 11 samples were above SST, too few to build a stable network independently (see "WGCNA FAQ" [3]). To address this, we constructed three types of networks: one using all samples to study differences in both traits (*complete network*), one using only sexualized cells (*small network*) to study differences related to mating type, and finally a collection of network (*bootstrap networks*) where we used all the 11 samples above SST and selected 10 different subsets of 11 samples below SST obtained with random resampling (bootstrap). Modules from the 10 bootstrap networks were compared using three external cluster evaluation metrics: *Adjust Rand index*, *variation information*, *V-measure* [62, 197, 247], all available in the R package `clevr` [182]. We retained networks that yielded unique modules or were representative of stable partition groups. To partially overcome the problem of incomplete knowledge of the system we opted to use two additional statistical tests together with the eigengene-trait correlation. The first one was to run for each cluster a PERMANOVA [160, 314] using Bray-Curtis distance [159], with predictors including the trait, the experiment, and their interaction. Modules in which the trait explained the largest and significant portion of variance were considered candidates. The second approach is a classification-inspired test. For each module, we reduced dimensionality with *multidimensional scaling* (MDS, `cmdscale` from package `stats` [233]), then clustered samples into two groups using *k-means* (function `kmeans` from package `stats` [233]). Agreement between these two groups with trait labels was assessed using classification accuracy and its 95% CI (binomial test, function `confusionMatrix` from package `caret` [151]). Modules with accuracy significantly above random expectation were flagged as informative. By combining three independent approaches (eigengene–trait correlation, PERMANOVA, and classification-inspired tests) we strengthened confidence in the identification of biologically meaningful, trait-associated modules.

3.3 Results

Raw read counts displayed a strong mean–variance dependency typical of RNA-seq data (Fig. 3.1a). Application of`vst` substantially reduced this dependency, producing

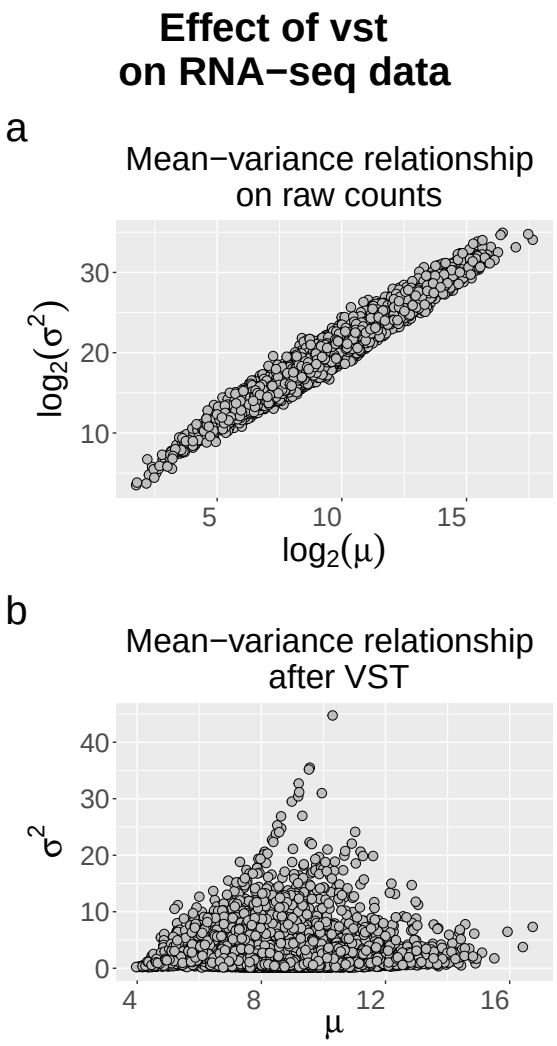


FIGURE 3.1: Effect of the application of the `vst` on the mean-variance relationship on raw counts. a: mean (μ) vs variance (σ^2) of raw counts on a log-log scale ; b: same relation on counts after `vst`, not on log scale.

approximately homoskedastic expression distributions with the exception of few genes (Fig. 3.1b). PCA on raw counts showed that samples clustered primarily according to their experiment of origin, indicating the presence of batch effects related to library preparation or sequencing run (Fig. 3.2a). Subsequent application of `ComBat` further removed the influence of the experimental factor (Fig. 3.2b). Variations in these transformed data should now be mainly driven by biological signals. However, in the following statistical analysis the experimental factor will still be considered.

3.3.1 Complete and small networks

We first describe the construction of the complete and small networks jointly, and then present a combined analysis of their modules, for reasons detailed below.

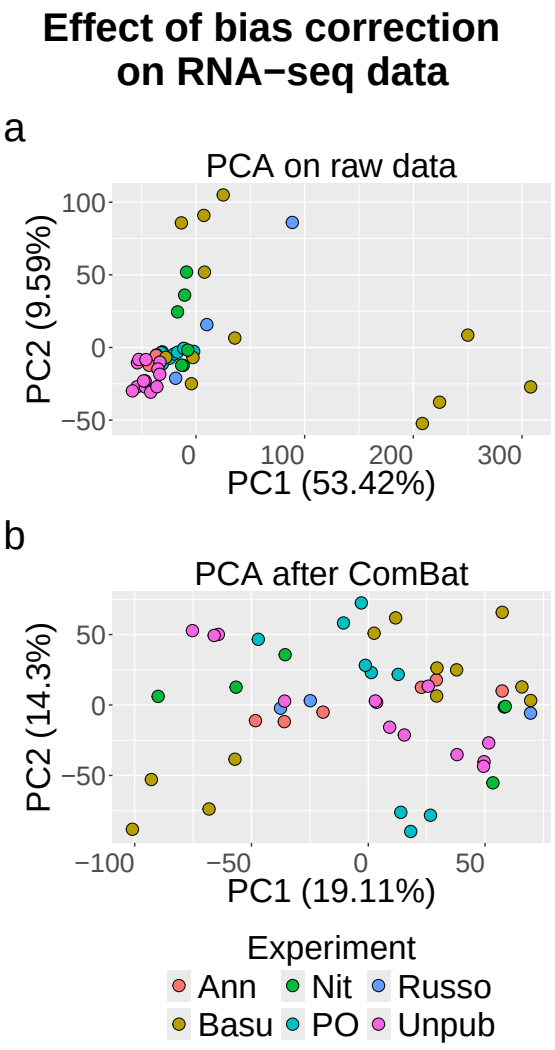


FIGURE 3.2: Effect of bias correction on the expression levels of RNA-seq data. a: first two PCA components of raw counts; b: first two PCA components of counts after the application of `vst` and `ComBat`. For legend labels refer to 3.3.

Of the 12008 annotated genes, 11925 showed non-zero expression in at least 25 samples and were retained for the complete network. Sample clustering identified one outlier (s13; Fig. B.1), which was removed, leaving 49 samples. For Spearman, Pearson, and biweight midcorrelation, the scale-free topology fit (R^2) plateaued for soft-threshold powers $\beta \in [10, 12]$, with all fits exceeding 0.9; we therefore selected biweight midcorrelation with $\beta = 11$ as a robust choice using the smallest sufficient β (Fig. B.2). `DynamicTreeCut` initially produced 27 modules (including grey), which merged into 19 meta-modules (hereafter, "modules"; Sec. 3.1.3). The results of the associations between modules and traits using the four statistics described above are summarized in Fig. 3.3 and Fig. 3.4.

The small network was built from the subset of 38 samples below SST. Here, 11558

genes passed the expression filter of having non-zero expression in less than 10 samples. No outliers were detected (Fig. B.3). As above, we used biweight midcorrelation with $\beta = 11$ (Fig. B.4). `DynamicTreeCut` yielded 30 modules which merged into 18 meta-modules. Trait associations for the small network are reported in Fig. 3.5.

The two partitions showed extensive agreement: Fisher's exact tests revealed statistically significant overlaps (FDR < 0.05) for many cross-network module pairs (Fig. B.5), which is expected because the small-network samples constitute $\sim$77% of the complete network. For this reason, small-network module names were remapped to the most similar modules in the complete network (Table 3.5), and the biological interpretation is presented jointly. Operationally, for each small-network module we identified its counterpart in the complete network as the module with a statistically significant overlap and the largest number of shared genes; if two small-network modules mapped to the same complete-network module, only the larger of the two (by gene count) was renamed.

Complete network	Small network	
Module names	Original names	New names
green	black	green
black	blue	black
cyan	brown	cyan
yellow	cyan	yellow
lightcyan	darkgreen	lightcyan
brown	darkorange	brown
turquoise	darkred	turquoise
magenta	darkturquoise	magenta
purple	green	purple
grey	grey	grey
grey60	grey60	grey60
midnightblue	lightgreen	midnightblue
pink	midnightblue	pink
darkturquoise	red	darkturquoise
tan	royalblue	tan
greenyellow	saddlebrown	greenyellow
darkgreen	skyblue	darkgreen
blue	turquoise	blue
not present	purple	purple
darkgrey	not present	not present
lightyellow	not present	not present
orange	not present	not present
red	not present	not present

TABLE 3.5: Mapping used to rename modules in small network as the most similar module in the complete network. First column lists names of modules of complete network, second column lists the original module names of the small network , and the third column lists the new names; "not present" means there is not a module with that name in one of the two networks.

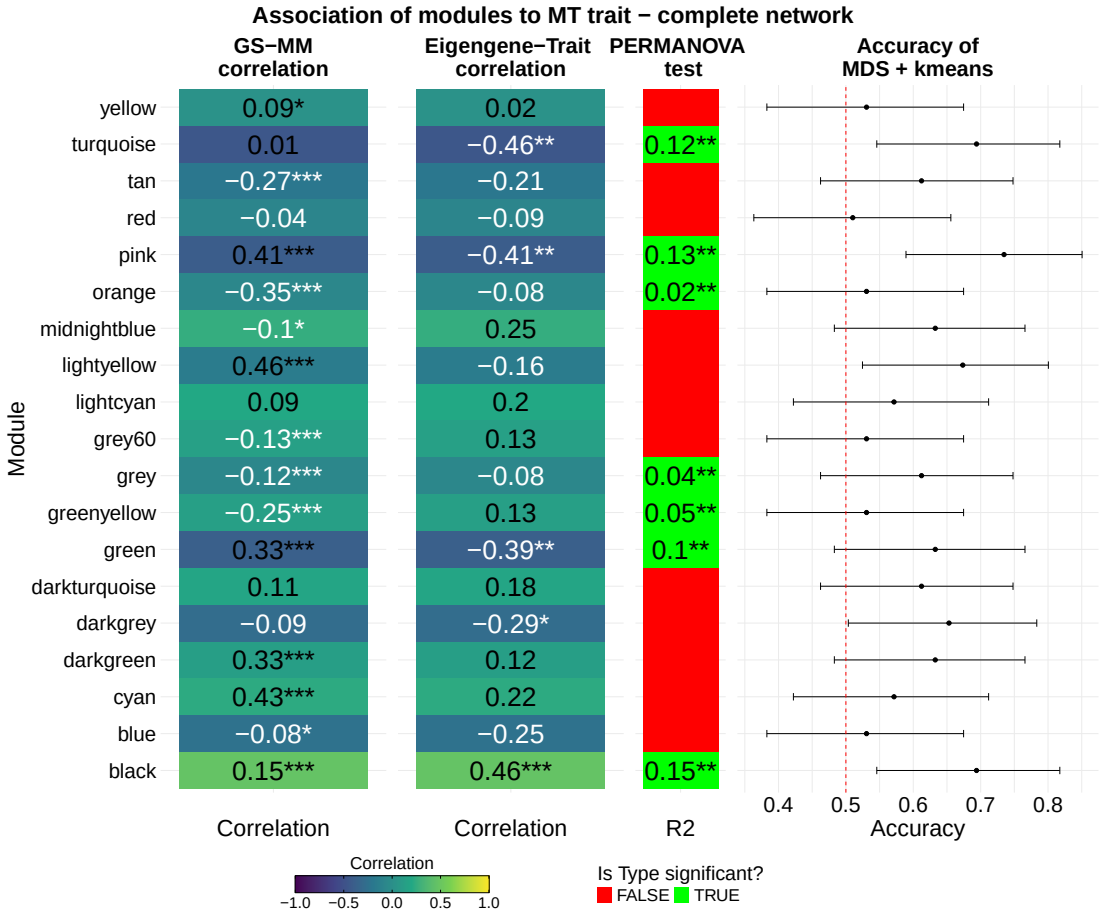


FIGURE 3.3: Different statistics used to determine which modules of the complete network can contain gene associated to MT: correlation between gene significance and module membership, eigengene-MT correlation, PERMANOVA test, accuracy computed between the MT and a partition created applying K-means to the first 28 MDS axes.

To ascribe biological functions to modules, we compared their gene content with differential expression signatures and curated process-level gene sets from *P. multistriata* studies [29, 13]. The resulting enrichments for both networks are shown in Fig. 3.6–3.7. In parallel, Gene Ontology (GO) over-representation analyses summarized in Fig. 3.8–3.9 were conducted to identify possible pathways associated to any module and trait.

The first thing to notice is that the four statistics were not always concordant. In particular, the GS–MM correlation appeared to be the weakest statistic, as several modules with a significant positive GS–MM correlation did not pass the other three tests. Examples of this for the MT trait are the cyan, darkgreen, lightyellow, and yellow modules of the complete network (Fig. 3.3), and the cyan, green, midnightblue, purple, and turquoise modules of the small network. We can expect that the modules bearing the strongest signal are those that pass the majority of tests. Accordingly, in the complete

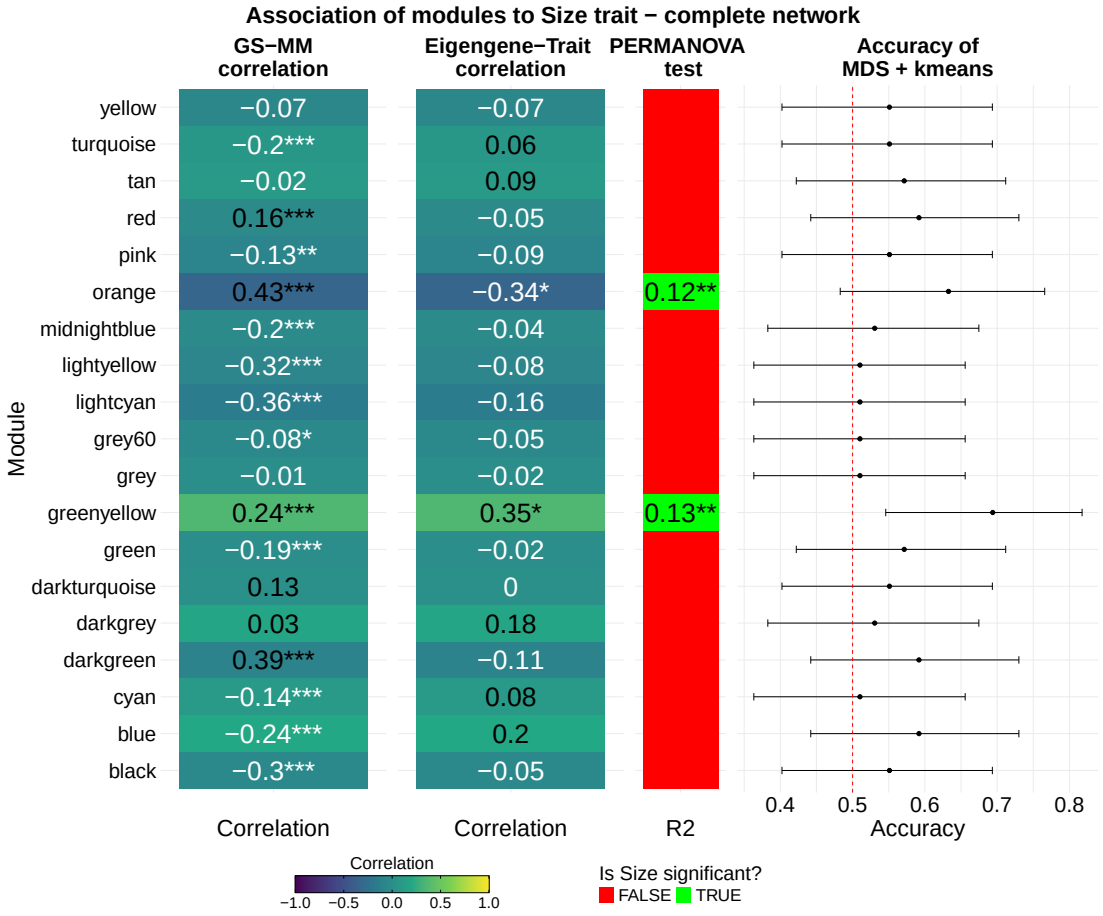


FIGURE 3.4: Different statistics used to determine which modules of the complete network can contain gene associated to size: correlation between gene significance and module membership, eigengene-Size correlation, PERMANOVA test, accuracy computed between the size and a partition created applying K-means to the first 28 MDS axes.

network, the black and pink modules are the most promising candidates for containing genes associated with MT, followed by green and turquoise (three out of four tests), then darkgrey (two out of four), and finally cyan, darkgreen, grey, lightyellow, orange, and yellow. For the small network, the most significant are black, greenyellow, and pink (three out of four), followed by cyan and green (two out of four), and finally lightcyan, midnightblue, purple, and turquoise. Regarding size, in the complete network the strongest associations were found for greenyellow, followed by orange (three), and finally darkgreen and red (one). From these lists, we can observe that the two networks share several clusters related to MT, as expected given that module names were adjusted according to the number of genes in common. A second observation is that there are more modules related to MT than to size; this will be further discussed in Sec. 3.4.

We next describe the modules considered most relevant, either because they show

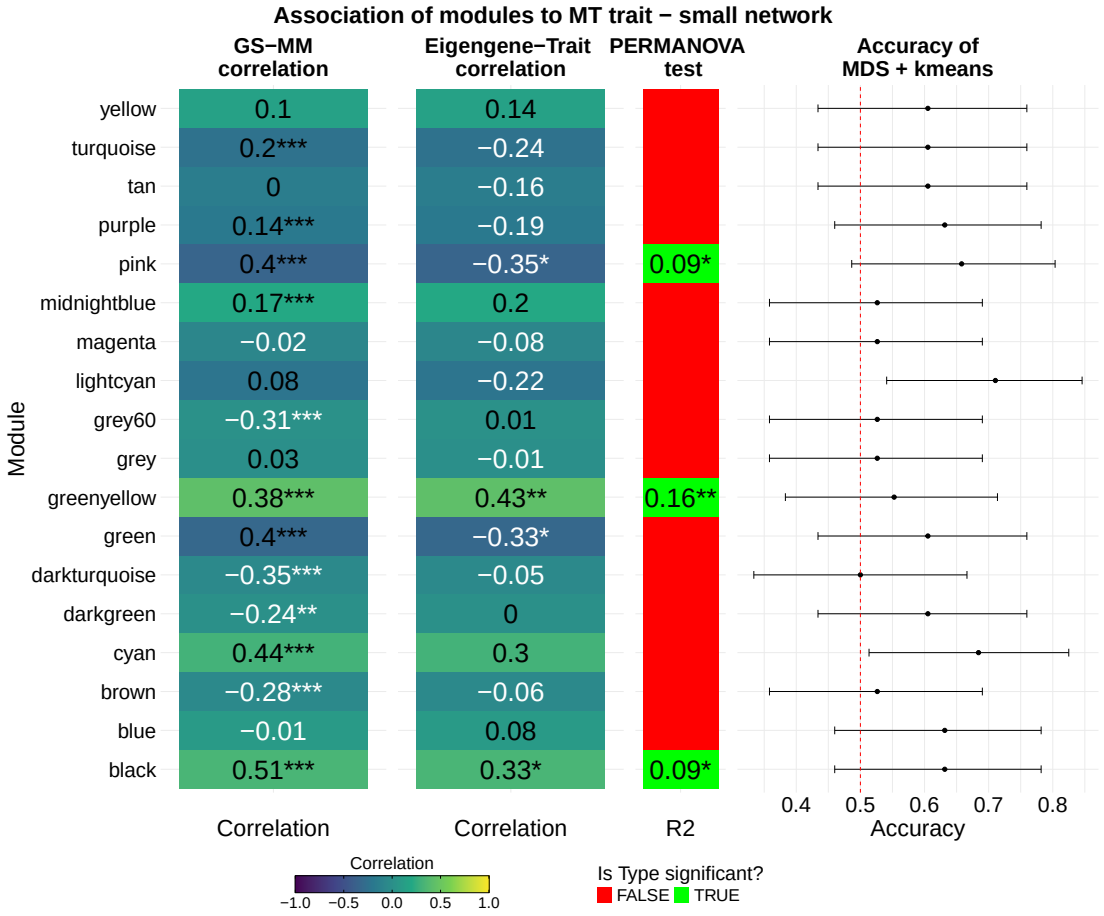


FIGURE 3.5: Different statistics used to determine which modules of the small network can contain genes associated to MT: correlation between gene significance and module membership, eigengene-MT correlation, PERMANOVA test, accuracy computed between the MT and a partition created applying K-means to the first 26 MDS axes.

strong associations with specific traits or because they significantly overlap with genes possessing particular GO functions or previously identified in other studies. The set of genes highlighted in [29, 252, 13] should be interpreted with caution, as those studies analyzed gene expression under conditions that differ slightly but fundamentally from ours. In those works, the differential gene expression analyses were designed to identify gene regulated in cells actively engaged in the sexual phase or responding to the presence of the opposite MT, conditions not represented in our dataset. Nevertheless, including these genes in our analysis provides valuable context for understanding their expression behavior in cells not engaged in sexual reproduction and for identifying other genes that are co-expressed with them under such conditions.

- **Black**: Exhibits one of the strongest and most consistent associations with mating type across all networks. It is enriched in genes involved in carbon assimilation, fatty-acid metabolism, ribosomal structure, and cofactor binding (iron, NAD, and pyridoxal phosphate). Expression levels of genes belonging to the fatty-acid and

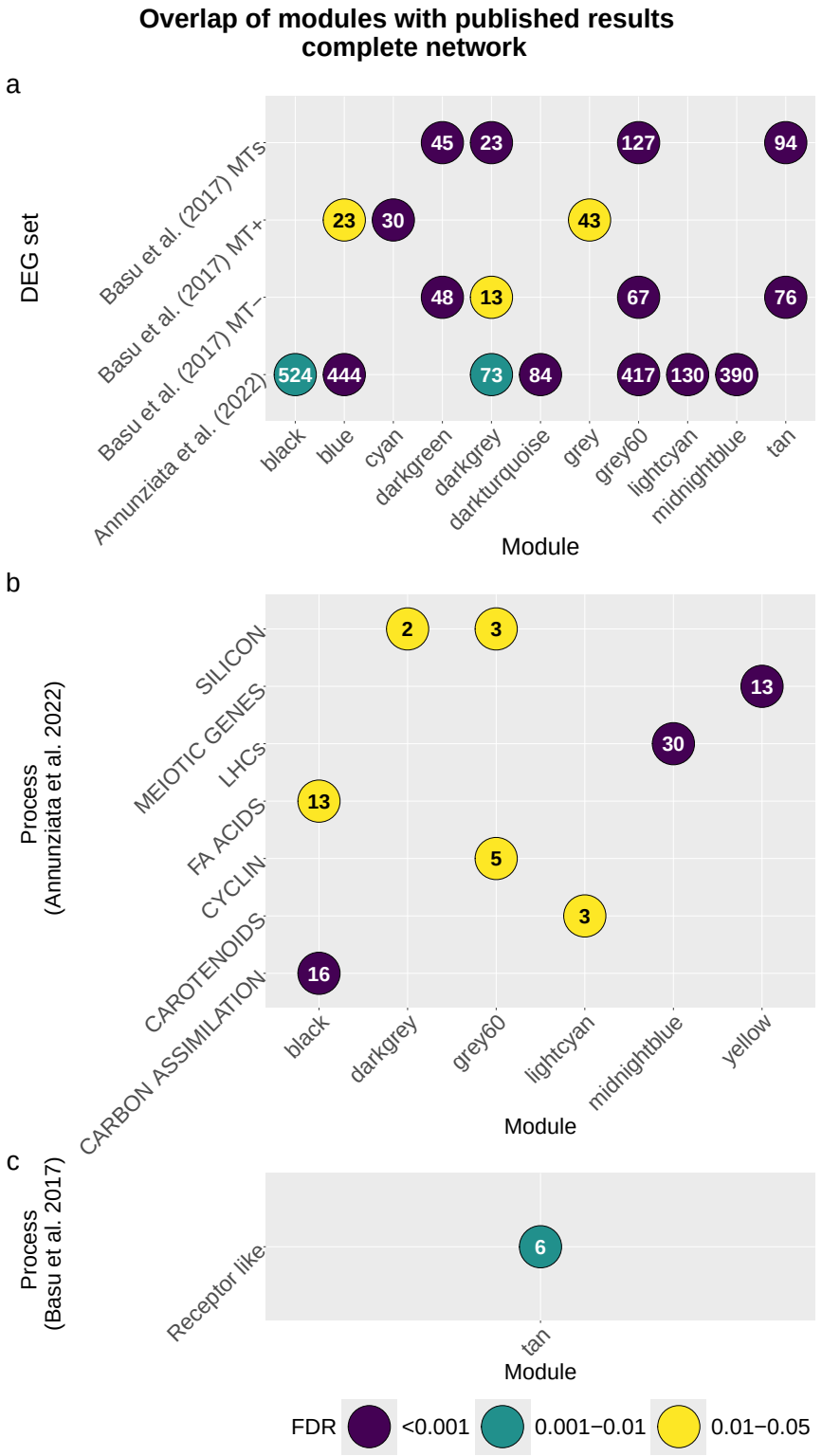


FIGURE 3.6: Overlap between modules in the complete network and groups of relevant genes found in published papaers about *P. multistriata*. (a) shows the overlap of the modules with sets of differentially expressed genes found in *Basu et al. (2017)* [29] and in *Annunziata et al. (2022)* [13]. (b) shows the enrichment of each module with genes belonging to various processes that were identified in *Annunziata et al. (2022)*. (c) shows the enrichment of each module with genes belonging to various processes that were identified in *Basu et al. (2017)*. The number inside each point is the amount of shared genes and colors indicate the range of the FDR.

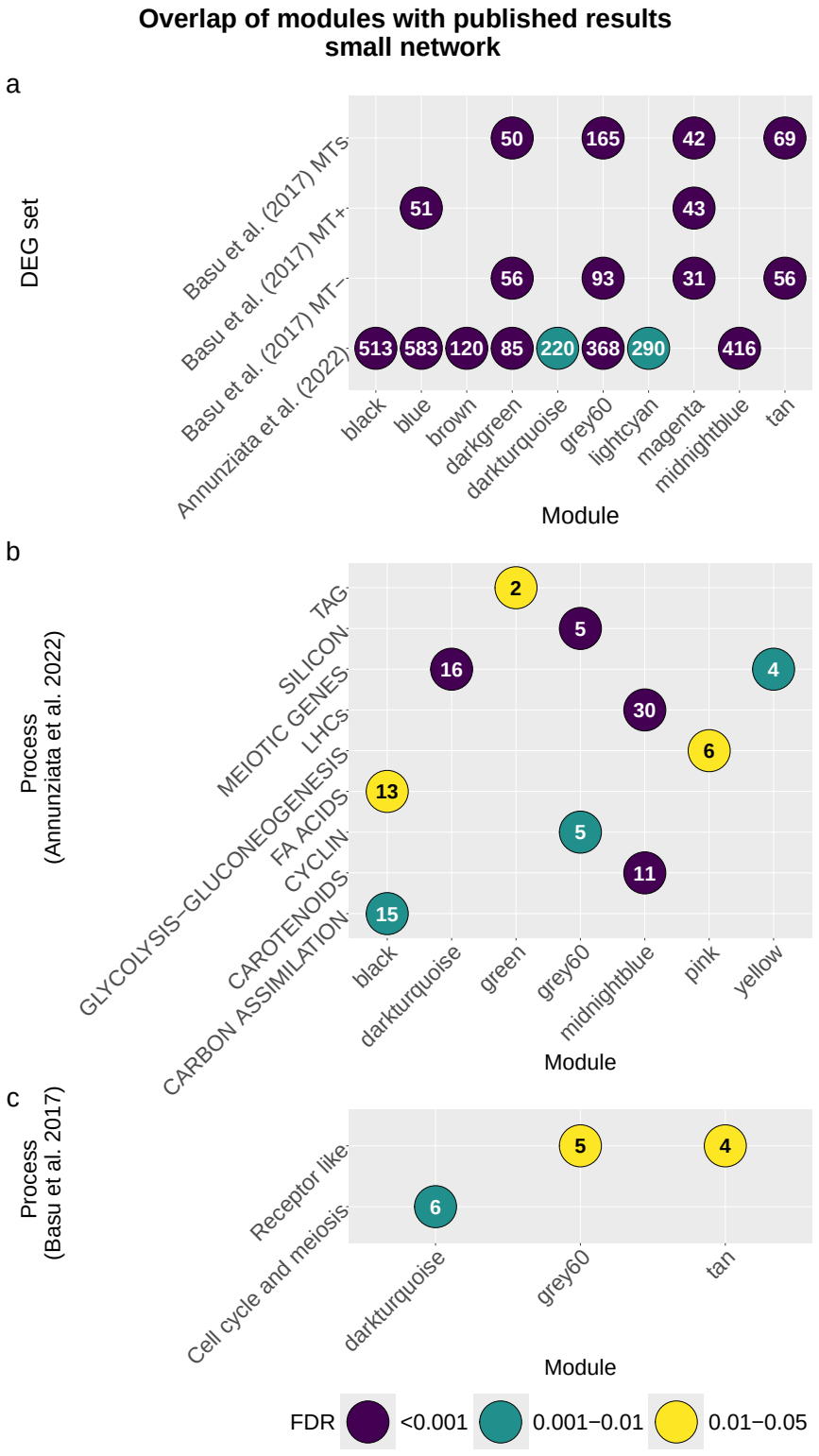
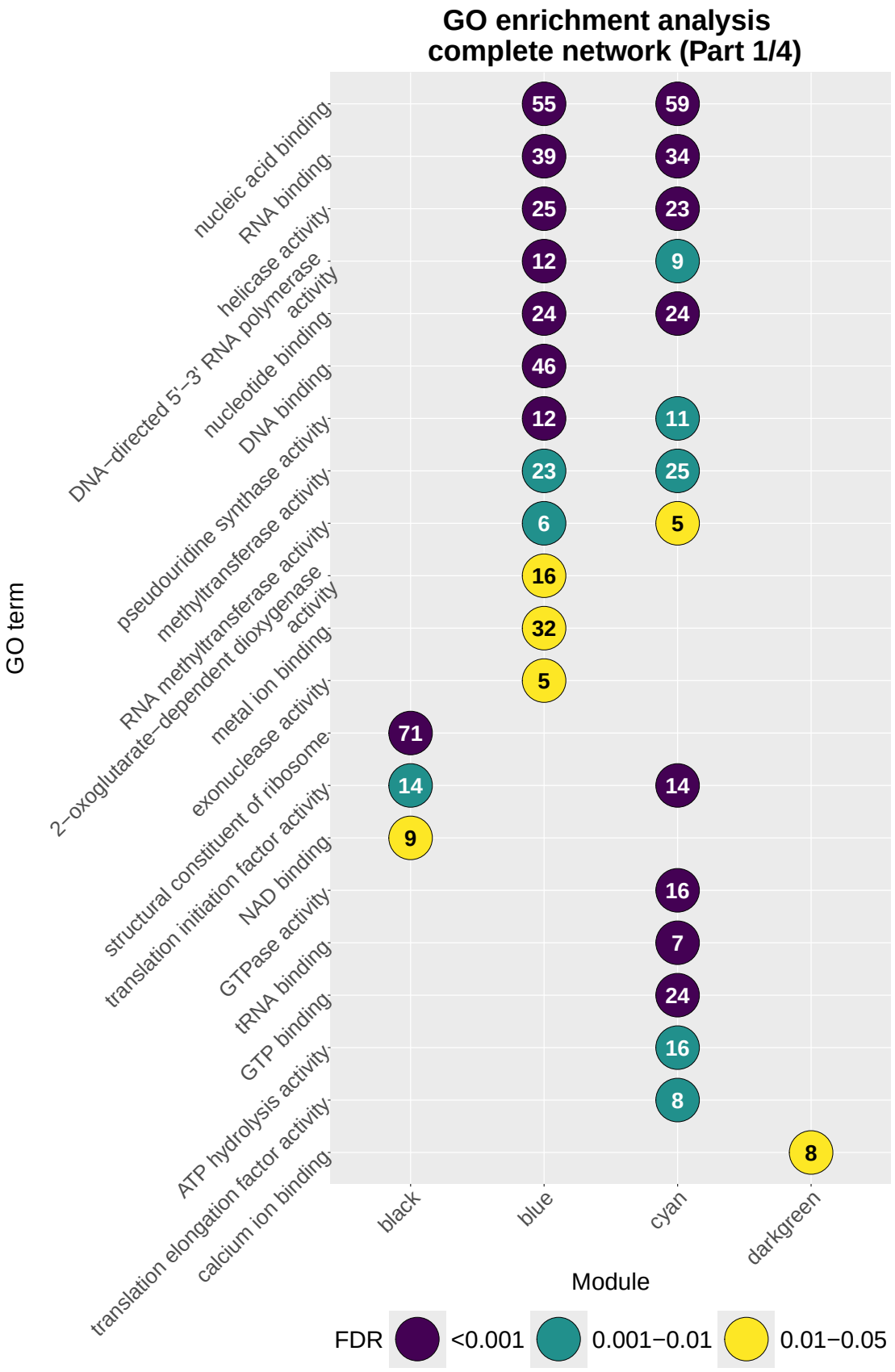
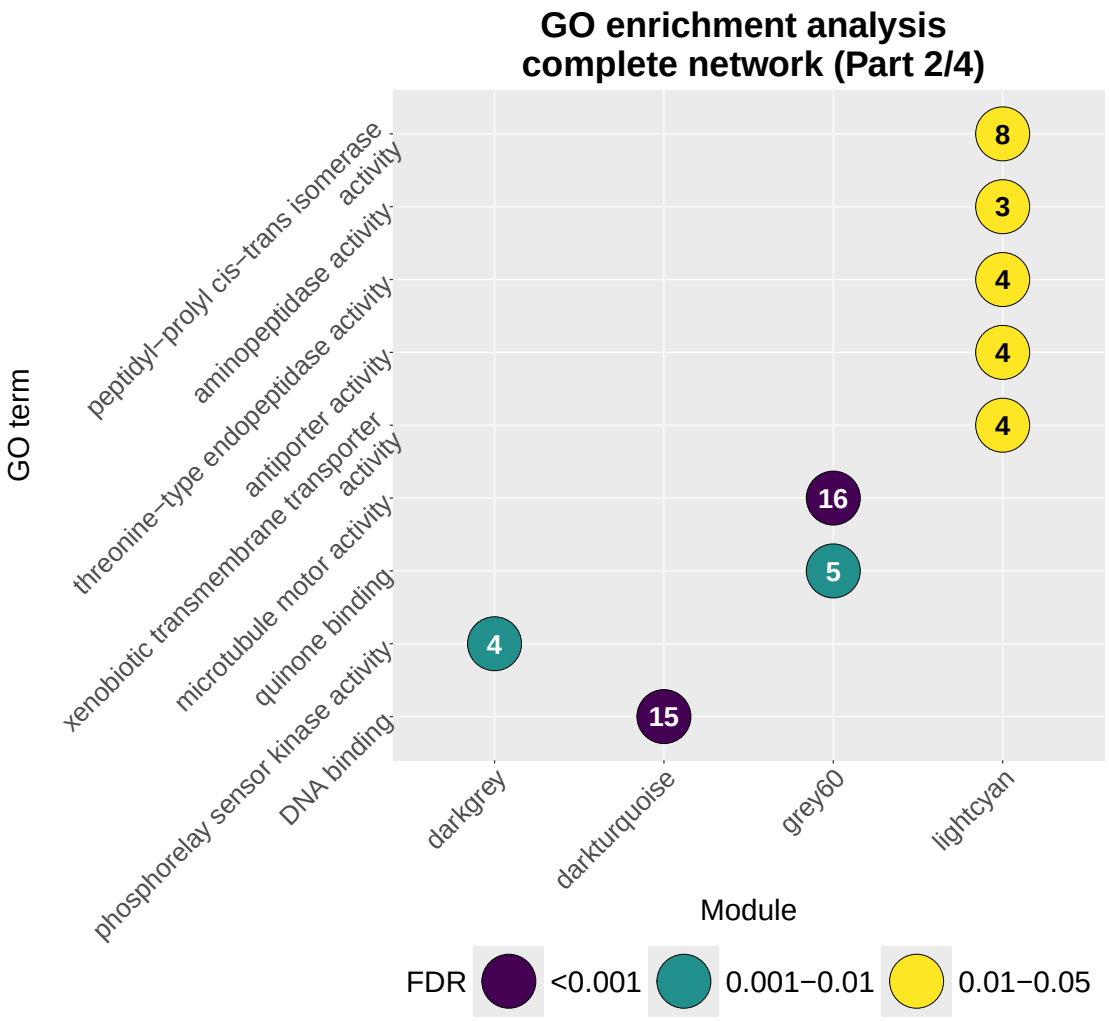


FIGURE 3.7: Overlap between modules in the small network and groups of relevant genes found in published papers about *P. multistriata*. Panels are analogous to Fig. 3.6.



(A) Enrichment analysis using GO for black, blue, cyan and darkgreen modules.

FIGURE 3.8: Enrichment analysis of the modules in the complete network using gene ontology (GO) information. (Continued)

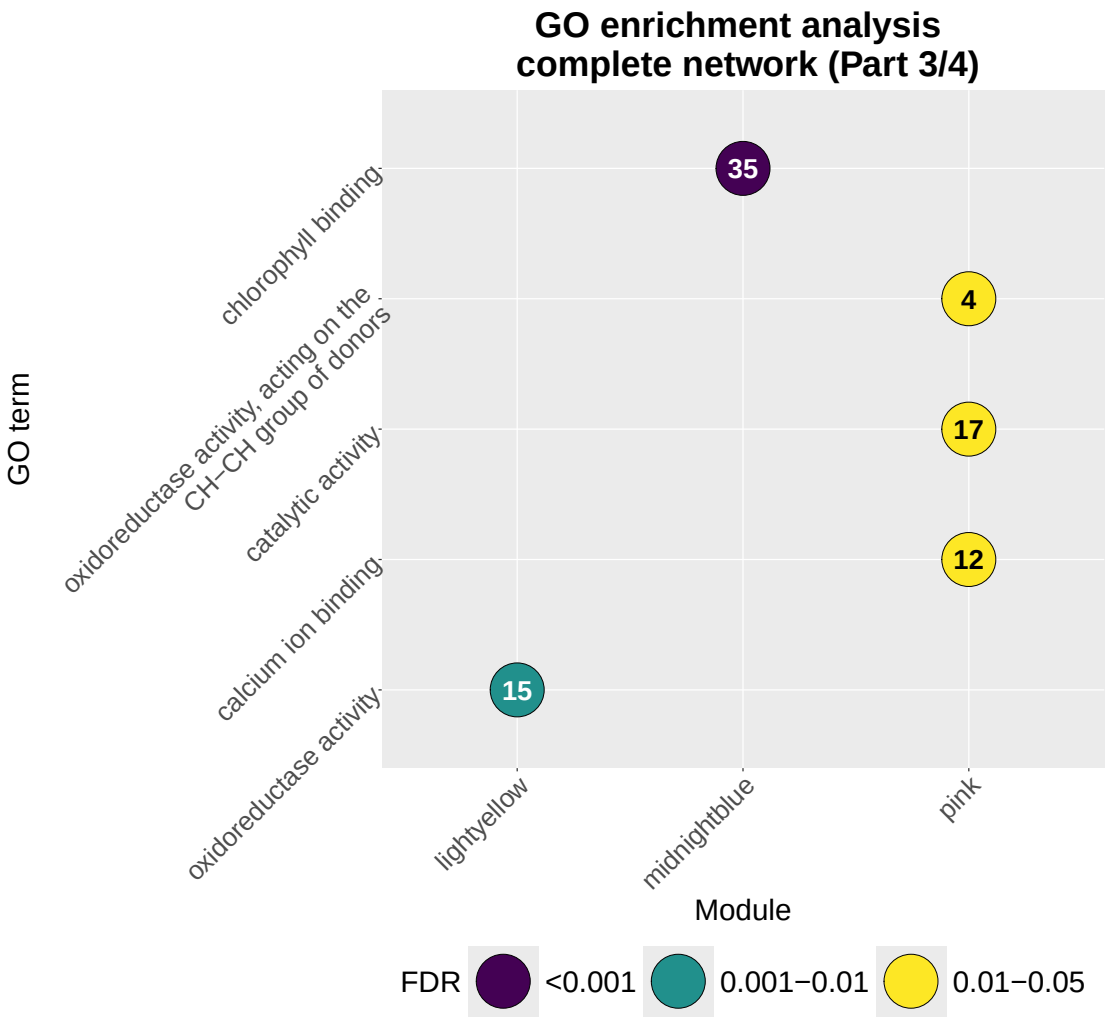


(B) Enrichment analysis using GO for darkgrey, darkturquoise, grey60 and lightcyan modules.

