Please cite this article in press as: Cleves et al., Unknown to Known: Advancing Knowledge of Coral Gene Function, Trends in Genetics (2019), https://doi.org/10.1016/j.tig.2019.11.001

CellPress REVIEWS

Review

Unknown to Known: Advancing Knowledge of Coral Gene Function

Phillip A. Cleves,[1,7] Alexander Shumaker,[2,7] JunMo Lee,[3,5,6,7] Hollie M. Putnam,[4] and Debashish Bhattacharya[3,6,@,*]

Given the catastrophic changes befalling coral reefs, understanding coral gene function is essential to advance reef conservation. This has proved challenging due to the paucity of genomic data and genetic tools available for corals. Recently, CRISPR/Cas9 gene editing was applied to these species; however, a major bottleneck is the identification and prioritization of candidate genes for manipulation. This issue is exacerbated by the many unknown ('dark') coral genes that may play key roles in the stress response. We review the use of gene coexpression networks that incorporate both known and unknown genes to identify targets for reverse genetic analysis. This approach also provides a framework for the annotation of dark genes in established interaction networks to improve our fundamental knowledge of coral gene function.

A Way Forward for Coral Genomics and Reverse Genetics

Reef-forming corals are foundational species that provide habitats for many marine organisms and offer a number of valuable services [1] for coastal communities, supporting biodiversity worldwide. However, corals are in global decline due to environmental stresses caused by anthropogenic climate change and other local human activities. These stressors can lead to the breakdown of the essential endosymbiosis with dinoflagellate algae in the family Symbiodiniaceae, leading to coral bleaching and death. Due to the ecological, cultural, and economic importance of coral reefs [2], there is an urgent need to understand the molecular and genetic bases of the coral stress response and bleaching [3]. Consequently, there has been an acceleration in the application of high-throughput single (or multi-) omics methods such as genomics, transcriptomics, proteomics, and epigenomics to study the coral animal and associated microbes (e.g., algal symbionts, bacteria, viruses) (Figure 1, Key Figure) [4,5]. The resulting data have led to many important hypotheses about the roles of genes and pathways involved in stress and bleaching. However, these hypotheses are generally based on correlational data and often derived from complex datasets that contain hundreds to thousands of genes and multiple pathways, making it difficult to identify and understand in detail the key components of the stress response [6–11]. Given this situation, several important questions remain in the field. First, which of these genes and pathways are involved in the coral stress response and bleaching? Second, what are the master genetic regulators of these responses? Third, and likely the most important, how will these genetic networks evolve in response to climate change? To address these questions, we need to capitalize on the power of multiomics data to identify candidate key genetic determinants of bleaching, which would then be functionally tested in corals using emerging genome editing tools. This final and key step, which may involve studying the outcomes of crosses or gene editing events, is hindered by the long lifespan of corals and the multiple years that are needed for juveniles to reach sexual maturity.

A technical hurdle for exploiting coral multiomics data is that they are often generated from diverse species and without consistent experimental conditions, making it impossible to compare datasets. Therefore, we advocate a model-systems approach to coral biology where the community selects a limited number of conditions and species to apply a range of multiomics approaches to the **holobiont** (see Glossary) from genome to metabolome, which will provide the data and capacity to identify key genes and pathways involved in bleaching. An excellent example of this approach is the emerging model system for coral biology, the sea anemone *Aiptasia* [12,13]. In this organism, there has been a concerted effort to standardize anemone and algal strains along with experimental conditions, making it possible to integrate omics research across the community [14–16]. Such an effort should be expended on the development of a small number of reef-building coral models that reflect the extant diversity of these species.

Highlights

Coral reefs are under threat from warming oceans. Understanding the basis of the thermal stress response is therefore critical to devising strategies to protect corals and the diverse ecosystem services they provide.

Developing a small number of coral model systems will be a necessary step to focus multiomics and functional genetics research to gain a mechanistic understanding of coral holobiont biology.

The recent development of CRISPR/Cas9 methods for gene editing in corals offers the opportunity to test hypotheses about coral gene function.

Along with studying orthologs of well-understood metazoan genes using reverse genetics, we advocate the use of gene coexpression networks to identify 'dark' genes of unknown function that occupy hub or peripheral network positions. Disruption of dark genes may offer novel insights into coral biology and identify species-specific adaptations.

[1]Department of Genetics, Stanford University, Stanford, CA 94305, USA

[2]Microbial Biology Graduate Program, Rutgers University, New Brunswick, NJ 08901, USA

[3]Department of Biochemistry and Microbiology, Rutgers University, New Brunswick, NJ 08901, USA

[4]Department of Biological Sciences, University of Rhode Island, Kingston, RI 02881, USA

[5]Current address: Department of Oceanography, Kyungpook National University, Daegu 41566, Korea

[6]Laboratory website: http://dblab.rutgers.edu/

[7]Equal contribution

[@]Twitter: @DebashB (D. Bhattacharya).

*Correspondence: d.bhattacharya@rutgers.edu

Please cite this article in press as: Cleves et al., Unknown to Known: Advancing Knowledge of Coral Gene Function, Trends in Genetics (2019), https://doi.org/10.1016/j.tig.2019.11.001

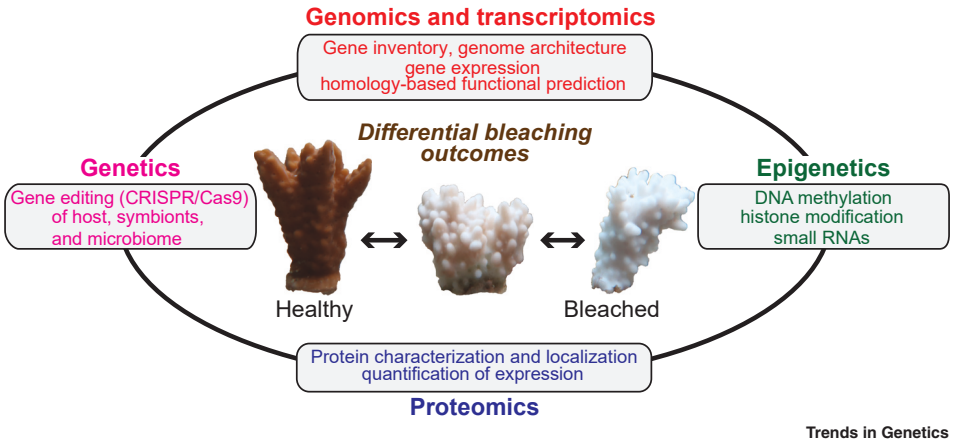

Figure 1. Multiomics and Functional Genetics Approaches to Be Used to Study the Coral (In This Case, *Montipora capitata*) Thermal Stress Response That Can Lead to Differential Bleaching Outcomes.
For example, outcomes range from healthy brown coral on the left to bleached white coral on the right.