FIGURE 3.8: Enrichment analysis of the modules in the complete network using gene ontology (GO) information. (Continued)

carbon assimilation process were lower in MT– than in MT+ cells, even after accounting for experimental effects (Table B.1). This module shares many genes with DEGs down-regulated during the sexual phase in [13]. These observations indicate that the module contains genes involved in the metabolic arrest and resource reallocation accompanying pheromone perception, with MT– showing an earlier and stronger down-regulation of the metabolism than MT+.

- **Blue**: Although not significantly correlated with either trait, this module is strongly enriched in functions related to nucleic-acid binding, RNA helicases, polymerase and methyltransferase activity, and metal-ion binding. Its overlap with the off-regulated gene set from [13] suggests that it represents the global transcriptional changes that follow mating induction.



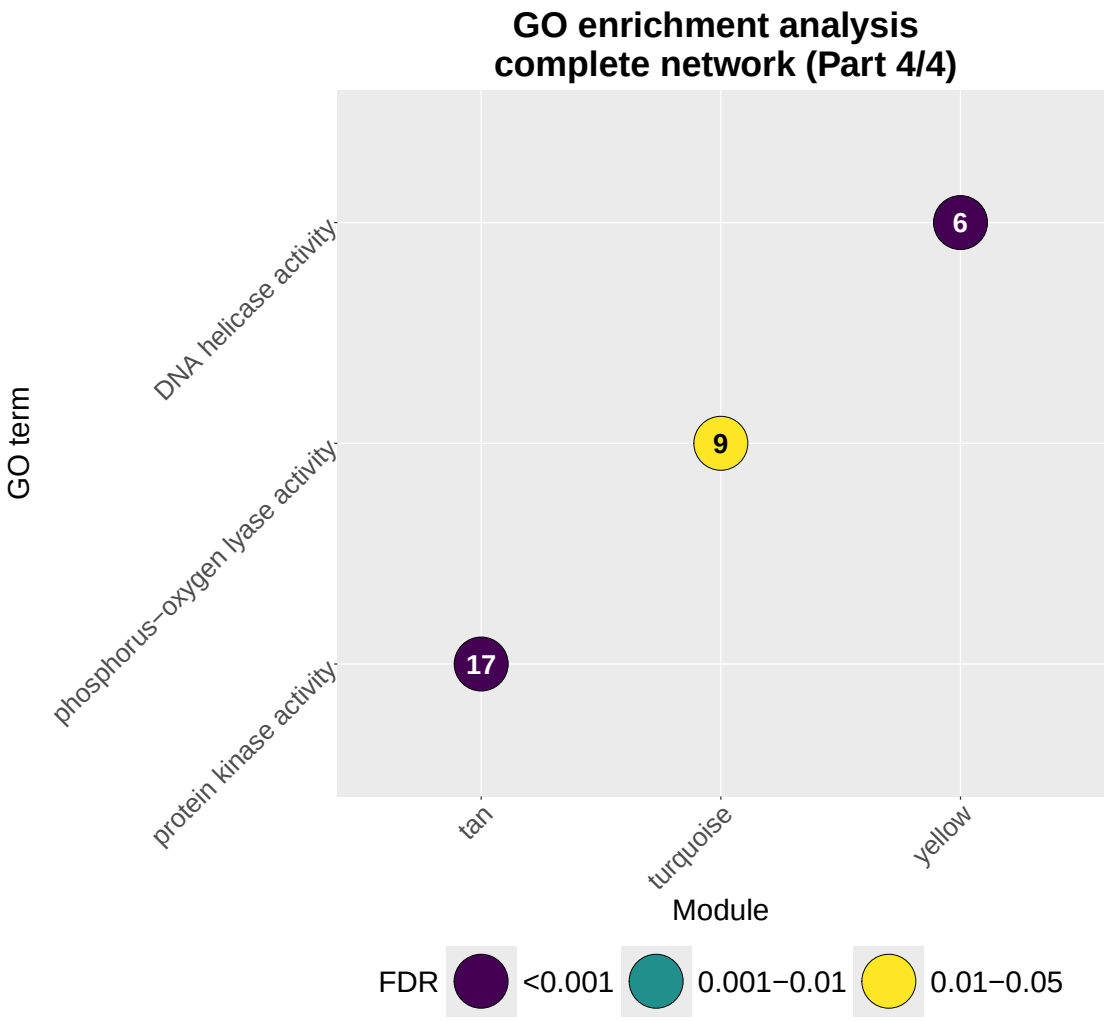
(C) Enrichment analysis using GO for lightyellow, midnightblue and pink modules.

FIGURE 3.8: Enrichment analysis of the modules in the complete network using gene ontology (GO) information. (Continued)

- **Cyan**: Weakly correlated with MT and containing genes up-regulated in MT+ as found in [29], probably related to exposure to MT– cues. Functional enrichment includes RNA and nucleotide binding, GTP-binding, methyltransferase, and translation-initiation factors, consistent with early modulation of the translational machinery during signal perception.

- **Darkgrey**: Present only in the complete network and weakly correlated with MT. Limited functional information is available: two genes belong to the "silicon" process, and the module is enriched in phosphorelay sensor kinase activity. This might indicate a weak or spurious signal. It is a small module, and nearly two-thirds of its genes are unannotated.

- **Darkturquoise**: Although not significantly associated with any specific trait, this module contains several meiosis-related genes from the "meiotic toolkit" and is



(D) Enrichment analysis using GO for tan, turquoise and yellow modules.

FIGURE 3.8: Enrichment analysis of the modules in the complete network using gene ontology (GO) information.

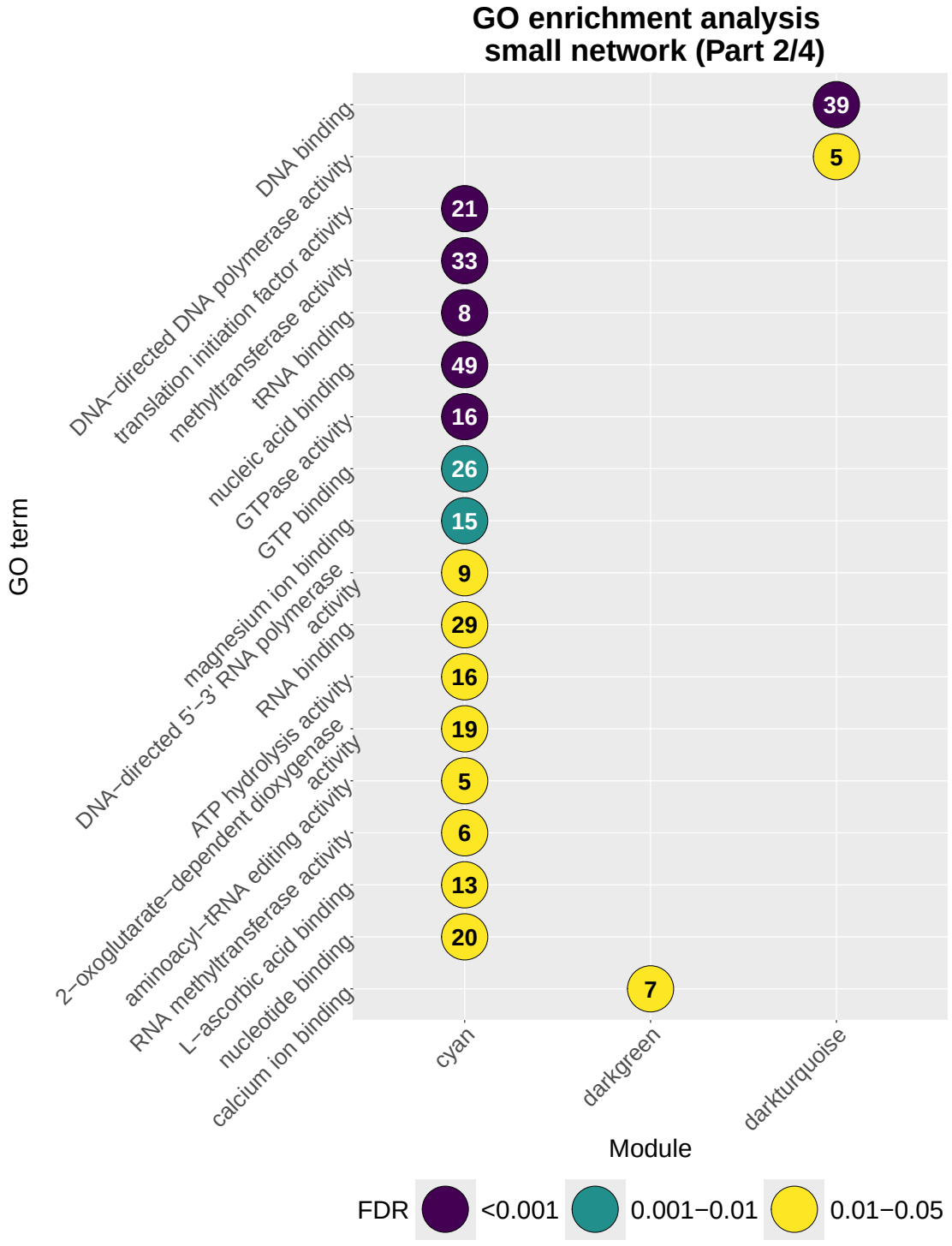
enriched in DNA-binding and recombination terms. This suggests that it captures the onset of meiotic gene expression occurring in a subset of cells after gametogenesis is triggered.

- **Green**: Related to MT but not associated with any particular group of DEGs. It is not enriched in the complete network and only enriched in hydrolase activity and hydrolyzing O-glycosyl compounds in the small network. It may contain genes not previously identified, potentially associated with differences across all cell sizes.

- **Greenyellow**: Appears associated with both size and MT but is not enriched in either network, nor associated with any set of DEGs or processes. It may include genes linked to these traits that do not respond to the onset of sexualization.

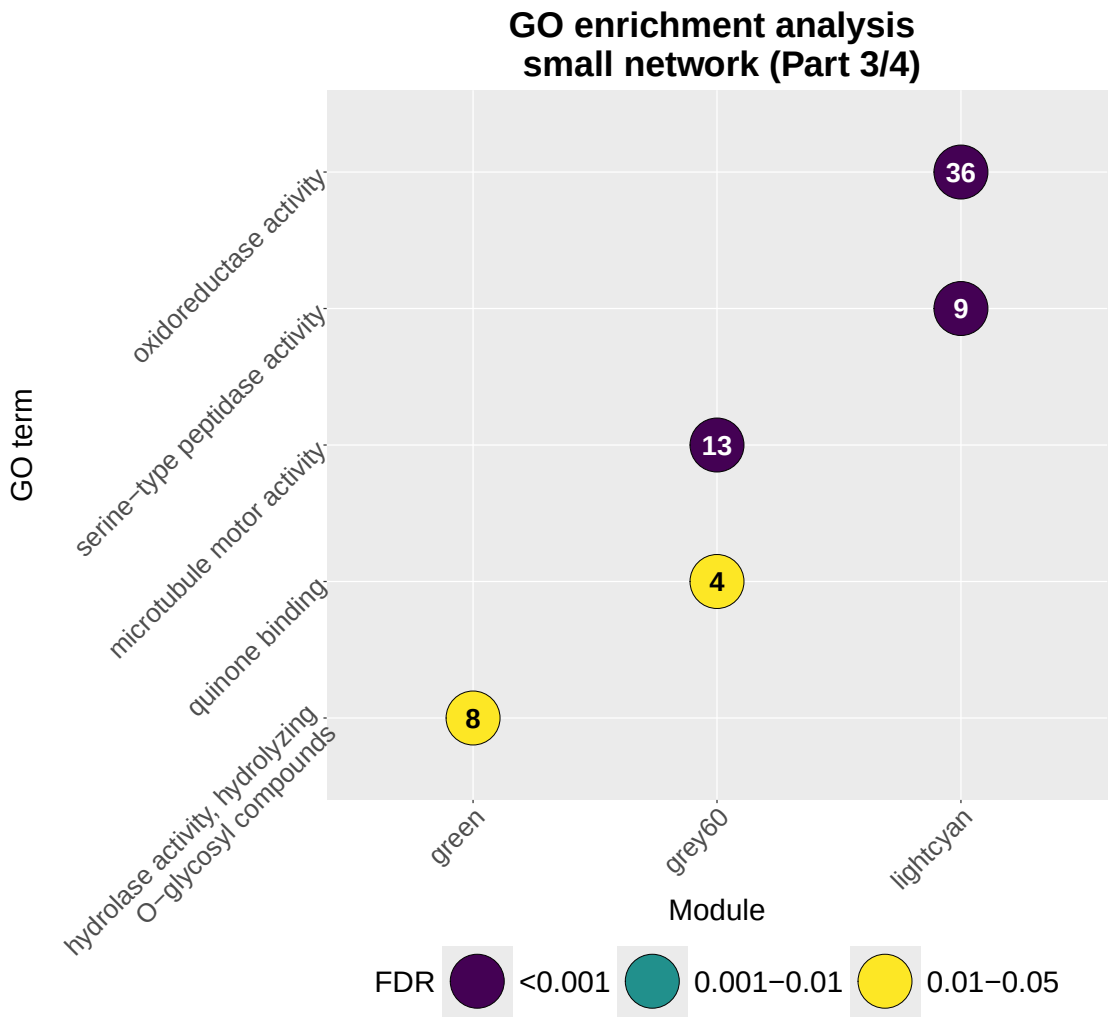
(A) Enrichment analysis using GO for black, blue and brown modules.

FIGURE 3.9: Enrichment analysis of the modules in the small network using gene ontology (GO) information. (Continued)



(B) Enrichment analysis using GO for cyan, darkgreen and darkturquoise modules.

FIGURE 3.9: Enrichment analysis of the modules in the small network using gene ontology (GO) information. (Continued)

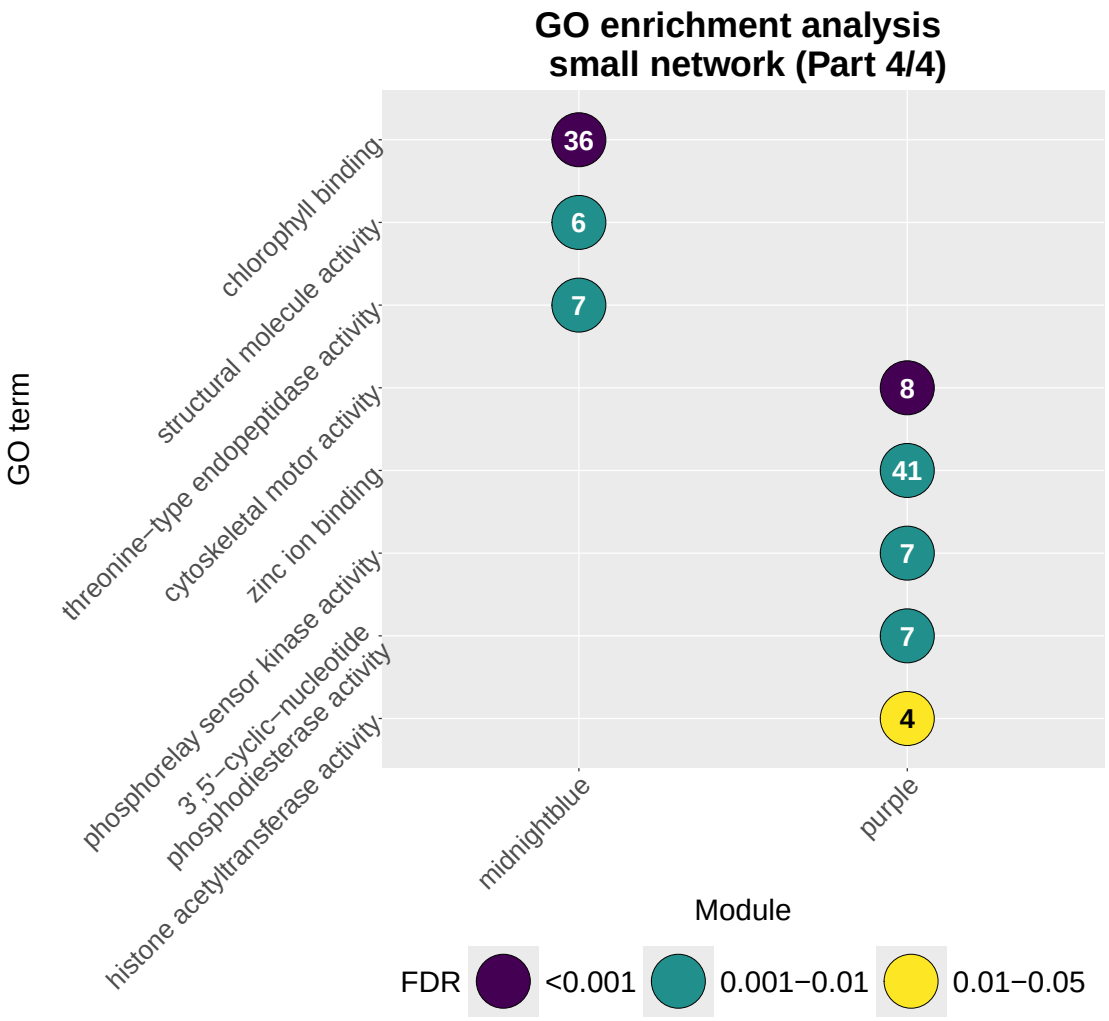


(C) Enrichment analysis using GO for green, grey60 and lightcyan modules.

FIGURE 3.9: Enrichment analysis of the modules in the small network using gene ontology (GO) information. (Continued)

- **Midnightblue**: Not associated with any trait but enriched in chlorophyll-binding genes. It shows strong overlap with the *LHCs* process and with genes from [13]. It likely contains genes activated below the SST, several of which are related to photosynthesis. Because cells in these samples were not sensing pheromones, these genes that are down-regulated during the perception of opposite MTs do not exhibit any significant association with MT in this context.

- **Orange**: Associated with size but shows no overlap with any process or enrichment in GO terms. The PERMANOVA test (B.2) indicates that most of the variation ($\sim$ 78%) is explained by experimental effects. Indeed, most genes in this module show the lowest expression levels across the s1–s14 samples: the association with size, although present, might be an artifact.



(D) Enrichment analysis using GO for midnightblue and purple modules.

FIGURE 3.9: Enrichment analysis of the modules in the small network using gene ontology (GO) information.

- **Pink**: Significantly correlates with MT in both the complete and small networks and is enriched for oxidoreductase and calcium-ion binding activities. It shows no significant overlap with sets of DEGs. As this module is strongly associated with MT in both networks but not with size, it may contain genes that behave differently between MTs below the SST but were not previously identified.

- **Turquoise**: Associated with MT in the complete but not in the small network. It is not associated with any set of DEGs and is slightly enriched in phosphorus–oxygen lyase activity. These genes may exhibit differences between MTs across the entire size spectrum and are more highly expressed in MT– above the SST.

3.3.2 bootstrap

Because the number of large-cell (i.e., above-SST) samples was limited, the statistical power to detect modules associated with cell size was expected to be low. To evaluate the robustness of the inferred co-expression structure under varying sample compositions, we constructed ten *bootstrap networks* by resampling subsets of below-SST samples (n = 11 per subset) together with all above-SST samples. Each resampled dataset was analyzed independently using the same WGCNA pipeline described above. Due to the smaller sample size, achieving a scale-free topology fit ($R^2 > 0.9$) would have required excessively sparse networks; therefore, we adopted a relaxed criterion of $R^2 \sim 0.8$ with soft-thresholding powers β between 9 and 12. Despite this adjustment, all networks exhibited well-defined modular structures, confirming that co-expression patterns are largely consistent across resampled datasets. To quantify the similarity among module partitions, three independent clustering comparison metrics—the Adjusted Rand Index (ARI), the Variation of Information (VI), and the V-measure—were calculated for all pairs of networks. The rationale was to identify robust clusters and to select a representative network for each cluster. Hierarchical clustering based on these similarity measures grouped the ten bootstrap networks into four major classes (Fig. 3.10). These four groups showed rather consistent patterns: networks 1 and 6 were always clustered together, network 4 consistently formed a separate branch, while networks 10 and 5, and 2, 8, and 7, displayed high mutual similarity (Fig. 3.10). Based on the ARI and V-measure results, we adopted the following partition: {8, 2, 7}, {1, 6}, {10, 3, 5}, and {4}. One representative network was selected from each class, prioritizing those with the smallest number of modules to facilitate interpretation: network 2 (19 modules), network 6 (21 modules), network 3 (19 modules), and network 4 (21 modules). Across all four representative networks, no clear module could be associated with the size trait. For this reason, we did not perform a renaming of module colors in these networks. Across these representative networks, the *black module* was consistently recovered and retained its characteristic enrichment in structural ribosomal proteins, NAD-binding enzymes, and pyridoxal-phosphate–dependent oxidoreductases (Figs. 3.11 to 3.14). This module repeatedly showed association with mating type and displayed strong overlap with the black module identified in both the complete and small networks. As observed previously, these bootstrap networks also yielded more modules related to MT than to size. In the following, we describe in detail networks 2 and 3, as additional discussion of networks 4 and 6 would not provide further insights into MT or size. We focus here primarily on the size trait, since associations with MT have already been extensively examined in previous analyses.

Regarding network 2 we have:

- **Darkorange**: Associated with MT. It contains genes down-regulated in MT– as reported in [29] (Fig. 3.15). No GO term enrichment was detected. These genes are likely related to pheromone sensing in MT–.
- **Darkred**: The eigengene correlates with size and is enriched in chlorophyll-binding genes. As observed previously, these genes may be involved in the down-regulation of photosynthesis. The PERMANOVA test (B.4) indicates that the experimental factor explains approximately $\sim 46\%$ of the variance, while size explains about $\sim 13\%$, suggesting that the size-related signal is likely genuine.
- **Lightcyan**: Correlated with MT, which explains $\sim 31\%$ of the variance (Fig. 3.11). The module overlaps with DEGs from [13] but is not enriched for any GO process. It may represent a candidate set for pheromone-related genes.
- **Royalblue**: PERMANOVA indicates that its genes are associated with both size and MT, with the latter explaining a larger portion of variance (Figs. 3.11 and 3.12). The module shows no enrichment for known processes. Its genes might be linked to metabolic changes occurring after cells pass the SST, as well as to pheromone sensing.

For network 3, only the module **bisque4** is noteworthy. It is associated with both MT and size, with the former explaining a larger proportion of variance (Figs. 3.13 and 3.14). Although not enriched in any GO process or overlapping with known DEG sets, this module contains several genes related to *chlorophyll binding*.

3.3.3 Expression of MT-related genes

In Sec. 3.1.1, we discussed five genes involved in the genetic program underlying heterothallism in *P. multistriata*. Figure 3.19 presents their normalized expression levels and the modules to which they belong in each of the four network configurations analyzed.
In the complete network, all five mating-type-related genes (*MRP1*, *MRP2*, *MRP3*, *MRM1* (both models), and *MRM2*) passed the initial filtering and were included in the network analyses. In the small network, *MRP3* was excluded, consistent with its strong upregulation in MT+ samples below SST, where most of the samples correspond to MT-. Of the two *MRM1* models, only one (id 41130) remained, while the other (id 8538) was removed. In contrast, in bootstrap networks 2 and 3, both *MRM1* models and *MRP3* were excluded.

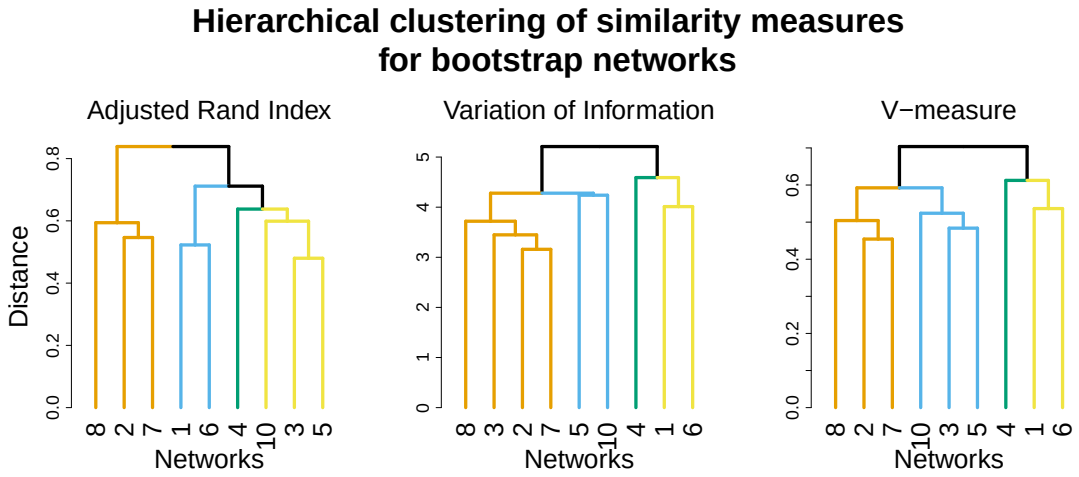


FIGURE 3.10: Hierarchical clustering (Ward linkage) for three association measures (adjust Rand index, variation of information, V-measure) between the modules obtained from the 10 bootstrap networks. Colors identify the 4 groups containing similar module structures.

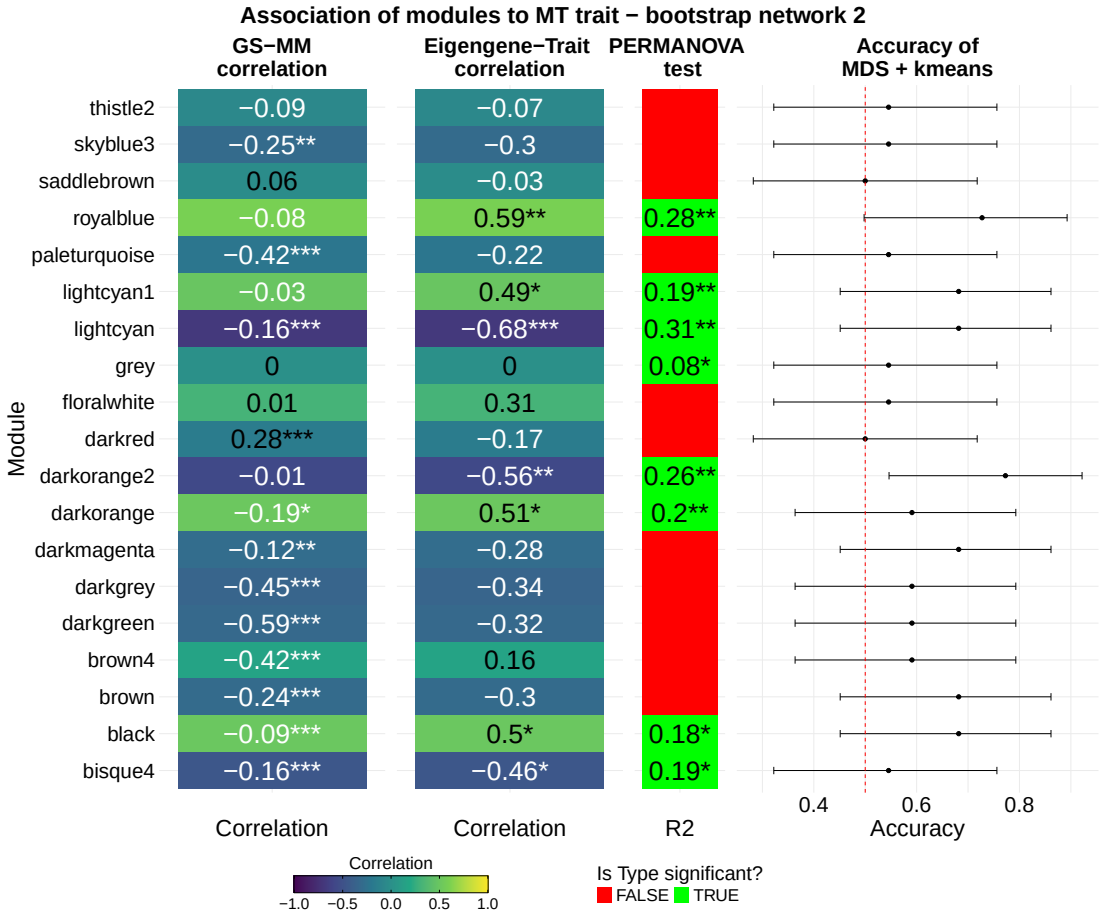


FIGURE 3.11: Different statistics used to determine which modules of the bootstrap network 2 can contain gene associated to MT: correlation between gene significance and module membership, eigengene-MT correlation, greatest amount of variance explained by MT in a PERMANOVA test, accuracy computed between the MT and a partition created applying Kmeans to the first 11 MDS axes.

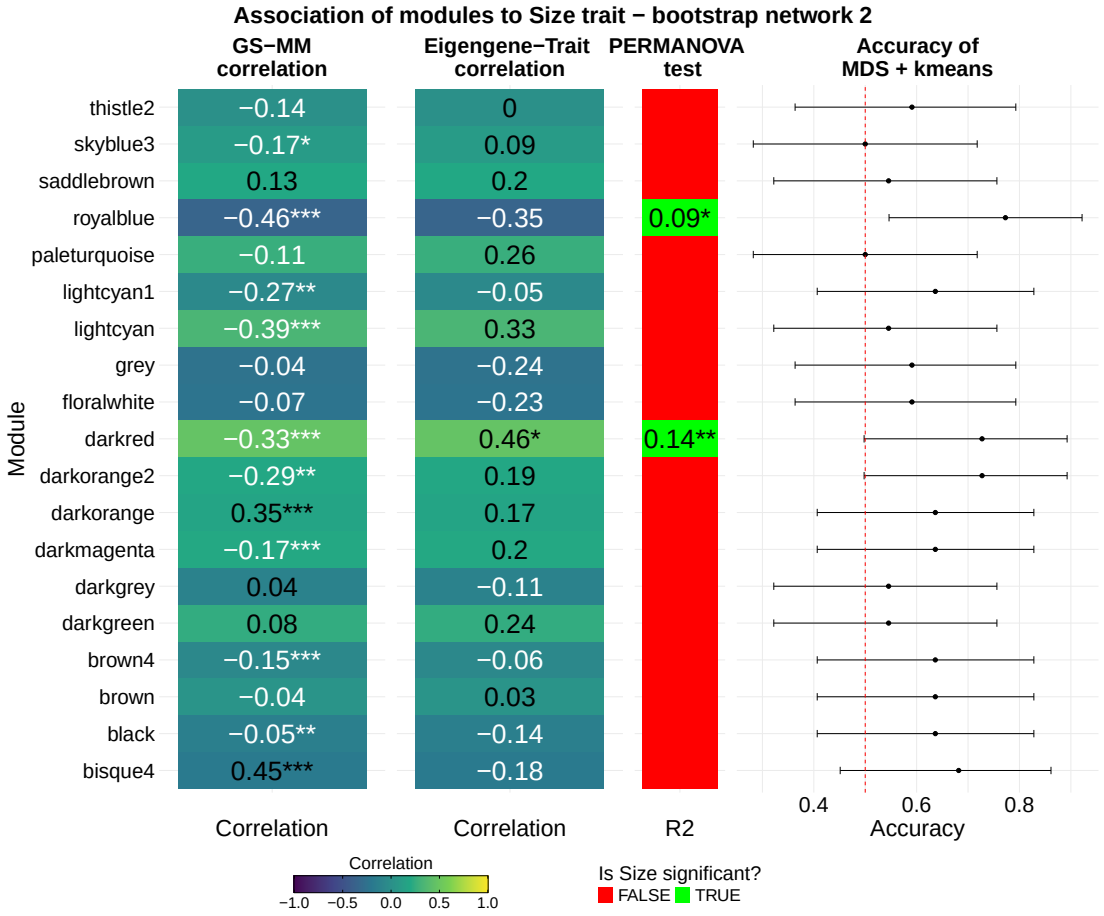


FIGURE 3.12: Different statistics used to determine which modules of the bootstrap network 2 can contain gene associated to size: correlation between gene significance and module membership, eigengene-MT correlation, greatest amount of variance explained by size in a PERMANOVA test, accuracy computed between the size and a partition created applying Kmeans to the first 11 MDS axes.

Regarding expression patterns, *MRP3* displayed distinctly higher expression in MT+ samples below SST and was almost always at its lowest levels in MT- samples. *MRP1* and *MRP2* showed more uniform expression across conditions, although their highest levels were also observed in MT+ samples below SST, in agreement with previous findings [252]. Both *MRM1* models exhibited generally low expression, with localized increases in a small subset of samples. *MRM2* also showed very low expression overall, peaking in a few samples where *MRM1* had also non-minimum expression; notably, the inverse was not observed (samples from lv193_T2_B downward to B939_C_T1).
Module assignment varied across networks. In the complete network, *MRP1* and *MRP2* were placed in the grey module, which typically contains genes that do not strongly correlate with other genes. Given their relatively uniform expression profiles, this assignment is reasonable and may suggest that heterothallism operates independently of the primary metabolic pathways represented in the other modules. *MRM2* was assigned to the orange module, which we previously noted may be associated with cell

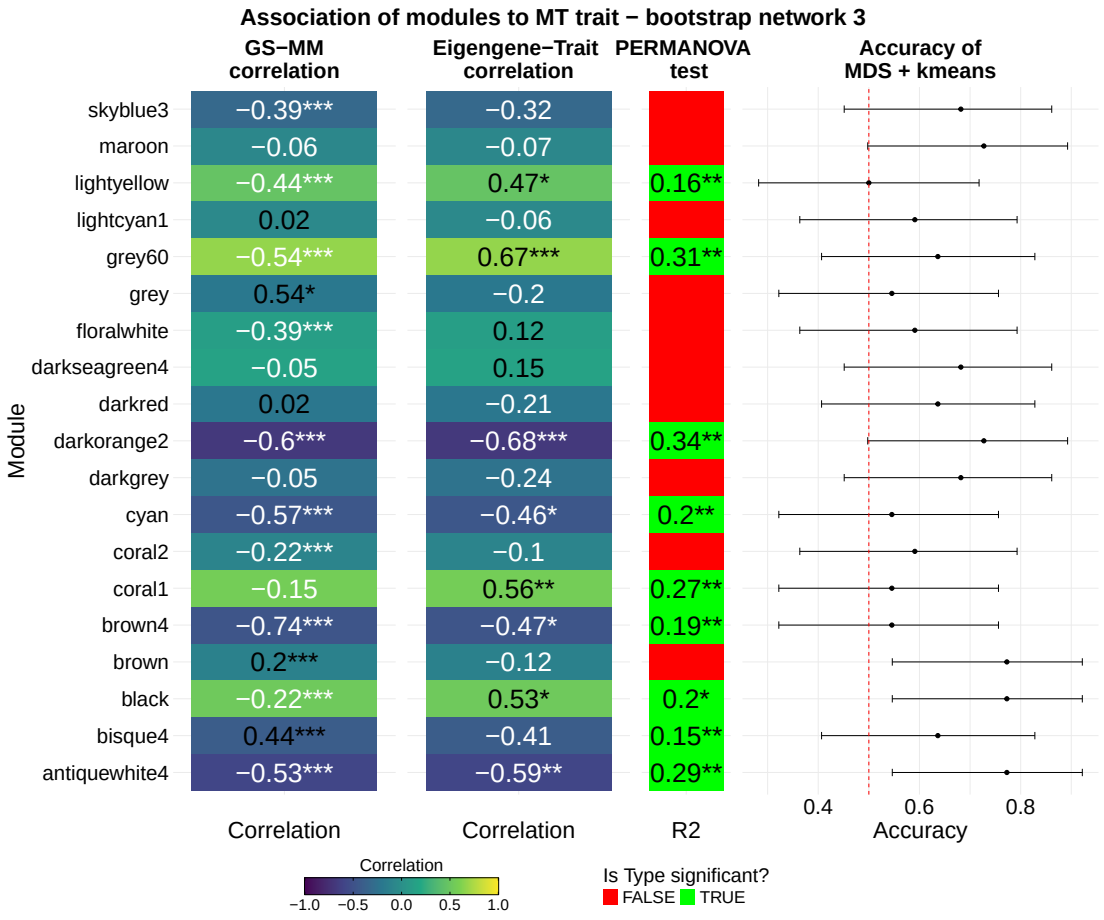


FIGURE 3.13: Different statistics used to determine which modules of the bootstrap network 3 can contain gene associated to MT: correlation between gene significance and module membership, eigengene-MT correlation, greatest amount of variance explained by MT in a PERMANOVA test, accuracy computed between the MT and a partition created applying Kmeans to the first 10 MDS axes.

size or possibly affected by batch effects. *MRM1* and *MRP3*, in contrast, were part of the greenyellow module, which is linked to MT; however, their expression profiles diverged except under PTX/noPTX conditions. In the small network, *MRP2* and *MRM1* were again assigned to the grey module, though in this case the inclusion of *MRM1* is less easily explained. *MRM2* was assigned to the turquoise module, which showed no strong association with MT, while *MRP1* belonged to the midnightblue module, which displayed a weak correlation with MT. In both bootstrap networks, *MRP1* and *MRM2* were assigned to the brown module, which was mildly correlated with MT in network 3. Finally, *MRP2* belonged to the lightcyan and antiquewhite4 modules, respectively in bootstrap network 2 and 3, and both module correlate to MT.

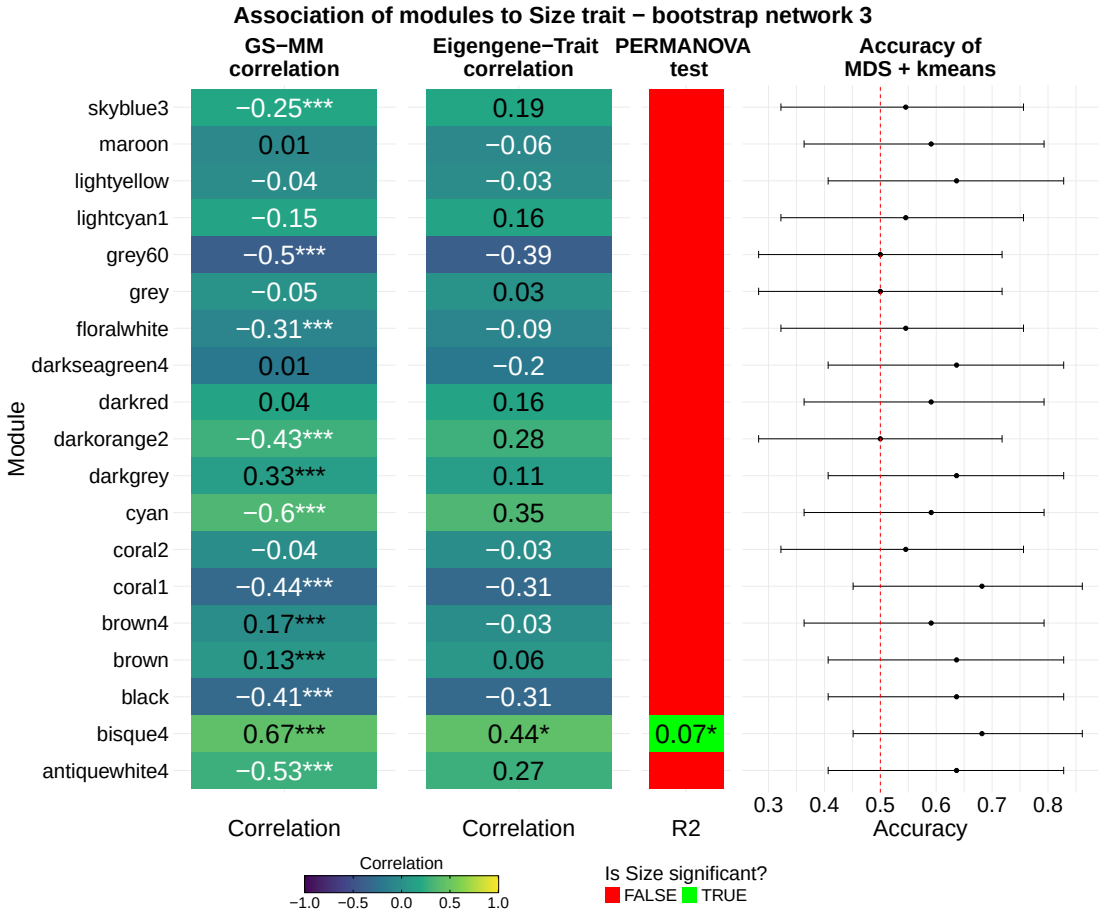


FIGURE 3.14: Different statistics used to determine which modules of the bootstrap network 3 can contain gene associated to size: correlation between gene significance and module membership, eigengene-MT correlation, greatest amount of variance explained by size in a PERMANOVA test, accuracy computed between the size and a partition created applying Kmeans to the first 10 MDS axes.

3.4 Discussion

The work presented in this thesis provides a system-level perspective on gene co-expression in *Pseudo-nitzschia multistriata*, focusing on the molecular programs that define mating-type identity and regulate the transition across the sexualization size threshold (SST). Previous studies have demonstrated that this diatom undergoes a profound metabolic reprogramming in preparation for sexual reproduction: nutrient uptake and photosynthetic activity are attenuated, pheromone production and perception are activated, and transcriptional profiles shift toward processes required for gametogenesis and zygote formation. On this basis, one expects marked transcriptional differences between small and large cells, reflecting their distinct physiological states, and between the two mating types (MTs) below the SST, as each experiences sexualization differently. Yet, because *P. multistriata* is unicellular and both MTs are morphologically

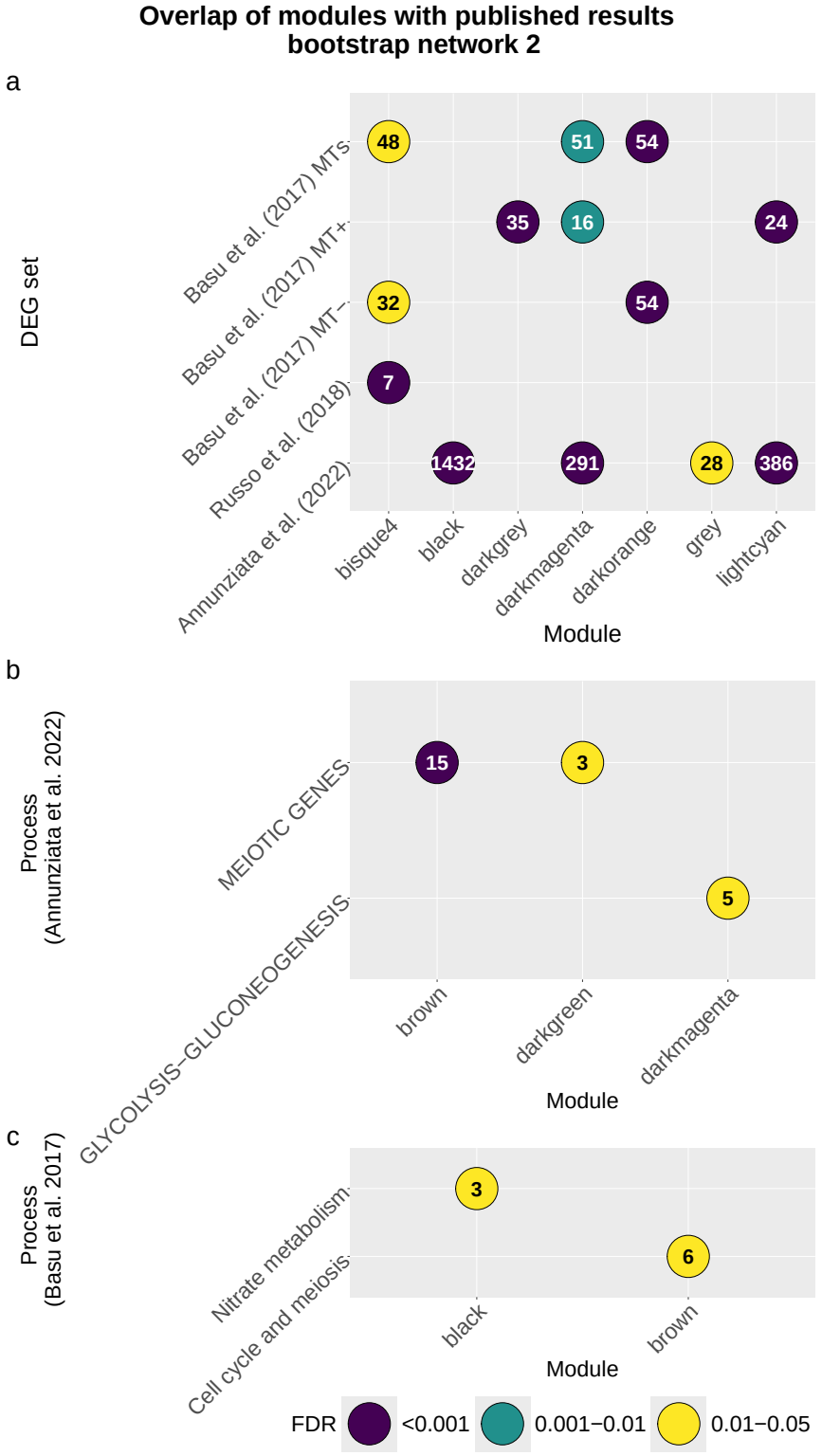


FIGURE 3.15: Overlap between modules in the bootstrap network 2 and groups of relevant genes found in published papaers about *P. multistriata* (see caption of Fig. 3.6 for details).

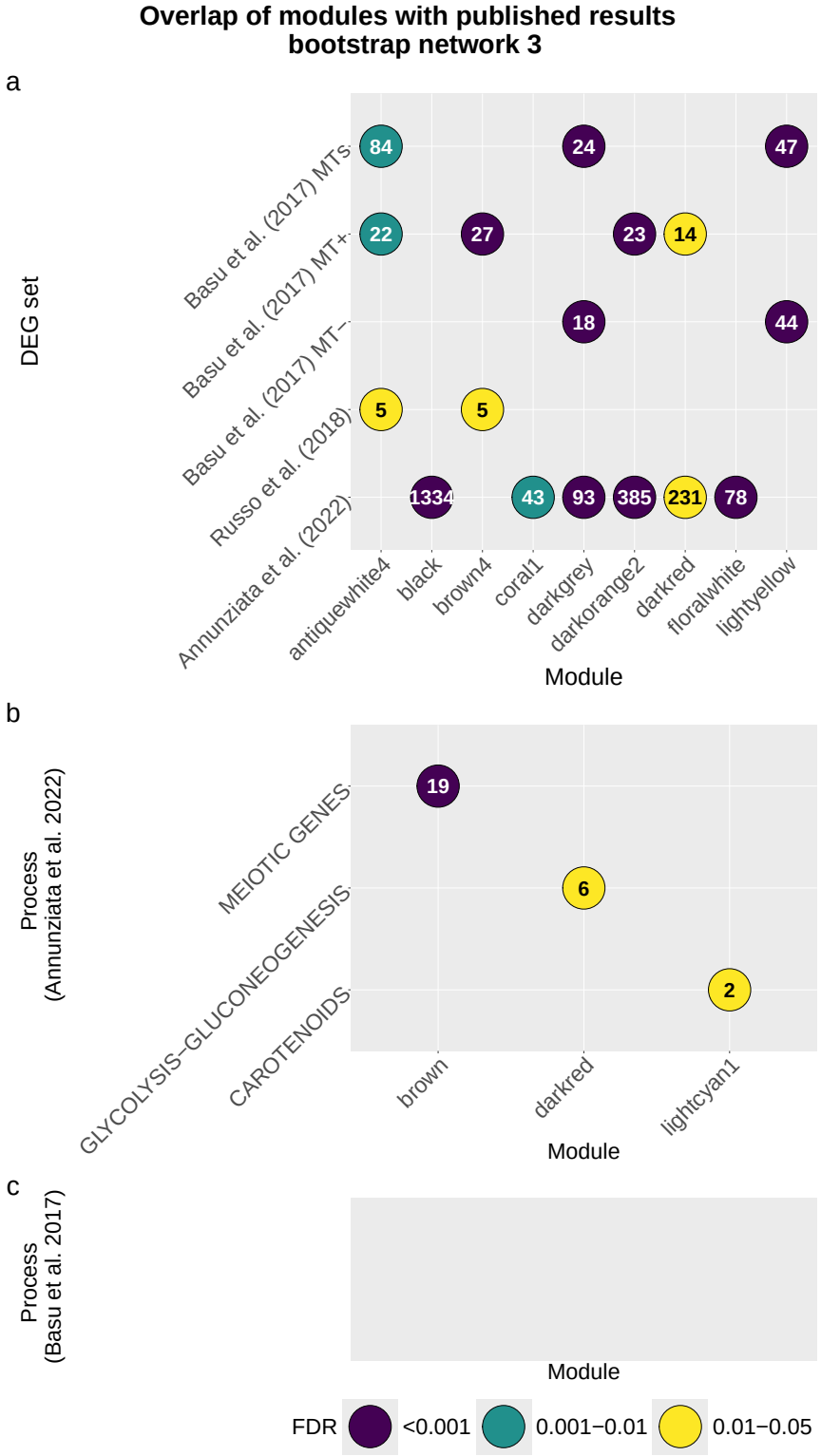
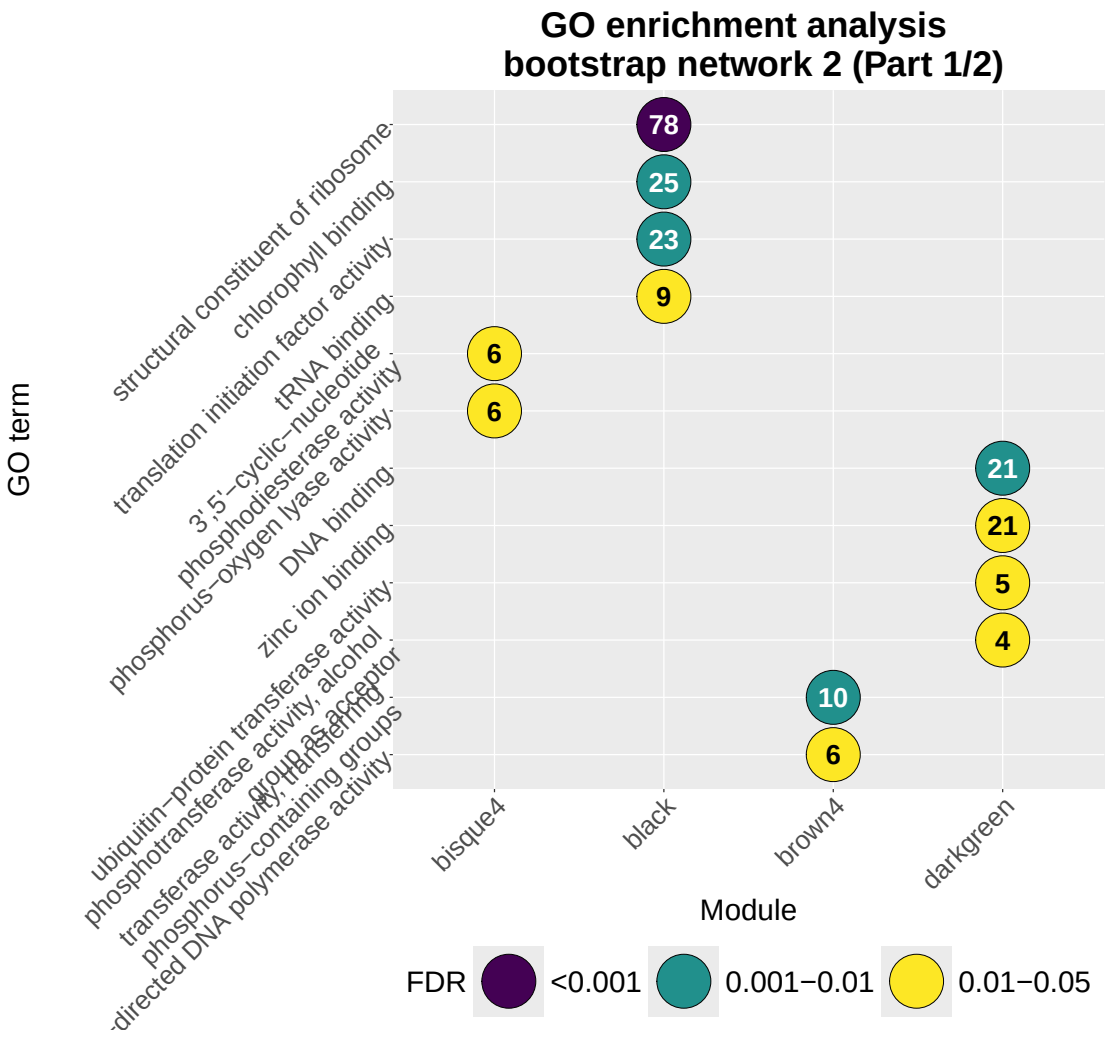


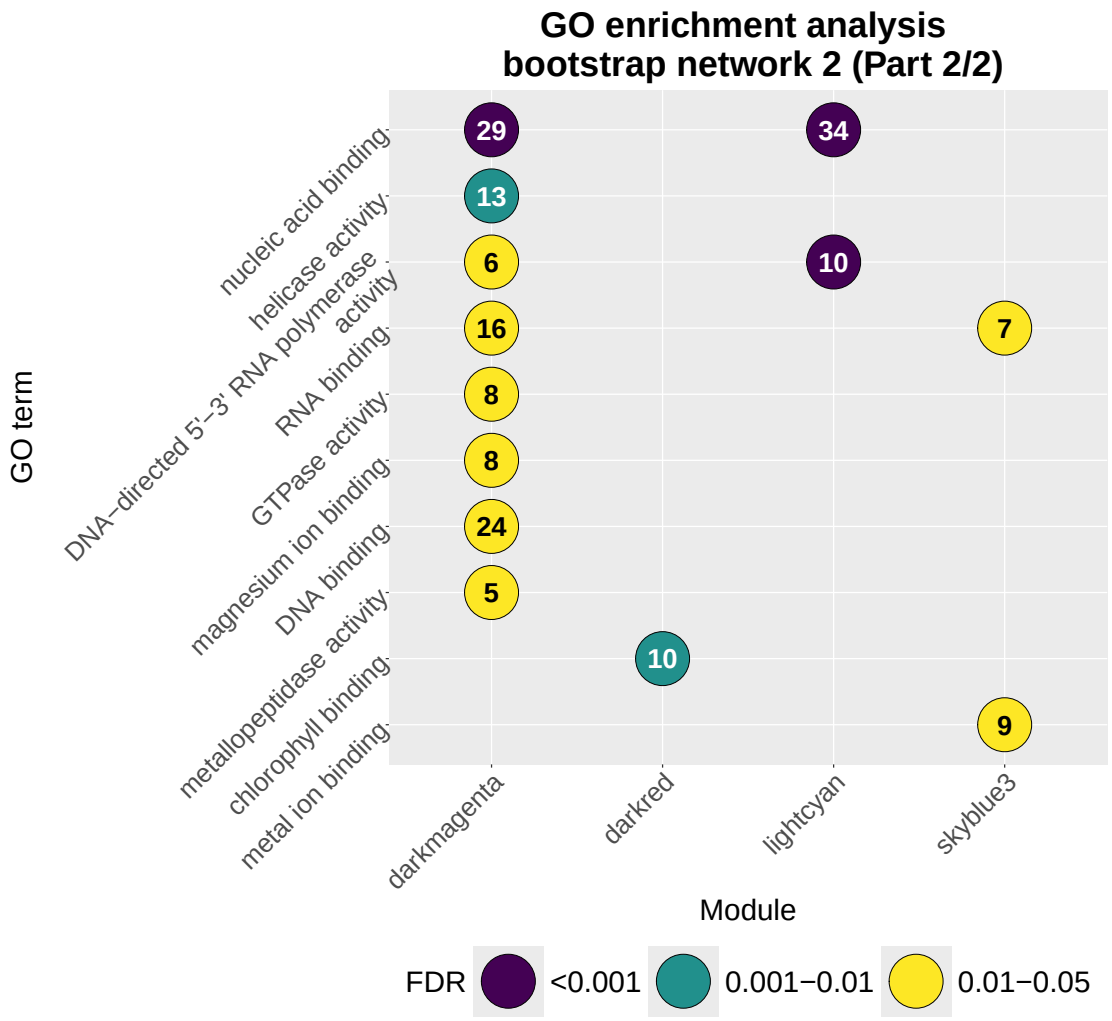
FIGURE 3.16: Overlap between modules in the bootstrap network 3 and groups of relevant genes found in published papaers about *P. multistriata* (see caption of Fig. 3.6 for details).



(A) Enrichment analysis using GO for bisque4, black, brown4 and darkgreen modules.

FIGURE 3.17: Enrichment analysis of the modules in the bootstrap network 2 using gene ontology (GO) information. (Continues on the following pages)

similar, differences during the vegetative phase are expected to be minimal. By applying weighted gene co-expression network analysis (WGCNA), we reconstructed modules of co-regulated genes and revealed an underlying architecture of transcriptional regulation that captures both conserved and asymmetric aspects of the diatom life cycle. The analysis highlights a pronounced asymmetry between MTs, a comparatively weaker signature of cell-size–dependent regulation, and the emergence of functionally distinct gene clusters that are broadly consistent with previous transcriptomic and experimental evidence.

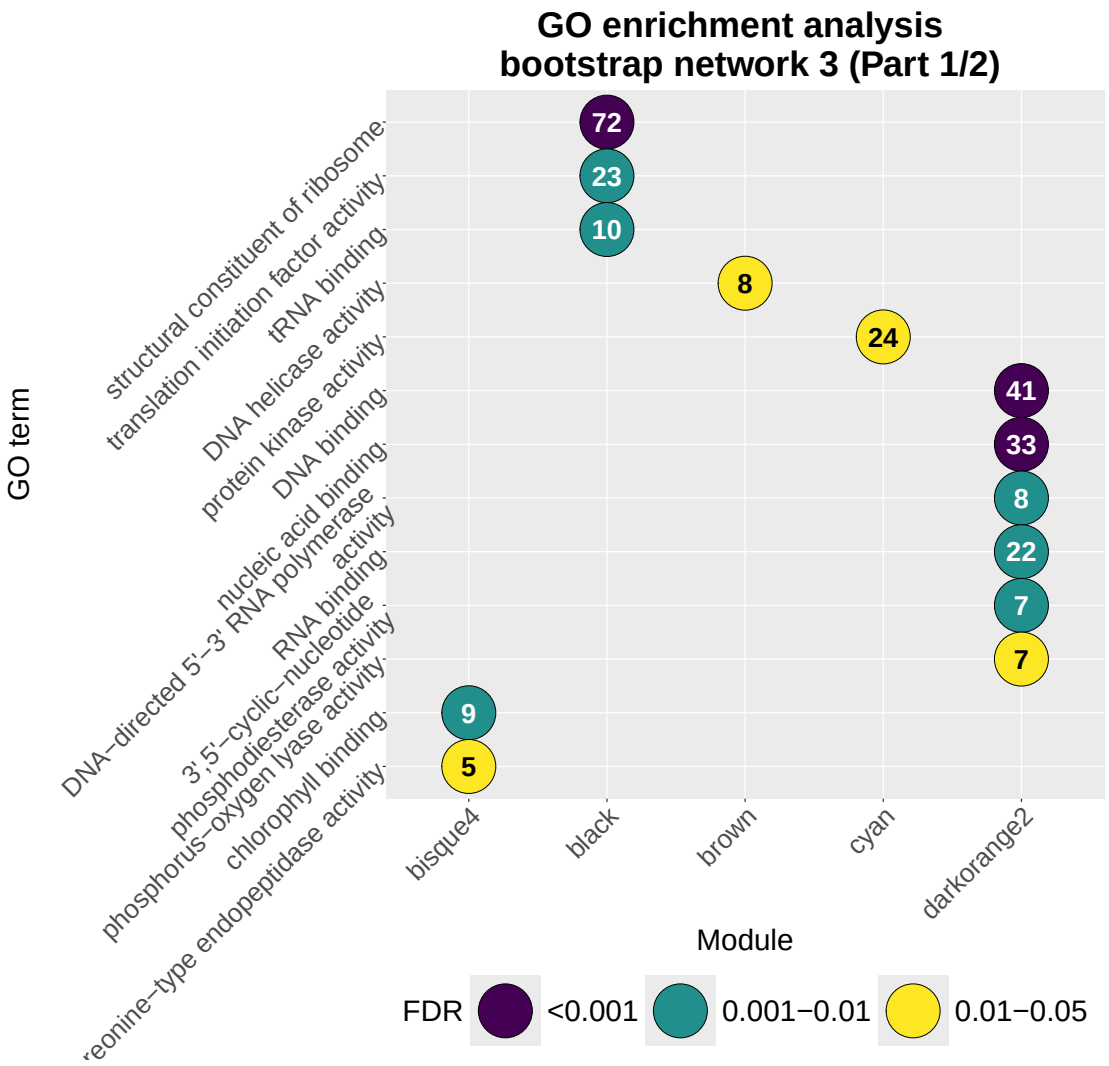


(B) Enrichment analysis using GO for darkmagenta, darkred, lightcyan and skyblue3 modules.

FIGURE 3.17: Enrichment analysis of the modules in the bootstrap network 2 using gene ontology (GO) information.

3.4.1 General considerations on WGCNA and dataset characteristics

One of the major advantages of co-expression network analysis is its ability to integrate data from multiple independent experiments. This strategy, successfully applied in other model diatoms such as *Phaeodactylum tricornutum* [50] and in metagenomic studies [119], allows the extraction of biological signals beyond specific experimental conditions. In the present case, this approach enabled us to leverage a diverse collection of RNA-seq datasets, each capturing a different aspect of the *P. multistriata* life cycle. However, the heterogeneity of experimental designs also introduces biases that can obscure true biological patterns. To minimize these effects, we employed a combination of variance-stabilizing transformation (`vst`)—which corrects for mean–variance

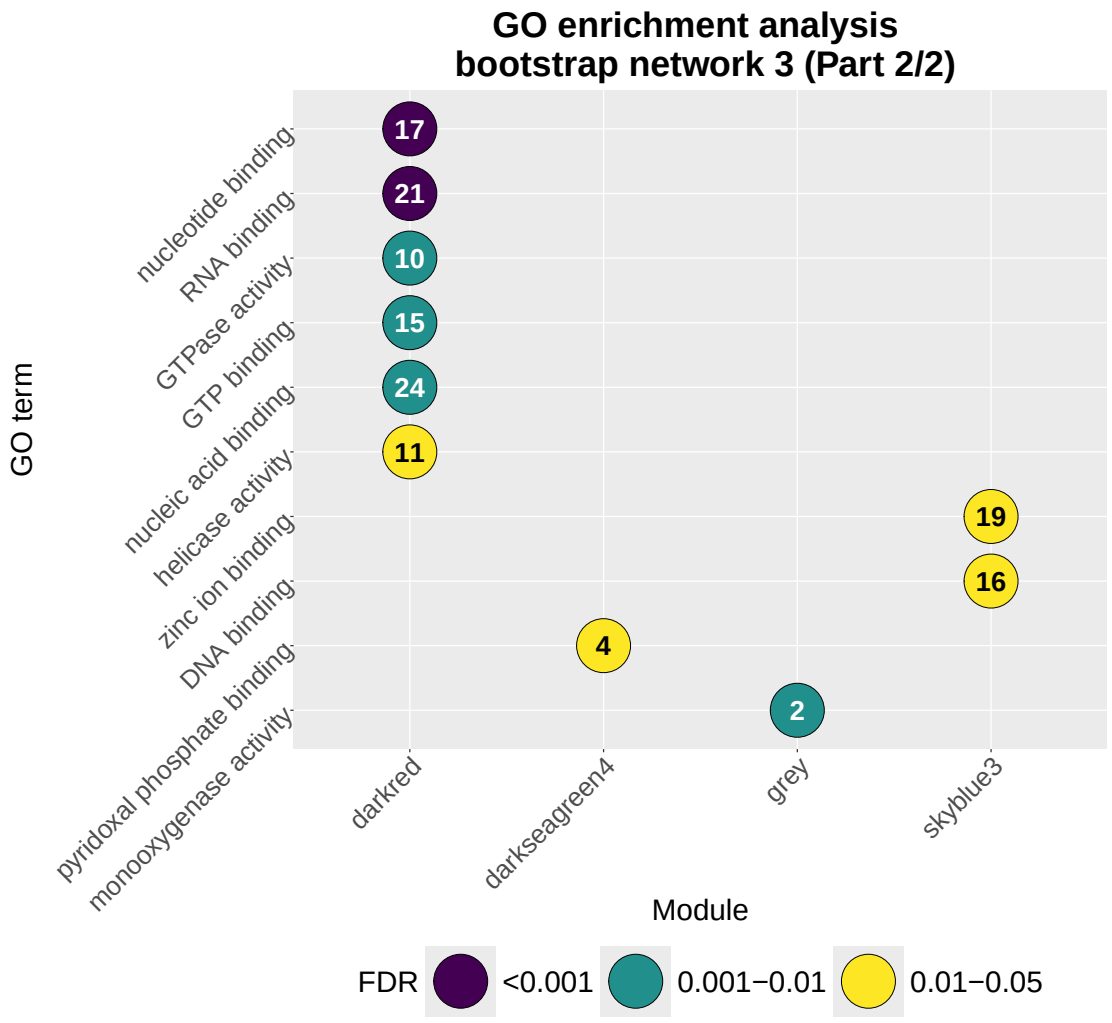


(A) Enrichment analysis using GO for bisque4, black, brown, cyan and darkorange2 modules.

FIGURE 3.18: Enrichment analysis of the modules in the small network using gene ontology (GO) information. (Continues on the following pages)

dependencies in RNA-seq data—and the `Combat` linear model, which effectively mitigates batch effects. Principal component analyses confirmed that these corrections substantially reduced inter-experimental variability, although subtle experiment-specific signals persisted. These likely reflect the fact that certain subsets of genes, rather than the whole transcriptome, are sensitive to specific experimental contexts.

A second challenge was the uneven representation of samples across life cycle stages: most libraries corresponded to below-SST cells, limiting the ability to detect transcriptional changes associated with the onset of sexualization. To circumvent this, we constructed several networks using distinct subsets of samples to disentangle overlapping sources of variation. Two main networks were produced—the "complete" network, including all samples, and the "small" network, restricted to below-SST cells. A third, above-SST network could not be built due to the low number of available replicates,



(B) Enrichment analysis using GO for darkred, darkseagreen4, grey and skyblue3 modules.

FIGURE 3.18: Enrichment analysis of the modules in the small network using gene ontology (GO) information.

since correlation-based methods such as WGCNA require a relatively large number of samples (typically > 20) to achieve statistical robustness. This limitation illustrates a general constraint when applying network-based methods to RNA-seq data, which are often designed with few replicates for differential expression studies of single gene rather than systems-level inference. To enhance the detection of size-dependent patterns, we adopted a bootstrap-inspired strategy: ten additional networks were generated by pairing all above-SST samples with different subsets of below-SST samples. Comparing these bootstrap networks allowed us to evaluate the stability and reproducibility of module–trait associations. The combination of the complete, small, and bootstrap networks thus should provide complementary insights, ensuring that the signals detected are not artifacts of specific sample configurations. Inevitably, since all networks derive from overlapping datasets, some redundancy remains, but the consistency of recurrent modules across independent subsamples should support their

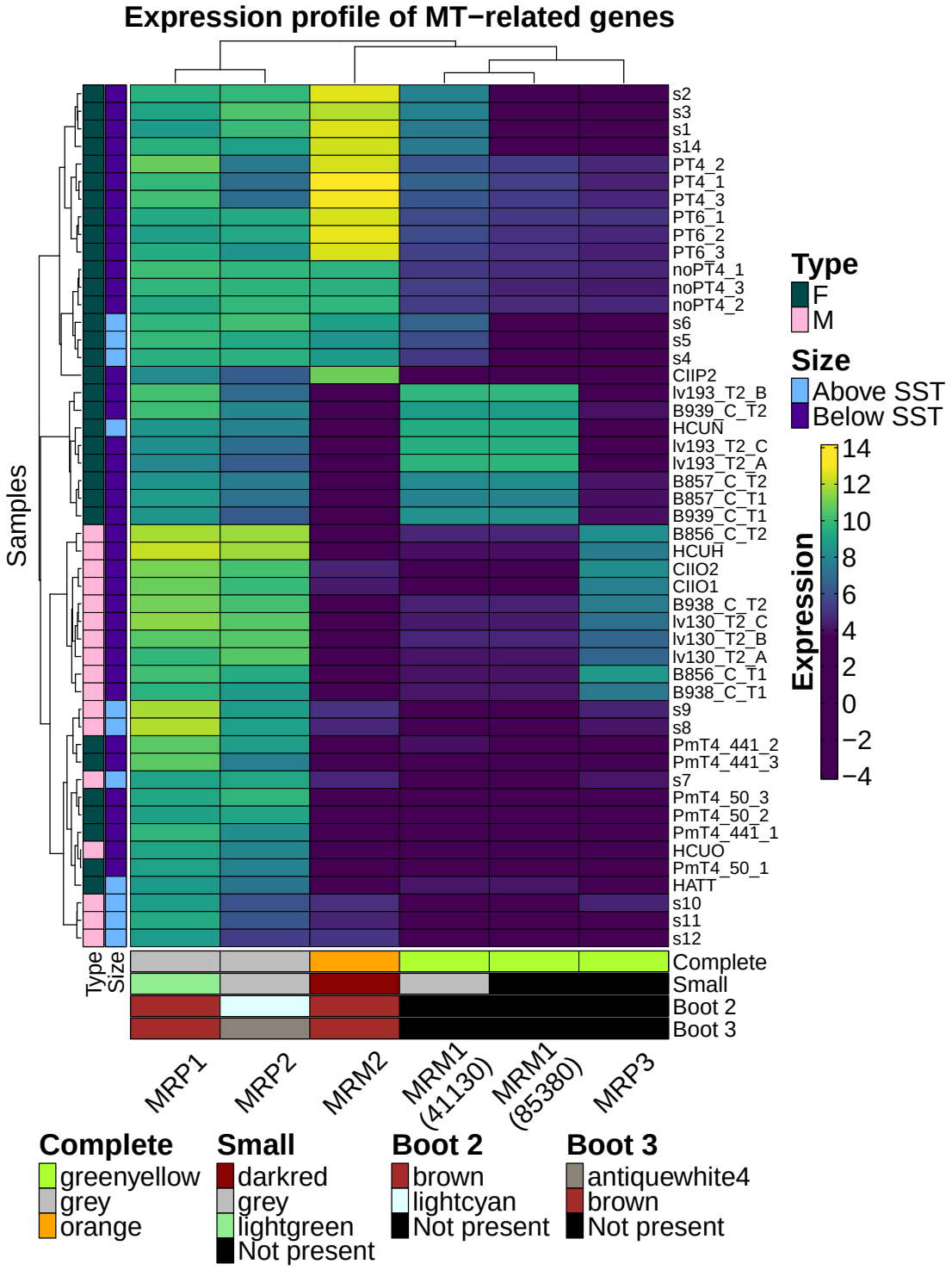


FIGURE 3.19: Normalized expression levels of the mating-type-related genes *MRP3*, *MRP2*, *MRP1*, *MRM2*, and *MRM1*. Columns represent genes and rows represent samples, both ordered according to hierarchical clustering. The two columns on the left indicate the MT and cell size for each sample. The row at the bottom shows the module to which each gene belongs, or "Not present" if the gene was filtered out during pre-processing.

biological significance. To reinforce the reliability of module–trait relationships, we complemented standard WGCNA correlations with additional statistical frameworks. The two canonical WGCNA metrics—the correlation between gene significance and module membership, and between module eigengenes and external traits—were accompanied by a PERMANOVA test, capable of quantifying the influence of multiple covariates on gene expression variance, and an accuracy assessment of a k-means clustering applyed on multidimensional scaling projections. Although full concordance among all four tests was not always achieved, agreement across at least two independent methods was considered a stringent and reliable criterion for module selection. A further source of uncertainty stems from class imbalance and expression filtering. Relevant lowly expressed genes may have been excluded, as exemplified by the omission of the MT+ markers *MRP1* and *MRP2* in the small network, which precluded insights into their co-expression partners. Similarly, the low number of large-cell samples reduced the statistical power to detect SST-related modules, while the lack of time-resolved sampling prevented the reconstruction of transcriptional dynamics during sexualization. More broadly, this work underscored the interpretative limitations of unsupervised clustering approaches such as WGCNA. The method efficiently groups genes based on expression similarity but does not inherently guarantee biological relevance. Thus, extensive contextual information—metadata, physiological measurements, and functional annotations—is essential to interpret clusters in a biologically meaningful way. In our study, while the main objective was to delineate genes associated with MT identity and sexual competence rather than to provide an exhaustive transcriptomic atlas, additional environmental or physiological metadata as nutrient availability could have refined the interpretation of some modules. Moreover, the high proportion of uncharacterized genes in diatom genomes remains a general obstacle to deep functional annotation. Despite these challenges, WGCNA proved to be an effective exploratory framework, guiding the identification of gene sets and pathways most promising for subsequent targeted analyses—an essential step when dealing with high-dimensional omics data.

3.4.2 Mating-Type Asymmetry

Across all network variants, a consistent pattern emerged: several modules were robustly associated with MT, whereas only a few correlated with cell size. This imbalance is somewhat unexpected, as the transition through the SST marks a major metabolic and physiological shift. The predominance of MT-related modules could, in principle, result from sampling bias, since the complete network largely comprises below-SST

samples, where mating-type differences are inherently stronger. However, this explanation does not fully apply to the bootstrap networks, in which the two size classes were balanced. It should be noted, though, that the same above-SST samples were used across all bootstrap iterations, meaning that any bias present in those data would persist throughout the analysis. The clearest transcriptional asymmetries in *P. multistriata* are known to occur during partner sensing and early sexualization—stages primarily represented by small cells in our dataset. These cells, although not yet actively engaged in sexual reproduction, may already display MT-specific transcriptional differences. Each MT fulfills distinct yet complementary roles in gametogenesis—emitting, perceiving, and responding to pheromonal cues. The persistence of MT-specific modules in networks not dominated by small-cell samples suggests that transcriptional divergence between MTs might extend beyond the canonical *MRP* and *MRM* genes, potentially continuing into the vegetative phase.

Among the identified modules, the **black** module in both the small and complete networks exhibited the strongest and most consistent association with MT. Its biological relevance is underscored by a significant overlap with MT-related modules from the bootstrap networks and by enrichment in metabolic and ribosomal functions. Notably, many of its genes coincide with differentially expressed genes reported in [29, 13], particularly those involved in the downregulation of anabolic processes in MT– cells. This aligns with the observation that MT– undergoes metabolic arrest earlier than MT+, suggesting a more rapid shift toward gametogenesis. Other modules, such as **pink** and **green**, also correlated with MT but lacked clear enrichment in known gene sets. These may represent uncharacterized branches of the MT regulatory network, potentially acting downstream of the master regulators *MRP3* and *MRM1/MRM2*. Their enrichment in oxidoreductase activity and calcium-ion binding suggests a connection to redox and ionic signaling pathways. In the benthic diatom *Seminavis robusta*, NADPH oxidase (NOX)—a key enzyme mediating oxygen reduction—was found to be more active in MT+ cells following exposure to the sex-inducing pheromone, and its activity appears to be calcium-regulated [42]. Comparable interactions between calcium fluxes and redox signaling have been documented during reproductive processes in other algal groups [288, 48] and in other eukaryotes [139, 156]. Together, these observations support the hypothesis that redox and calcium signaling constitute conserved components of gametogenic control in diatoms. The **cyan** module contains genes that are down-regulated in MT+ relative to control cells [29], suggesting a link to processes specific to MT+ during sexualization, potentially in pheromone biosynthesis or signaling. Conversely, the **blue** module, which shows no strong association with either MT or size, encompasses genes unresponsive under standard conditions but enriched in regulated genes during sexual activation [13]. Its content of DNA- and RNA-binding proteins, including helicases, mirrors transcriptional features observed in below-SST

cells of both MTs in *Cylindrotheca closterium*, suggesting conserved regulatory strategies across pennate diatoms.

3.4.3 Modules related to meiosis and cell-size–dependent regulation

In contrast to the strong MT signatures, size-dependent regulation was comparatively weak. Only a few modules—most notably **greenyellow** and **orange**—showed a correlation with the SST. Both lacked strong functional enrichment, and the **orange** module was influenced by experimental artifacts. The **greenyellow** module also displayed a correlation with MT, indicating that it may include genes jointly regulated by sexual maturity and mating type. The scarcity of clearly size-related modules is surprising given the extensive physiological changes that are expected to occur at the onset of sexual reproduction. THowever, this outcome may result from both the limited number of above-SST samples and the inherent asynchrony of cell populations around the SST, where only a fraction of individuals are truly sexually competent [29]. The **darkturquoise** module, though not directly correlated with size, was enriched in canonical meiotic genes, including several of the conserved "meiotic toolkit" components identified in earlier studies. The co-expression of these genes with uncharacterized, diatom-specific genes points to novel components possibly linked to meiosis.

3.4.4 Consistency and robustness of the inferred co-expression structure

The bootstrap networks confirmed the stability and reproducibility of the main network features. A module enriched in ribosomal and NAD-binding enzymes was consistently detected across all bootstrap iterations, maintaining its strong association with MT. Additional modules—such as **darkred** and **royalblue** in bootstrap network 2 and **bisque4** in bootstrap network 3—emerged as promising candidates for future exploration, particularly in the context of transitions across the SST. The absence of clearly size-associated modules across bootstrap analyses might suggest that transcriptional variation linked to cell size is modest relative to the robust MT-specific signature. Nonetheless, the repeated use of identical above-SST samples may have limited sensitivity to detect such effects.