Another complication of multiomics data is that a large number of coral and algal genes have no known function (i.e., are 'dark'). For example, ~33% of dinoflagellate genes lack an annotation and only ~1.4% of these unknown proteins contain a known domain [17]. Similarly, in the genome of the rice coral *Montipora capitata* [18], ~11.7% of the predicted gene models do not have a significant ortholog (BLAST *e*-value cut-off ≤ 1e−10 against nonredundant NCBI database). Interestingly, many coral dark genes are differentially expressed under stress (see below). In corals, dark genes restricted to single species are relatively small in number but may encode important and potentially coral-specific functions that remain to be studied. For example, with 8279 **orthogroups (OGs)** shared by all currently available coral genomes, we calculated the numbers of species-specific dark gene groups to be 84 in *Acropora digitifera*, 479 in *M. capitata*, and 72 in *Pocillopora damicornis*. Therefore, although targeting genes with conserved functions [e.g., heat shock proteins, master regulators of the immune response or the unfolded protein response (UPR)] derived from a variety of established model organisms is a key step in validating hypotheses about coral genes, ultimately there is a need to understand the function of unannotated dark genes that may be coral-specific innovations or environmental adaptations.

Here, we begin by briefly surveying what is known about genes involved in the coral thermal stress response and bleaching using standard omics approaches that provide correlational data (a more exhaustive handling is available in [3]). In these sections, we also highlight approaches where future research can facilitate the integration of multiomics data both within and between species and across studies. To illustrate the importance of including unannotated genes in multiomics analyses, we present a simple example using existing gene coexpression data including both known and dark genes. Following this section, we discuss the development of CRISPR/Cas9 genome editing, a **reverse genetic** tool in corals, potential avenues for prioritizing genes for manipulation, and the need to move beyond multiomics datasets toward rigorous functional testing of hypotheses regarding coral stress and bleaching. We end with a discussion of the significant challenges facing the field of coral conservation, even with the availability of powerful molecular tools to address this problem.

Contribution of Traditional Omics Methods to Coral Science

Limited proteomic and **epigenetic** studies of host–symbiont interactions and development in coral have been undertaken. Proteomes have been generated under bleaching stress in the cosmopolitan genus *Acropora*. A study in *A. microphthalma* demonstrated that proteins thought to combat reactive oxygen species (ROS) produced by symbionts under damaging thermal conditions were not detected in the coral host proteome, but rather were detected in the proteome of the alga [19]. These

Glossary

Epigenetics: the study of heritable changes in phenotype that do not rely on DNA sequence alteration. An example is DNA modification by the addition of methyl groups to DNA that can modulate gene expression.
Gene knockdown and knockout: an experimental method that reduces (knockdown) or abolishes (knockout) the function of one or more genes in an organism. CRISPR/Cas9 offers the possibility of generating permanent gene knockouts because the edited DNA sites may be inherited in offspring.
Holobiont: a group of organisms that coexist as a defined functional unit, such as the coral holobiont that includes the cnidarian animal, the dinoflagellate algal symbiont, and the remaining microbiome, comprising of microeukaryotes (e.g., fungi, chromerids), prokaryotes, and viruses.
Morpholinos: a type of chemically modified oligonucleotide used in reverse genetics research, whereby the modified DNA binds to sites on RNA to obstruct processes such as translation, splicing, miRNA binding, and ribozyme activity.
Nonhomologous end joining: a pathway that results in the repair of DNA double-strand breaks. This process is referred to as 'nonhomologous' because the repair occurs through direct DNA ligation and does not require a homologous region to act as a template.
Orthogroup (OG): the collection of genes in the genomes of living species that are descended from a single gene in the last common ancestor of the group under study.
Reverse genetics: an approach that elucidates gene function by disrupting a target gene and studying the resulting phenotype; as opposed to forward genetics, where a set of random mutants are generated and then screened for a phenotype of interest (e.g., traditional genetic screens).
Scleractinia: stony or hard corals in the animal phylum Cnidaria that are biomineralizing. The clonal polyps build colonies that, together, comprise the backbone of coral reefs. The Scleractinia

Please cite this article in press as: Cleves et al., Unknown to Known: Advancing Knowledge of Coral Gene Function, Trends in Genetics (2019), https://doi.org/10.1016/j.tig.2019.11.001

authors did not identify effector proteins of apoptosis or autophagy in the host, which contrasts with the common notion that the coral host mediates bleaching via destruction of endosymbionts. Instead, they reported expression of host synaptotagmin, vSNARE, and tSNARE proteins, and posited that *A. microphthalma* may employ an exocytotic mechanism to eject the symbionts [19]. These results are consistent with the observation in *Aiptasia* that the vast majority of heat-induced bleaching involves the expulsion of intact algal cells [20]. Ricaurte and colleagues [21] conducted a proteomic study of *Acropora palmata* during a bleaching event, assessing both bleached and unbleached colonies, and identified 38 proteins exhibiting significant differences. Among expressed and differentially expressed proteins were participants in UV, thermal, and oxidative stress responses, as well as proapoptotic proteins. It remains an open question whether any or all of these genes have a role in bleaching or survival during heat stress.

In animals, epigenetic factors, including DNA methylation state, contribute significantly to the control of gene expression through modifying the transcriptional accessibility of the genome [22]. Epigenetic changes (e.g., in the gene body or promoter regions of genes) can regulate transcription to produce more or fewer protein products in response to the environment – conferring a degree of phenotypic plasticity – and can be heritable [23]. In coral, this plasticity may provide a capacity to acclimate to changes in the environment, but the interactions between plasticity and evolutionary adaptation are not clear and thus more investigations into the role of epigenetic mechanisms in enhancing coral resilience are needed (reviewed in [24]).