3.4.5 Behavoiur of MT-related gene

The five genes previously identified as components of the mating-type regulatory program (*MRP1*, *MRP2*, *MRP3*, *MRM1*, and *MRM2*) display distinct expression patterns

and network associations. Among them, *MRP3* shows the most pronounced and specific expression profile, being strongly upregulated in MT+ samples below SST and almost completely repressed in MT- samples. Its consistent exclusion from several bootstrap networks likely reflects its highly condition-specific expression, emphasizing its role as a tightly regulated determinant rather than a broadly co-expressed gene.
MRP1 and *MRP2*, by contrast, exhibit more uniform expression across conditions, with only modest increases in MT+ below SST. Their assignment to the grey module in the full network suggests that they are not tightly co-regulated with other gene clusters, reinforcing the idea that they may function independently or in pathways peripheral to the major co-expression structures. This observation further supports the hypothesis that certain components of the heterothallism program operate in specialized contexts, separate from the primary metabolic processes.
MRM1 and *MRM2* present overall low expression levels, with peaks restricted to a small subset of samples. The co-occurrence of their expression peaks might indicate a degree of functional linkage or shared regulation, which could be linked to the sexualization process, since *MRM2* is a candidate for being a protein involved in the receptor/adhesion of pheromones [252].

3.4.6 Conclusion and Future prospects

Taken together, these findings highlight that transcriptional regulation in *P. multistriata* is strongly shaped by mating-type identity, with metabolic and signaling asymmetries persisting beyond the immediate sexual phase. In contrast, size-dependent transitions appear to involve fewer and more transient transcriptional adjustments. This reinforces the concept that diatom life-cycle progression is governed by a tight interplay between mating-type–specific regulatory programs and conditional responses to size. The modular framework revealed by WGCNA provides a foundation for identifying candidate regulators, elucidating functional networks, and ultimately linking molecular mechanisms to ecological outcomes in this and related species. Looking ahead, expanding both the diversity and number of available transcriptomes will allow the construction of more balanced and statistically robust co-expression networks. As discussed, the limited representation of above-SST samples likely introduced biases, and increasing their number would help mitigate spurious correlations. At the time of writing, a new set of libraries from above-SST cells has been produced, which will likely enable the construction of a network comprising exclusively vegetative-stage samples. In parallel, generating networks from cells actively engaged in sexual reproduction, specifically after pheromone perception and pairing, would be particularly valuable to capture the

transcriptional signatures directly associated with mating competence and gamete differentiation. The emergence of these needs underscores another important contribution of broad-scale exploratory studies such as this one: they can inform the design of future experiments aimed at investigating aspects of the system that were not previously considered or prioritized. By revealing unexpected patterns and gaps in current knowledge, such analyses serve as a foundation for more targeted, hypothesis-driven research. The relatively weak signal observed for size-related regulation may also reflect that not all cells were not actively undergoing sexualization. Subjecting them to controlled stress or mating-inducing conditions could elicit stronger, more informative transcriptional responses. Another limitation concerns the availability of external annotation resources to complement unsupervised clustering analyses. Progress in this area will benefit from integrating pathway-based annotation systems. The Kyoto Encyclopedia of Genes and Genomes (KEGG) provides a curated framework that organizes genes into metabolic and signaling pathways based on biochemical knowledge [142, 143]. This resource can complement GO enrichment analyses by offering mechanistic insights into how co-expressed genes interact within defined biochemical routes. Combining KEGG and GO data will improve the functional characterization of modules and facilitate the identification of key regulatory nodes within complex gene networks. A promising methodological extension would be the implementation of differential co-expression analysis (DCA). Unlike traditional differential expression approaches that focus on changes in transcript abundance of individual genes, DCA examines how the correlation structure between genes varies across conditions [263]. It thus captures rewiring of gene–gene relationships rather than absolute expression shifts, uncovering subtle changes in the regulatory architecture. By comparing network topology between conditions, DCA can pinpoint modules or specific gene pairs whose connectivity changes during the life cycle—highlighting transient regulatory events that may not involve strong fold changes yet play crucial roles in coordinating developmental transitions. Applied to *P. multistriata*, DCA could reveal key regulatory switches and provide a dynamic view of how transcriptional interactions are reorganized as cells move from vegetative growth to sexual reproduction.

Chapter 4

Conclusions and Outlook

This thesis adopted a data-driven perspective to understand patterns and mechanisms operating at complementary levels of biological organization in the ocean. The first part of the work addressed large-scale ecological questions through the analysis of field observations obtained via microscopy-based monitoring of coastal phytoplankton. The second part explored molecular mechanisms in the pennate diatom *Pseudo-nitzschia multistriata*, dissecting the gene expression programs that regulate sexual competence and mating type identity. Although these two case studies operate across distinct spatial and biological scales, they are unified by a common quantitative and data-science approach, collectively advancing our understanding of marine phytoplankton ecology and physiology.

In Chap. 2, we analyzed a large, heterogeneous dataset of coastal phytoplankton observations collected along the Italian coastline under the Marine Strategy Framework Directive. Through descriptive and statistical modeling approaches, which included ordination techniques in the form of constrained analyses, we disentangled the spatial and seasonal structuring of phytoplankton communities across basins characterized by contrasting hydrographic regimes. Despite variability in sampling design and taxonomic resolution, the analyses revealed consistent large-scale patterns: community composition and abundance were primarily structured by spatial gradients linked to hydrography, riverine nutrient inputs, and circulation features, with regional contrasts overriding seasonal variability. Richness estimates and indicator value (IndVal) analyses identified basin-specific assemblages, while constrained ordinations indicated that spatial effects and unmeasured drivers outweighed direct environmental control. Together, these results demonstrate how data harmonization, dimensionality reduction,

and multivariate modeling can extract robust ecological signals from complex monitoring data, supporting a quantitative assessment of biogeographic and temporal dynamics at the national scale.

In Chap. 3, we employed a systems-biology framework integrating multiple RNA-seq datasets and constructing weighted gene co-expression networks (WGCNA) for *P. multistriata*. After variance stabilization, batch correction, and bootstrap-based network inference, we identified stable gene modules associated with mating type (MT) and with sexual competence triggered by passing the sexualization size threshold (SST). Modules correlated with MT were enriched in translational, oxidoreductive, and calcium-signaling processes, while canonical meiotic genes formed coherent subnetworks, suggesting a conserved regulatory backbone for sexualization. The analysis confirmed known asymmetries in canonical mating type genes (*MRP1–3*, *MRM1–2*) and highlighted new candidate regulators potentially involved in the sexual phase. These findings provide a first quantitative view of the regulatory architecture underlying mating-type identity and sexual competence in a pennate diatom.

4.1 Novelty and contribution of this thesis

The work presented in Chap. 2 constitutes the first large-scale, quantitative characterization of phytoplankton communities along the entire Italian coastline. By integrating and harmonizing microscopy-based data from over 2,000 samples collected across 162 stations and 54 transects, the study generated a unified assessment of community composition and abundance. This synthesis revealed both recurrent biogeographic patterns and basin-specific peculiarities that were previously inaccessible through regional studies. In particular, the identification of characteristic taxa and basin-specific indicator genera provides the first baseline for national-scale biodiversity assessments of coastal phytoplankton.

A key methodological contribution lies in the standardization of an heterogeneous monitoring data. Reconciling inconsistent taxonomic nomenclature through the World Register of Marine Species (WoRMS), treating detection limits in chemical data, and harmonizing sampling effort across regions transformed a monitoring-oriented database, characterized by irregular temporal coverage and variable resolution, into a coherent dataset suitable for macroecological inference. This demonstrates how national monitoring programs, despite intrinsic heterogeneity, can be repurposed for ecological research through rigorous data treatment and statistical modeling.

By constructing standardized species–accumulation curves and applying IndVal analyses, this study delivers the first quantitative comparison of phytoplankton richness

and characteristic taxa across Italian marine basins. The results reveal a pronounced spatial structuring of communities that had not been previously documented, thereby establishing a new baseline for regional biodiversity assessment. The consistent detection of diatom–dinoflagellate seasonal successions across much of the Italian coastline, together with their attenuation in oligotrophic sectors, reflects the classical dynamics of temperate coastal systems, where the alternation between r- and K-strategist species is modulated by basin productivity and hydrographic structure. These findings point to a strong coupling between trophic status and community turnover: sharp and well-defined in eutrophic systems such as the northern Adriatic, but weaker or more irregular in oligotrophic basins like the Ionian and Sardinian coasts. This integrative perspective elucidates how regional-scale hydrography shapes the phenological trajectories of phytoplankton communities within the Mediterranean context.

Given the wide spatial extant of the study area, ordination models such as RDA and CCA, aimed at determining the relation between community structure and external covariates, needed to include spatial autocorrelation term, which in our case took the form of the Moran's Eigenvector Maps (MeMs). It was possible to separate the effects of environmental gradients, spatial connectivity, and seasonal forcing, quantifying their relative contributions to community organization. The finding that spatial structure explains the largest fraction of variance, while environment and season contribute smaller but significant shares, provides new empirical evidence of the dominant role of basin-scale hydrography in shaping phytoplankton distribution. This work also help in giving an idea of the potial and limitation of these kind of monitoring programs and by highliting the limitation help in the menagement of the resources for future campaign at being more effective.

To account for the broad spatial extent of the study area and the resulting non-independence among samples, ordination models such as RDA and CCA, designed to explore community–environment relationships, were complemented with a spatial autocorrelation term derived from Moran's Eigenvector Maps (MeMs). This approach enabled the separation of the effects of environmental gradients, spatial connectivity, and seasonal forcing, allowing their relative contributions to community organization to be quantified. The finding that spatial structure explains the largest share of variance while environment and season contribute smaller but significant fractions provides robust empirical evidence for the dominant influence of basin-scale hydrography on phytoplankton distribution. Beyond its ecological insights, this work also underscores the potential and limitations of large-scale monitoring programs: by highlighting sources of uncertainty and bias, it informs the design of future campaigns and the allocation of monitoring resources toward more efficient ecological assessments.

In Chap. 3, we provided the first systems-level reconstruction of the gene co-expression architecture in the pennate diatom *Pseudo-nitzschia multistriata*, revealing the regulatory organization underlying mating type identity and the transition toward sexual competence. This work integrated multiple independent RNA-seq datasets into a single coherent analytical framework. These datasets had previously been analyzed individually to address specific aspects of sexualization, such as pheromone signaling or nutrient-dependent regulation, using differential expression analyses under tightly controlled experimental conditions. In contrast, our approach sought to extract system-level regularities across studies. To achieve this, we implemented rigorous normalization and bias-correction procedures to transform heterogeneous raw data into a statistically comparable dataset suitable for co-expression analysis.
Following the example of prior work on *Phaeodactylum tricornutum*, we applied WGCNA to characterize the transcriptional organization of *P. multistriata* throughout its life cycle, which alternates between vegetative and sexual phases. The analysis revealed a modular structure of co-regulated genes, uncovering the global organization of transcriptional regulation and the relationships among modules and phenotypic traits such as mating type and sexual competence.
To address limitations in sample representation and to test the robustness of the inferred co-expression patterns, we employed a combination of complementary network and statistical approaches. Specifically, the two configurations “complete” and “small” networks provided both a general overview of co-regulation across the species and a more detailed view of metabolic organization below SST. To enhance the reliability differences inferred between samples above and below the SST, we constructed multiple bootstrap networks by resampling subsets of below-SST samples together with all above-SST ones. Furthermore, we extended the classical correlation-based WGCNA framework by incorporating additional statistical evaluations to verify module–trait associations: variance partitioning via PERMANOVA and a classification-based accuracy test using multidimensional scaling and k-means clustering. This multi-test strategy increased the robustness of the analysis and strengthened the confidence in the resulting biological interpretations.

Modules associated with mating type were more numerous and consistent than those linked to cell size, possibly suggesting that MT-specific regulation extends beyond previously known master regulators. Incorporating external information, such as gene ontology annotations and results from prior differential expression analyses, allowed more refined biological interpretation of co-expression modules. These gene sets provide a resource for future experimental validation and hypothesis testing on the molecular control of sexualization.

Beyond its specific findings, this chapter reinforces the value of network-based transcriptomics for diatom biology. Whereas previous studies on this species identified lists of differentially expressed genes, the present work uncovered the organizational logic of the transcriptome, translating gene-level data into a structured map of regulatory relationships. The resulting pre-processed expression matrix and co-expression network form a reusable analytical foundation for future studies on *P. multistriata* and related species. The analytical pipeline can be easily extended to new datasets, including those addressing active sexualization, nutrient stress, or environmental perturbations. Ultimately, this work contributes conceptually to our understanding of diatom regulatory topology and provides a scalable framework for comparative and integrative omics.

4.2 Limitations

Despite the advances achieved through large-scale data integration, the analysis of phytoplankton communities along the Italian coastline remains constrained by several methodological and structural limitations. The heterogeneity of sampling effort among regions, uneven temporal coverage, and reliance on microscopy-based identification introduce uncertainty in the quantitative comparison of diversity and abundance patterns. Differences in analytical sensitivity, the treatment of values below detection limits, and the absence of key environmental variables (e.g., irradiance, stratification) limit the interpretability of species–environment relationships. The focus on surface samples and the aggregation of data into broad seasonal categories may obscure fine-scale or vertical variability. While spatial modeling and harmonization mitigate these issues, they cannot entirely compensate for missing drivers or inconsistent protocols; results should therefore be viewed as indicative rather than exhaustive representations of coastal phytoplankton diversity. The limited explanatory power of ordination models likely reflects the omission of critical drivers such as upwelling, riverine discharge, and episodic meteorological events. Although some proxies (e.g., salinity, season) partially capture these influences, a more comprehensive environmental characterization is required for predictive ecological modeling.

Similarly, the transcriptomic analysis of *P. multistriata* faces limitations arising from dataset heterogeneity, experimental imbalance, and methodological assumptions. Integrating independent RNA-seq experiments conducted under diverse laboratory conditions introduces residual batch effects that may confound biological signals. The limited number of samples representing sexually competent cells reduces statistical power,

while asynchronous sexualization in vegetative cultures may dilute transient transcriptional signatures. Furthermore, WGCNA captures correlations rather than causation, and module inference depends on parameter choices and filtering thresholds. The incomplete functional annotation of diatom genes further constrains biological interpretation. Thus, while the networks identified here offer valuable hypotheses on regulatory architecture, their mechanistic validation will require targeted experimental studies and comparative analyses across related species and environmental conditions.

4.3 Future Prospects

Building on the present findings, future work should advance toward more quantitative analyses aimed at disentangling the environmental and ecological processes that drive community variability across space and time. In this study, regional richness was examined and basins characterized by high and low taxonomic turnover were identified. However, a more robust quantification of diversity along the Italian coasts can be achieved through the application of formal α- and β-diversity metrics, which can be more effectively linked to environmental gradients and drivers, thereby providing a rigorous assessment of spatial and temporal heterogeneity in community structure [216, 215]. The chemico-physical variables currently available exhibited limited explanatory power for phytoplankton community structure, suggesting that a broader characterization of the environmental context is required. As direct measurements are unavailable, oceanographic models such as those provided by the Copernicus Marine Service [74] could help fill this gap. Nevertheless, their applicability remains uncertain, as even state-of-the-art oceanographic models have been shown to struggle in capturing the full complexity of biogeochemical processes in coastal areas [25].
In parallel, a similarly detailed analysis of the zooplankton community (available within the original MSFD dataset) will be essential to evaluate its spatial organization, diversity patterns, and relationships with environmental gradients. Quantifying zooplankton diversity through analogous α- and β-diversity indices will enable direct comparison with phytoplankton assemblages, providing insights into the degree of coupling between trophic levels. Such an approach will also support the exploration of both bottom-up and top-down controls, thereby assessing the extent to which nutrient availability and grazing pressure jointly regulate phytoplankton dynamics. Beyond pairwise associations, adopting a network-based analytical framework could substantially enhance our understanding of food-web architecture and the relative importance of species interactions. Network approaches have proven highly effective in disentangling complex ecological dependencies and identifying key taxa or functional groups that underpin ecosystem structure [101, 167, 297, 251]. Applying such methods to the

MSFD dataset would bridge community-level observations with ecosystem-level processes, contributing to a more mechanistic understanding of trophic linkages, energy transfer, and ecosystem resilience in coastal waters.

Regarding the molecular work, integrating transcriptomic data with proteomics and metabolomics will be critical to achieve a comprehensive view of molecular regulation throughout the diatom life cycle. Proteomics, the large-scale study of the complete protein complement of a cell, aims to quantify not only protein abundance but also post-translational modifications such as phosphorylation or acetylation, which can alter protein stability, localization, and function [5, 316]. Combining proteomic and transcriptomic data allows disentangling transcriptional from translational or post-translational control; a key aspect during gametogenesis, when translational regulation ensures rapid activation of meiosis- and differentiation-related factors [13, 42]. Metabolomics, in turn, provides a snapshot of small-molecule intermediates and end products of metabolism, reflecting the physiological state and energy allocation of the cell [209, 104]. Because metabolic pathways represent the functional output of gene and protein activity, metabolomic data directly inform on the organism's biochemical and ecological strategies. In diatoms, metabolomics has already been instrumental in revealing metabolic reprogramming during nutrient limitation, silicification, and sexual reproduction [105]. Together, these omics layers—transcriptomics, proteomics, and metabolomics—will enable the reconstruction of regulatory cascades spanning gene expression to metabolic output, offering a holistic and mechanistic perspective on how diatom cells orchestrate transitions between vegetative and sexual phases. Such integration will also help identify regulatory checkpoints acting post-transcriptionally or post-translationally, processes that are known to play central roles in reproductive transitions across eukaryotes [211, 23]. From a comparative perspective, extending co-expression analyses to other pennate diatoms will provide evolutionary context for the regulatory networks characterized here. Cross-species comparisons could uncover conserved modules corresponding to ancestral reproductive circuits and identify lineage-specific innovations, such as the emergence of the sex-determining gene MRP3 [252]. Mapping orthologous components (i.e., finding groups genes deriving from a common ancestor that differentiate through speciation [95]) of co-expression modules across species will help distinguish universal features of pennate diatom reproduction from species-specific adaptations to ecological niches.

Altogether, the work presented in this thesis, alongside the proposed future directions, may contribute to advancing our understanding of the metabolism and regulatory complexity of *P. multistriata*. Integrating transcriptomic, proteomic, and metabolomic

information within an evolutionary framework will enable linking molecular regulation to the ecological strategies that shape diatom life cycles. Ultimately, elucidating the regulatory networks that govern transitions between vegetative growth, sexualization, and resting stages will not only deepen our understanding of diatom reproductive biology but also shed light on the mechanisms underpinning the ecological success and resilience of pennate diatoms in marine environments.

Appendix A

Additional resources for characterizing phytoplankton community

We report here a few additional resources used to support statements made in the main manuscript.

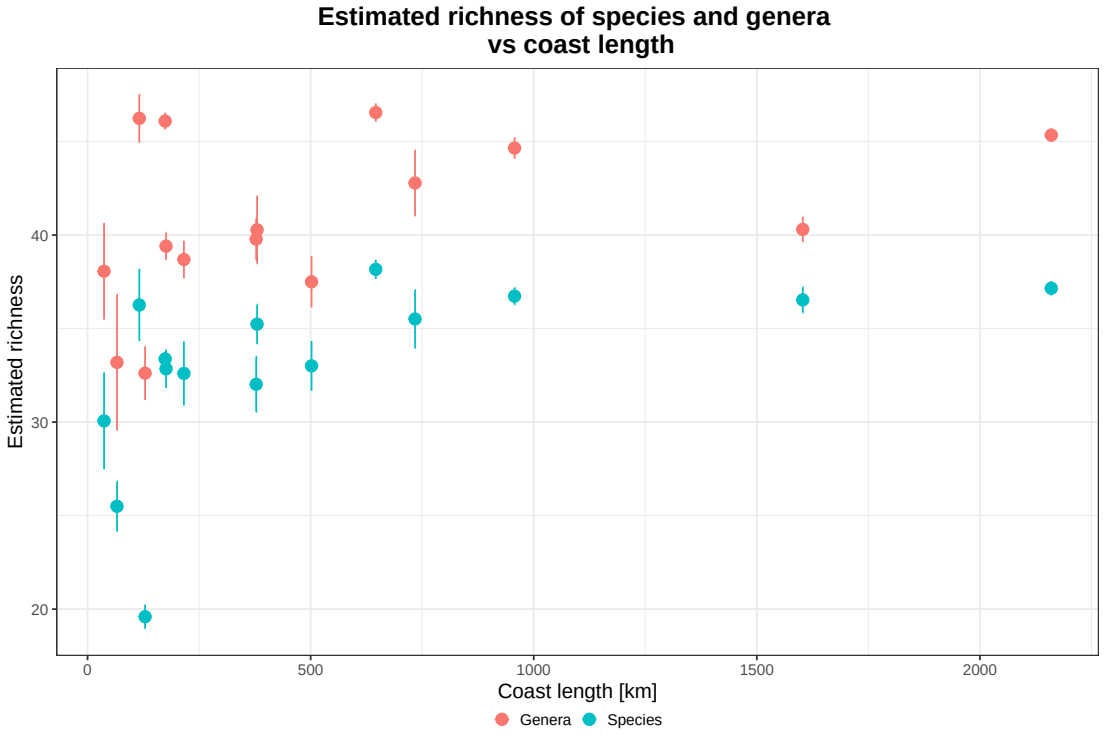


FIGURE A.1: Dependence of estimated species (light blue dots) and genus (red dots) richness on coastline length.

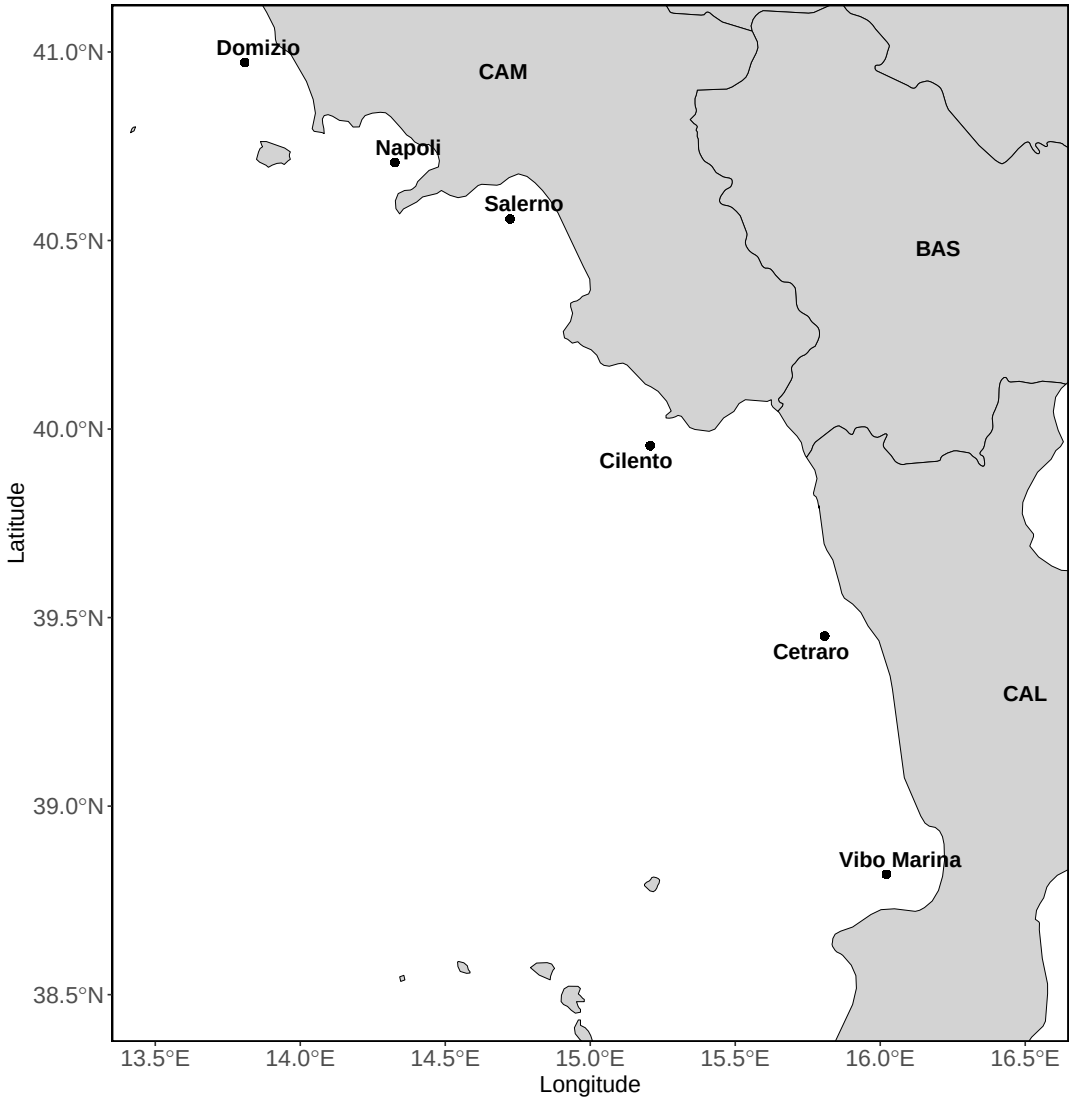


FIGURE A.2: Transects of Calabria and Campania region in Southern Tyrrhenian.

In Fig. A.4 we report the estimated coefficients obtained from multiple linear models. For each model, the regression formula was (following the R convention):

$$\log(\text{Abundance}) \sim \text{DO} + \log(\text{NH}_4) + \log(\text{NO}_3) + \log(\text{PO}_4) + \log(\text{SiO}_4) + \text{Salinity} + T \tag{A.1}$$

Here, *Abundance* represents the sample abundance (regardless of community composition), and a logarithmic transformation was applied to the nutrient variables to reduce skewness in their distributions. The main difference among models lies in the spatial extent of the data used: one model (*Italy*) was built using all samples from the dataset, while the others were constructed separately for each basin as defined in the study. Each model underwent stepwise regression to identify the most relevant covariates.
We note a clear heterogeneity both in the selected covariates across models and in the

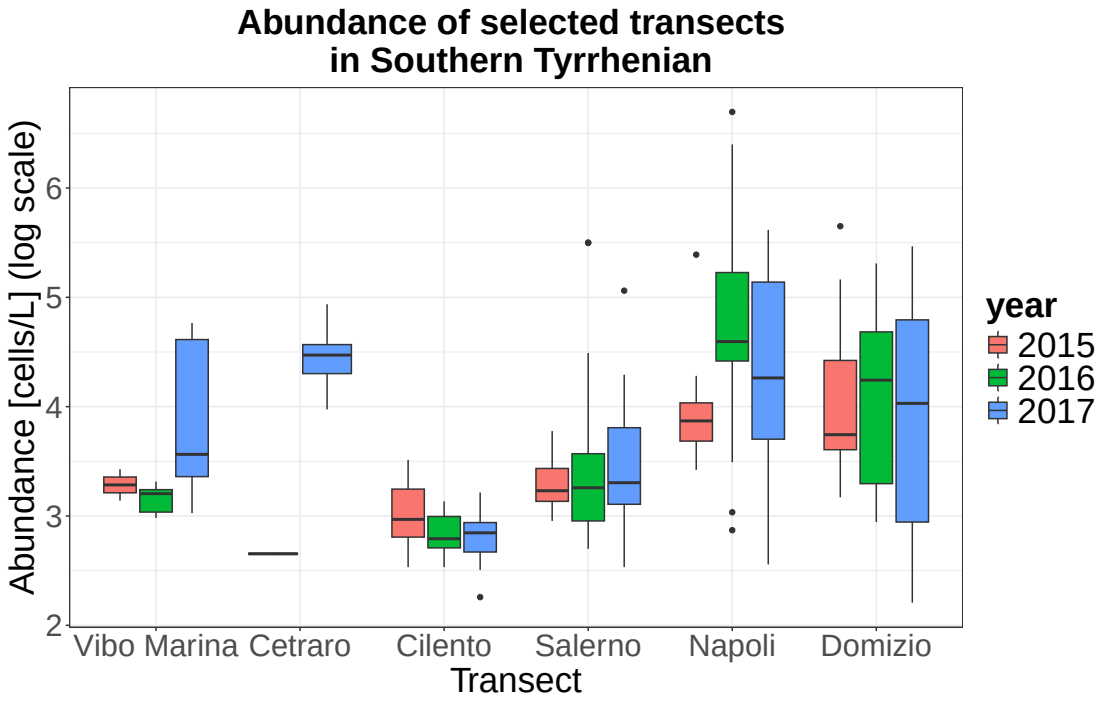


FIGURE A.3: Log-transformed sample abundances along transects in the Calabria and Campania regions, with colours indicating the sampling year.

direction (sign) of their effects. In particular, for $\log(NH_4)$, $\log(SiO_4)$, and T, it is not straightforward to infer a general response of phytoplankton abundance to these variables. This variability may arise because communities from different basins exhibit distinct ecological responses to nutrient and temperature conditions. Alternatively, for the basin-scale models, the range of measured environmental variables might be too narrow to adequately capture the relationships between specific covariates and phytoplankton abundance. This is especially true if the basin has an oligotrophic character, such as LIG and ST, or of there was a lack of proper coverage of the gradient due to the issues regarding the sampling campaign.

FIGURE A.4: Coefficients of statistically significant predictors identified via stepwise selection in linear models with log-transformed abundance as the response variable. Blue and red bars represent positive and negative coefficients, respectively. The x-axis indicates the different models: one combining all samples (Italy) and one for each basin individually. Each panel corresponds to a different predictor. The absence of a bar indicates that the variable was either not statistically significant or excluded during stepwise selection.

Appendix B

Additional results of WGCNA

Here we report any output, from plots to statistical tests, that was not included in the main body of the thesis because was not considered strictly relevant for the discussion.

B.1 Network construction

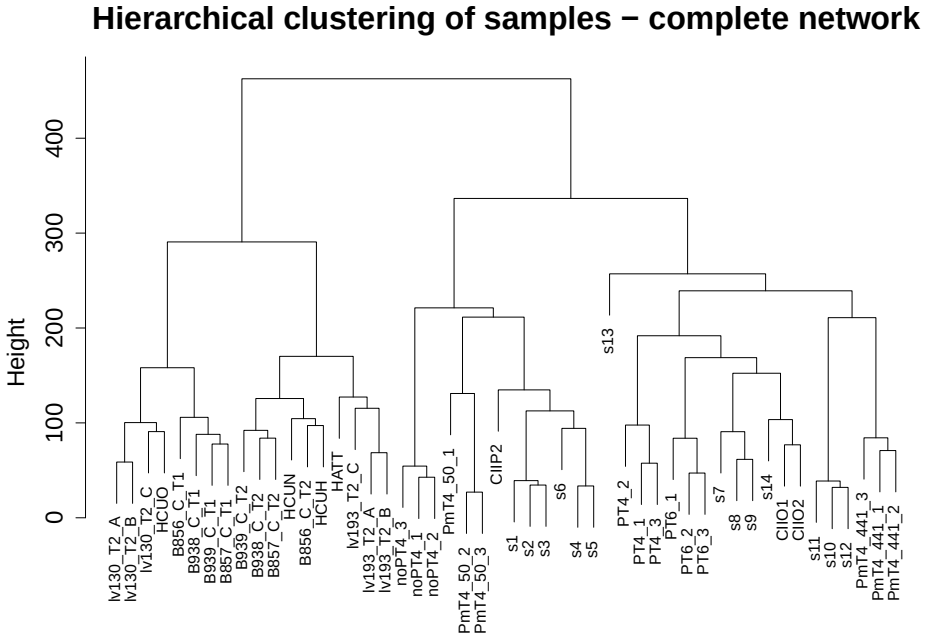


FIGURE B.1: Hierarchical clustering of samples for the complete network using Ward linkage. s13 is the most dissimilar sample, so it was removed.

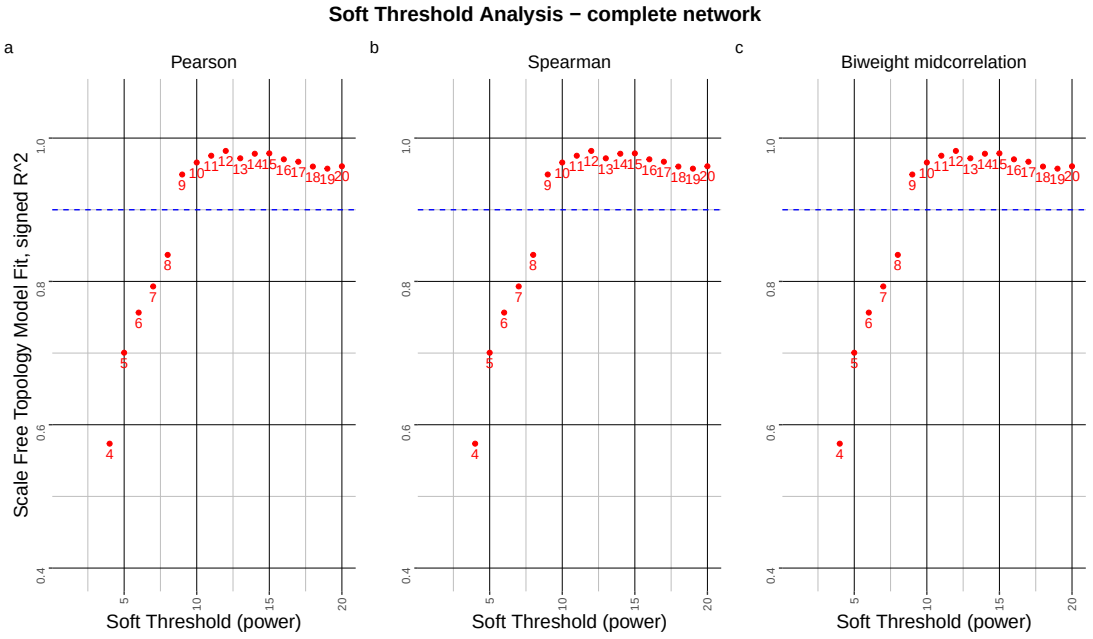


FIGURE B.2: R^2 vs β for the complete network for the three correlation considered: pearson (a), spearman (b), bi-weight (c)

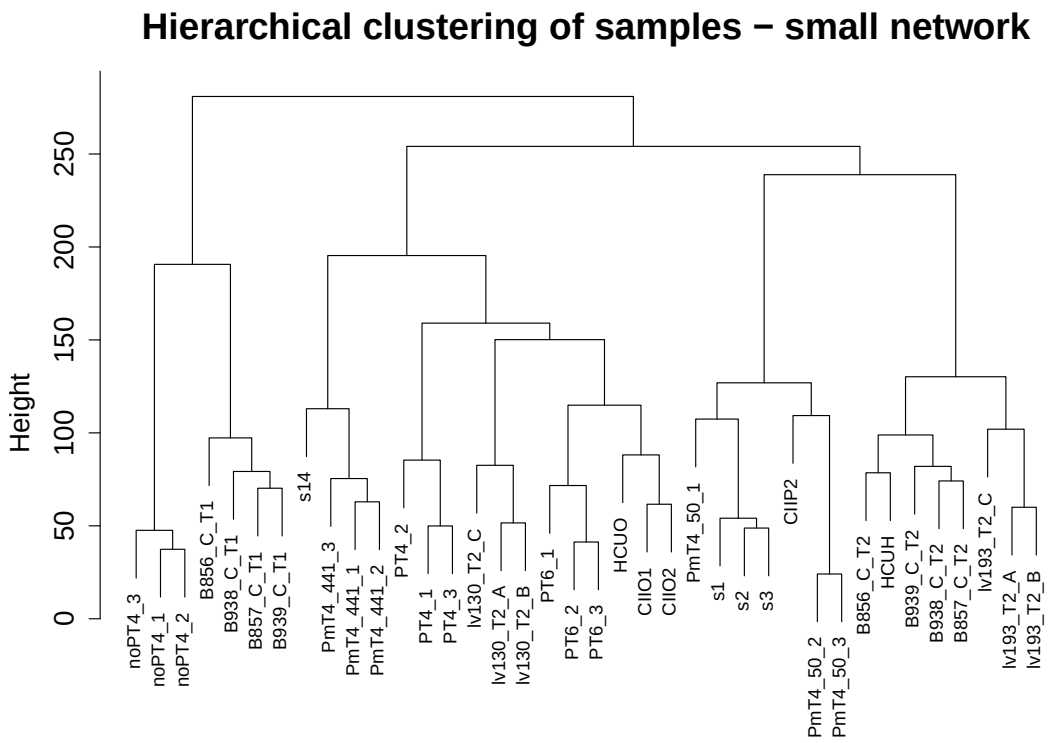


FIGURE B.3: Hierarchical clustering of samples for the small network using Ward linkage. No sample was removed.

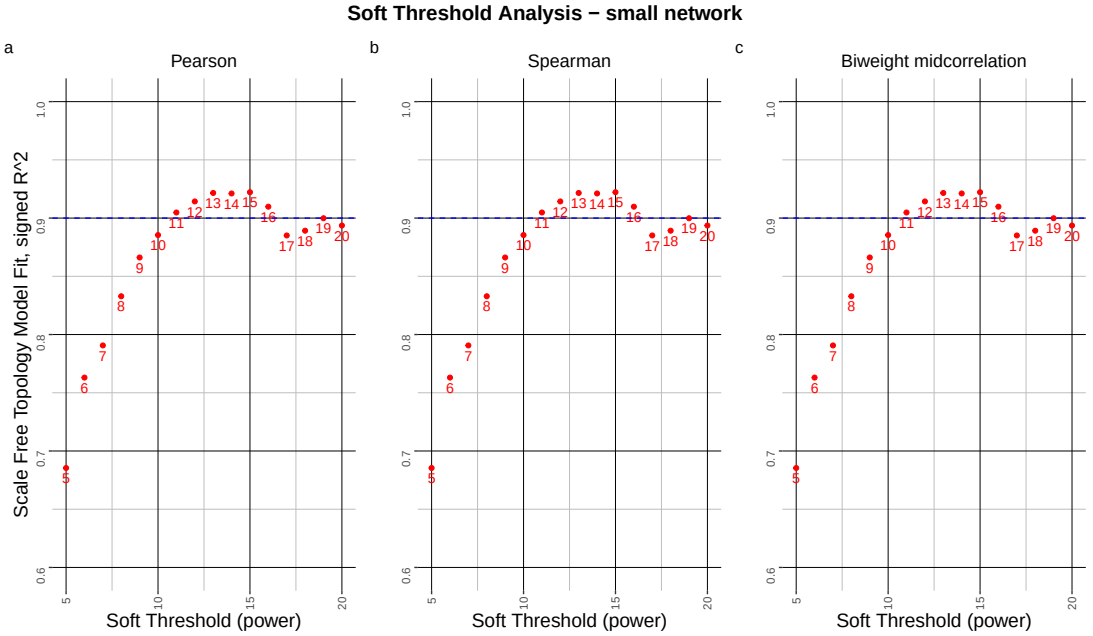


FIGURE B.4: R^2 vs β for the small network for the three correlation considered: pearson (a), spearman (b), bi-weight (c)

B.2 Properties of modules

	Df	SumOfSqs	R^2	F-value	p-value
MT	1	0.009410	0.11320	4.1343	0.042
Expr	6	0.007598	0.09140	0.5563	0.807
MT:Expr	2	0.002389	0.02873	0.5247	0.642

TABLE B.1: PERMANOVA test for FA acids and carbon assimilation genes in black module of small network

	Df	SumOfSqs	R^2	F-value	p-value
MT	1	0.02611	0.02387	32.545	0.001
Size	1	0.17373	0.15882	216.582	0.001
Expr	6	0.86084	0.78697	178.862	0.001
MT:Size	1	0.00195	0.00178	2.433	0.119
MT:Expr	3	0.00236	0.00215	0.979	0.446

TABLE B.2: PERMANOVA test of genes in greenyellow module of complete network

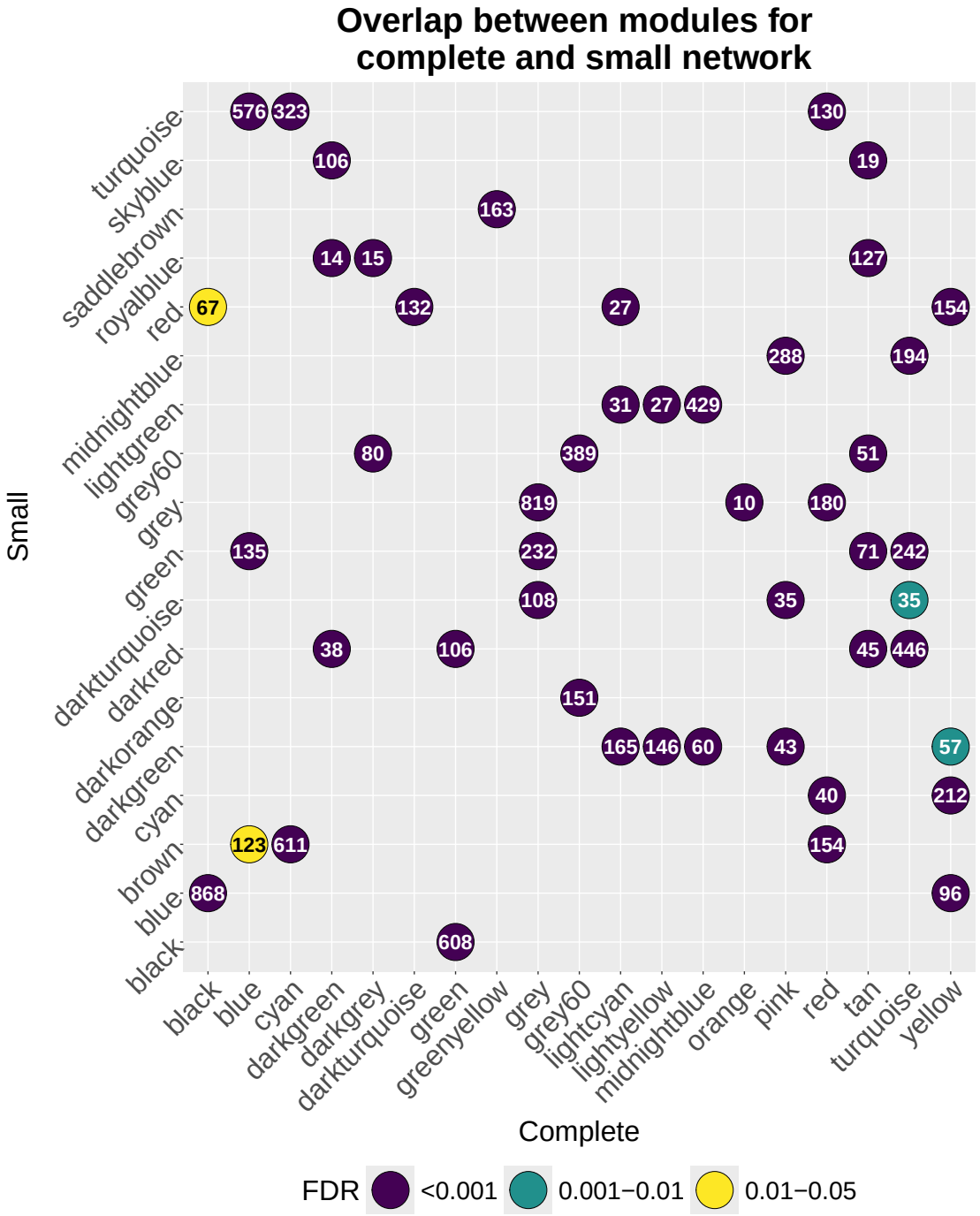


FIGURE B.5: Number of genes shared by any two pair of modules obtained from the complete and small network. Number inside each point is the number of genes and colors indicate range of the FDR. For each module of the small network, the new name was taken to be the name of the module with which it shared the greatest number of genes. It two modules were to be mapped to the same name, only the module with greater amount of shared genes had its name changed.

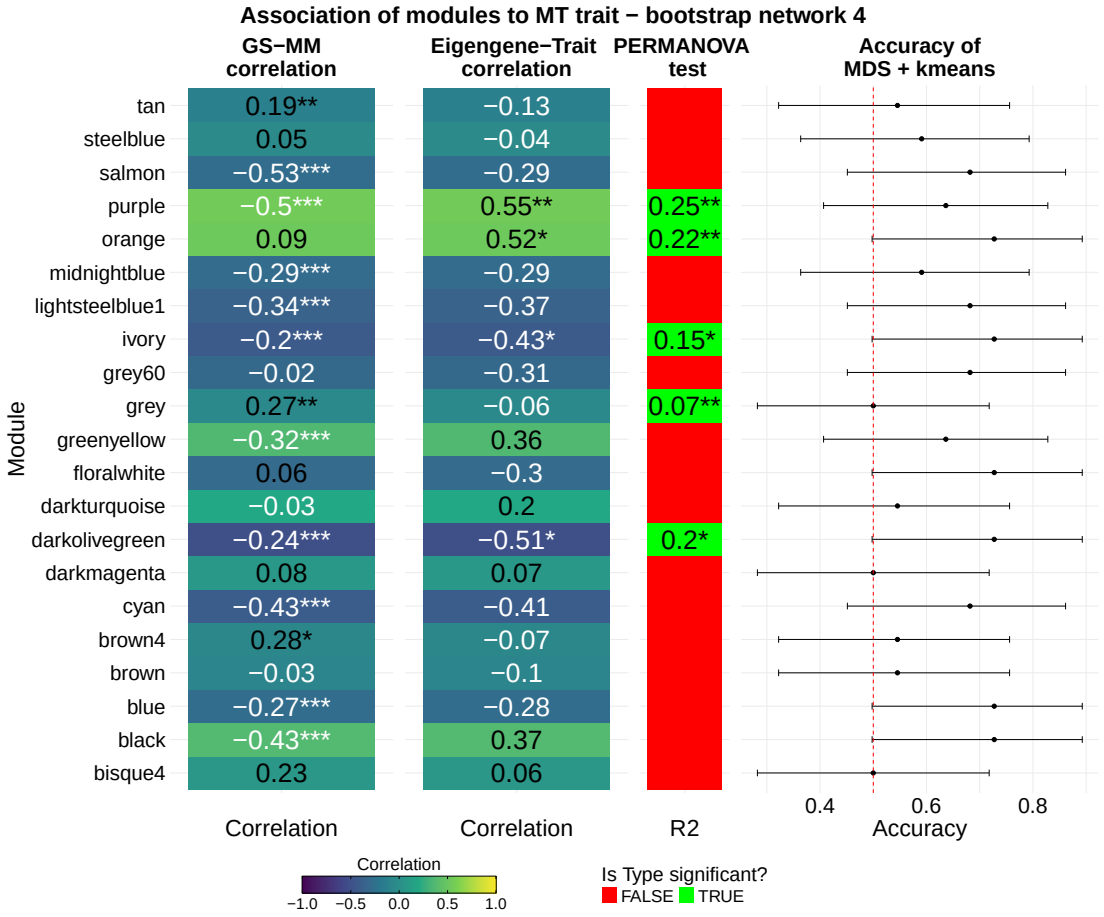


FIGURE B.6: Different statistics used to determine which modules of the bootstrap network 4 can contain gene associated to MT: correlation between gene significance and module membership, eigengene-MT correlation, greatest amount of variance explained by MT in a PERMANOVA test, accuracy computed between the MT and a partition created applying Kmeans to the first 10 MDS axes.

	Df	SumOfSqs	R^2	F-value	p-value
MT	1	0.003716	0.10253	1.4734	0.235
Size	1	-0.000127	-0.00350	-0.0503	1.000
MT:Size	5	0.001835	0.05063	0.1455	0.993
MT:Expr	1	0.000195	0.00539	0.0775	0.942

TABLE B.3: PERMANOVA test of enriched GO terms in black module of bootstrap network 2

	Df	SumOfSqs	R^2	F-value	p-value
Size	1	0.009517	0.13663	4.9733	0.009
Expr	5	0.032004	0.45948	3.3449	0.001
Size:Expr	1	0.001342	0.01927	0.7012	0.563

TABLE B.4: PERMANOVA test genes in darkred module of bootstrap network 2

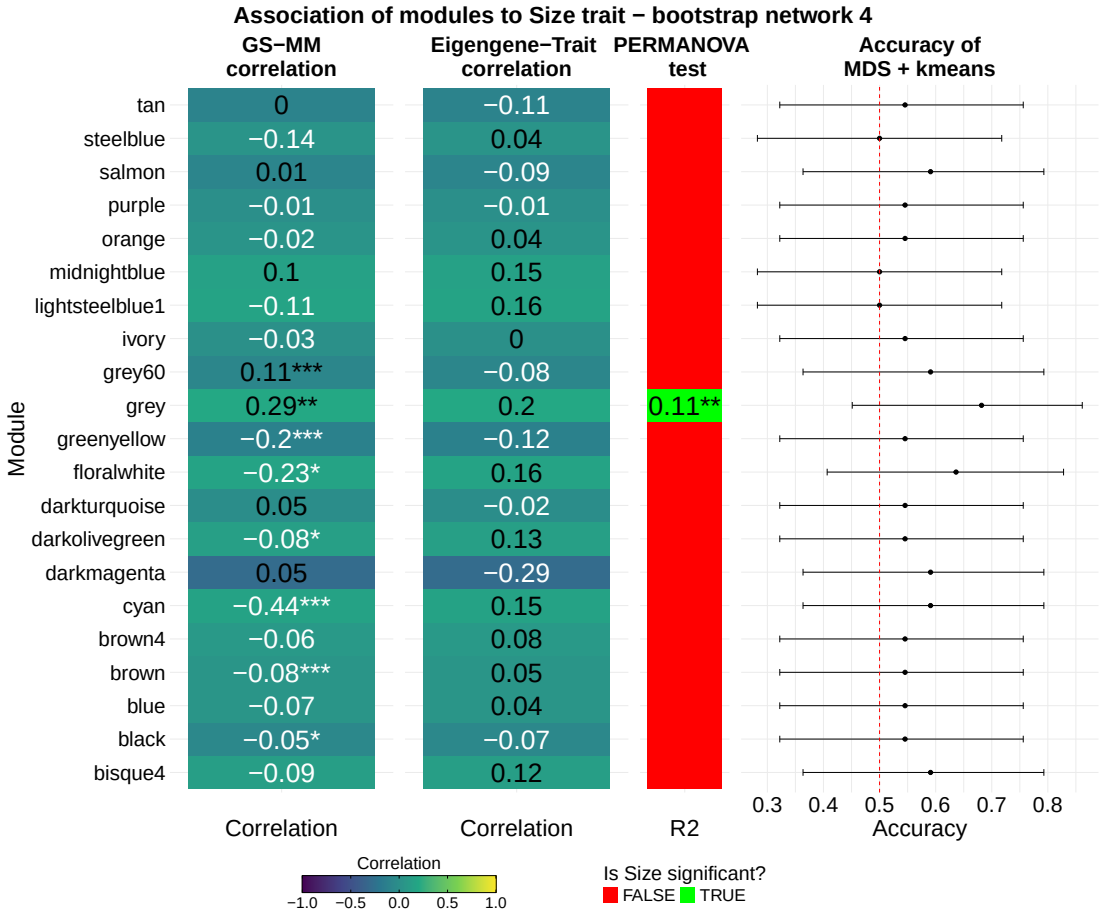


FIGURE B.7: Different statistics used to determine which modules of the bootstrap network 4 can contain gene associated to size: correlation between gene significance and module membership, eigengene-MT correlation, greatest amount of variance explained by size in a PERMANOVA test, accuracy computed between the size and a partition created applying Kmeans to the first 10 MDS axes.

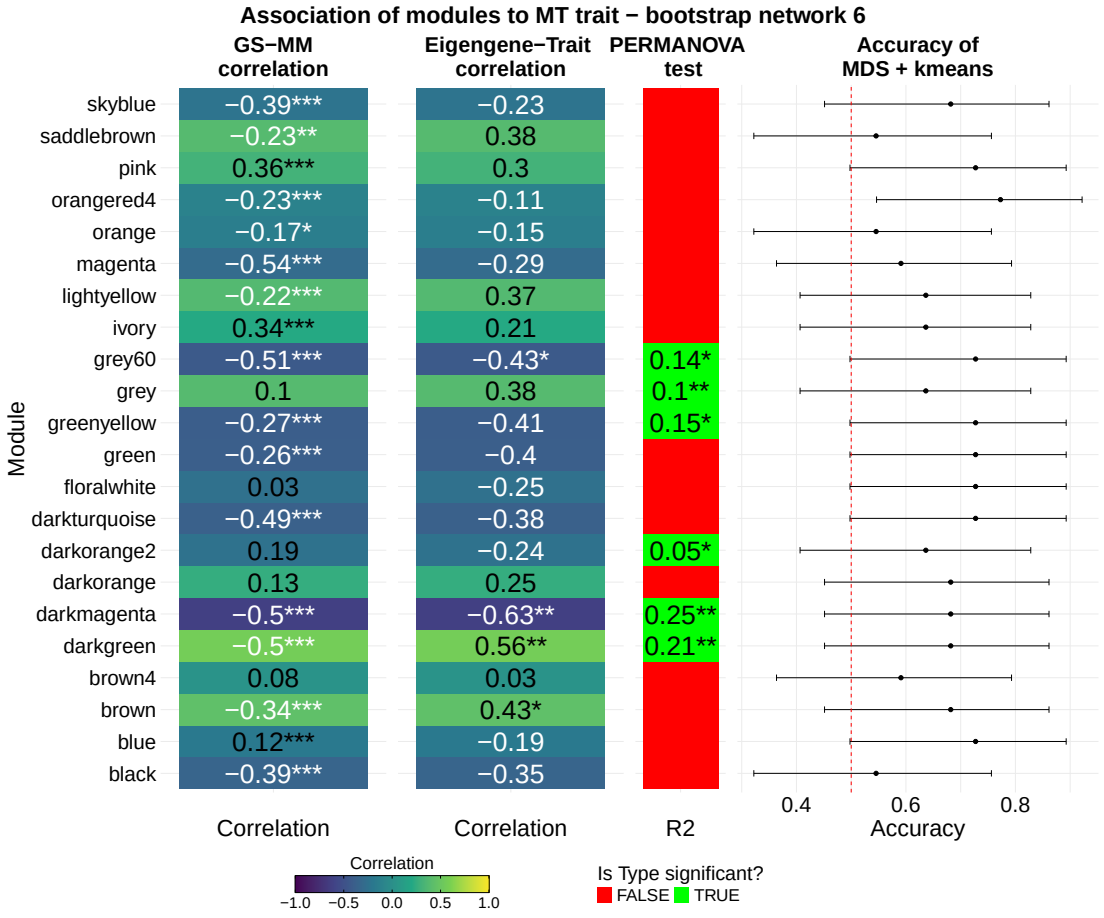


FIGURE B.8: Different statistics used to determine which modules of the bootstrap network 6 can contain gene associated to MT: correlation between gene significance and module membership, eigengene-MT correlation, greatest amount of variance explained by MT in a PERMANOVA test, accuracy computed between the MT and a partition created applying Kmeans to the first 10 MDS axes.

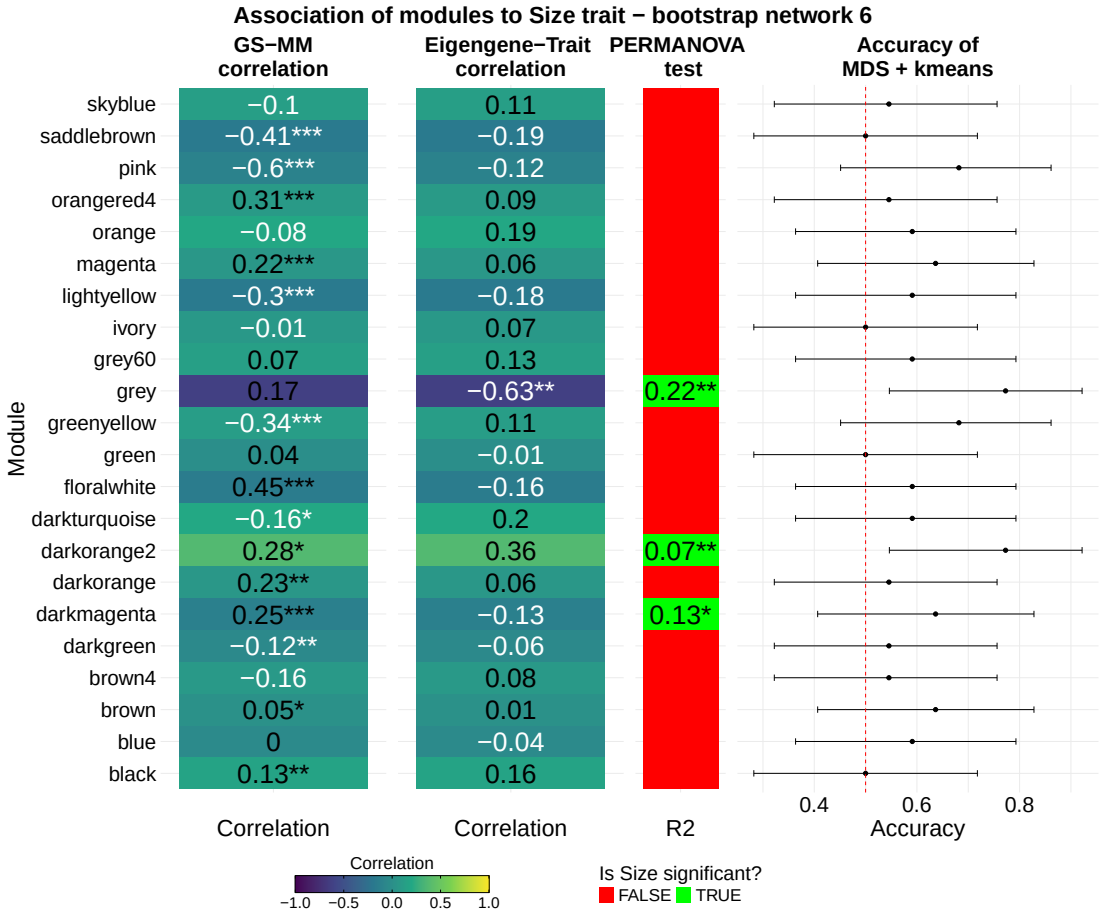


FIGURE B.9: Different statistics used to determine which modules of the bootstrap network 6 can contain gene associated to size: correlation between gene significance and module membership, eigengene-MT correlation, greatest amount of variance explained by size in a PERMANOVA test, accuracy computed between the size and a partition created applying Kmeans to the first 10 MDS axes.

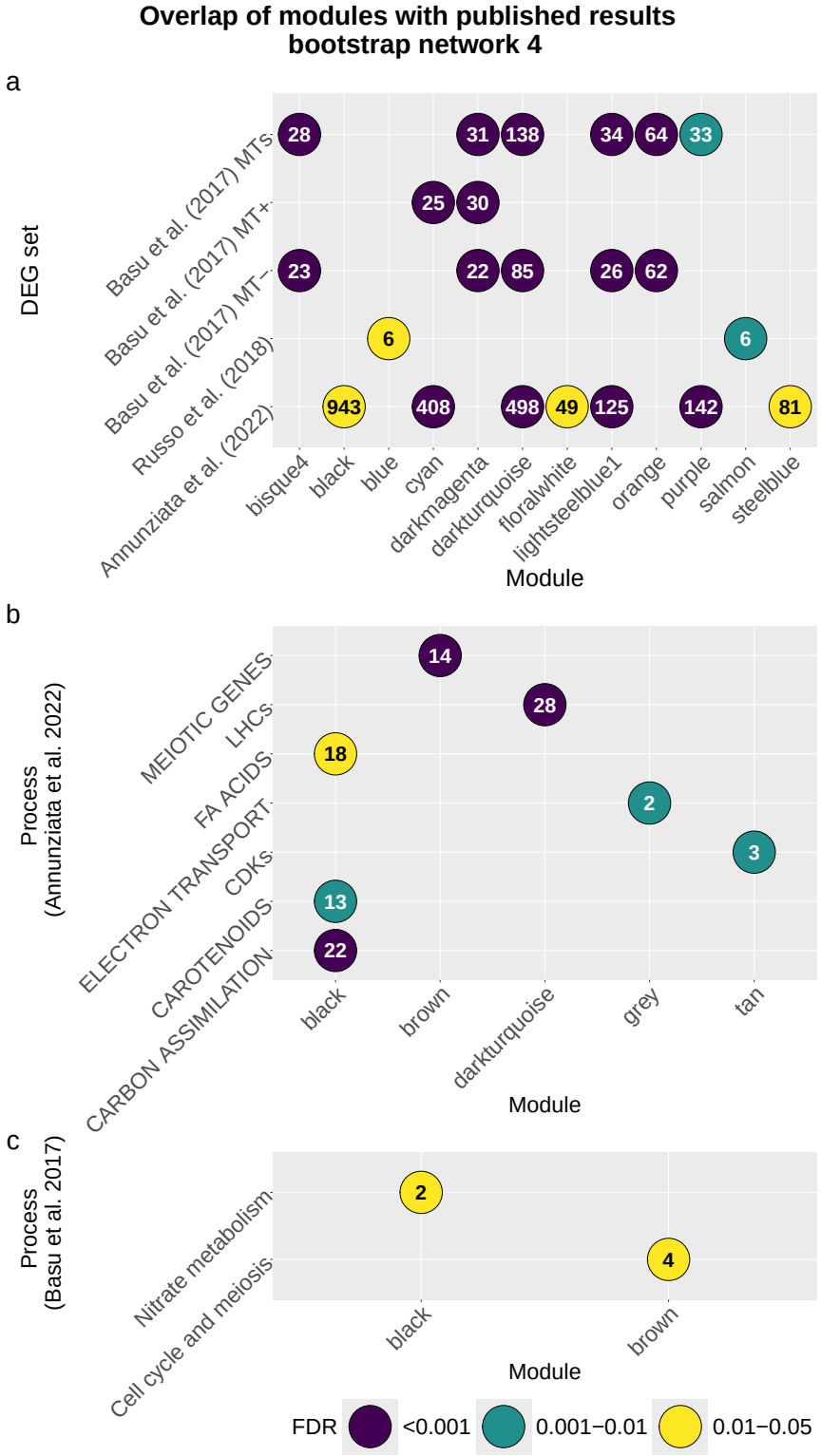


FIGURE B.10: Overlap between modules in the bootstrap network 4 and groups of relevant genes found in published papaers about *P. multistriata* (see caption of Fig. 3.6 for details).

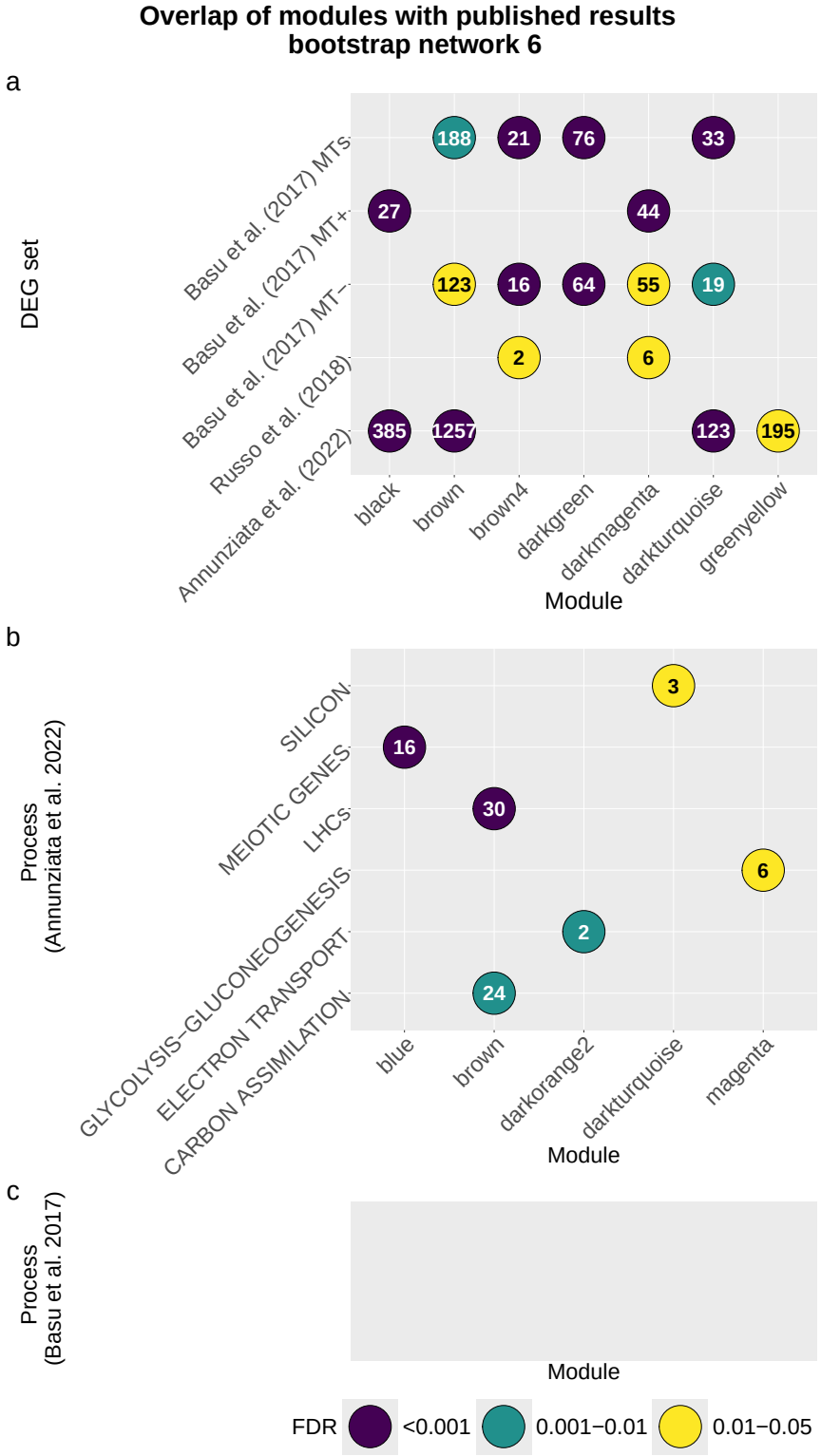
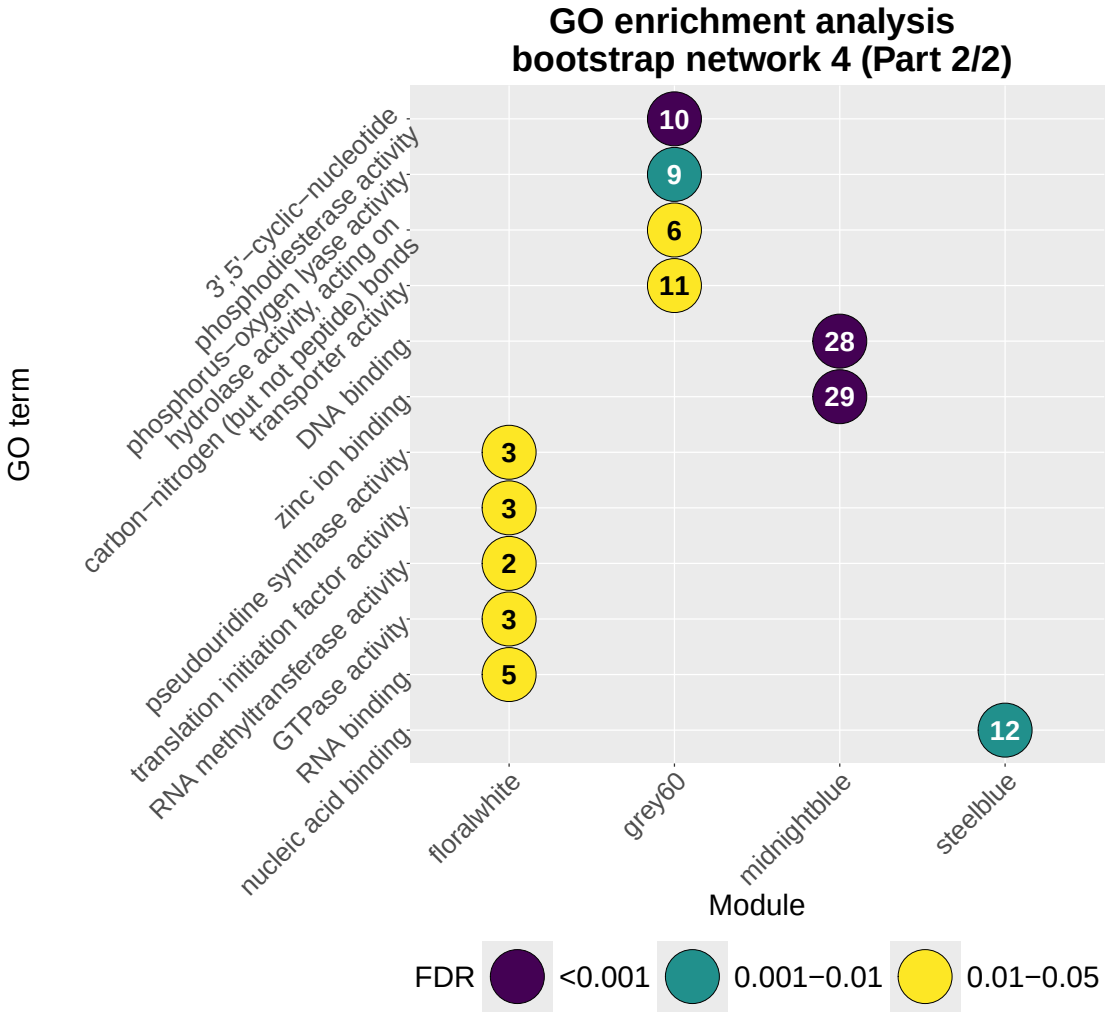


FIGURE B.11: Overlap between modules in the bootstrap network 6 and groups of relevant genes found in published papaers about *P. multistriata* (see caption of Fig. 3.6 for details).

(A) Enrichment analysis using GO for black, brown, cyan and darkturquoise modules.

FIGURE B.12: Enrichment analysis of the modules in the bootstrap network 4 using gene ontology (GO) information. (Continues on the following pages)

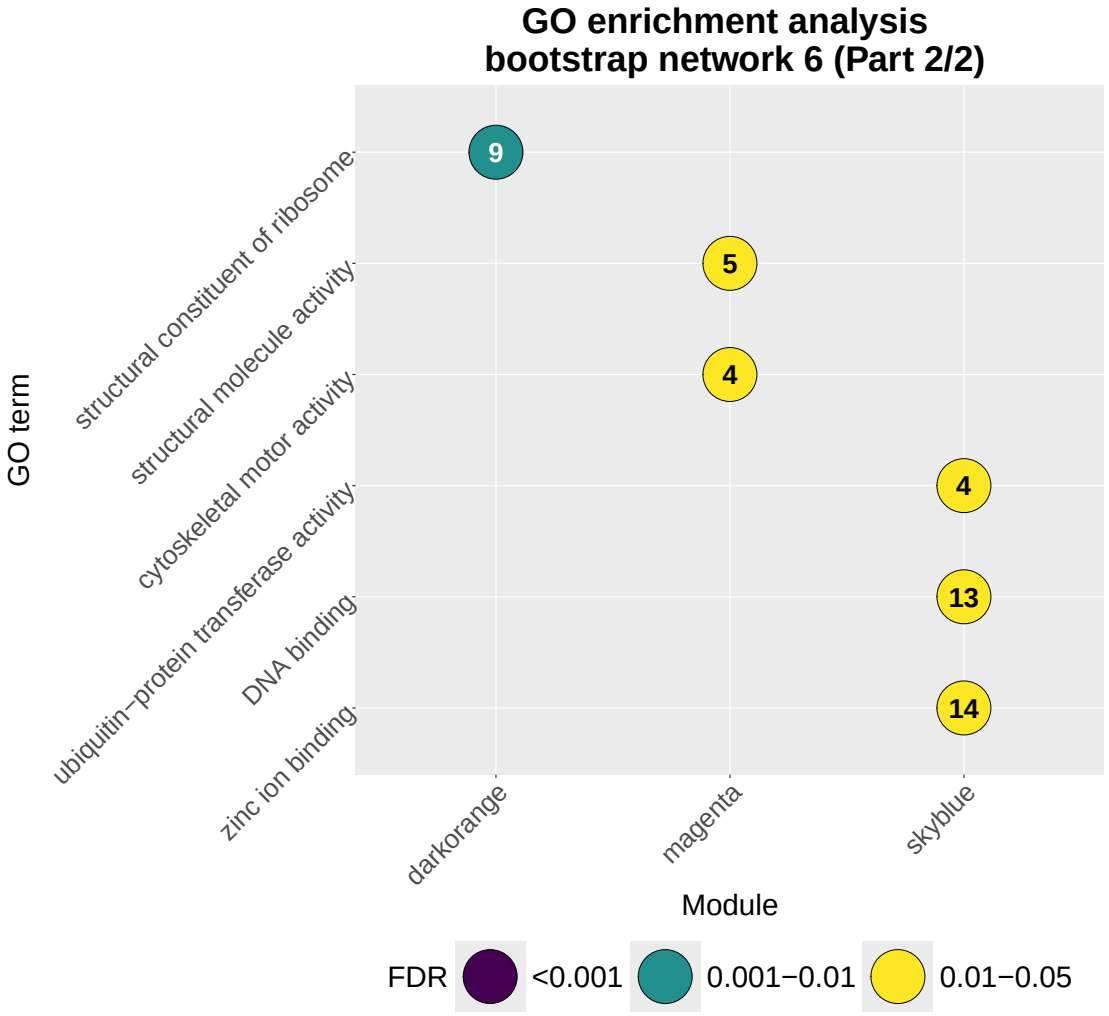


(B) Enrichment analysis using GO for floralwhite, grey60, midnightblue, steelblue modules.

FIGURE B.12: Enrichment analysis of the modules in the bootstrap network 4 using gene ontology (GO) information.

(A) Enrichment analysis using GO for black, blue, brown and darkmagenta modules.

FIGURE B.13: Enrichment analysis of the modules in the small network using gene ontology (GO) information. (Continues on the following pages)



(B) Enrichment analysis using GO for darkorange, magenta and skyblue modules.

FIGURE B.13: Enrichment analysis of the modules in the small network using gene ontology (GO) information.

Appendix C

Comparing Pseudonitzschia multistriata and Pseudonitzschia arenysensis

As discussed in Chapter 3, comparative analyses across multiple species can provide valuable insights into both species-specific traits and, more importantly, conserved molecular characteristics that may reveal fundamental biological functions. In this appendix, we present preliminary results from an RNA-seq experiment conducted on *Pseudonitzschia arenysensis*, aimed at comparing its transcriptional response with that previously characterized in *P. multistriata*.

This work forms part of a collaborative study with Andrea Broccoli, PhD candidate at the *Open University*, XXIV cycle, *School of Life, Health and Chemical Sciences*. These analyses are included in his doctoral thesis, entitled *Evolution of the Sex Determination Mechanisms in the Diatom Genus Pseudo-nitzschia* (Chapter 6: *Demographic and Transcriptional Dynamics During the Sexual Phase of P. arenysensis*), which was under revision at the time of writing. A. Broccoli was responsible for the experimental design, execution, and generation of the raw data, while we carried out the bioinformatic analyses reported in this appendix.

The experimental design closely followed that described in [13]. Briefly, two monocultures of opposite MTs and one cross-culture containing both MTs were established. RNA was extracted from the cross-culture and the two monocultures (used as controls) at three successive time points, in order to capture the temporal dynamics of the transcriptional response. All RNA-seq data were pre-processed following the workflow described in Section 3.1.2, using the parameters reported in Table C.1.

Software	Parameters
trimmomatic	ILLUMINACLIP::2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36 HEADCROP:10
STAR	--outSAMtype BAM SortedByCoordinate --outSAMunmapped Within --outFilterMultimapNmax 5 --seedMultimapNmax 1000 --seedSearchStartLmax 25 --outFilterMismatchNoverLmax 0.01 --outFilterMatchNminOverLread 0.9 --outFilterIntronMotifs RemoveNoncanonical --outSAMstrandField intronMotif --outSJfilterCountUniqueMin 5 5 5 5 --outFilterType BySJout

TABLE C.1: List of non-default parameters used in `Trimmomatic` and `STAR` software to generate raw counts from the RNA-seq libraries of *Pseudonitzschia multistriata* and *P. arenysensis*

To compute the differential expressed genes (DGEs) we used the R package `EdgeR` and conditions for being differentially expressed where False Discovery Rate (FDR) < 0.5, $|log(FC)| > 1$ and q-value < 0.2. for each of the three time points, we analyzed these contrasts: MT+ control against cross culture, MT- control against cross-culture, both controls against cross-culture. DEGs were identified using the R package `EdgeR`. Genes were considered significantly differentially expressed if they satisfied the following criteria: False Discovery Rate (FDR) < 0.05, absolute log fold change $|log_2FC| > 1$, and q-value < 0.2. For each of the three time points, differential expression analyses were performed for the following contrasts: (i) control MT+ versus cross-culture, (ii) control MT- versus cross-culture, and (iii) both controls versus cross-culture; Tab C.2 reports the results.

Overall, the number of DEGs is lowest at T1 and increases monotonically across subsequent time points in all cases, except for the number of up-regulated genes obtained when the MT+ and MT- moncultures are used as controls, where the maximum occurs at T2. The cross-culture condition exhibits a markedly higher number of up-regulated than down-regulated genes when compared against MT+ and MT- controls at T1 and T2; however, this trend is reversed at T3. When the cross-culture is compared against both monocultures combined (unsex), a greater number of down-regulated genes is

Control	Regulation	T1	T2	T3
MT+	Down	315	607	1904
	Up	994	2077	1718
MT-	Down	603	1319	3353
	Up	1054	3117	2876
Unsex	Down	62	304	1990
	Up	38	1191	1369

TABLE C.2: Number of genes differentially expressed at time points T1, T2, and T3, divided according to the control used (MT+, MT-, unsex) and the direction of regulation (up- or down-regulated).

observed at T1, although the overall counts remain low relative to the other comparisons. se results already suggest a highly dynamic transcriptional response in the cross-cultures, particularly given the rapid shifts in the numbers of up- and down-regulated genes over time. To better characterize this dynamic behavior, it is instructive to examine the UpSet plots for each time point (Figs. C.1, C.2, C.3). These plots illustrate the number of DEGs unique to each group as well as those shared among two or more groups. They are conceptually similar to Venn diagrams but offer a clearer visualization when more than four groups are involved.