The trend emerging from experiments that bisulfite sequence corals and other invertebrates is that they have comparatively sparse methylation relative to vertebrates. For example, whole-genome bisulfite sequencing (WGBS) performed by Liew and colleagues on *Stylophora pistillata* revealed that approximately 7% of all CpG sites were methylated, with the majority of these positions located in intronic regions of gene bodies [9]. Notably, this study combined methylation and transcriptomic data to reveal that gene-body methylation was positively correlated with gene expression and that methylated genes displayed reduced transcriptional variability [9]. By contrast, a combination of methylation and expression data was used in *Acropora millepora* to reveal links between gene expression plasticity and the extent of gene-body methylation [6]. The presence of epimutations as drivers of divergent phenotypes has also been investigated in the Caribbean acroporid *A. palmata*. Here, a relatively small portion of the phenotypic variance was assigned to variation in DNA methylation linked to microenvironmental differences [25]. The differences between these studies could be due to differences across species, in experimental design, or both. To further test these hypotheses, it should be possible to experimentally modify the epigenetic state of corals using CRISPR/Cas9 by **knocking out** chromatin modifiers and/or directly modifying the epigenetic state with dCas9 fused to chromatin modifiers [26].

Although studies documenting changes in DNA methylation patterns in response to changes in environmental conditions exist [24,27], only recently has evidence for transmission of such modifications between parent and offspring been examined in corals (in this case, in the brain coral *Platygyra daedalea*) [28]. Whether similar transmission may occur in other coral species, and the extent to which the transmitted modifications prove advantageous, requires additional physiological and ecological investigation paired with epigenetics. Plasticity that has been transmitted across generations [i.e., cross-generational plasticity (CGP)] has been documented in the larvae [29] and juveniles [30] of the brooding coral *Pocillipora acuta*. This species is expected to exhibit CGP based on the close connection between parents and offspring during brooding and the high correlation between parent and offspring environment during the rapid settling process [31], However, to date no epigenetic data are available. It remains a largely open question whether and to what degree coral adaptation to climate changes will involve epigenetic changes. This question has been difficult to answer due in part to the high cost of WGBS. This high cost is driving a search for alternative sequencing approaches, such as methylation-enrichment MeDIP and MBD-seq [24]. In addition, improvements in Nanopore and PacBio sequencing technologies may have the potential to avoid the harsh conditions and high cost of bisulfite treatment while retaining single-base-pair resolution [32]. In the future, the application of coral epigenetic data to conservation efforts may involve 'environmental hardening'

originated in the Middle Triassic at least 240 million years ago.
Single-guide RNA (sgRNA): an RNA that contains a sequence (approximately 20 bases) that recognizes the gene of interest to be modified along with a tracRNA scaffold that is required to complex with the Cas9 nuclease to confer activity.

Please cite this article in press as: Cleves et al., Unknown to Known: Advancing Knowledge of Coral Gene Function, Trends in Genetics (2019), https://doi.org/10.1016/j.tig.2019.11.001

[27,30]. This process involves the identification (using epigenetics) and cultivation of well-acclimatized individuals under stressful conditions. These hardened corals could be introduced into natural populations for the long-term goal of propagating selected epigenotypes.

The most widely used omics method in corals has been high-throughput transcriptomics. The first attempts at transcriptomic profiling in corals used microarrays and were followed by a flurry of RNA-seq experiments aimed at identifying gene expression correlates of thermal stress and other aspects of coral biology (recently reviewed in [33]). These studies typically use a generalized linear model to identify differentially expressed genes (DEGs) between control and various test conditions to characterize the animal response to various environmental conditions [7,18,34–36]. This approach has led to many insights into coral biology *vis-à-vis* the response to thermal stress [35] and the creation of transcriptome databases to spur coral science [37]. Major DEGs associated with the heat and bleaching responses are transcription factors (e.g., heat shock transcription factor 1 that regulates heat-shock protein expression [18]), genes involved in the UPR [7,38], genes that putatively respond to ROS production presumably by the algal symbiont or the coral host [39], and genes in various signaling pathways implicated in stress, such as the Ras (small GTPases), mitogen-activated protein kinase (MAPK), and phosphatidylinositol 3-kinase and Akt/protein kinase B (PI3K–Akt) families [18].

To date, a major goal in coral omics is to identify core stress response genes and understand how differences in this response relate to evolved differences in stress tolerance within and between species. Unlike proteomic and epigenomic studies, there is an abundance of transcriptomic data in coral. These existing transcriptomic datasets make it theoretically possible to conduct metagenomic studies aimed at identifying core pathways. However, even with the existing transcriptomic datasets, it is still often challenging to compare across studies due to differences in experimental design, methods, and/or species studied. One solution would be a concerted community effort to generate multiomics datasets that are as comparable as possible (e.g., by using standardized heat-stress protocols) in a small number of model coral species that reflect the biodiversity on reefs. This type of resource should identify similarities and differences between model corals and will serve as a solid foundation for comparing other coral populations and species.

Network Analysis of Coral Transcriptome Data

More recently, network approaches have been developed to reduce the inherent complexity and size of transcriptomic data to enable the identification of broad patterns of gene expression [40,41]. A common approach is weighted gene coexpression network analysis (WGCNA), which quantifies the coexpression patterns among genes in a dataset and can be used to identify clusters (modules) of highly correlated genes [42]. It is also possible to identify regulatory points or hubs in the network that can be assessed with parameters of centrality. These include degree (i.e., the number of connections of one node with other nodes) and betweenness, which is the connectivity of a node between other unconnected node pairs (i.e., the capacity to act as a link), to identify hubs that may act as regulatory components of transcriptional networks [43]. The networks that result from WGCNA are often highly complex, suggesting that they are shaped primarily by natural selection. However, recent analyses suggest that evolutionary drift can also generate high complexity and convergence in transcriptional networks (e.g., mutations in short *cis*-regulatory sequences can create novel transcription factor binding sites of master regulators across divergent species) [44]. This observation suggests that coral gene coexpression networks, when compared across different species, need to be interpreted with caution because shared circuits may not solely be explained by vertical inheritance of an ancestral plan (i.e., are the result of convergence).

Rose and colleagues [45] used coexpression networks to study the response to experimental heat stress of *Acropora hyacinthus* colonies that had been transplanted between two differing reef environments. These experiments identified modules of coexpressed genes where some of which correlated strongly with the bleaching outcomes of individual colonies. Ruiz-Jones and Palumbi [46] identified modules of genes activated during environmental fluctuations that included many genes associated with the UPR and suggested that the UPR represents an important defense response

Please cite this article in press as: Cleves et al., Unknown to Known: Advancing Knowledge of Coral Gene Function, Trends in Genetics (2019), https://doi.org/10.1016/j.tig.2019.11.001

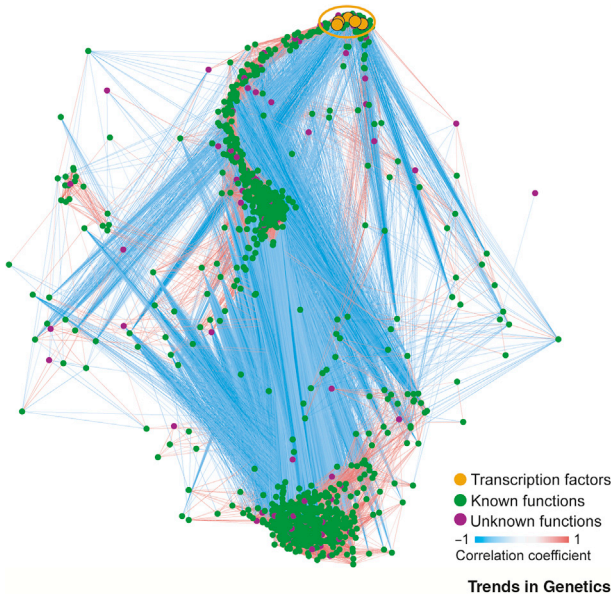

Figure 2. Weighted Gene Coexpression Network of *Montipora capitata* after 1 h of Thermal Stress.
Nodes represent genes that are connected by edges based on coexpression coefficients above the weighted gene coexpression network analysis (WGCNA) default. Edge colors indicate correlation coefficient of expression between each pair of genes (see legend). Node color indicates known (green) and unknown, dark (magenta) genes. Orange nodes represent the selected transcription factors inspected in Figure 3. Adapted from [18].

against environmental stress. Bay and Palumbi [47] monitored the survival and growth of *A. hyacinthus* colonies that were transplanted from highly variable (HV) or moderately variable (MV) reef environments and identified genes in coexpression modules that were highly correlated with survival in the HV environment and growth in the MV environment. These genes could be potential biomarkers for predicting coral survival under stress.

WGCNA analyses have also revealed several potential adaptive mechanisms; namely, transcriptional 'frontloading' [35] (i.e., the constitutive higher baseline expression of stress response genes) or gene expression plasticity. Kenkel and Matz [48] transplanted *Porites astreoides* colonies between inshore and offshore reefs and monitored gene expression and symbiont density. They noted that larger expression shifts in inshore-native corals transplanted to the offshore site than in offshore-native corals transplanted to the inshore site, and suggested that the offshore corals had less capacity for gene expression plasticity than the inshore corals because the inshore reefs experienced more frequent and higher levels of environmental stressors. The WGCNA analytical framework was recently applied to *M. capitata* in Hawaii and identified an array of transcription factors that may prove useful for downstream analysis of short-term (1 h or 6 h) thermal stress (i.e., 29.8°C compared with the control of 27.4°C; for details see [18]) response in this coral. These results are presented below in the analysis of dark gene expression patterns.

Dark Genes in Coral Gene Coexpression Networks

We previously defined dark genes as 'unknown functional genes' that are often coexpressed with genes of known function as a response to environmental (i.e., thermal) stress. Therefore, dark genes are potentially involved in environmental stress-response pathways as enzymes or regulatory factors. We refer to the stress-induced subnetworks comprising both known and dark genes as 'modules' that are identified using gene coexpression analysis. As an example of how gene coexpression networks can be used to identify both known and dark genes for genetic manipulation, we use the RNA-seq results generated by Shumaker *et al.* [18] in which *M. capitata* triplicate samples were either

Please cite this article in press as: Cleves et al., Unknown to Known: Advancing Knowledge of Coral Gene Function, Trends in Genetics (2019), https://doi.org/10.1016/j.tig.2019.11.001

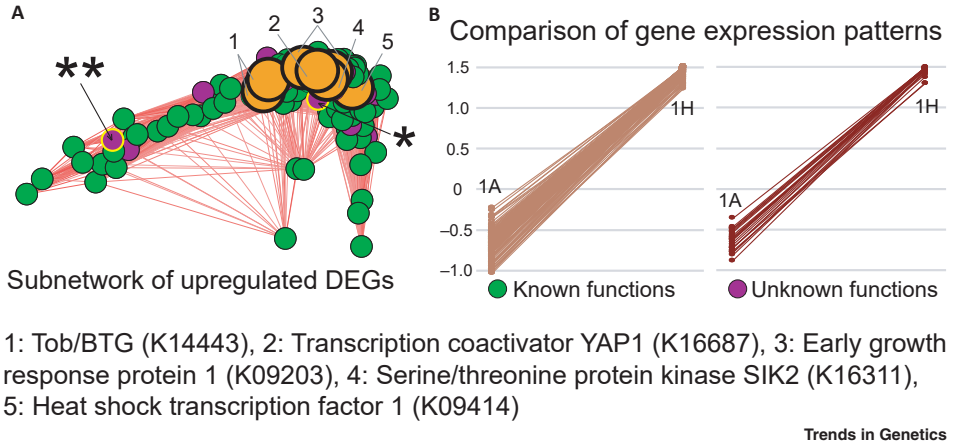

Figure 3. Subnetwork (A) of Upregulated Genes (Circles), Including Some Transcription Factors (Orange), Responding to 1 h of Thermal Stress (versus Ambient) That Are Marked with an Orange Oval in Figure 2. The expression patterns of known (green nodes) and unknown (magenta nodes) genes are summarized using normalized values [*z*-score: ('Expression counts' – 'Average expression count of all conditions in each gene')/'Standard deviation of all conditions in each gene']. The unknown, coral-specific genes in *Montipora capitata* were identified using BLASTp against the NCBI nonredundant database (*e*-value cutoff = 1*e*−05), whereby all hits were only to other coral species. Examples, with asterisks in the subnetwork, were selected for phylogenetic analysis in Figures 4 and 5. Comparison of normalized expression values (y-axis) of known and unknown genes (B) shows an equivalent pattern of coexpression across these two categories.