The

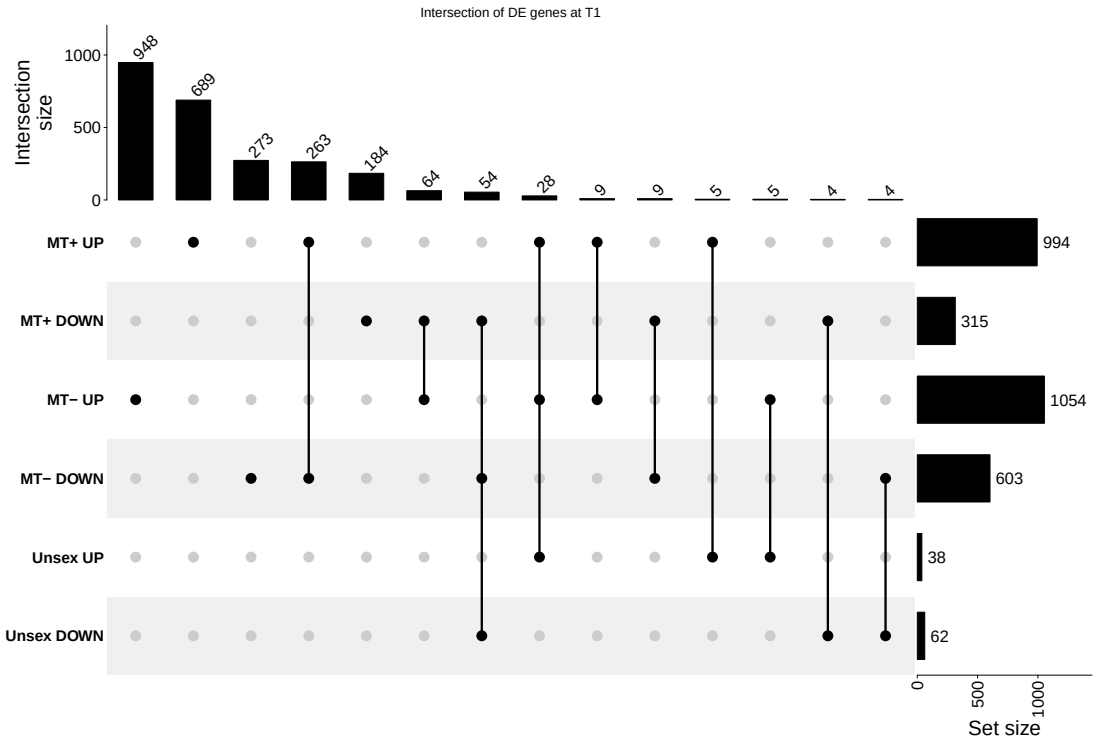


FIGURE C.1: UpSet plot showing the number of differentially expressed genes at T1 shared among groups defined by control status and direction of regulation (up- or down-regulated).

At T1 (Fig. C.1), we can see that many DEGs obtained when testing against MT- are unique to this case: 948 out of the 1054 genes up-regulated unique to this case and analogously 273 out of the 603 down-regulated genes. Similarly, 689 out of the 994 genes up-regulated and 184 out of 315 genes down-regulated when testing against MT+ are unique to these cases.At T1 (Fig. C.1), many DEGs identified when testing against MT- are unique to this comparison: 948 out of the 1054 up-regulated genes and 273 out of the 603 down-regulated genes are exclusive to this contrast. Similarly, when testing against MT+, 689 out of 994 up-regulated genes and 184 out of 315 down-regulated genes are unique to this case. A substantial fraction of genes (263) are up-regulated when MT+ is used as the control but down-regulated whenMT- is used, whereas 64 genes show the opposite pattern.

In contrast, genes differentially expressed in the unsex comparison display a more consistent pattern with the other contrasts: 54 of its down-regulated genes exhibit the same behavior when using MT+ and MT- as controls, and 28 up-regulated genes show similar consistency. The remaining genes are also coherently regulated with either MT+ or MT-. The presence of genes showing opposite regulation between the MT+ and MT- contrasts suggests the occurrence of mating-type-specific transcriptional responses. Conversely, the complete concordance observed in the unsex case is expected,

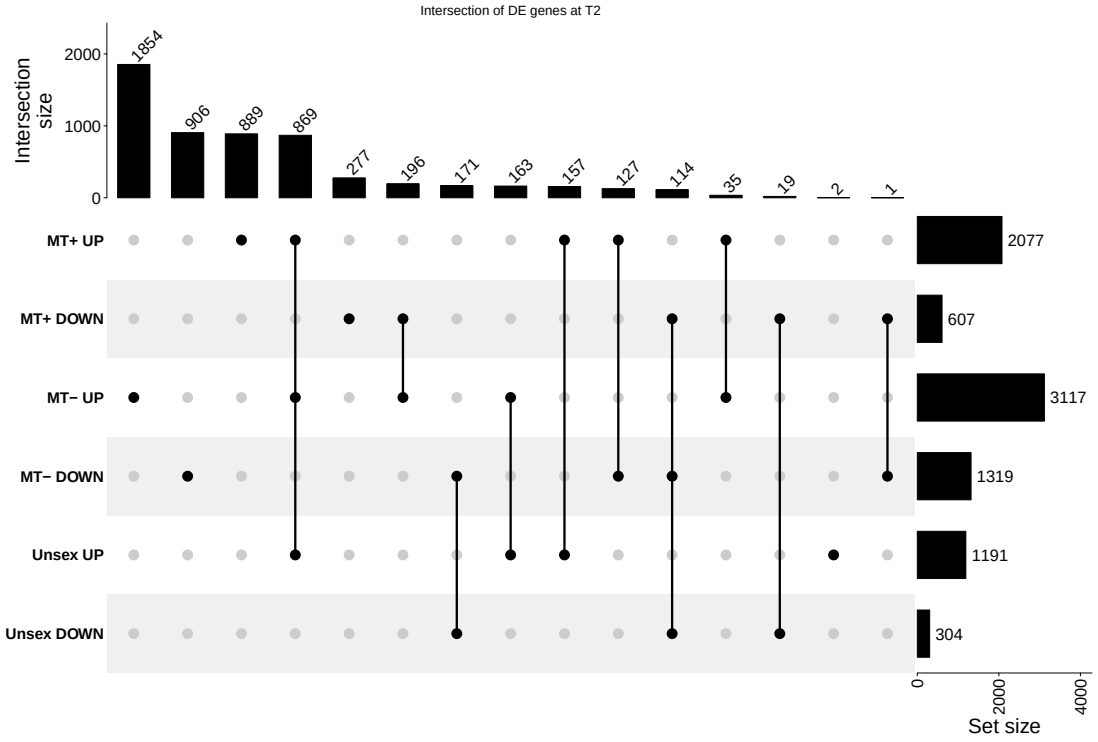


FIGURE C.2: UpSet plot showing the number of differentially expressed genes at T2 shared among groups defined by control status and direction of regulation (up- or down-regulated).

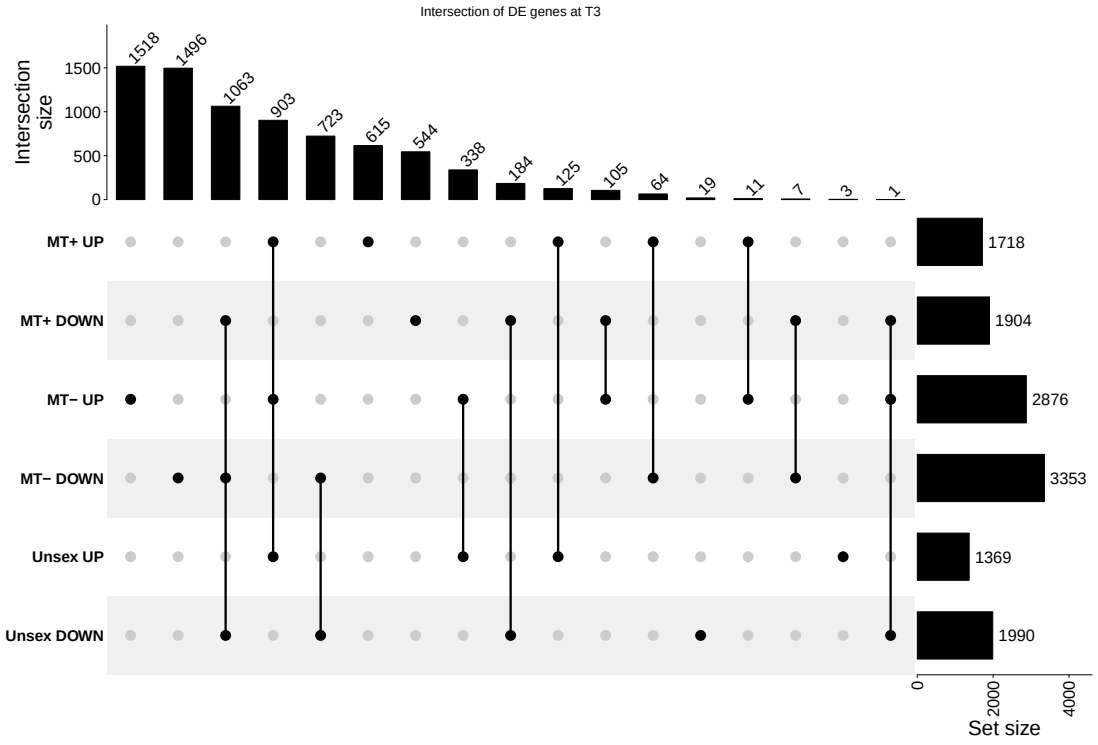


FIGURE C.3: UpSet plot showing the number of differentially expressed genes at T3 shared among groups defined by control status and direction of regulation (up- or down-regulated).

		Arenysensis			
		UP	DOWN	No DE	No orthologue
Multistriata	UP	9	13	510	372
	DOWN	10	2	793	311
	No DE	50	56	0	0
	No orthologue	55	78	0	0

TABLE C.3: Distribution of DEGs at T1 between *P. arenysensis* and *P. multistriata*. Both rows and columns indicate the transcriptional status of genes in one-to-one orthologous relationships between the two species, classified as "UP", "DOWN", or "No DE" (not differentially expressed). Entries in the "No orthologue" row or column correspond to genes that are differentially expressed in one species but lack an identified orthologue in the other.

as both monocultures were used as controls; consequently, only genes consistently regulated relative to both MT+ and MT- could emerge from this analysis.
A similar pattern is observed at T2 and T3 (Figs. C.2, C.3), differing only in the specific numbers of genes involved. Notably, across all time points, a considerable fraction of DEGs remain unique to the MT- contrast.

Among the many features provided by `OrthoFinder`, we focus here only on those relevant to this analysis. Specifically, we were interested in identifying orthologous relationships between the two species, which can take several forms: one-to-one (a single gene from species A corresponds uniquely to a single gene from species B), one-to-many (a gene from species A corresponds to multiple genes from species B, or vice versa), and many-to-many (multiple genes from species A correspond to multiple genes from species B). In the present study, only one-to-one orthologous pairs were considered, in order to establish unambiguous gene relationships between the two species. Given the similarity between this experiment and that described in [13], we compared the sets of DEGs obtained using both monocultures as controls. Tables C.3, C.4, and C.5 report, for T1, T2, and T3 respectively, the distribution of DEGs for the two species. Our aim was to determine the extent to which the transcriptional responses are conserved between them. At all time points, despite the increasing number of DEGs, most genes differentially expressed in one species are either not differentially expressed or lack an identified orthologue in the other. The highest level of concordance is observed at T2, where 313 genes are up-regulated and 68 are down-regulated in both species, whereas the number of discordant cases (where one gene is up-regulated and its orthologue is down-regulated, or vice versa) is considerably lower.

To conclude, [13] compiled several lists of DEGs identified during sexualization, with each list corresponding to genes involved in a specific biological process. We performed

		Arenysensis			
		UP	DOWN	No DE	No orthologue
Multistriata	UP	313	10	977	696
	DOWN	42	68	973	606
	No DE	321	97	0	0
	No orthologue	734	202	0	0

TABLE C.4: Distribution of DEGs at T2 between *P. arenysensis* and *P. multistriata*. Both rows and columns indicate the transcriptional status of genes in one-to-one orthologous relationships between the two species, classified as "UP", "DOWN", or "No DE" (not differentially expressed). Entries in the "No orthologue" row or column correspond to genes that are differentially expressed in one species but lack an identified orthologue in the other.

		Arenysensis			
		UP	DOWN	No DE	No orthologue
Multistriata	UP	72	283	621	553
	DOWN	89	106	598	449
	No DE	465	1203	0	0
	No orthologue	918	874	0	0

TABLE C.5: Distribution of DEGs at T3 between *P. arenysensis* and *P. multistriata*. Both rows and columns indicate the transcriptional status of genes in one-to-one orthologous relationships between the two species, classified as "UP", "DOWN", or "No DE" (not differentially expressed). Entries in the "No orthologue" row or column correspond to genes that are differentially expressed in one species but lack an identified orthologue in the other.

the same analysis by identifying the orthologues of these genes from *P. multistriata* and examining their $\log_2$ fold-change values in *P. arenysensis*, as shown in Fig. C.4. Consistent with the observations reported in the previous tables, most orthologues are not differentially expressed in *P. arenysensis* at one or more time points, and in many cases not at all. Nevertheless, certain patterns resemble those previously described in *P. multistriata*: genes associated with meiosis are up-regulated at T2, while several genes involved in nutrient uptake (e.g., nitrogen and phosphate transport) are down-regulated at T2 and T3.

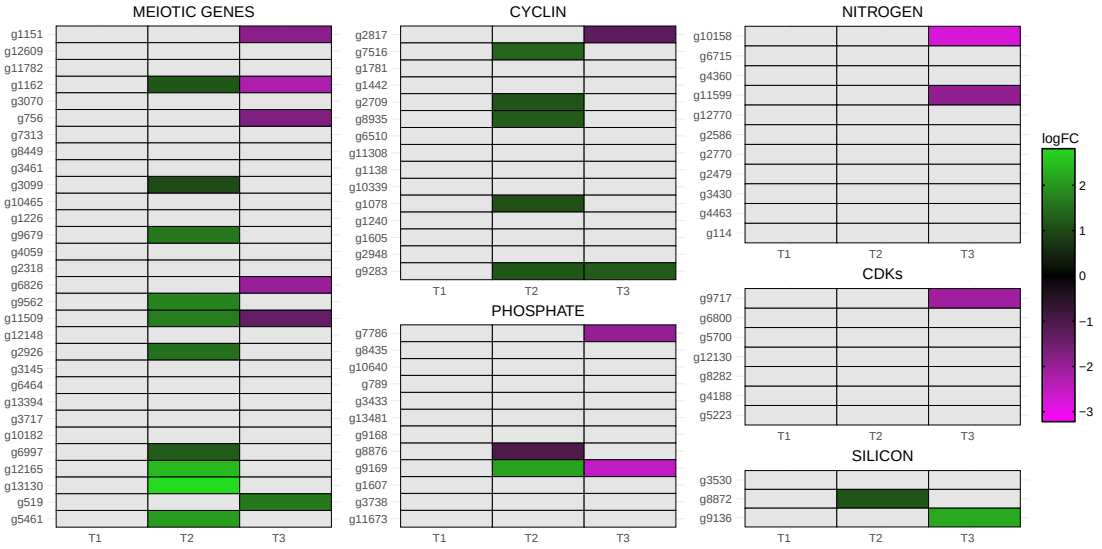


FIGURE C.4: Log fold-change of genes involved in processes influenced by partner perception during the sexualization phase in *P. arenysensis*. These processes were identified in *P. multistriata* by [13], and *P. arenysensis* genes were selected based on orthology analysis. Grey panels indicate the gene was not differentially expressed at the specific timepoint.

Bibliography

[1] https://archive.jgi.doe.gov/csp-2021-100-diatom-genomes/. [Online; last accessed 22-September-2025].

[2] https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. [Online; last accessed 22-September-2025].

[3] https://bioinformatics.stackexchange.com/questions/21885/where-to-access-the-wgcna-tutorial-documents-horvath-lab-site-down. [Online; last accessed 19-September-2025].

[4] Abreu, P. et al. "Short- and Long-Term Chlorophyll a Variability in the Shallow Microtidal Patos Lagoon Estuary, Southern Brazil". In: *Estuaries and Coasts* 33 (Mar. 2010), pp. 554–569. DOI: 10.1007/s12237-009-9181-9.

[5] Aebersold, R. and Mann, M. "Mass-spectrometric exploration of proteome structure and function". In: *Nature* 537 (Sept. 2016), pp. 347–355. DOI: 10.1038/nature19949.

[6] Ait-Mohamed, O. et al. "PhaeoNet: A Holistic RNAseq-Based Portrait of Transcriptional Coordination in the Model Diatom Phaeodactylum tricornutum". In: *Frontiers in Plant Science* Volume 11 - 2020 (2020). ISSN: 1664-462X. DOI: 10.3389/fpls.2020.590949. URL: https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2020.590949.

[7] Alexander, H. et al. "Functional group-specific traits drive phytoplankton dynamics in the oligotrophic ocean". In: *Proceedings of the National Academy of Sciences of the United States of America* 112.44 (Nov. 3, 2015), E5972–E5979. ISSN: 0027-8424. DOI: 10.1073/pnas.1518165112. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC4640801/.

[8] Allen, J. et al. "Characterization of DNA from dinoflagellate Crypthecodinium cohnii and implications for nuclear organization". In: *Cell* 6 (Nov. 1975), pp. 161–9. DOI: 10.1016/0092-8674(75)90006-9.

[9] Alster, A. et al. "Analyzing semiquantitative phytoplankton counts". In: *Hydrobiologia* 851.4 (Feb. 2024), pp. 1079–1090. ISSN: 0018-8158, 1573-5117. DOI: 10.1007/s10750-023-05391-4. URL: https://link.springer.com/10.1007/s10750-023-05391-4.

[10] Altwegg, R. et al. "Emerging Topics and New Directions in Statistical Ecology". In: *Journal of Statistical Theory and Practice* 19.3 (Sept. 2025), p. 44. ISSN: 1559-8608, 1559-8616. DOI: 10.1007/s42519-025-00460-4. URL: https://link.springer.com/10.1007/s42519-025-00460-4.

[11] Amato, A., Lüdeking, A., and Kooistra, W. "Intracellular domoic acid production in Pseudo-nitzschia multistriata isolated from the Gulf of Naples (Tyrrhenian Sea, Italy)". In: *Toxicon : official journal of the International Society on Toxinology* 55 (Jan. 2010), pp. 157–61. DOI: 10.1016/j.toxicon.2009.07.005.

[12] Anders, S. and Huber, W. "Differential expression analysis for sequence count data". In: *Nature Precedings* 5 (Apr. 2010). DOI: 10.1038/npre.2010.4282.2.

[13] Annunziata, R. et al. "Trade-off between sex and growth in diatoms: Molecular mechanisms and demographic implications". In: *Science advances* 8 (Jan. 2022), eabj9466. DOI: 10.1126/sciadv.abj9466.

[14] Armbrust, E. et al. "The Genome of the Diatom Thalassiosira Pseudonana: Ecology, Evolution, and Metabolism". In: *Science (New York, N.Y.)* 306 (Nov. 2004), pp. 79–86. DOI: 10.1126/science.1101156.

[15] Armbrust, E. V. "The life of diatoms in the world's oceans". In: *Nature* 459.7244 (May 2009), pp. 185–192. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature08057. URL: https://www.nature.com/articles/nature08057.

[16] Arrigo, K. R. "Marine microorganisms and global nutrient cycles". In: *Nature* 437.7057 (Sept. 15, 2005), pp. 349–355. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature04159. URL: https://www.nature.com/articles/nature04159.

[17] Artegiani, A. et al. "The Adriatic Sea General Circulation. Part I: Air–Sea Interactions and Water Mass Structure". In: *Journal of Physical Oceanography* 27.8 (Aug. 1997), pp. 1492–1514. ISSN: 0022-3670, 1520-0485. DOI: 10.1175/1520-0485(1997)027<1492:TASGCP>2.0.CO;2. URL: http://journals.ametsoc.org/doi/10.1175/1520-0485(1997)027%3C1492:TASGCP%3E2.0.CO;2.

[18] Ascioti, F. A. et al. "Is there chaos in plankton dynamics?" In: *Journal of Plankton Research* 15.6 (1993), pp. 603–617. ISSN: 0142-7873, 1464-3774. DOI: 10.1093/plankt/15.6.603. URL: https://academic.oup.com/plankt/article-lookup/doi/10.1093/plankt/15.6.603.

[19] Ashworth, J. et al. “Genome-wide diel growth state transitions in the diatom ¡i¿Thalassiosira pseudonana¡/i¿”. In: *Proceedings of the National Academy of Sciences* 110.18 (2013), pp. 7518–7523. DOI: 10.1073/pnas.1300962110. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1300962110. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1300962110.

[20] Aubry, F. et al. “Phytoplankton succession in a coastal area of the NW Adriatic, over a 10-year sampling period (1990–1999)”. In: *Continental Shelf Research* 24.1 (Jan. 2004), pp. 97–115. ISSN: 02784343. DOI: 10.1016/j.csr.2003.09.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S0278434303001821.

[21] Azam, F. et al. “The Ecological Role of Water-Column Microbes in the Sea”. In: *Marine Ecology Progress Series* 10 (1983), pp. 257–263. ISSN: 0171-8630, 1616-1599. DOI: 10.3354/meps010257. URL: http://www.int-res.com/articles/meps/10/m010p257.pdf.

[22] Azzaro, F. et al. “Seasonal variability of phytoplankton fluorescence in relation to the Straits of Messina (Sicily) tidal upwelling”. In: *Ocean Science* 3.4 (2007), pp. 451–460. DOI: 10.5194/os-3-451-2007. URL: https://os.copernicus.org/articles/3/451/2007/.

[23] Badia-i-Mompel, P. et al. “Gene regulatory network inference in the era of single-cell multi-omics”. In: *Nature Reviews Genetics* 24.11 (Nov. 2023). Publisher: Nature Publishing Group, pp. 739–754. ISSN: 1471-0064. DOI: 10.1038/s41576-023-00618-5. URL: https://www.nature.com/articles/s41576-023-00618-5.

[24] Baek, S. H. et al. “Ecological behavior of the dinoflagellate Ceratium furca in Jangmok harbor of Jinhae Bay, Korea”. In: *Journal of Plankton Research* 33 (Nov. 2011), pp. 1842–1846. DOI: 10.1093/plankt/fbr075.

[25] Bandelj, V. et al. *Rapporto tecnico scientifico conclusivo, Convenzione Strategia Marina 2021, Accordo di collaborazione scientifica tra ISPRA e OGS nell’ambito dello studio sulle tematiche della Strategia Marina ai fini del conseguimento degli obiettivi prioritari della direttiva 2008/56/CE, come recepita dal D.Lgs. 190/2010 (CUP F85B18006490001, Atto OGS n. 30 ADW del 21/07/2021), Parte 1, Descrittore 5 - Eutrofizzazione*. 2023. URL: https://ricerca.ogs.it/handle/20.500.14083/35204.

[26] Bandelj, V. et al. “Fuzziness and Heterogeneity of Benthic Metacommunities in a Complex Transitional System”. In: *PLOS ONE* 7.12 (Dec. 2012), pp. 1–15. DOI: 10.1371/journal.pone.0052395. URL: https://doi.org/10.1371/journal.pone.0052395.

[27] Barabasi, A.-L. "Scale-Free Networks: A Decade and Beyond". In: *Science (New York, N.Y.)* 325 (Aug. 2009), pp. 412–3. DOI: 10.1126/science.1173299.

[28] Barkai, N., Rose, M., and Wingreen, N. "Protease helps yeast find mating partners". In: *Nature* 396 (Jan. 1999), pp. 422–3. DOI: 10.1038/24760.

[29] Basu, S. et al. "Finding a partner in the ocean: Molecular and evolutionary bases of the response to sexual cues in a planktonic diatom". In: *New Phytologist* 215 (Apr. 2017). DOI: 10.1111/nph.14557.

[30] Bates, S. S. et al. "Pseudo-nitzschia, Nitzschia, and domoic acid: New research since 2011". In: *Harmful Algae* 79 (2018). Domoic acid 30 years on, pp. 3–43. ISSN: 1568-9883. DOI: https://doi.org/10.1016/j.hal.2018.06.001. URL: https://www.sciencedirect.com/science/article/pii/S156898831830091X.

[31] Bazzoni, A. et al. "Spatial distribution and multiannual trends of potentially toxic microalgae in shellfish farms along the Sardinian coast (NW Mediterranean Sea)". In: *Environmental Monitoring and Assessment* 187 (Mar. 2015). DOI: 10.1007/s10661-014-4250-3.

[32] Behrenfeld, M. J. "Abandoning Sverdrup's Critical Depth Hypothesis on phytoplankton blooms". In: *Ecology* 91.4 (2010), pp. 977–989. ISSN: 1939-9170. DOI: 10.1890/09-1207.1. URL: https://onlinelibrary.wiley.com/doi/abs/10.1890/09-1207.1.

[33] Behrenfeld, M. J. and Falkowski, P. G. "Photosynthetic rates derived from satellite-based chlorophyll concentration". In: *Limnology and Oceanography* 42.1 (Jan. 1997), pp. 1–20. ISSN: 0024-3590, 1939-5590. DOI: 10.4319/lo.1997.42.1.0001. URL: https://aslopubs.onlinelibrary.wiley.com/doi/10.4319/lo.1997.42.1.0001.

[34] Belišová, D. et al. "Molecular fingerprints of cell size sensing and mating type differentiation in pennate diatoms". In: *New Phytologist* 245.4 (2025), pp. 1625–1639. DOI: https://doi.org/10.1111/nph.20334. URL: https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/nph.20334.

[35] Berdalet, E. "EFFECTS OF TURBULENCE ON THE MARINE DINOFLAGELLATE GYMNODINIUM NELSONII". In: *Journal of Phycology* 28.3 (1992), pp. 267–272. DOI: https://doi.org/10.1111/j.0022-3646.1992.00267.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0022-3646.1992.00267.x.

[36] Bergamasco, A. and Malanotte-Rizzoli, P. "The circulation of the Mediterranean Sea: a historical review of experimental investigations". In: *Advances in Oceanography and Limnology* 1.1 (June 2010), pp. 11–28. ISSN: 1947-5721, 1947-573X. DOI: 10.1080/19475721.2010.491656. URL: http://www.tandfonline.com/doi/abs/10.1080/19475721.2010.491656.

[37] Berline, L. et al. "Long-distance particle transport to the central Ionian Sea". In: *Biogeosciences* 18.24 (Dec. 14, 2021), pp. 6377–6392. ISSN: 1726-4189. DOI: 10.5194/bg-18-6377-2021. URL: https://bg.copernicus.org/articles/18/6377/2021/.

[38] Bernardi Aubry, F. et al. "Plankton communities in the northern Adriatic Sea: Patterns and changes over the last 30 years". In: *Estuarine Coastal and Shelf Science* 115 (Dec. 2012), pp. 125–137. DOI: 10.1016/j.ecss.2012.03.011.

[39] Bernardi Aubry, F. et al. "Differences and similarities in the phytoplankton communities of two coupled transitional and marine ecosystems (the Lagoon of Venice and the Gulf of Venice - Northern Adriatic Sea)". In: *Frontiers in Marine Science* 9 (Sept. 2, 2022), p. 974967. ISSN: 2296-7745. DOI: 10.3389/fmars.2022.974967. URL: https://www.frontiersin.org/articles/10.3389/fmars.2022.974967/full.

[40] Bertrand, E. M. et al. "Phytoplankton–bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge". In: *Proceedings of the National Academy of Sciences* 112.32 (Aug. 11, 2015), pp. 9938–9943. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1501615112. URL: https://pnas.org/doi/full/10.1073/pnas.1501615112.

[41] Bilcke, G. and Kamakura, S. "Scaling the invisible wall: Molecular acclimation of a salinity-tolerant diatom to freshwater". In: *Molecular Ecology* 32.11 (June 2023), pp. 2692–2694. ISSN: 0962-1083, 1365-294X. DOI: 10.1111/mec.16971. URL: https://onlinelibrary.wiley.com/doi/10.1111/mec.16971.

[42] Bilcke, G. et al. "Mating type specific transcriptomic response to sex inducing pheromone in the pennate diatom Seminavis robusta". In: *The ISME Journal* 15.2 (Oct. 2020), pp. 562–576. ISSN: 1751-7362. DOI: 10.1038/s41396-020-00797-7. eprint: https://academic.oup.com/ismej/article-pdf/15/2/562/55256939/41396_2020_article_797.pdf. URL: https://doi.org/10.1038/s41396-020-00797-7.

[43] Boldrin, A. et al. "Effects of bora wind on physical and biogeochemical properties of stratified waters in the northern Adriatic". In: *Journal of Geophysical Research: Oceans* 114 (C8 Aug. 2009), 2008JC004837. ISSN: 0148-0227. DOI: 10.1029/

2008JC004837. URL: https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2008JC004837.

[44] Bolger, A. M., Lohse, M., and Usadel, B. "Trimmomatic: a flexible trimmer for Illumina sequence data". In: *Bioinformatics* 30.15 (Apr. 2014), pp. 2114–2120. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btu170. URL: https://doi.org/10.1093/bioinformatics/btu170.

[45] Bolinesi, F. et al. "Spatial and temporal variation of phytoplankton community structure in a coastal marine system subjected to human pressure". In: *Regional Studies in Marine Science* 35 (2020), p. 101198. ISSN: 2352-4855. DOI: https://doi.org/10.1016/j.rsma.2020.101198. URL: https://www.sciencedirect.com/science/article/pii/S2352485519308564.

[46] Bonanno, A. et al. "Variability of water mass properties in the Strait of Sicily in summer period of 1998–2013". In: *Ocean Science* 10.5 (2014), pp. 759–770. DOI: 10.5194/os-10-759-2014. URL: https://os.copernicus.org/articles/10/759/2014/.

[47] Borcard, D., Gillet, F., and Legendre, P. *Numerical Ecology with R*. Jan. 2018. ISBN: 978-3-319-71403-5. DOI: 10.1007/978-3-319-71404-2.

[48] Borg, M. et al. "Red macroalgae in the genomic era". In: *New Phytologist* 240.2 (2023), pp. 471–488. DOI: https://doi.org/10.1111/nph.19211. eprint: https://nph.onlinelibrary.wiley.com/doi/pdf/10.1111/nph.19211. URL: https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/nph.19211.

[49] Bowler, C., Vardi, A., and Allen, A. E. "Oceanographic and biogeochemical insights from diatom genomes". In: *Annual review of marine science* 2 (2010), pp. 333–365.

[50] Bowler, C. et al. "The Phaeodactylum genome reveals the evolutionary history of diatom genomes". In: *Nature* 456 (Nov. 2008), pp. 239–44. DOI: 10.1038/nature07410.

[51] Boyd, P. W. et al. "Mesoscale Iron Enrichment Experiments 1993-2005: Synthesis and Future Directions". In: *Science* 315.5812 (Feb. 2, 2007), pp. 612–617. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1131669. URL: https://www.science.org/doi/10.1126/science.1131669.

[52] Brainerd, K. E. and Gregg, M. C. "Surface mixed and mixing layer depths". In: *Deep Sea Research Part I: Oceanographic Research Papers* 42.9 (Sept. 1995), pp. 1521–1543. ISSN: 09670637. DOI: 10.1016/0967-0637(95)00068-H. URL: https://linkinghub.elsevier.com/retrieve/pii/096706379500068H.

[53] Brunson, J. et al. "Molecular forecasting of domoic acid during a pervasive toxic diatom bloom". In: *Proceedings of the National Academy of Sciences of the United States of America* 121 (Sept. 2024), e2319177121. DOI: 10.1073/pnas.2319177121.

[54] Buffett, G. G. et al. "Seismic Oceanography in the Tyrrhenian Sea: Thermohaline Staircases, Eddies, and Internal Waves". In: *Journal of Geophysical Research: Oceans* 122.11 (Nov. 2017), pp. 8503–8523. ISSN: 2169-9275, 2169-9291. DOI: 10.1002/2017JC012726. URL: https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2017JC012726.

[55] Carlson, M. et al. "Gene Connectivity, Function, and Sequence Conservation: Predictions from Modular Yeast Co-expression Networks". In: *BMC genomics* 7 (Feb. 2006), p. 40. DOI: 10.1186/1471-2164-7-40.

[56] Caroppo, C., Turicchia, S., and Margheri, M. "Phytoplankton assemblages in coastal waters of the northern Ionian Sea (eastern Mediterranean), with special reference to cyanobacteria". In: *Journal of the Marine Biological Association of the United Kingdom* 86 (Oct. 2006), pp. 927–937. DOI: 10.1017/S0025315406013889.

[57] Caroppo, C. et al. "Phytoplankton and Bacterial Communities' Patterns in a Highly Dynamic Ecosystem (Central Mediterranean Sea)". In: *Water* 14.13 (2022). ISSN: 2073-4441. DOI: 10.3390/w14132057. URL: https://www.mdpi.com/2073-4441/14/13/2057.

[58] Carstensen, J., Klais-Peets, R., and Cloern, J. "Phytoplankton Blooms in Estuarine and Coastal Waters: Seasonal Patterns and Key Species". In: *Estuarine, Coastal and Shelf Science* 162 (May 2015). DOI: 10.1016/j.ecss.2015.05.005.

[59] Casabianca, S. et al. "A phytoplankton time series in the Northwestern Adriatic Sea: Structure and dynamics of the assemblages in a coastal ecosystem". In: *Estuarine, Coastal and Shelf Science* 278 (2022), p. 108109. ISSN: 0272-7714. DOI: https://doi.org/10.1016/j.ecss.2022.108109. URL: https://www.sciencedirect.com/science/article/pii/S0272771422003675.

[60] Casotti, R. et al. "Composition and dynamics of the phytoplankton of the Ionian Sea (eastern Mediterranean)". In: *Journal OF GEOPHYSICAL RESEARCH -OCEANS* 108 (Sept. 2003). DOI: 10.1029/2002JC001541.

[61] Cescon, B. et al. "Processes affecting upwelling and primary productivity of the Messina Straits". In: *Bollettino di Geofisica Teorica ed Applicata* 38 (Jan. 1997), pp. 137–147.

[62] Chacón, J. and Rastrojo, A. "Minimum adjusted Rand index for two clusterings of a given size". In: *Advances in Data Analysis and Classification* 17 (Feb. 2022). DOI: 10.1007/s11634-022-00491-w.

[63] Chao, A. et al. “Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies”. In: *Ecological Monographs* 84 (2014), pp. 45–67.

[64] Chen, C. et al. “A search engine to identify pathway genes from expression data on multiple organisms”. In: *BMC systems biology* 1 (Feb. 2007), p. 20. DOI: 10.1186/1752-0509-1-20.

[65] Chen, Y. et al. “edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets”. In: *Nucleic Acids Research* 53.2 (Jan. 2025), gkaf018. ISSN: 1362-4962. DOI: 10.1093/nar/gkaf018. URL: https://doi.org/10.1093/nar/gkaf018.

[66] Chepurnov, V. et al. “In search of new tractable diatoms for experimental biology”. In: *BioEssays : news and reviews in molecular, cellular and developmental biology* 30 (July 2008), pp. 692–702. DOI: 10.1002/bies.20773.

[67] Chepurnov, V. A. et al. “Experimental Studies on Sexual Reproduction in Diatoms”. In: vol. 237. International Review of Cytology. Academic Press, 2004, pp. 91–154. DOI: https://doi.org/10.1016/S0074-7696(04)37003-8. URL: https://www.sciencedirect.com/science/article/pii/S0074769604370038.

[68] Chepurnov, V. A. et al. “Experimental Studies on Sexual Reproduction in Diatoms”. In: *International Review of Cytology*. Vol. 237. Elsevier, 2004, pp. 91–154. ISBN: 978-0-12-364641-5. DOI: 10.1016/S0074-7696(04)37003-8. URL: https://linkinghub.elsevier.com/retrieve/pii/S0074769604370038.

[69] Cianelli, D. et al. “Dynamics of a Very Special Mediterranean Coastal Area: The Gulf of Naples”. In: *Mediterranean Ecosystems: Dynamics, Management & Conservation*. NOVA SCIENCE PUBLISHERS, 2012.

[70] Civitarese, G. et al. “On the impact of the Bimodal Oscillating System (BiOS) on the biogeochemistry and biology of the Adriatic and Ionian Seas (Eastern Mediterranean)”. In: *Biogeosciences* 7.12 (Dec. 15, 2010), pp. 3987–3997. ISSN: 1726-4189. DOI: 10.5194/bg-7-3987-2010. URL: https://bg.copernicus.org/articles/7/3987/2010/.

[71] Claflin, N. et al. “Impact of Pulse Disturbances on Phytoplankton: How Four Storms of Varying Magnitude, Duration, and Timing Altered Community Responses”. In: *Environments* 11.10 (Oct. 4, 2024), p. 218. ISSN: 2076-3298. DOI: 10.3390/environments11100218. URL: https://www.mdpi.com/2076-3298/11/10/218.

[72] Cloern, J. E., Foster, S. Q., and Kleckner, A. E. "Phytoplankton primary production in the world's estuarine-coastal ecosystems". In: *Biogeosciences* 11.9 (2014), pp. 2477–2501. DOI: 10.5194/bg-11-2477-2014. URL: https://bg.copernicus.org/articles/11/2477/2014/.

[73] Coles, V. et al. "Ocean biogeochemistry modeled with emergent trait-based genomics". In: *Science* 358 (Dec. 2017), pp. 1149–1154. DOI: 10.1126/science.aan5712.

[74] *Copernicus Marine Servive*. Accessed online. URL: https://marine.copernicus.eu/.

[75] Cullen, J. J. "The Deep Chlorophyll Maximum: Comparing Vertical Profiles of Chlorophyll *a*". In: *Canadian Journal of Fisheries and Aquatic Sciences* 39.5 (May 1, 1982), pp. 791–803. ISSN: 0706-652X, 1205-7533. DOI: 10.1139/f82-108. URL: http://www.nrcresearchpress.com/doi/10.1139/f82-108.

[76] D'Ortenzio, F. and Ribera d'Alcalà, M. "On the trophic regimes of the Mediterranean Sea: a satellite analysis". In: *Biogeosciences* 6.2 (Feb. 5, 2009), pp. 139–148. ISSN: 1726-4189. DOI: 10.5194/bg-6-139-2009. URL: https://bg.copernicus.org/articles/6/139/2009/.

[77] De Cáceres, M. and Legendre, P. "Associations between species and groups of sites: indices and statistical inference". In: *Ecology* 90 (2009), pp. 3566–3574. DOI: 10.1890/08-1823.1.

[78] De Cáceres, M., Legendre, P., and Moretti, M. "Improving indicator species analysis by combining groups of sites". In: *Oikos* 119.10 (2010), pp. 1674–1684. DOI: https://doi.org/10.1111/j.1600-0706.2010.18334.x. eprint: https://nsojournals.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1600-0706.2010.18334.x. URL: https://nsojournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0706.2010.18334.x.

[79] De La Vara, A. et al. "Climate change signal in the ocean circulation of the Tyrrhenian Sea". In: *Earth System Dynamics* 13.1 (Feb. 2, 2022), pp. 303–319. ISSN: 2190-4987. DOI: 10.5194/esd-13-303-2022. URL: https://esd.copernicus.org/articles/13/303/2022/.

[80] De Martino, A. et al. "De Martino A, Meichenin A, Shi J, Pan KH, Bowler C.. Genetic and phenotypic characterization of Phaeodactylum tricornutum (Bacillariophyceae) accessions. J Phycol 43: 992-1009". In: *Journal of Phycology* 43 (Sept. 2007), pp. 992–1009. DOI: 10.1111/j.1529-8817.2007.00384.x.

[81] De Ruggiero, P. et al. "Modelling the marine circulation of the Campania coastal system (Tyrrhenian Sea) for the year 2016: Analysis of the dynamics". In: *Journal of Marine Systems* 210 (Oct. 2020), p. 103388. ISSN: 09247963. DOI: 10.1016/j.jmarsys.2020.103388. URL: https://linkinghub.elsevier.com/retrieve/pii/S0924796320300841.