maintained in control ambient conditions or exposed to 1 h or 6 h of thermal stress. The 2335 significant DEGs identified in these comparisons were used to calculate Pearson correlations of gene expression (Figure 2). The coral animal genes with a significant correlation coefficient (positive or negative) with *P*-value ≤ 0.05 were used to build the network (see [18] for details). A module (marked with the orange oval in Figure 2) that was previously [18] found includes a Tob/BTG transcription factor regulator, the transcription coactivator YAP1, early growth response protein 1, the serine/threonine protein kinase SIK2, and heat shock transcription factor 1, which are all extensively upregulated after 1 h of thermal stress in *M. capitata* and in many other coral animal hosts, is shown in Figure 3. Any one of these genes is a potential target for **gene knockdown**. Specifically, an interesting animal dark gene embedded in the center of this module is marked with an asterisk in Figure 3 and analyzed using phylogenetics in Figure 4. This tree shows the ancient duplication of this coral-specific dark gene, OG.2649 [18] in the **Scleractinia** ancestor. Furthermore, several species-specific amino acid changes occur in the targeted conserved region that could confer adaptive traits in different lineages (e.g., amino acids in red text in Figure 4). These specific changes in a thermal-stress-tolerant coral species such as *M. capitata* may be regarded as the products of selection to potentially confer this important trait. We consider this hypothesis to be worth testing using functional tools. This approach to identify targets for genetic manipulation relies on natural variation within conserved gene families [49–51]. A second example of a coral animal dark gene is shown in Figure 5 that represents a more peripheral member of the transcription factor subnetwork (marked by double asterisks in Figure 3; OG.3855) with a complex history evolution among stony corals. This relatively less-conserved protein provides a target for genetic manipulation that may provide important insights into thermal stress tolerance in coral hosts.

It was shown that node degree in a coexpression network is one of the major determinants of transcriptional noise (variance). Although genes in a module may be coregulated, central genes show less expression variance across individual cells, which is likely to propagate less noise throughout the network. Central genes also tend to have more conserved functions due to strong purifying selection. Older genes tend to be central and display more deterministic expression patterns compared with more recently derived, noisy genes [52]. However, younger genes may be targets

Please cite this article in press as: Cleves et al., Unknown to Known: Advancing Knowledge of Coral Gene Function, Trends in Genetics (2019), https://doi.org/10.1016/j.tig.2019.11.001

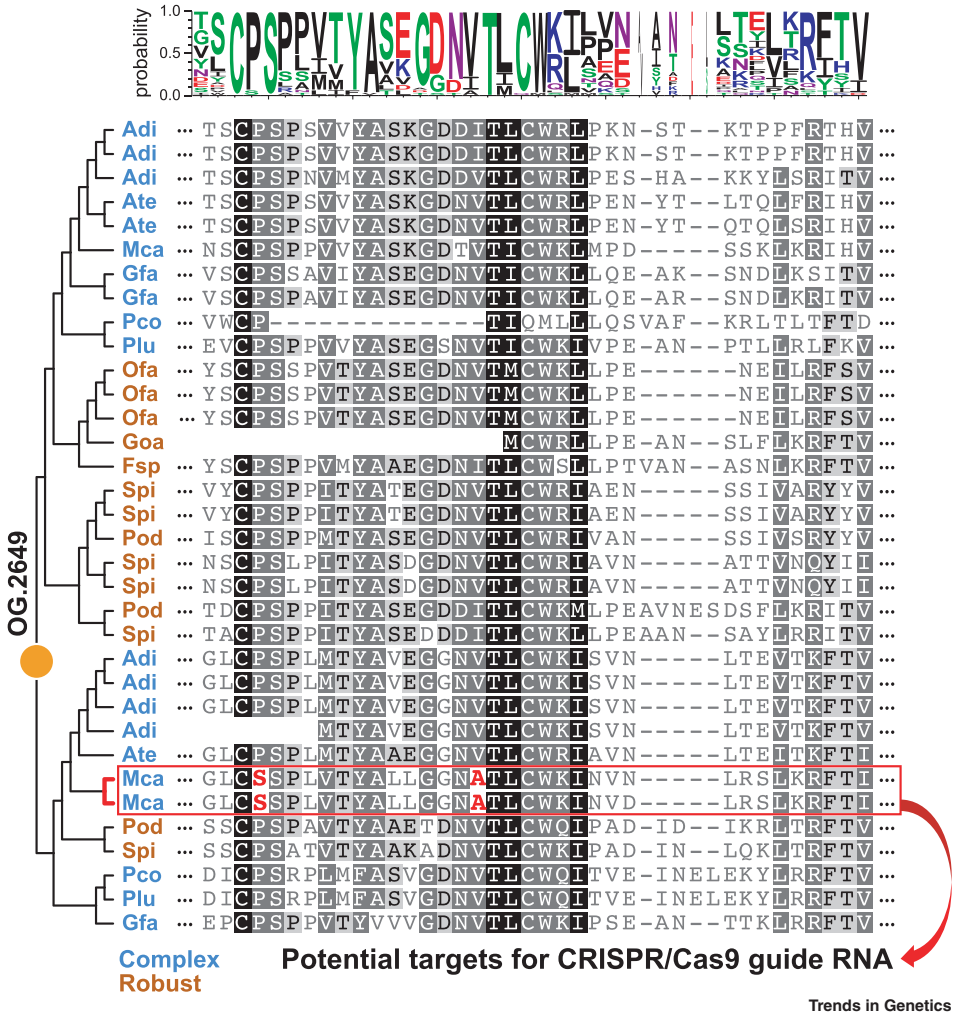

Figure 4. Phylogeny of the Anciently Duplicated (See Orange Filled Circle) Coral-Specific (i.e., Not Detected in Other Taxa), Hub Dark Gene Family OG.2649 (Encoded by the Query *Montipora capitata* Gene Model adi2mcaRNA33759_R0) Inferred Using the Maximum Likelihood Method (IQ-TREE v1.6.11 with 1000 Ultrafast Bootstrap Replicates [65]).

This gene is marked with a single asterisk in Figure 3. The tree is rooted on the branch that putatively represents the common ancestor of complex and robust corals. The partial alignment shown above the tree was generated using Geneious (v6.1.8; https://www.geneious.com) and the composition of amino acids in the alignment was drawn with WebLogo [66] with probability-based size differences and the 'Chemistry' color scheme. The coral species that are divided phylogenetically into complex (blue text) and robust (brown text) forms are as follows: Adi, *Acropora digitifera*; Ate, *Acropora tenuis*; Mca, *M. capitata*; Spi, *Stylophora pistillata*; Pod, *Pocillopora damicornis*; Pco, *Porites compressa*; Plu, *Porites lutea*; Ofa, *Orbicella faveolata*; Fsp, *Fungia* spp.; Goa, *Goniastrea aspera*; Gfa, *Galaxea fascicularis*.

for selection because their unpredictable expression patterns are responsive to fluctuating conditions, providing a form of bet-hedging [53]. In the context of prioritizing genes for functional studies, it may be important to take into account the position of these target genes within coexpression networks. For example, genes near the nodes might have more severe phenotypes when mutated due to pleiotropic effects compared with peripheral genes. In summary, it is possible not only to identify potential coral-specific targets for genome editing that are relevant to coral stress responses, but also to position dark genes within known interaction networks to aid in their functional annotation. In the next

Please cite this article in press as: Cleves et al., Unknown to Known: Advancing Knowledge of Coral Gene Function, Trends in Genetics (2019), https://doi.org/10.1016/j.tig.2019.11.001

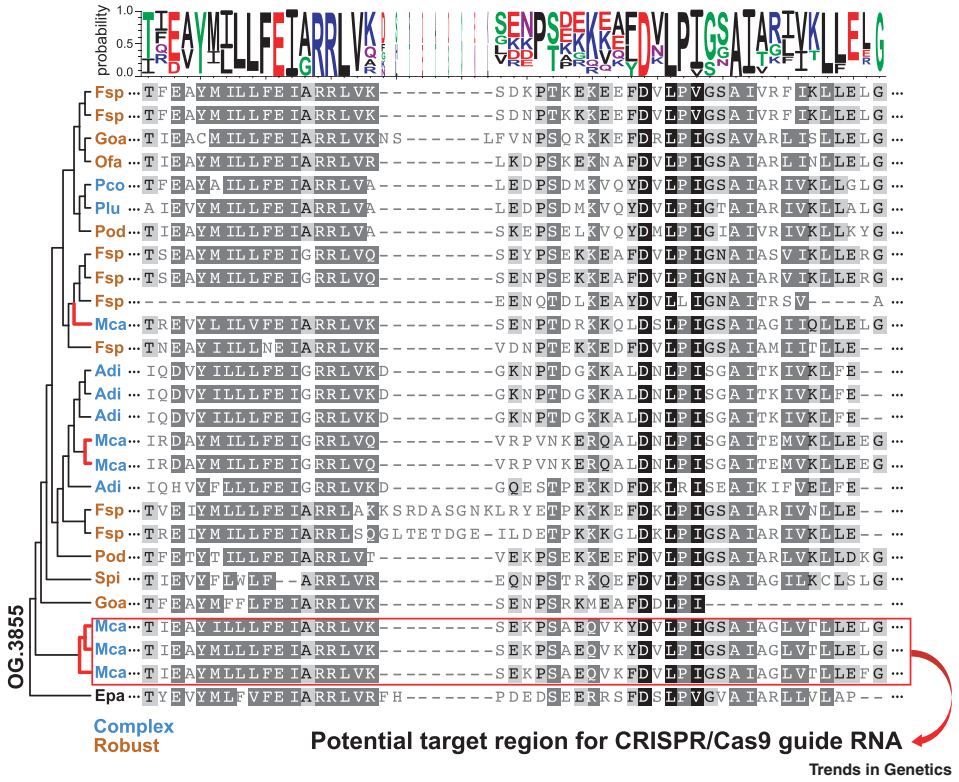

Figure 5. IQ-TREE (Built as Described in Figure 4) of the Duplicated Coral-Specific Dark Gene Family OG.3855 (Encoded by the Query *Montipora capitata* Gene Model augustus.g21614.t1) That Occupies a Peripheral Position (Marked with Two Asterisks) in the Gene Coexpression Subnetwork Shown in Figure 3.
The partial alignment shown above the tree was generated using Geneious and the composition of amino acids in the alignment was drawn with WebLogo as in Figure 4. The coral species that are divided phylogenetically into complex (blue text) and robust (brown text) forms are as follows: Adi, *Acropora digitifera*; Ate, *Acropora tenuis*; Mca, *M. capitata*; Spi, *Stylophora pistillata*; Pod, *Pocillopora damicornis*; Pco, *Porites compressa*; Plu, *Porites lutea*; Ofa, *Orbicella faveolata*; Fsp, *Fungia* spp.; Goa, *Goniastrea aspera*. The outgroup (Epa) is the sea anemone *Exaiptasia pallida*.

section, we review the current state of CRISPR/Cas9 work in corals that can be used to study the functions of both known and unknown genes and what the future holds for this important approach in coral conservation.

Reverse Genetic Tools in Corals

A deeper understanding of the molecular, cellular, and developmental bases of ecologically important aspects of coral biology, such as symbiosis and bleaching, will both help to predict the effects of climate change on wild coral populations and guide potential conservation actions. However, a major limitation toward this understanding has been the lack of genetic tools to functionally test candidate genes and pathways. Despite the difficulties of implementing such genetic tools, the first successful gene knockdown was performed using microinjection of **morpholinos** into one-cell-stage zygotes of *A. digitifera* [54]. This strategy was used to study the role of the transcription factor Brachyury during early coral embryogenesis. Although morpholinos are efficient tools to understand gene function at the early embryonic or larval stages, they are of limited use for genes acting later in development because their effects are transient due to dilution and/or degradation during sequential cell divisions [55]. Despite this limitation, this study was an important step forward in coral reverse genetics.