[82] De Vargas, C. et al. "Eukaryotic plankton diversity in the sunlit ocean". In: *Science* 348 (May 22, 2015), p. 1261605. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1261605. URL: https://www.science.org/doi/10.1126/science.1261605.

[83] Decembrini, F., Caroppo, C., and Bergamasco, A. "Influence of lateral advection on phytoplankton size-structure and composition in a Mediterranean coastal area". In: *Continental Shelf Research* 209 (Dec. 2020), p. 104216. ISSN: 02784343. DOI: 10.1016/j.csr.2020.104216. URL: https://linkinghub.elsevier.com/retrieve/pii/S0278434320301710.

[84] Decembrini, F. et al. "Linking Microbial Functioning and Trophic Pathways to Ecological Status in a Coastal Mediterranean Ecosystem". In: *Water* 13.9 (2021). ISSN: 2073-4441. DOI: 10.3390/w13091325. URL: https://www.mdpi.com/2073-4441/13/9/1325.

[85] Denman, K. L. and Gargett, A. E. "Biological-Physical Interactions in the Upper Ocean: The Role of Vertical and Small Scale Transport Processes". In: *Annual Review of Fluid Mechanics* 27.1 (Jan. 1995), pp. 225–256. ISSN: 0066-4189, 1545-4479. DOI: 10.1146/annurev.fl.27.010195.001301. URL: https://www.annualreviews.org/doi/10.1146/annurev.fl.27.010195.001301.

[86] Deshpande, D. et al. "RNA-seq data science: From raw data to effective interpretation". In: *Frontiers in Genetics* Volume 14 - 2023 (2023). ISSN: 1664-8021. DOI: 10.3389/fgene.2023.997383. URL: https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2023.997383.

[87] Devlin, M. et al. "Extending the phytoplankton tool kit for the UK Water Framework Directive: indicators of phytoplankton community structure". In: *Hydrobiologia* 633.1 (Oct. 2009), pp. 151–168. ISSN: 0018-8158, 1573-5117. DOI: 10.1007/s10750-009-9879-5. URL: http://link.springer.com/10.1007/s10750-009-9879-5.

[88] Diciotti, R. et al. *Seasonal dynamic and distribution of the invasive species Mnemiopsis leidyi Agassiz (Ctenophora, Bolinopsidae) in three lagoons of the Gulf of Oristano (western Sardinia, Mediterranean Sea)*. Sept. 2017. DOI: 10.13140/RG.2.2.27327.00168.

[89] Directive, S. F. "Directive 2008/56/EC of the European Parliament and of the Council". In: *Journal). Council Decision of* (2008).

[90] Dobin, A. et al. "STAR: ultrafast universal RNA-seq aligner". In: *Bioinformatics* 29.1 (Oct. 2012), pp. 15–21. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bts635. eprint: https://academic.oup.com/bioinformatics/article-pdf/29/1/15/49060422/bioinformatics_29_1_15.pdf. URL: https://doi.org/10.1093/bioinformatics/bts635.

[91] Dorrell, R. G. et al. "Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome". In: *eLife* 6 (May 2017). Ed. by Bhattacharya, D., e23717. ISSN: 2050-084X. DOI: 10.7554/eLife.23717. URL: https://doi.org/10.7554/eLife.23717.

[92] Dougan, K. et al. "Multi-omics analysis reveals the molecular response to heat stress in a "red tide" dinoflagellate". In: *Genome Biology* 24 (Nov. 2023). DOI: 10.1186/s13059-023-03107-4.

[93] Ducklow, H., Steinberg, D., and Buesseler, K. "Upper Ocean Carbon Export and the Biological Pump". In: *Oceanography* 14.4 (2001), pp. 50–58. ISSN: 10428275. DOI: 10.5670/oceanog.2001.06. URL: https://tos.org/oceanography/article/upper-ocean-carbon-export-and-the-biological-pump.

[94] Dufrene, M. and Legendre, P. "Species Assemblages and Indicator Species: The Need for a Flexible Asymmetrical Approach". In: *Ecological monographs* 67 (Aug. 1997), pp. 345–366. DOI: 10.2307/2963459.

[95] Emms, D. M. and Kelly, S. "OrthoFinder: phylogenetic orthology inference for comparative genomics". In: *Genome Biology* 20.1 (Nov. 14, 2019), p. 238. ISSN: 1474-760X. DOI: 10.1186/s13059-019-1832-y. URL: https://doi.org/10.1186/s13059-019-1832-y.

[96] European Environment Agency. *MSFD Art. 17 Reporting (2018)*. Accessed online. URL: https://cdr.eionet.europa.eu/it/eu/msfd_art17/2018reporting/textreport/envxbdazg/.

[97] Falciatore, A. et al. "Diatom Molecular Research Comes of Age: Model Species for Studying Phytoplankton Biology and Diversity". In: *The Plant Cell* 32 (Dec. 2019), tpc.00158.2019. DOI: 10.1105/tpc.19.00158.

[98] Falkowski, P. G., Barber, R. T., and Smetacek, V. "Biogeochemical Controls and Feedbacks on Ocean Primary Production". In: *Science* 281.5374 (July 10, 1998), pp. 200–206. DOI: 10.1126/science.281.5374.200. URL: https://www.science.org/doi/10.1126/science.281.5374.200.

[99] Falkowski, P. G. and Raven, J. A. *Aquatic photosynthesis*. Second edition. Princeton: Princeton University Press, 2010. 1 p. ISBN: 978-0-691-11551-1 978-0-691-11550-4 978-1-4008-4972-7.

[100] Fanelli, E. et al. "Spatial changes in community composition and food web structure of mesozooplankton across the Adriatic basin (Mediterranean Sea)". In: *Biogeosciences* 19.6 (Apr. 1, 2022), pp. 1833–1851. ISSN: 1726-4189. DOI: 10.5194/bg-19-1833-2022. URL: https://bg.copernicus.org/articles/19/1833/2022/.

[101] Faust, K. and Raes, J. "Microbial interactions: from networks to models". In: *Nature Reviews Microbiology* 10.8 (Aug. 2012), pp. 538–550. ISSN: 1740-1526, 1740-1534. DOI: 10.1038/nrmicro2832. URL: https://www.nature.com/articles/nrmicro2832.

[102] Fennel, K. and Boss, E. "Subsurface maxima of phytoplankton and chlorophyll: Steady-state solutions from a simple model". In: *Limnology and Oceanography* 48.4 (July 2003), pp. 1521–1534. ISSN: 0024-3590, 1939-5590. DOI: 10.4319/lo.2003.48.4.1521. URL: https://aslopubs.onlinelibrary.wiley.com/doi/10.4319/lo.2003.48.4.1521.

[103] Ferrante, M. I., Broccoli, A., and Montresor, M. "The pennate diatom Pseudo-nitzschia multistriata as a model for diatom life cycles, from the laboratory to the sea". In: *Journal of Phycology* 59.4 (2023), pp. 637–643. DOI: https://doi.org/10.1111/jpy.13342. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jpy.13342.

[104] Fiehn, O. "Metabolomics – the link between genotypes and phenotypes". In: *Plant Molecular Biology* 48.1 (Jan. 2002), pp. 155–171. ISSN: 0167-4412, 1573-5028. DOI: 10.1023/A:1013713905833. URL: https://link.springer.com/10.1023/A:1013713905833.

[105] Fiorini, F. et al. "A Metabolomics Exploration of the Sexual Phase in the Marine Diatom Pseudo-nitzschia multistriata". In: *Marine Drugs* 18.6 (2020). ISSN: 1660-3397. DOI: 10.3390/md18060313. URL: https://www.mdpi.com/1660-3397/18/6/313.

[106] Flynn, R. F. et al. "Nanoplankton: The dominant vector for carbon export across the Atlantic Southern Ocean in spring". In: *Science Advances* 9.48 (Dec. 2023), eadi3059. ISSN: 2375-2548. DOI: 10.1126/sciadv.adi3059. URL: https://www.science.org/doi/10.1126/sciadv.adi3059.

[107] Follows, M. J. and Dutkiewicz, S. "Modeling Diverse Communities of Marine Microbes". In: *Annual Review of Marine Science* 3.1 (Jan. 15, 2011), pp. 427–451. ISSN: 1941-1405, 1941-0611. DOI: 10.1146/annurev-marine-120709-142848.

URL: https://www.annualreviews.org/doi/10.1146/annurev-marine-120709-142848.

[108] Forabosco, P. et al. "Insights into TREM2 biology by network analysis of human brain gene expression data". In: *Neurobiology of Aging* 34.12 (2013), pp. 2699–2714. ISSN: 0197-4580. DOI: https://doi.org/10.1016/j.neurobiolaging.2013.05.001. URL: https://www.sciencedirect.com/science/article/pii/S0197458013001991.

[109] Fuller, T. et al. "Weighted Gene Co-expression Network Analysis Strategies Applied to Mouse Weight". In: *Mammalian genome : official journal of the International Mammalian Genome Society* 18 (Aug. 2007), pp. 463–72. DOI: 10.1007/s00335-007-9043-3.

[110] Furnas, M. J. "DIEL SYNCHRONIZATION OF SPERM FORMATION IN THE DIATOM CHAETOCEROS CURVISETUM CLEVE". In: *Journal of Phycology* 21.4 (1985), pp. 667–671. DOI: https://doi.org/10.1111/j.0022-3646.1985.00667.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0022-3646.1985.00667.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0022-3646.1985.00667.x.

[111] Gangrade, S. and Mangolte, I. "Patchiness of plankton communities at fronts explained by Lagrangian history of upwelled water parcels". In: *Limnology and Oceanography* 69.9 (Sept. 2024), pp. 2123–2137. ISSN: 0024-3590, 1939-5590. DOI: 10.1002/lno.12654. URL: https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.12654.

[112] Gianuca, A. T. et al. "Effects of dispersal and environmental heterogeneity on the replacement and nestedness components of β-diversity". In: *Ecology* 98.2 (2017), pp. 525–533. DOI: https://doi.org/10.1002/ecy.1666. URL: https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/ecy.1666.

[113] Gillard, J. et al. "Metabolomics Enables the Structure Elucidation of a Diatom Sex Pheromone". In: *Angewandte Chemie International Edition* 52.3 (2013), pp. 854–857. DOI: https://doi.org/10.1002/anie.201208175. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.201208175. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201208175.

[114] "Globally Consistent Quantitative Observations of Planktonic Ecosystems". In: *Frontiers in Marine Science* 6 (Apr. 25, 2019), p. 196. ISSN: 2296-7745. DOI: 10.3389/fmars.2019.00196. URL: https://www.frontiersin.org/article/10.3389/fmars.2019.00196/full.

[115] Godhe, A., Kremp, A., and Montresor, M. "Genetic and Microscopic Evidence for Sexual Reproduction in the Centric Diatom Skeletonema marinoi". In: *Protist* 165.4 (2014), pp. 401–416. ISSN: 1434-4610. DOI: https://doi.org/10.1016/j.protis.2014.04.006. URL: https://www.sciencedirect.com/science/article/pii/S1434461014000479.

[116] Gower, J. F. R., Denman, K. L., and Holyer, R. J. "Phytoplankton patchiness indicates the fluctuation spectrum of mesoscale oceanic structure". In: *Nature* 288.5787 (Nov. 1980), pp. 157–159. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/288157a0. URL: https://www.nature.com/articles/288157a0.

[117] Grilli, F. et al. "Circulation and horizontal fluxes in the northern Adriatic Sea in the period June 1999–July 2002. Part II: Nutrients transport". In: *Science of The Total Environment* 353.1 (2005). Mucilages in the Adriatic and Tyrrhenian Seas, pp. 115–125. ISSN: 0048-9697. DOI: https://doi.org/10.1016/j.scitotenv.2005.09.011. URL: https://www.sciencedirect.com/science/article/pii/S0048969705006285.

[118] Guadayol, Ò. et al. "Episodic meteorological and nutrient-load events as drivers of coastal planktonic ecosystem dynamics: A time-series analysis". In: *Marine Ecology-progress Series - MAR ECOL-PROGR SER* 381 (Apr. 2009), pp. 139–155. DOI: 10.3354/meps07939.

[119] Guidi, L. et al. "Plankton networks driving carbon export in the oligotrophic ocean". In: *Nature* 532.7600 (Apr. 2016), pp. 465–470. ISSN: 1476-4687. DOI: 10.1038/nature16942. URL: https://www.nature.com/articles/nature16942.

[120] Guillot, G. et al. "Editorial: Advances in Statistical Ecology: New Methods and Software". In: *Frontiers in Ecology and Evolution* 9 (Jan. 17, 2022), p. 828919. ISSN: 2296-701X. DOI: 10.3389/fevo.2021.828919. URL: https://www.frontiersin.org/articles/10.3389/fevo.2021.828919/full.

[121] Guiry, M. D. "How many species of algae are there? A reprise. Four kingdoms, 14 phyla, 63 classes and still growing". In: *Journal of Phycology* 60.2 (Apr. 2024), pp. 214–228. ISSN: 0022-3646, 1529-8817. DOI: 10.1111/jpy.13431. URL: https://onlinelibrary.wiley.com/doi/10.1111/jpy.13431.

[122] Hastings, A. et al. "Chaos in Ecology: Is Mother Nature a Strange Attractor?" In: *Annual Review of Ecology and Systematics* 24.1 (Nov. 1993), pp. 1–33. ISSN: 0066-4162. DOI: 10.1146/annurev.es.24.110193.000245. URL: https://www.annualreviews.org/doi/10.1146/annurev.es.24.110193.000245.

[123] Hildebrand, M. et al. "Nanoscale control of silica morphology and three-dimensional structure during diatom cell wall formation". In: *Journal of Materials Research* 21 (Oct. 2006), pp. 2689–2698. DOI: 10.1557/jmr.2006.0333.

[124] Hillebrand, H. et al. "BIOVOLUME CALCULATION FOR PELAGIC AND BENTHIC MICROALGAE". In: *Journal of Phycology* 35.2 (Apr. 1999), pp. 403–424. ISSN: 0022-3646, 1529-8817. DOI: 10.1046/j.1529-8817.1999.3520403.x. URL: https://onlinelibrary.wiley.com/doi/10.1046/j.1529-8817.1999.3520403.x.

[125] Hingsamer, P., Peeters, F., and Hofmann, H. "The Consequences of Internal Waves for Phytoplankton Focusing on the Distribution and Production of Planktothrix rubescens". In: *PLoS ONE* 9.8 (Aug. 7, 2014). Ed. by Ban, S., e104359. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0104359. URL: https://dx.plos.org/10.1371/journal.pone.0104359.

[126] Hinnebusch, A. et al. "An evaluation of the phylogenetic position of the dinoflagellate Crypthecodinium cohnii based on 5S rRNA characterization". In: *Journal of molecular evolution* 17 (Feb. 1981), pp. 334–7. DOI: 10.1007/BF01734355.

[127] Ho, T.-Y. et al. "The elemental composition of some phytoplankton". In: *Journal of Phycology* 39 (Nov. 2003), pp. 1145–1159. DOI: 10.1111/j.0022-3646.2003.03-090.x.

[128] Horvath, S. *Weighted Network Analysis: Applications in Genomics and Systems Biology*. Springer New York, 2011. ISBN: 9781441988201.

[129] Hsieh, T. C., Ma, K. H., and Chao, A. *iNEXT: Interpolation and Extrapolation for Species Diversity*. R package version 3.0.1. 2024. URL: http://chao.stat.nthu.edu.tw/wordpress/software_download/.

[130] Huppert, A., Blasius, B., and Stone, L. "A Model of Phytoplankton Blooms". In: *The American Naturalist* 159.2 (Feb. 2002), pp. 156–171. ISSN: 0003-0147, 1537-5323. DOI: 10.1086/324789. URL: https://www.journals.uchicago.edu/doi/10.1086/324789.

[131] Hutchins, D. and Bruland, K. "Iron-limited diatom growth and Si:N uptake ratios in a coastal upwelling regime". In: *Nature* 393 (June 1998), pp. 561–564. DOI: 10.1038/31203.

[132] Hutchins, D. et al. "Phytoplankton iron limitation in the Humboldt Current and Peru Upwelling". In: *Limnology and Oceanography* 47 (July 2002), pp. 997–1011.

[133] Hutchinson, G. E. "The Paradox of the Plankton". In: *The American Naturalist* 95.882 (May 1961), pp. 137–145. ISSN: 0003-0147, 1537-5323. DOI: 10.1086/282171. URL: https://www.journals.uchicago.edu/doi/10.1086/282171.

[134] Iacono, R. et al. "The Tyrrhenian Sea Circulation: A Review of Recent Work". In: *Sustainability* 13.11 (June 3, 2021), p. 6371. ISSN: 2071-1050. DOI: 10.3390/su13116371. URL: https://www.mdpi.com/2071-1050/13/11/6371.

[135] Ibarbalz, F. M. et al. "Global Trends in Marine Plankton Diversity across Kingdoms of Life". In: *Cell* 179.5 (2019), 1084–1097.e21. ISSN: 0092-8674. DOI: https://doi.org/10.1016/j.cell.2019.10.008. URL: https://www.sciencedirect.com/science/article/pii/S0092867419311249.

[136] ISPRA. *ISPRA Marine Strategy Data Consultation Portal*. Accessed online. URL: http://www.db-strategiamarina.isprambiente.it/app/#/data_consultation.

[137] ISPRA. *ISPRA official website*. Accessed online. URL: https://www.snpambiente.it/.

[138] Jeong, H. et al. "The Large-Scale Organization of Metabolic Networks". In: *Nature* 407 (Nov. 2000), pp. 651–4. DOI: 10.1038/35036627.

[139] Jiménez-Quesada, M. J., Traverso, J. Á., and Alché, J. d. D. "NADPH Oxidase-Dependent Superoxide Production in Plant Reproductive Tissues". In: *Frontiers in Plant Science* Volume 7 - 2016 (2016). ISSN: 1664-462X. DOI: 10.3389/fpls.2016.00359. URL: https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2016.00359.

[140] Johansson, O. et al. "Skeletonema marinoi as a new genetic model for marine chain-forming diatoms". In: *Scientific Reports* 9 (Apr. 2019). DOI: 10.1038/s41598-019-41085-5.

[141] Johnson, W., Li, C., and Rabinovic, A. "Adjusting batch effects in microarray expression data using empirical Bayes methods". In: *Biostatistics (Oxford, England)* 8 (Jan. 2007), pp. 118–27. DOI: 10.1093/biostatistics/kxj037.

[142] Kanehisa, M. and Goto, S. "KEGG: Kyoto Encyclopedia of Genes and Genomes". In: *Nucleic Acids Research* 28.1 (Jan. 2000), pp. 27–30. ISSN: 0305-1048. DOI: 10.1093/nar/28.1.27. eprint: https://academic.oup.com/nar/article-pdf/28/1/27/9895154/280027.pdf. URL: https://doi.org/10.1093/nar/28.1.27.

[143] Kanehisa, M. et al. "KEGG for taxonomy-based analysis of pathways and genomes". In: *Nucleic Acids Research* 51.D1 (Oct. 2022), pp. D587–D592. ISSN: 0305-1048. DOI: 10.1093/nar/gkac963. eprint: https://academic.oup.com/nar/article-pdf/51/D1/D587/48440889/gkac963.pdf. URL: https://doi.org/10.1093/nar/gkac963.

[144] Karl, D. M. et al. "Seasonal-to-decadal scale variability in primary production and particulate matter export at Station ALOHA". In: *Progress in Oceanography* 195 (July 2021), p. 102563. ISSN: 00796611. DOI: 10.1016/j.pocean.2021.102563. URL: https://linkinghub.elsevier.com/retrieve/pii/S0079661121000501.

[145] Karlusich, J. J. P., Ibarbalz, F. M., and Bowler, C. "Exploration of marine phytoplankton: from their historical appreciation to the omics era". In: *Journal of Plankton Research* (Oct. 30, 2020). Ed. by Lisa, C., fbaa049. ISSN: 0142-7873, 1464-3774. DOI: 10.1093/plankt/fbaa049. URL: https://academic.oup.com/plankt/advance-article/doi/10.1093/plankt/fbaa049/5943115.

[146] Keeling, P. J. et al. "The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing". In: *PLOS Biology* 12.6 (June 2014), pp. 1–6. DOI: 10.1371/journal.pbio.1001889. URL: https://doi.org/10.1371/journal.pbio.1001889.

[147] Klapper, F. et al. "Pheromone Mediated Sexual Reproduction of Pennate Diatom Cylindrotheca closterium". In: *Journal of Chemical Ecology* 47 (June 2021). DOI: 10.1007/s10886-021-01277-8.

[148] Kooistra, W. et al. "The Origin and Evolution of the Diatoms: Their Adaptation to a Planktonic Existence". In: *Evolution of Primary Producers in the Sea* (Dec. 2007), pp. 207–249. DOI: 10.1016/B978-012370518-1/50012-6.

[149] Korpelainen, E. et al. *RNA-seq Data Analysis: A Practical Approach*. Sept. 2014. ISBN: 9780429169205. DOI: 10.1201/b17457.

[150] Kress, N. et al. "Continuing influence of the changed thermohaline circulation in the eastern Mediterranean on the distribution of dissolved oxygen and nutrients: Physical and chemical characterization of the water masses". In: *Journal of Geophysical Research: Oceans* 108 (C9 Sept. 2003), 2002JC001397. ISSN: 0148-0227. DOI: 10.1029/2002JC001397. URL: https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2002JC001397.

[151] Kuhn and Max. "Building Predictive Models in R Using the caret Package". In: *Journal of Statistical Software* 28.5 (2008), pp. 1–26. DOI: 10.18637/jss.v028.i05. URL: https://www.jstatsoft.org/index.php/jss/article/view/v028i05.

[152] Lafuente, J. G. et al. "Hydrographic phenomena influencing early life stages of the Sicilian Channel anchovy". In: *Fisheries Oceanography* 11.1 (2002), pp. 31–44. DOI: https://doi.org/10.1046/j.1365-2419.2002.00186.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1365-2419.2002.00186.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2419.2002.00186.x.

[153] Lalli, C. M. and Parsons, T. R. *Biological oceanography: an introduction*. 2nd ed. Open University oceanography series. Oxford [England]: Butterworth Heinemann, 1997. ISBN: 978-0-7506-3384-0.

[154] Langfelder, P. and Horvath, S. "WGCNA: an R package for weighted correlation network analysis". In: *BMC Bioinformatics* (2008), 9:559. DOI: 10.1186/1471-2105-9-559.

[155] Langfelder, P., Zhang, B., and Horvath, S. "Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R". In: *Bioinformatics* 24.5 (Nov. 2007), pp. 719–720. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btm563. eprint: https://academic.oup.com/bioinformatics/article-pdf/24/5/719/49051558/bioinformatics_24_5_719.pdf. URL: https://doi.org/10.1093/bioinformatics/btm563.

[156] Lara-Ortíz, T., Riveros-Rosas, H., and Aguirre, J. "Reactive oxygen species generated by microbial NADPH oxidase NoxA regulate sexual development in Aspergillus nidulans". In: *Molecular Microbiology* 50.4 (2003), pp. 1241–1255. DOI: https://doi.org/10.1046/j.1365-2958.2003.03800.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1365-2958.2003.03800.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2958.2003.03800.x.

[157] Leek, J. et al. "Tackling the widespread and critical impact of batch effects in high-throughput data". In: *Nature reviews. Genetics* 11 (Oct. 2010), pp. 733–9. DOI: 10.1038/nrg2825.

[158] Leek, J. T. et al. *sva: Surrogate Variable Analysis*. R package version 3.54.0. 2024. DOI: 10.18129/B9.bioc.sva. URL: https://bioconductor.org/packages/sva.

[159] Legendre, L. and Legendre, P. *Numerical ecology*. OCLC: 873894148. Amsterdam: Elsevier, 2012. ISBN: 978-0-444-53868-0.

[160] Legendre, P. and Anderson, M. J. "DISTANCE-BASED REDUNDANCY ANALYSIS: TESTING MULTISPECIES RESPONSES IN MULTIFACTORIAL ECOLOGICAL EXPERIMENTS". In: *Ecological Monographs* 69.1 (1999), pp. 1–24. DOI: https://doi.org/10.1890/0012-9615(1999)069[0001:DBRATM]2.0.CO;2.

[161] Legendre, P. and Gallagher, E. "Ecologically meaningful transformations for ordination of species data". In: *Oecologia* 129 (Oct. 2001), pp. 271–280. DOI: 10.1007/s004420100716.

[162] Lehahn, Y. et al. "Dispersion/dilution enhances phytoplankton blooms in low-nutrient waters". In: *Nature Communications* 8.1 (Mar. 31, 2017), p. 14868. ISSN: 2041-1723. DOI: 10.1038/ncomms14868. URL: https://www.nature.com/articles/ncomms14868.

[163] Levine, N. M. et al. "Microbial Ecology to Ocean Carbon Cycling: From Genomes to Numerical Models". In: *Annual Review of Earth and Planetary Sciences* 53.Volume 53, 2025 (2025), pp. 595–624. ISSN: 1545-4495. DOI: https://doi.org/10.1146/annurev-earth-040523-020630. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-earth-040523-020630.

[164] Lévy, M., Franks, P. J. S., and Smith, K. S. "The role of submesoscale currents in structuring marine ecosystems". In: *Nature Communications* 9.1 (Nov. 12, 2018), p. 4758. ISSN: 2041-1723. DOI: 10.1038/s41467-018-07059-3. URL: https://www.nature.com/articles/s41467-018-07059-3.

[165] Liao, Y., Smyth, G., and Shi, W. "FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features". In: *Bioinformatics (Oxford, England)* 30 (Nov. 2013). DOI: 10.1093/bioinformatics/btt656.

[166] Liao, Y., Smyth, G. K., and Shi, W. "The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads". In: *Nucleic Acids Research* 47.8 (Feb. 2019), e47–e47. ISSN: 0305-1048. DOI: 10.1093/nar/gkz114. eprint: https://academic.oup.com/nar/article-pdf/47/8/e47/28534862/gkz114.pdf. URL: https://doi.org/10.1093/nar/gkz114.

[167] Lima-Mendez, G. et al. "Determinants of community structure in the global plankton interactome". In: *Science* 348.6237 (2015), p. 1262073. DOI: 10.1126/science.1262073. URL: https://www.science.org/doi/abs/10.1126/science.1262073.

[168] Litchman, E. et al. "The role of functional traits and trade-offs in structuring phytoplankton communities: scaling from cellular to ecosystem level". In: *Ecology Letters* 10.12 (Dec. 2007), pp. 1170–1181. ISSN: 1461-023X, 1461-0248. DOI: 10.1111/j.1461-0248.2007.01117.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1461-0248.2007.01117.x.

[169] Liu, X. et al. "SLC24A-mediated calcium exchange as an indispensable component of the diatom cell density-driven signaling pathway". In: *The ISME Journal* 18.1 (Mar. 2024), wrae039. ISSN: 1751-7362. DOI: 10.1093/ismejo/wrae039. eprint: https://academic.oup.com/ismej/article-pdf/18/1/wrae039/57125927/wrae039.pdf. URL: https://doi.org/10.1093/ismejo/wrae039.

[170] Lomas, M. et al. "Biogeochemical responses to late-winter storms in the Sargasso Sea. IV. Rapid succession of major phytoplankton groups". In: *Deep Sea Research Part I: Oceanographic Research Papers* 56.6 (June 2009), pp. 892–908. ISSN: 09670637. DOI: 10.1016/j.dsr.2009.03.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0967063709000533.

[171] Lommer, M. et al. “Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation”. In: *Genome biology* 13 (July 2012), R66. DOI: 10.1186/gb-2012-13-7-r66.

[172] Longhurst, A. “Seasonal cycles of pelagic production and consumption”. In: *Progress in Oceanography* 36.2 (1995), pp. 77–167. ISSN: 0079-6611. DOI: https://doi.org/10.1016/0079-6611(95)00015-1. URL: https://www.sciencedirect.com/science/article/pii/0079661195000151.

[173] Lopes Dos Santos, A. et al. “Phytoplankton diversity and ecology through the lens of high throughput sequencing technologies”. In: *Advances in Phytoplankton Ecology*. Elsevier, 2022, pp. 353–413. ISBN: 978-0-12-822861-6. DOI: 10.1016/B978-0-12-822861-6.00020-0. URL: https://linkinghub.elsevier.com/retrieve/pii/B9780128228616000200.

[174] Love, M. I., Huber, W., and Anders, S. “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2”. In: *Genome Biology* 15.12 (Dec. 5, 2014), p. 550. ISSN: 1474-760X. DOI: 10.1186/s13059-014-0550-8. URL: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8.

[175] Lovejoy, S. “Universal multifractals and ocean patchiness: phytoplankton, physical fields and coastal heterogeneity”. In: *Journal of Plankton Research* 23.2 (Feb. 1, 2001), pp. 117–141. ISSN: 14643774. DOI: 10.1093/plankt/23.2.117. URL: https://academic.oup.com/plankt/article-lookup/doi/10.1093/plankt/23.2.117.

[176] Lucas, A. J. et al. “The influence of diurnal winds on phytoplankton dynamics in a coastal upwelling system off southwestern Africa”. In: *Deep Sea Research Part II: Topical Studies in Oceanography* 101 (Mar. 2014), pp. 50–62. ISSN: 09670645. DOI: 10.1016/j.dsr2.2013.01.016. URL: https://linkinghub.elsevier.com/retrieve/pii/S0967064513000350.

[177] Lund, J. W. G., Kipling, C., and Le Cren, E. D. “The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting”. In: *Hydrobiologia* 11.2 (Apr. 1958), pp. 143–170. ISSN: 0018-8158, 1573-5117. DOI: 10.1007/BF00007865. URL: http://link.springer.com/10.1007/BF00007865.

[178] Ma, L. et al. “Responses of phytoplankton communities to the effect of internal wave-powered upwelling”. In: *Limnology and Oceanography* 66.4 (Apr. 2021), pp. 1083–1098. ISSN: 0024-3590, 1939-5590. DOI: 10.1002/lno.11666. URL: https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.11666.

[179] Magaldi, M. G. et al. "On the response of a turbulent coastal buoyant current to wind events: the case of the Western Adriatic Current". In: *Ocean Dynamics* 60.1 (Feb. 2010), pp. 93–122. ISSN: 1616-7341, 1616-7228. DOI: 10.1007/s10236-009-0247-9. URL: http://link.springer.com/10.1007/s10236-009-0247-9.

[180] Mahadevan, A. "The Impact of Submesoscale Physics on Primary Productivity of Plankton". In: *Annual Review of Marine Science* 8.1 (Jan. 3, 2016), pp. 161–184. ISSN: 1941-1405, 1941-0611. DOI: 10.1146/annurev-marine-010814-015912. URL: https://www.annualreviews.org/doi/10.1146/annurev-marine-010814-015912.

[181] Malviya, S. et al. "Insights into global diatom distribution and diversity in the world's ocean". In: *Proceedings of the National Academy of Sciences* 113.11 (Mar. 15, 2016). ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1509523113. URL: https://pnas.org/doi/full/10.1073/pnas.1509523113.

[182] Marchant, N. and Steorts, R. *clevr: Clustering and Link Prediction Evaluation in R*. R package version 0.1.2. 2023. URL: https://CRAN.R-project.org/package=clevr.

[183] Marchese, C. et al. "Analysis of Chlorophyll-*a* and Primary Production Dynamics in North Tyrrhenian and Ligurian Coastal–Neritic and Oceanic Waters". In: *Journal of Coastal Research* 313 (May 6, 2015), pp. 690–701. ISSN: 0749-0208, 1551-5036. DOI: 10.2112/JCOASTRES-D-13-00210.1. URL: http://www.bioone.org/doi/10.2112/JCOASTRES-D-13-00210.1.

[184] Margiotta, F. et al. "Do plankton reflect the environmental quality status? The case of a post-industrial Mediterranean Bay". In: *Marine Environmental Research* 160 (Apr. 2020), p. 104980. DOI: 10.1016/j.marenvres.2020.104980.

[185] Marinchel, N. et al. "Structural variability of protist assemblages in surface sediments across Italian Mediterranean marine subregions". In: *Frontiers in Marine Science* Volume 11 - 2024 (2024). ISSN: 2296-7745. DOI: 10.3389/fmars.2024.1427357. URL: https://www.frontiersin.org/journals/marine-science/articles/10.3389/fmars.2024.1427357.

[186] Marini, M. et al. "Seasonal variability and Po River plume influence on biochemical properties along western Adriatic coast". In: *Journal of Geophysical Research: Oceans* 113 (C5 May 2008), 2007JC004370. ISSN: 0148-0227. DOI: 10.1029/2007JC004370. URL: https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2007JC004370.

[187] Marotta, P. et al. “Mate Perception and Gene Networks Regulating the Early Phase of Sex in Pseudo-nitzschia multistriata”. In: *Journal of Marine Science and Engineering* 10.12 (2022). ISSN: 2077-1312. DOI: 10.3390/jmse10121941. URL: https://www.mdpi.com/2077-1312/10/12/1941.

[188] Martellucci, R. et al. “Coastal phytoplankton bloom dynamics in the Tyrrhenian Sea: Advantage of integrating in situ observations, large-scale analysis and forecast systems”. In: *Journal of Marine Systems* 218 (2021), p. 103528. ISSN: 0924-7963. DOI: https://doi.org/10.1016/j.jmarsys.2021.103528. URL: https://www.sciencedirect.com/science/article/pii/S0924796321000269.

[189] Martellucci, R. et al. “Recent changes of the dissolved oxygen distribution in the deep convection cell of the southern Adriatic Sea”. In: *Journal of Marine Systems* 245 (Sept. 2024), p. 103988. ISSN: 09247963. DOI: 10.1016/j.jmarsys.2024.103988. URL: https://linkinghub.elsevier.com/retrieve/pii/S0924796324000265.

[190] Martellucci, R. et al. “On the presence of coastal upwelling along the northeastern Tyrrhenian coast”. In: EGU General Assembly Conference Abstracts. Apr. 1, 2017, p. 13767. URL: https://ui.adsabs.harvard.edu/abs/2017EGUGA..19.13767M.

[191] Martellucci, R. et al. “Physical and Biological Water Column Observations during Summer Sea/Land Breeze Winds in the Coastal Northern Tyrrhenian Sea”. In: *Water* 10.11 (Nov. 16, 2018), p. 1673. ISSN: 2073-4441. DOI: 10.3390/w10111673. URL: https://www.mdpi.com/2073-4441/10/11/1673.

[192] Martin, A. “Phytoplankton patchiness: the role of lateral stirring and mixing”. In: *Progress in Oceanography* 57.2 (May 2003), pp. 125–174. ISSN: 00796611. DOI: 10.1016/S0079-6611(03)00085-5. URL: https://linkinghub.elsevier.com/retrieve/pii/S0079661103000855.

[193] Mason, M. et al. “Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells”. In: *BMC genomics* 10 (Aug. 2009), p. 327. DOI: 10.1186/1471-2164-10-327.

[194] Masotti, I. et al. “The Influence of River Discharge on Nutrient Export and Phytoplankton Biomass Off the Central Chile Coast (33°–37°S): Seasonal Cycle and Interannual Variability”. In: *Frontiers in Marine Science* Volume 5 - 2018 (2018). ISSN: 2296-7745. DOI: 10.3389/fmars.2018.00423. URL: https://www.frontiersin.org/journals/marine-science/articles/10.3389/fmars.2018.00423.

[195] Mazzocchi, M. et al. “Regional and seasonal characteristics of epipelagic mesozooplankton in the Mediterranean Sea based on an artificial neural network analysis”. In: *Journal of Marine Systems* 135 (2014). Assessing and modelling ecosystem changes in the Mediterranean and the Black Sea pelagic ecosystem - SESAME, pp. 64–80. ISSN: 0924-7963. DOI: https://doi.org/10.1016/j.jmarsys.2013.04.009. URL: https://www.sciencedirect.com/science/article/pii/S0924796313000997.

[196] McGillicuddy, D. J. “Mechanisms of Physical-Biological-Biogeochemical Interaction at the Oceanic Mesoscale”. In: *Annual Review of Marine Science* 8.1 (Jan. 3, 2016), pp. 125–159. ISSN: 1941-1405, 1941-0611. DOI: 10.1146/annurev-marine-010814-015606. URL: https://www.annualreviews.org/doi/10.1146/annurev-marine-010814-015606.

[197] Meilă, M. “Comparing clusterings—an information based distance”. In: *Journal of Multivariate Analysis* 98.5 (2007), pp. 873–895. ISSN: 0047-259X. DOI: https://doi.org/10.1016/j.jmva.2006.11.013. URL: https://www.sciencedirect.com/science/article/pii/S0047259X06002016.

[198] Menden-Deuer, S. and Lessard, E. J. “Carbon to volume relationships for dinoflagellates, diatoms, and other protist plankton”. In: *Limnology and Oceanography* 45.3 (May 2000), pp. 569–579. ISSN: 0024-3590, 1939-5590. DOI: 10.4319/lo.2000.45.3.0569. URL: https://aslopubs.onlinelibrary.wiley.com/doi/10.4319/lo.2000.45.3.0569.

[199] Menkes, C. E. et al. “Global impact of tropical cyclones on primary production”. In: *Global Biogeochemical Cycles* 30.5 (May 2016), pp. 767–786. ISSN: 0886-6236, 1944-9224. DOI: 10.1002/2015GB005214. URL: https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2015GB005214.

[200] Mesman, J. P. et al. “Drivers of phytoplankton responses to summer wind events in a stratified lake: A modeling study”. In: *Limnology and Oceanography* 67.4 (Apr. 2022), pp. 856–873. ISSN: 0024-3590, 1939-5590. DOI: 10.1002/lno.12040. URL: https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.12040.

[201] Meyer, B., Wulf, M., and Håkansson, H. “PHENOTYPIC VARIATION OF LIFE-CYCLE STAGES IN CLONES OF THREE SIMILAR CYCLOTELLA SPECIES AFTER INDUCED AUXOSPORE PRODUCTION”. In: *Diatom Research* 16.2 (2001), pp. 343–361. DOI: 10.1080/0269249X.2001.9705525. eprint: https://doi.org/10.1080/0269249X.2001.9705525. URL: https://doi.org/10.1080/0269249X.2001.9705525.

[202] Millot, C. and Taupier-Letage, I. "Circulation in the Mediterranean Sea". In: *The Mediterranean Sea*. Ed. by Saliot, A. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 29–66. DOI: 10.1007/b107143. URL: https://doi.org/10.1007/b107143.

[203] Ministero dell'Ambiente e della Sicurezza Energetica. *MSFD Protocol D1 (Programmi di Monitoraggio 2021–2026)*. Accessed online. URL: https://www.mase.gov.it/sites/default/files/archivio/allegati/strategia_marina/d1_programmi_monitoraggio_2021-2026_giugno_2020.pdf.

[204] Mock, T. et al. "Evolutionary genomics of the cold-adapted diatom Fragilariopsis cylindrus". In: *Nature* 541.7638 (2017), pp. 536–540.

[205] Moeys, S. et al. "A sex-inducing pheromone triggers cell cycle arrest and mate attraction in the diatom Seminavis robusta". In: *Scientific Reports* 6 (Jan. 2016). DOI: 10.1038/srep19252.

[206] Moon-van der Staay, S. Y., De Wachter, R., and Vaulot, D. "Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity". In: *Nature* 409.6820 (2001), pp. 607–610.

[207] Moreno, C. et al. "Molecular physiology of Antarctic diatom natural assemblages and bloom event reveal insights into strategies contributing to their ecological success". In: *mSystems* 9 (Feb. 2024), e0130623. DOI: 10.1128/msystems.01306-23.