Please cite this article in press as: Cleves et al., Unknown to Known: Advancing Knowledge of Coral Gene Function, Trends in Genetics (2019), https://doi.org/10.1016/j.tig.2019.11.001

Box 1. Multiomics Data Integration

Given the wealth of data that already exists for a variety of coral species and the identification of a class of potential targets for genetic manipulation (e.g., UPR, heat shock factors, signaling pathways), there still exists the need to integrate these data across different biological scales of response, through the use of various omics platforms to identify specific genes of interest. In this regard, multiomics are rapidly developing bioinformatic tools that are used to integrate multiple lines of high-throughput genomic data and other types of output such as protein–protein interaction networks to explain phenotype [67,68]. A common approach is to use shared network hubs built using different omics datasets to identify genes that underlie a phenotype of interest (e.g., as recently done for mouse [69]).

Multiomics is yet to be widely applied in corals and will require extensive testing under field conditions to move beyond the generation of vast data archives [70], to provide useful data for the identification of genes and pathways of interest. Once better developed, it will be possible to integrate between gene expression and methylation data, for example, by using correlation analyses between average methylation and the quantiles of expression for coral genes [71]. Thereafter, WGCNA data can be used to correlate the separately generated gene co-methylation and coexpression network [72]. This can be done via pairwise correlations between the representative methylation patterns (i.e., eigengene vectors) of the methylation modules and eigengene vectors of the expression modules. Ultimately, it should be possible to study independently derived networks from each individual omics analysis to identify the hub (or peripheral) genes that co-occur across the different datasets. The metabolome of the coral is likely to be the most informative omics data yet challenging to interpret because, in the absence of coral animal tissue lines, the holobiont will need to be used for such an approach. The current methods will provide a 'metabolic phenotype' across different control and test conditions to assess the coral stress response.

In recent years, CRISPR/Cas9 genome editing technologies have facilitated reverse genetics in many organisms [56]. These methods are derived from naturally occurring defense mechanisms that bacteria use to protect against invading viruses [57,58]. The most commonly used CRISPR/Cas9 genome editing system uses a target-specific **single-guide RNA (sgRNA)** that directs the nuclease Cas9 to a specific site in the genome, which then creates a double-stranded break at that position [59,60]. These double-stranded breaks are often repaired using **nonhomologous end joining**, which is prone to errors and results in small insertion and deletion mutations at the target site. Importantly, unlike morpholinos, CRISPR/Cas9 can result in permanent and heritable genetic changes in the organism of interest. Recently, Cleves and colleagues generated CRISPR/Cas9-mediated mutations in several genes by microinjecting ribonucleoprotein complexes into fertilized *A. millepora* eggs [61]. Although the initial mutation frequency was low, it should be possible to improve the existing methods to generate sufficient mutations in *A. millepora* and other broadcast-spawning coral species to perform functional genetics.

The prospect of using CRISPR/Cas9 to understand the genetic bases of ecologically relevant traits is exciting. However, because there are over 25 000 genes in the typical coral genome, it is challenging to prioritize genes for functional analysis. This prioritization is particularly important due to the limited availability of coral gametes for experimentation, restricted by their natural spawning times. One potential avenue for choosing targets is to select genes with known and likely conserved functions based on studies in other organisms (e.g., known members of the UPR or heat stress response). This strategy increases the likelihood of generating mutations that result in expected phenotypes but creates a bias against the identification of unknown genes with important roles in coral biology (e.g., coral-specific regulators of the stress response or symbiosis). It is, however, difficult to anticipate the phenotypic consequences of knocking out genes with unknown functions. An alternate avenue is to use the integrated multiomics approach (Box 1) to select known or unknown genes based on their positions in coexpression networks (as described above). This method builds strong correlational support that a particular gene, agnostic to it having a specific annotation, is involved in a particular biological process. For example, prioritizing genes in the nodes of genetic networks that respond to thermal stress may increase the likelihood of identifying key genetic regulators of the stress response. Additionally, this approach will become more powerful as the integration of several types of multiomics datasets becomes possible.

Outstanding Questions

How can we best identify and overcome the challenges of working with a holobiont comprising a cnidarian host, a photosynthetic dinoflagellate, and a prokaryotic microbiome? It is clear that coral resilience and stress responses rely on interactions that involve all of these biotic levels.

Can epigenetic modifications be a substrate for coral evolution and adaptation to climate change?

What types of genes provide the best targets for the elucidation of coral biology? Node genes in coexpression networks that act as master genetic regulators might have negative pleiotropic effects if mutated. However, more downstream genes may have less dramatic fitness advantages but may have few such effects and be more transcriptionally noisy. Therefore, it is important to design each experiment based on extensive gene coexpression data as well as knowledge of the evolutionary history and level of target-gene-sequence constraint among stony corals.

Can we generate coral cell lines and/or control coral spawning timing to increase the opportunities for functional genetics experiments? Early attempts to generate long-lived coral cell lines have been thwarted; however, the potential for 'immortal' cells in cnidarians exists for *Hydra*. Reports of controlled coral spawning have come from technically sophisticated aquaria, but the multimonth generation time of gametes remains an impediment to establishing continuously and regularly spawning coral.

How does the coral stress response change during exposure to elevated temperatures? Are there early, late, and/or constitutively expressed genes involved in the stress response? The existence of a time-series-dependent response to stress remains unclear in corals, as does knowledge about how it is regulated at the transcriptional and epigenetic levels. These data

Please cite this article in press as: Cleves et al., Unknown to Known: Advancing Knowledge of Coral Gene Function, Trends in Genetics (2019), https://doi.org/10.1016/j.tig.2019.11.001

Looking toward the future of reverse genetics in corals, it will be important to apply the state of the art to coral biology as current genome editing methods are expanded and improved. Therefore, upcoming efforts should focus on using microinjection or other delivery methods to help usher a new era of genetically tractable coral research.

would be valuable in choosing target genes for genetic manipulation.