[208] Neri, F. et al. "Phytoplankton communities in a coastal and offshore stations of the northern Adriatic Sea approached by network analysis and different statistical descriptors". In: *Estuarine, Coastal and Shelf Science* 282 (Mar. 2023), p. 108224. ISSN: 02727714. DOI: 10.1016/j.ecss.2023.108224. URL: https://linkinghub.elsevier.com/retrieve/pii/S0272771423000148.

[209] Nicholson, J., Lindon, J., and Holmes, E. "'Metabonomics': Understanding the Metabolic Responses of Living Systems to Pathophysiological Stimuli Via Multivariate Statistical Analysis of Biological Nmr Spectroscopic Data". In: *Xenobiotica; the fate of foreign compounds in biological systems* 29 (Dec. 1999), pp. 1181–9. DOI: 10.1080/004982599238047.

[210] Nightingale, P. et al. "A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization experiment in the equatorial Pacific Ocean". In: *Nature* 383 (Jan. 1996).

[211] Nützmann, H.-W., Scazzocchio, C., and Osbourn, A. "Metabolic Gene Clusters in Eukaryotes". In: *Annual Review of Genetics* 52.Volume 52, 2018 (2018), pp. 159–183. ISSN: 1545-2948. DOI: https://doi.org/10.1146/annurev-genet-120417-031237. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-genet-120417-031237.

[212] Occhipinti, G. et al. "Plankton Communities Behave Chaotically Under Seasonal or Stochastic Temperature Forcings". In: *Ecology and Evolution* 15.8 (Aug. 2025), e71930. ISSN: 2045-7758, 2045-7758. DOI: 10.1002/ece3.71930. URL: https://onlinelibrary.wiley.com/doi/10.1002/ece3.71930.

[213] Oldham, M., Horvath, S., and Geschwind, D. "Conservation and Evolution of Gene Co-expression Networks in Human and Chimpanzee Brains". In: *Proceedings of the National Academy of Sciences of the United States of America* 103 (Dec. 2006), pp. 17973–8. DOI: 10.1073/pnas.0605938103.

[214] Oliver, S. "Proteomics: Guilt-by-association goes global". In: *Nature* 403 (Feb. 2000), pp. 601–603. DOI: 10.1038/35001165.

[215] Olli, K., Nyman, E., and Tamminen, T. "Half-century trends in alpha and beta diversity of phytoplankton summer communities in the Helsinki Archipelago, the Baltic Sea". In: *Journal of Plankton Research* 45.1 (Jan. 1, 2023), pp. 146–162. ISSN: 0142-7873. DOI: 10.1093/plankt/fbac029. URL: https://doi.org/10.1093/plankt/fbac029.

[216] Olli, K., Paerl, H. W., and Klais, R. "Diversity of coastal phytoplankton assemblages – Cross ecosystem comparison". In: *Estuarine, Coastal and Shelf Science*. Special Issue: Global Patterns of Phytoplankton Dynamics in Coastal Ecosystems 162 (Sept. 5, 2015), pp. 110–118. ISSN: 0272-7714. DOI: 10.1016/j.ecss.2015.03.015. URL: https://www.sciencedirect.com/science/article/pii/S0272771415000840.

[217] Olson, R. J., Zettler, E. R., and DuRand, M. D. "Phytoplankton Analysis Using Flow Cytometry". In: *Handbook of Methods in Aquatic Microbial Ecology*. Num Pages: 12. CRC Press, 1993.

[218] Orlic, M., Gacic, M., and Laviolette, P. "The currents and circulation of the adriatic sea". In: *Oceanologica Acta* 15.2 (1992). Ed. by Gauthier-Villars, pp. 109–124. URL: https://archimer.ifremer.fr/doc/00100/21145/.

[219] Orlić, M. et al. "Wind-driven upwelling in the Adriatic Sea: Coastal vs. open-sea processes". In: *Progress in Oceanography* 234 (June 2025), p. 103470. ISSN: 00796611. DOI: 10.1016/j.pocean.2025.103470. URL: https://linkinghub.elsevier.com/retrieve/pii/S0079661125000588.

[220] Orsini, L. et al. "Toxic Pseudo-nitzschia multistriata (Bacillariophyceae) from the Gulf of Naples: Morphology, toxin analysis and phylogenetic relationships with other Pseudo-nitzschia species". In: *European Journal of Phycology* 37 (May 2002), pp. 247–257. DOI: 10.1017/S0967026202003608.

[221] Ovens, K., Eames, B. F., and McQuillan, I. "Comparative Analyses of Gene Co-expression Networks: Implementations and Applications in the Study of Evolution". In: *Frontiers in Genetics* Volume 12 - 2021 (2021). ISSN: 1664-8021. DOI: 10.3389/fgene.2021.695399. URL: https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2021.695399.

[222] Paliy, O. and Shankar, V. "Application of multivariate statistical techniques in microbial ecology". In: *Molecular Ecology* 25.5 (Mar. 2016), pp. 1032–1057. ISSN: 1365-294X. DOI: 10.1111/mec.13536.

[223] Patti, B., Torri, M., and Cuttitta, A. "Interannual summer biodiversity changes in ichthyoplankton assemblages of the Strait of Sicily (Central Mediterranean) over the period 2001–2016". In: *Frontiers in Marine Science* Volume 9 - 2022 (2022). ISSN: 2296-7745. DOI: 10.3389/fmars.2022.960929. URL: https://www.frontiersin.org/journals/marine-science/articles/10.3389/fmars.2022.960929.

[224] Patti, B. et al. "Role of physical forcings and nutrient availability on the control of satellite-based chlorophyll a concentration in the coastal upwelling area of the Sicilian Channel". In: *Scientia Marina* 74.3 (Sept. 2010), pp. 577–588. DOI: 10.3989/scimar.2010.74n3577. URL: https://scientiamarina.revistas.csic.es/index.php/scientiamarina/article/view/1185.

[225] Penna, A. et al. "Potentially harmful microalgal distribution in an area of the NW Adriatic coastline: Sampling procedure and correlations with environmental factors". In: *Estuarine, Coastal and Shelf Science* 70.1 (Oct. 2006), pp. 307–316. ISSN: 02727714. DOI: 10.1016/j.ecss.2006.06.023. URL: https://linkinghub.elsevier.com/retrieve/pii/S0272771406002836.

[226] Penna, A. et al. "Analysis of phytoplankton assemblage structure in the Mediterranean Sea based on high-throughput sequencing of partial 18S rRNA sequences". In: *Marine Genomics* 36 (June 2017). DOI: 10.1016/j.margen.2017.06.001.

[227] Picheral, M. et al. "The Underwater Vision Profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton". In: *Limnology and Oceanography: Methods* 8.9 (Sept. 2010), pp. 462–473. ISSN: 1541-5856, 1541-5856. DOI: 10.4319/lom.2010.8.462. URL: https://aslopubs.onlinelibrary.wiley.com/doi/10.4319/lom.2010.8.462.

[228] Pierella Karlusich, J. J. et al. "Patterns and drivers of diatom diversity and abundance in the global ocean". In: *Nature Communications* 16.1 (Apr. 11, 2025), p. 3452. ISSN: 2041-1723. DOI: 10.1038/s41467-025-58027-7. URL: https://www.nature.com/articles/s41467-025-58027-7.

[229] Pierella Karlusich, J. J., Ibarbalz, F. M., and Bowler, C. "Phytoplankton in the Tara Ocean". In: *Annual Review of Marine Science* 12.Volume 12, 2020 (2020), pp. 233–265. ISSN: 1941-0611. DOI: https://doi.org/10.1146/annurev-marine-010419-010706. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-marine-010419-010706.

[230] Placenti, F. et al. "Hydrological and Biogeochemical Patterns in the Sicily Channel: New Insights From the Last Decade (2010-2020)". In: *Frontiers in Marine Science* Volume 9 - 2022 (2022). ISSN: 2296-7745. DOI: 10.3389/fmars.2022.733540. URL: https://www.frontiersin.org/journals/marine-science/articles/10.3389/fmars.2022.733540.

[231] Prairie, J. C. et al. "Biophysical interactions in the plankton: A cross-scale review". In: *Limnology and Oceanography: Fluids and Environments* 2.1 (Apr. 2012), pp. 121–145. ISSN: 2157-3689, 2157-3689. DOI: 10.1215/21573689-1964713. URL: https://aslopubs.onlinelibrary.wiley.com/doi/10.1215/21573689-1964713.

[232] Pulina, S. et al. "Seasonal variations of phytoplankton size structure in relation to environmental variables in three Mediterranean shallow coastal lagoons". In: *Estuarine, Coastal and Shelf Science* 212 (July 2018), pp. 95–104. DOI: 10.1016/j.ecss.2018.07.002.

[233] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2024. URL: https://www.R-project.org/.

[234] Ramaswami, G. et al. "Integrative genomics identifies a convergent molecular subtype that links epigenomic with transcriptomic differences in autism". In: *Nature communications* 11 (Sept. 2020), p. 4873. DOI: 10.1038/s41467-020-18526-1.

[235] Ramette, A. "Multivariate analyses in microbial ecology". In: *FEMS microbiology ecology* 62.2 (Nov. 2007), pp. 142–160. ISSN: 0168-6496. DOI: 10.1111/j.1574-6941.2007.00375.x.

[236] Reale, M. et al. "The Regional Earth System Model RegCM-ES: Evaluation of the Mediterranean Climate and Marine Biogeochemistry". In: *Journal of Advances in Modeling Earth Systems* 12.9 (2020). e2019MS001812 2019MS001812, e2019MS001812. DOI: https://doi.org/10.1029/2019MS001812. eprint: https://agupubs.

onlinelibrary.wiley.com/doi/pdf/10.1029/2019MS001812. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001812.

[237] Régimbeau, A. et al. "Unveiling the link between phytoplankton molecular physiology and biogeochemical cycling via genome-scale modeling". In: *Science Advances* 11.23 (2025), eadq3593. DOI: 10.1126/sciadv.adq3593. eprint: https://www.science.org/doi/pdf/10.1126/sciadv.adq3593. URL: https://www.science.org/doi/abs/10.1126/sciadv.adq3593.

[238] Reichert, K. et al. "Species richness estimation: Estimator performance and the influence of rare species". In: *Limnology and Oceanography: Methods* 8.6 (2010), pp. 294–303. DOI: https://doi.org/10.4319/lom.2010.8.294. eprint: https://aslopubs.onlinelibrary.wiley.com/doi/pdf/10.4319/lom.2010.8.294. URL: https://aslopubs.onlinelibrary.wiley.com/doi/abs/10.4319/lom.2010.8.294.

[239] Reynolds, C. "Ecology of phytoplankton". In: *Cambridge University Press* (Jan. 2006).

[240] Ribera d'Alcalà, M. et al. "Seasonal patterns in plankton communities in a pluriannual time series at a coastal Mediterranean site (Gulf of Naples): an attempt to discern recurrences and trends". In: *Scientia Marina* 68 (S1 Apr. 2004), pp. 65–83. DOI: 10.3989/scimar.2004.68s165. URL: https://scientiamarina.revistas.csic.es/index.php/scientiamarina/article/view/385.

[241] Rinaldi, E. et al. "Chlorophyll distribution and variability in the Sicily Channel (Mediterranean Sea) as seen by remote sensing data". In: *Continental Shelf Research* 77 (Apr. 2014), pp. 61–68. ISSN: 02784343. DOI: 10.1016/j.csr.2014.01.010. URL: https://linkinghub.elsevier.com/retrieve/pii/S0278434314000235.

[242] Rinaldi, S. and Solidoro, C. "Chaos and Peak-to-Peak Dynamics in a Plankton–Fish Model". In: *Theoretical Population Biology* 54.1 (Aug. 1998), pp. 62–77. ISSN: 00405809. DOI: 10.1006/tpbi.1998.1368. URL: https://linkinghub.elsevier.com/retrieve/pii/S0040580998913685.

[243] Robinson, A. et al. "The Atlantic Ionian Stream". In: *Journal of Marine Systems* 20.1 (Apr. 1999), pp. 129–156. ISSN: 09247963. DOI: 10.1016/S0924-7963(98)00079-7. URL: https://linkinghub.elsevier.com/retrieve/pii/S0924796398000797.

[244] Robinson, K. L. et al. "Big or small, patchy all: Resolution of marine plankton patch structure at micro- to submesoscales for 36 taxa". In: *Science Advances* 7.47 (Nov. 19, 2021), eabk2904. ISSN: 2375-2548. DOI: 10.1126/sciadv.abk2904. URL: https://www.science.org/doi/10.1126/sciadv.abk2904.

[245] Rogers, T. L., Johnson, B. J., and Munch, S. B. "Chaos is not rare in natural ecosystems". In: *Nature Ecology & Evolution* 6.8 (June 27, 2022), pp. 1105–1111. ISSN: 2397-334X. DOI: 10.1038/s41559-022-01787-y. URL: https://www.nature.com/articles/s41559-022-01787-y.

[246] Rogers, T. L. et al. "Intermittent instability is widespread in plankton communities". In: *Ecology Letters* 26.3 (Mar. 2023), pp. 470–481. ISSN: 1461-023X, 1461-0248. DOI: 10.1111/ele.14168. URL: https://onlinelibrary.wiley.com/doi/10.1111/ele.14168.

[247] Rosenberg, A. and Hirschberg, J. "V-Measure: A Conditional Entropy-Based External Cluster Evaluation Measure." In: Jan. 2007, pp. 410–420.

[248] Roshchin, A. M. and Khauilov, K. *Zhiznennye tsikly diatomovykh vodorosleui*. Naukova dumka, 1994.

[249] Roy, S. and Chattopadhyay, J. "Towards a resolution of 'the paradox of the plankton': A brief overview of the proposed mechanisms". In: *Ecological Complexity* 4.1 (Mar. 2007), pp. 26–33. ISSN: 1476945X. DOI: 10.1016/j.ecocom.2007.02.016. URL: https://linkinghub.elsevier.com/retrieve/pii/S1476945X07000165.

[250] Rusch, D. B. et al. "The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific". In: *PLOS Biology* 5.3 (Mar. 2007), pp. 1–34. DOI: 10.1371/journal.pbio.0050077. URL: https://doi.org/10.1371/journal.pbio.0050077.

[251] Russo, L. et al. "Trophic hierarchy in a marine community revealed by network analysis on co-occurrence data". In: *Food Webs* 32 (2022), e00246. ISSN: 2352-2496. DOI: https://doi.org/10.1016/j.fooweb.2022.e00246. URL: https://www.sciencedirect.com/science/article/pii/S2352249622000283.

[252] Russo, M. et al. "MRP3 is a sex determining gene in the diatom Pseudo-nitzschia multistriata". In: *Nature Communications* 9 (Nov. 2018). DOI: 10.1038/s41467-018-07496-0.

[253] Russo, S. et al. "Unveiling the Relationship Between Sea Surface Hydrographic Patterns and Tuna Larval Distribution in the Central Mediterranean Sea". In: *Frontiers in Marine Science* Volume 8 - 2021 (2021). ISSN: 2296-7745. DOI: 10.3389/fmars.2021.708775. URL: https://www.frontiersin.org/journals/marine-science/articles/10.3389/fmars.2021.708775.

[254] Sabetta, L. et al. "Body size-abundance distributions of nano-/micro-phytoplankton guilds in coastal marine ecosystems". In: *Estuarine, Coastal and Shelf Science* 63 (June 2005), pp. 645–663. DOI: 10.1016/j.ecss.2005.01.009.

[255] Sabetta, L. et al. "Body size–abundance distributions of nano- and micro-phytoplankton guilds in coastal marine ecosystems". In: *Estuarine, Coastal and Shelf Science* 63.4 (2005), pp. 645–663. ISSN: 0272-7714. DOI: https://doi.org/10.1016/j.ecss.2005.01.009. URL: https://www.sciencedirect.com/science/article/pii/S0272771405000181.

[256] Sabine, C. L. et al. "The Oceanic Sink for Anthropogenic CO2". In: *Science* 305.5682 (July 16, 2004), pp. 367–371. DOI: 10.1126/science.1097403. URL: https://www.science.org/doi/10.1126/science.1097403.

[257] Salazar, G. et al. "Gene Expression Changes and Community Turnover Differentially Shape the Global Ocean Metatranscriptome". In: *Cell* 179.5 (2019), 1068–1083.e21. ISSN: 0092-8674. DOI: https://doi.org/10.1016/j.cell.2019.10.014. URL: https://www.sciencedirect.com/science/article/pii/S009286741931164X.

[258] Salon, S. et al. "Novel metrics based on Biogeochemical Argo data to improve the model uncertainty evaluation of the CMEMS Mediterranean marine ecosystem forecasts". In: *Ocean Science* 15.4 (2019), pp. 997–1022. DOI: 10.5194/os-15-997-2019. URL: https://os.copernicus.org/articles/15/997/2019/.

[259] Samarasekara, S. et al. "Measurements and analysis of nitrogen and phosphorus in oceans: Practice, frontiers, and insights". In: *Heliyon* 10 (Mar. 2024). DOI: 10.1016/j.heliyon.2024.e28182.

[260] Sani, T. et al. "Evolution of Freshwater Runoff in the Western Adriatic Sea over the Last Century". In: *Environments* 11.1 (Jan. 20, 2024), p. 22. ISSN: 2076-3298. DOI: 10.3390/environments11010022. URL: https://www.mdpi.com/2076-3298/11/1/22.

[261] Sarmiento, J. L. and Gruber, N. *Ocean biogeochemical dynamics*. Princeton: Princeton university press, 2006. ISBN: 978-0-691-01707-5.

[262] Sato, S. et al. "Novel Sex Cells and Evidence for Sex Pheromones in Diatoms". In: *PLOS ONE* 6.10 (Oct. 2011), pp. 1–15. DOI: 10.1371/journal.pone.0026923. URL: https://doi.org/10.1371/journal.pone.0026923.

[263] Savino, A., Provero, P., and Poli, V. "Differential Co-Expression Analyses Allow the Identification of Critical Signalling Pathways Altered during Tumour Transformation and Progression". In: *International Journal of Molecular Sciences* 21 (Dec. 2020), p. 9461. DOI: 10.3390/ijms21249461.

[264] Saxena, A. et al. "A Review of Clustering Techniques and Developments". In: *Neurocomputing* 267 (July 2017). DOI: 10.1016/j.neucom.2017.06.053.

[265] Scalco, E. et al. “The sexual phase of the diatom Pseudo-nitzschia multistriata: cytological and timelapse cinematography characterization”. In: *Protoplasma* (Nov. 2016). DOI: 10.1007/s00709-015-0891-5.

[266] Scheffer, M. et al. “Why plankton communities have no equilibrium: solutions to the paradox”. In: *Hydrobiologia* 491.1 (Jan. 2003), pp. 9–18. ISSN: 0018-8158, 1573-5117. DOI: 10.1023/A:1024404804748. URL: https://link.springer.com/10.1023/A:1024404804748.

[267] Sellner, K. G. and Fonda-Umani, S. “Dinoflagellate blooms and mucilage production”. In: *Coastal and Estuarine Studies*. Ed. by Malone, T. C. et al. Vol. 55. Washington, D. C.: American Geophysical Union, 1999, pp. 173–206. ISBN: 978-0-87590-269-2. DOI: 10.1029/CE055p0173. URL: http://www.agu.org/books/ce/v055/CE055p0173/CE055p0173.shtml.

[268] Serin, E. A. R. et al. “Learning from Co-expression Networks: Possibilities and Challenges”. In: *Frontiers in Plant Science* Volume 7 - 2016 (2016). ISSN: 1664-462X. DOI: 10.3389/fpls.2016.00444. URL: https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2016.00444.

[269] Seuront, L., Schmitt, F., and Lagadeuc, Y. “Turbulence intermittency, small-scale phytoplankton patchiness and encounter rates in plankton: where do we go from here?” In: *Deep Sea Research Part I: Oceanographic Research Papers* 48.5 (May 2001), pp. 1199–1215. ISSN: 09670637. DOI: 10.1016/S0967-0637(00)00089-3. URL: https://linkinghub.elsevier.com/retrieve/pii/S0967063700000893.

[270] Shaltout, M. and Omstedt, A. “Modelling the water and heat balances of the Mediterranean Sea using a two-basin model and available meteorological, hydrological, and Ocean data”. In: *Oceanologia* 44 (Nov. 2014). DOI: 10.1016/j.oceano.2014.11.001.

[271] Shendure, J. and Ji, H. “Shendure J, Ji H.. Next-generation DNA sequencing. Nat Biotechnol 26: 1135-1145”. In: *Nature biotechnology* 26 (Nov. 2008), pp. 1135–45. DOI: 10.1038/nbt1486.

[272] Shi, H. et al. “Impact of Tropical Cyclone on Coastal Phytoplankton Blooms and Underlying Mechanisms”. In: *Journal of Hydrology: Regional Studies* 59 (June 2025), p. 102389. ISSN: 22145818. DOI: 10.1016/j.ejrh.2025.102389. URL: https://linkinghub.elsevier.com/retrieve/pii/S2214581825002149 (visited on 10/17/2025).

[273] Sieburth, J. M., Smetacek, V., and Lenz, J. “Pelagic ecosystem structure: Heterotrophic compartments of the plankton and their relationship to plankton size fractions 1”. In: *Limnology and Oceanography* 23.6 (Nov. 1978), pp. 1256–1263.

ISSN: 0024-3590, 1939-5590. DOI: 10.4319/lo.1978.23.6.1256. URL: https://aslopubs.onlinelibrary.wiley.com/doi/10.4319/lo.1978.23.6.1256.

[274] Šizling, A. L. et al. "Between Geometry and Biology: The Problem of Universality of the Species-Area Relationship." In: *The American Naturalist* 178.5 (2011), pp. 602–611. DOI: 10.1086/662176. URL: https://doi.org/10.1086/662176.

[275] Smayda, T. and Reynolds, C. "Community Assembly in Marine Phytoplankton: Application of Recent Models to Harmful Dinoflagellate Blooms". In: *Journal of Plankton Research* 23 (May 2001). DOI: 10.1093/plankt/23.5.447.

[276] Smith, H., Kim, H., and Walker, S. "Scarcity of scale-free topology is universal across biochemical networks". In: *Scientific Reports* 11 (Mar. 2021). DOI: 10.1038/s41598-021-85903-1.

[277] Solidoro, C. et al. "Current state, scales of variability, and trends of biogeochemical properties in the northern Adriatic Sea". In: *Journal of Geophysical Research: Oceans* 114 (C7 July 2009), 2008JC004838. ISSN: 0148-0227. DOI: 10.1029/2008JC004838. URL: https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2008JC004838.

[278] Sommer, U. et al. "Beyond the Plankton Ecology Group (PEG) Model: Mechanisms Driving Plankton Succession". In: *Annual Review of Ecology, Evolution, and Systematics* 43.Volume 43, 2012 (2012), pp. 429–448. DOI: https://doi.org/10.1146/annurev-ecolsys-110411-160251. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-ecolsys-110411-160251.

[279] Sournia, A., Chrdtiennot-Dinet, M.-J., and Ricard, M. "Marine phytoplankton: how many species in the world ocean?" In: *Journal of Plankton Research* 13.5 (1991), pp. 1093–1099. ISSN: 0142-7873, 1464-3774. DOI: 10.1093/plankt/13.5.1093. URL: https://academic.oup.com/plankt/article-lookup/doi/10.1093/plankt/13.5.1093.

[280] Spatharis, S. and Tsirtsis, G. "Ecological quality scales based on phytoplankton for the implementation of Water Framework Directive in the Eastern Mediterranean". In: *Ecological Indicators* 10.4 (July 2010), pp. 840–847. ISSN: 1470160X. DOI: 10.1016/j.ecolind.2010.01.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S1470160X10000063.

[281] Stabili, L., Caroppo, C., and Cavallo, R. A. "Monitoring of a Coastal Mediterranean Area: Culturable Bacteria, Phytoplankton, Environmental Factors and their Relationships in the Southern Adriatic Sea". In: *Environmental Monitoring and Assessment* 121.1 (Oct. 2006), pp. 303–325. ISSN: 0167-6369, 1573-2959. DOI: 10.1007/s10661-005-9124-2. URL: http://link.springer.com/10.1007/s10661-005-9124-2.

[282] Stec, K. F. et al. "Modelling plankton ecosystems in the meta-omics era. Are we ready?" In: *Marine Genomics* 32 (2017), pp. 1–17. ISSN: 1874-7787. DOI: https://doi.org/10.1016/j.margen.2017.02.006. URL: https://www.sciencedirect.com/science/article/pii/S1874778716301544.

[283] Stockwell, J. D. et al. "Storm impacts on phytoplankton community dynamics in lakes". In: *Global Change Biology* 26.5 (May 2020), pp. 2756–2784. ISSN: 1354-1013, 1365-2486. DOI: 10.1111/gcb.15033. URL: https://onlinelibrary.wiley.com/doi/10.1111/gcb.15033.

[284] Stoecker, D. K. et al. "Mixotrophy in the Marine Plankton". In: *Annual Review of Marine Science* 9.1 (Jan. 3, 2017), pp. 311–335. ISSN: 1941-1405, 1941-0611. DOI: 10.1146/annurev-marine-010816-060617. URL: https://www.annualreviews.org/doi/10.1146/annurev-marine-010816-060617.

[285] Stosch, H. "Manipulierung der Zellgröße von Diatomeen im experiment". In: *Phycologia* 5.1 (1965), pp. 21–44.

[286] Stuart, J. et al. "A Gene-Coexpression Network for Global Discovery of Conserved Genetic Modules". In: *Science (New York, N.Y.)* 302 (Nov. 2003), pp. 249–55. DOI: 10.1126/science.1087447.

[287] Subramanian, A. et al. "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles". In: *Proceedings of the National Academy of Sciences* 102.43 (2005), pp. 15545–15550. DOI: 10.1073/pnas.0506580102. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.0506580102. URL: https://www.pnas.org/doi/abs/10.1073/pnas.0506580102.

[288] Sun, Q. et al. "Role of reactive oxygen species (ROS) signaling pathway in the wound-induced spore formation of Pyropia yezoensis". In: *Algal Research* 79 (2024), p. 103457. ISSN: 2211-9264. DOI: https://doi.org/10.1016/j.algal.2024.103457. URL: https://www.sciencedirect.com/science/article/pii/S2211926424000699.

[289] Sunagawa, S. et al. "Structure and function of the global ocean microbiome". In: *Science* 348.6237 (May 22, 2015), p. 1261359. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1261359. URL: https://www.science.org/doi/10.1126/science.1261359.

[290] Thompson, P. A. and Carstensen, J. "Global observing for phytoplankton? A perspective". In: *Journal of Plankton Research* 45.1 (Feb. 1, 2023). Ed. by Campbell, L., pp. 221–234. ISSN: 0142-7873, 1464-3774. DOI: 10.1093/plankt/fbab090. URL: https://academic.oup.com/plankt/article/45/1/221/6520712.

[291] Thompson, P. A. et al. "Tropical cyclones: what are their impacts on phytoplankton ecology?" In: *Journal of Plankton Research* 45.1 (Feb. 1, 2023), pp. 180–204. ISSN: 0142-7873, 1464-3774. DOI: 10.1093/plankt/fbac062. URL: https://academic.oup.com/plankt/article/45/1/180/6835332.

[292] Tobler, W. "A Computer Movie Simulating Urban Growth in the Detroit Region". In: *Economic Geography* 46 (June 1970). DOI: 10.2307/143141.

[293] Tornero, V. and Ribera d'Alcalà, M. "Contamination by hazardous substances in the Gulf of Naples and nearby coastal areas: A review of sources, environmental levels and potential impacts in the MSFD perspective". In: *Science of The Total Environment* 466-467 (2014), pp. 820–840. ISSN: 0048-9697. DOI: https://doi.org/10.1016/j.scitotenv.2013.06.106. URL: https://www.sciencedirect.com/science/article/pii/S0048969713007584.

[294] Totti, C. et al. "Phytoplankton communities in the northwestern Adriatic Sea: Interdecadal variability over a 30-years period (1988–2016) and relationships with meteoclimatic drivers". In: *Journal of Marine Systems* 193 (May 2019), pp. 137–153. ISSN: 09247963. DOI: 10.1016/j.jmarsys.2019.01.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S0924796318302471.

[295] Tréguer, P. et al. "Influence of diatom diversity on the ocean biological carbon pump". In: *Nature Geoscience* 11.1 (Jan. 2018), pp. 27–37. ISSN: 1752-0894, 1752-0908. DOI: 10.1038/s41561-017-0028-x. URL: https://www.nature.com/articles/s41561-017-0028-x.

[296] Trigueros, M. and Orive, E. "Seasonal variations of diatoms and dinoflagellates in a shallow, temperate estuary, with emphasis on neritic assemblages". In: *Hydrobiologia* 444 (Feb. 2001), pp. 119–133. DOI: 10.1023/A:1017563031810.

[297] Trombetta, T. et al. "Co-occurrence networks reveal the central role of temperature in structuring the plankton community of the Thau Lagoon". In: *Scientific Reports* 11.1 (Sept. 3, 2021), p. 17675. ISSN: 2045-2322. DOI: 10.1038/s41598-021-97173-y. URL: https://www.nature.com/articles/s41598-021-97173-y.

[298] Trudnowska, E. et al. "Cells of matter and life – towards understanding the structuring of particles and plankton patchiness in the Arctic fjords". In: *Frontiers in Marine Science* 9 (Aug. 18, 2022), p. 909457. ISSN: 2296-7745. DOI: 10.3389/fmars.2022.909457. URL: https://www.frontiersin.org/articles/10.3389/fmars.2022.909457/full.

[299] Tsirtsis, G., Spatharis, S., and Karydis, M. "Application of the lognormal equation to assess phytoplankton community structural changes induced by marine eutrophication". In: *Hydrobiologia* 605.1 (June 2008), pp. 89–98. ISSN: 0018-8158, 1573-5117. DOI: 10.1007/s10750-008-9307-2. URL: http://link.springer.com/10.1007/s10750-008-9307-2.

[300] Turk Dermastia, T. et al. "Ecological time series and integrative taxonomy unveil seasonality and diversity of the toxic diatom Pseudo-nitzschia H. Peragallo in the northern Adriatic Sea". In: *Harmful Algae* 93 (Mar. 2020), p. 101773. ISSN: 15689883. DOI: 10.1016/j.hal.2020.101773. URL: https://linkinghub.elsevier.com/retrieve/pii/S1568988320300536.

[301] Twining, B. and Baines, S. "The Trace Metal Composition of Marine Phytoplankton". In: *Annual review of marine science* 5 (Dec. 2011). DOI: 10.1146/annurev-marine-121211-172322.

[302] Utermöhl, H. "Zur Vervollkommnung der quantitativen Phytoplankton-Methodik". In: 1958. URL: https://api.semanticscholar.org/CorpusID:132193798.

[303] Vanormelingen, P. et al. "Heterothallic sexual reproduction in the model diatom Cylindrotheca". In: *European Journal of Phycology* 48.1 (2013), pp. 93–105. DOI: 10.1080/09670262.2013.772242. URL: https://doi.org/10.1080/09670262.2013.772242.

[304] Vara, A. D. L. et al. "Role of atmospheric resolution in the long-term seasonal variability of the Tyrrhenian Sea circulation from a set of ocean hindcast simulations (1997–2008)". In: *Ocean Modelling* 134 (Feb. 2019), pp. 51–67. ISSN: 14635003. DOI: 10.1016/j.ocemod.2019.01.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S1463500318301215.

[305] Varghese, D. et al. "Tiber River-Driven Chlorophyll-a and Total Suspended Matter Dynamics and Their Impacts along the Central Tyrrhenian Sea Coast: A Sentinel-2 Approach". In: *International Journal of Applied Earth Observation and Geoinformation* 134 (Nov. 2024), p. 104161. ISSN: 15698432. DOI: 10.1016/j.jag.2024.104161. URL: https://linkinghub.elsevier.com/retrieve/pii/S156984322400517X.

[306] Vascotto, I., Mozetič, P., and Francé, J. "Phytoplankton Time-Series in a LTER Site of the Adriatic Sea: Methodological Approach to Decipher Community Structure and Indicative Taxa". In: *Water* 13 (July 2021), p. 2045. DOI: 10.3390/w13152045.

[307] Verri, G. et al. "River runoff influences on the Central Mediterranean overturning circulation". In: *Climate Dynamics* 50.5 (Mar. 2018), pp. 1675–1703. ISSN: 0930-7575, 1432-0894. DOI: 10.1007/s00382-017-3715-9. URL: http://link.springer.com/10.1007/s00382-017-3715-9.

[308] Villamaña, M. et al. "Role of internal waves on mixing, nutrient supply and phytoplankton community structure during spring and neap tides in the upwelling ecosystem of Ría de Vigo (NW Iberian Peninsula)". In: *Limnology and Oceanography* 62.3 (May 2017), pp. 1014–1030. ISSN: 0024-3590, 1939-5590. DOI: 10.1002/lno.10482. URL: https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.10482.

[309] Villamaña, M. et al. "The role of mixing in controlling resource availability and phytoplankton community composition". In: *Progress in Oceanography* 178 (Sept. 2019), p. 102181. DOI: 10.1016/j.pocean.2019.102181.

[310] Volk, T. and Hoffert, M. I. "Ocean Carbon Pumps: Analysis of Relative Strengths and Efficiencies in Ocean-Driven Atmospheric CO_2 Changes". In: *Geophysical Monograph Series*. Ed. by Sundquist, E. and Broecker, W. Washington, D. C.: American Geophysical Union, Mar. 18, 2013, pp. 99–110. ISBN: 978-1-118-66432-2 978-0-87590-060-5. DOI: 10.1029/GM032p0099. URL: http://doi.wiley.com/10.1029/GM032p0099.

[311] Vollenweider, R. A. "Coastal marine eutrophication: principles and control". In: *Marine Coastal Eutrophication*. Elsevier, 1992, pp. 1–20. ISBN: 978-0-444-89990-3. DOI: 10.1016/B978-0-444-89990-3.50011-0. URL: https://linkinghub.elsevier.com/retrieve/pii/B9780444899903500110.

[312] Volpe, G. et al. "Seasonal to interannual phytoplankton response to physical processes in the Mediterranean Sea from satellite observations". In: *Remote Sensing of Environment* 117 (Feb. 2012), pp. 223–235. DOI: 10.1016/j.rse.2011.09.020.

[313] Ward, B. A. et al. "The Size Dependence of Phytoplankton Growth Rates: A Trade-Off between Nutrient Uptake and Metabolism". In: *The American Naturalist* 189.2 (Feb. 2017), pp. 170–177. ISSN: 0003-0147, 1537-5323. DOI: 10.1086/689992. URL: https://www.journals.uchicago.edu/doi/10.1086/689992.

[314] Warton, D. I., Wright, S. T., and Wang, Y. "Distance-based multivariate analyses confound location and dispersion effects". In: *Methods in Ecology and Evolution* 3.1 (2012), pp. 89–101. DOI: https://doi.org/10.1111/j.2041-210X.2011.00127.x.

[315] Weston, D. et al. “Connecting Genes, Coexpression Modules, and Molecular Signatures to Environmental Stress Phenotypes in Plants”. In: *BMC systems biology* 2 (Feb. 2008), p. 16. DOI: 10.1186/1752-0509-2-16.

[316] Wilhelm, M. et al. “Mass-spectrometry-based draft of the human proteome”. In: *Nature* 509 (May 2014), pp. 582–7. DOI: 10.1038/nature13319.

[317] Williams, R. and Follows, M. *Ocean Dynamics and the Carbon Cycle: Principles and Mechanisms*. June 2012. ISBN: 9780521843690. DOI: 10.1017/CBO9780511977817.

[318] Wood, S. N. *Generalized Additive Models: An Introduction with R*. 2nd ed. Chapman and Hall/CRC, May 18, 2017. ISBN: 978-1-315-37027-9. DOI: 10.1201/9781315370279. URL: https://www.taylorfrancis.com/books/9781498728348.

[319] Worden, A. Z. et al. “Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes”. In: *Science* 347.6223 (Feb. 13, 2015), p. 1257594. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1257594. URL: https://www.science.org/doi/10.1126/science.1257594.

[320] WoRMS. *World Register of Marine Species (WoRMS)*. Accessed online. URL: https://www.marinespecies.org/.

[321] Xu, D. and Tian, Y. “A Comprehensive Survey of Clustering Algorithms”. In: *Annals of Data Science* 2 (Aug. 2015). DOI: 10.1007/s40745-015-0040-1.

[322] Yamaguchi, M. et al. “Effects of temperature and salinity on the growth of the red tide flagellate Heterocapsa circularisquama (Dinophyceae) and Chattonella verruculosa (Raphidophyceae)”. In: *Journal of Plankton Research* 19 (Aug. 1997). DOI: 10.1093/plankt/19.8.1167.

[323] Yang, Y. et al. “Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types”. In: *Nature communications* 5 (Feb. 2014), p. 3231. DOI: 10.1038/ncomms4231.

[324] Yip, A. and Horvath, S. “Gene network interconnectedness and the generalized topological overlap measure”. In: *BMC bioinformatics* 8 (Feb. 2007), p. 22. DOI: 10.1186/1471-2105-8-22.

[325] Zhang, B. and Horvath, S. “A General Framework For Weighted Gene Co-Expression Network Analysis”. In: *Statistical applications in genetics and molecular biology* 4 (Feb. 2005), Article17. DOI: 10.2202/1544-6115.1128.

[326] Zhang, M. et al. “The underlying causes and effects of phytoplankton seasonal turnover on resource use efficiency in freshwater lakes”. In: *Ecology and Evolution* 11.13 (2021), pp. 8897–8909. DOI: https://doi.org/10.1002/ece3.7724. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.7724.

[327] Zhou, L. et al. “Transcriptomic and metabolic signatures of diatom plasticity to light fluctuations”. In: *Plant Physiology* 190.4 (Sept. 2022), pp. 2295–2314. ISSN: 0032-0889. DOI: 10.1093/plphys/kiac455. eprint: https://academic.oup.com/plphys/article-pdf/190/4/2295/47382630/kiac455.pdf. URL: https://doi.org/10.1093/plphys/kiac455.

[328] Zingone, A. et al. “Assembly rules vary seasonally in stable phytoplankton associations of the Gulf of Naples (Mediterranean Sea)”. In: *Marine Ecology* (Dec. 2022). DOI: 10.1111/maec.12730.

[329] Zogopoulos, V. L. et al. “Approaches in Gene Coexpression Analysis in Eukaryotes”. In: *Biology* 11.7 (2022). ISSN: 2079-7737. DOI: 10.3390/biology11071019. URL: https://www.mdpi.com/2079-7737/11/7/1019.

[330] Zufia, J. A., Legrand, C., and Farnelid, H. “Seasonal dynamics in picocyanobacterial abundance and clade composition at coastal and offshore stations in the Baltic Sea”. In: *Scientific Reports* 12.1 (Aug. 22, 2022), p. 14330. ISSN: 2045-2322. DOI: 10.1038/s41598-022-18454-8. URL: https://www.nature.com/articles/s41598-022-18454-8.

[331] Zuur, A. F., Ieno, E. N., and Smith, G. M. *Analysing Ecological Data*. Red. by Gail, M. et al. Statistics for Biology and Health. New York, NY: Springer New York, 2007. ISBN: 978-0-387-45967-7 978-0-387-45972-1. DOI: 10.1007/978-0-387-45972-1. URL: http://link.springer.com/10.1007/978-0-387-45972-1.

[332] Zuur, A. F. et al. *Mixed effects models and extensions in ecology with R*. Statistics for Biology and Health. New York, NY: Springer New York, 2009. ISBN: 978-0-387-87457-9 978-0-387-87458-6. DOI: 10.1007/978-0-387-87458-6. URL: https://link.springer.com/10.1007/978-0-387-87458-6.