What are the genetic and/or chemical signals for bleaching? Can we modulate these signals artificially (genetically or chemically) in coral tissues independent of external stress? This approach will help us uncover the molecular and cell biological basis of bleaching.

Concluding Remarks

Accompanying increasing genomic resources for reef-forming corals [4,5,18] and calls for interventional strategies for their conservation and preservation [3,62–64] is an increased need for studies integrating multiple lines of experimental evidence to better elucidate the mechanisms underlying coral resilience. The combination of multiomics and reverse genetic approaches, challenges notwithstanding, may be best suited to uncovering these mechanisms, although many open questions remain to be answered (see Outstanding Questions).

A wealth of comparative genomic studies, stress-based transcriptomic studies, and analyses of coral proteomes and epigenomics provide a list of genes involved in the stress response, development, the immune response, and biomineralization that will require further study. As described above, multiomics approaches have and will aid in our understanding of the evolution, function, and regulation of the coral gene inventory, although the lack of robust genetic systems for functional studies has hitherto been a confounding factor. Recent efforts to implement such tools in corals are encouraging. However, a major limitation of coral genetics is the 3–7 years required for offspring to reach sexual maturity. This long generation time makes rearing CRISPR/Cas9-induced mutant coral lines challenging. Although it should be possible to conduct such experiments, current research should focus on tractable phenotypes that are apparent during the early larval or polyp stages. Despite this major technical issue, there are many important aspects of coral biology (e.g., symbiont acquisition, stress tolerance, biomineralization) that can be studied without crosses if the induced mutation rate is sufficiently high to generate completely mutant animals. Furthermore, introducing genetically manipulated (i.e., transgenic) corals into the environment is not likely to be a realistic solution for conservation of the massive populations present worldwide. Rather, these emerging tools are better suited to the discovery of gene functions. These data can provide the basis for more realistic countermeasures against coral bleaching in marine environments. Despite these major issues, coral research must move forward to provide tools, even when using less powerful correlational data, that promote reef preservation. Here we have discussed a variety of omics approaches and some interesting genes (of both known and unknown functions) and pathways that may prove to be useful targets for genetic manipulation. Our approach highlights the integration of transcriptomic data and protein divergence to facilitate gene choice for CRISPR/Cas9 gene editing technology.

Acknowledgments

The authors acknowledge the input of Arik Harel regarding the WGCNA analysis. D.B. and H.M.P. were supported by NSF-OCE 1756616 and NSF-OCE 1756623. P.A.C. was supported by a LIFE#336932 grant from the Simons Foundation to John R. Pringle.

References

1. Costanza, R. *et al.* (2014) Changes in the global value of ecosystem services. *Global Environ. Chang.* 26, 152–158
2. Costanza, R. *et al.* (1997) The value of the world's ecosystem services and natural capital. *Nature* 387, 253–260
3. National Academies of Sciences, Engineering, and Medicine (2019) *A Research Review of Interventions to Increase the Persistence and Resilience of Coral Reefs*, National Academies Press
4. Voolstra, C.R. *et al.* (2015) The ReFuGe 2020 consortium – using 'omics' approaches to explore the adaptability and resilience of coral holobionts to environmental change. *Front. Mar. Sci.* 2, 68
5. Bhattacharya, D. *et al.* (2016) Comparative genomics explains the evolutionary success of reef-forming corals. *eLife* 5, e13288
6. Dixon, G.B. *et al.* (2014) Bimodal signatures of germline methylation are linked with gene expression plasticity in the coral *Acropora millepora*. *BMC Genomics* 15, 1109
7. Hou, J. *et al.* (2018) RNA-seq reveals extensive transcriptional response to heat stress in the stony coral. *Front. Genet.* 9, 37
8. Jin, Y.K. *et al.* (2016) Genetic markers for antioxidant capacity in a reef-building coral. *Sci. Adv.* 2, e1500842

Please cite this article in press as: Cleves et al., Unknown to Known: Advancing Knowledge of Coral Gene Function, Trends in Genetics (2019), https://doi.org/10.1016/j.tig.2019.11.001

9. Liew, Y.J. *et al.* (2018) Epigenome-associated phenotypic acclimatization to ocean acidification in a reef-building coral. *Sci. Adv.* 4, eaar8028
10. Palumbi, S.R. *et al.* (2014) Mechanisms of reef coral resistance to future climate change. *Science* 344, 895–898
11. Rose, N.H. *et al.* (2018) Polygenic evolution drives species divergence and climate adaptation in corals. *Evolution* 72, 82–94
12. Baumgarten, S. *et al.* (2015) The genome of *Aiptasia*, a sea anemone model for coral symbiosis. *Proc. Natl Acad. Sci. U. S. A.* 112, 11893–11898
13. Oakley, C.A. *et al.* (2017) Thermal shock induces host proteostasis disruption and endoplasmic reticulum stress in the model symbiotic cnidarian *Aiptasia*. *J. Proteome Res.* 16, 2121–2134
14. Lehnert, E.M. *et al.* (2014) Extensive differences in gene expression between symbiotic and aposymbiotic cnidarians. *G3 (Bethesda)* 4, 277–295
15. Cui, G. *et al.* (2019) Host-dependent nitrogen recycling as a mechanism of symbiont control in *Aiptasia*. *PLoS Genet.* 15, e1008189
16. Cziesielski, M.J. *et al.* (2018) Multi-omics analysis of thermal stress response in a zooxanthellate cnidarian reveals the importance of associating with thermotolerant symbionts. *Proc. Biol. Sci.* 285, 1877
17. Stephens, T.G. *et al.* (2018) Core genes in diverse dinoflagellate lineages include a wealth of conserved dark genes with unknown functions. *Sci. Rep.* 8, 17175
18. Shumaker, A. *et al.* (2019) Genome analysis of the rice coral *Montipora capitata*. *Sci. Rep.* 9, 2571
19. Weston, A.J. *et al.* (2015) Proteomics links the redox state to calcium signaling during bleaching of the scleractinian coral *Acropora microphthalma* on exposure to high solar irradiance and thermal stress. *Mol. Cell. Proteomics* 14, 585–595
20. Bieri, T. *et al.* (2016) Relative contributions of various cellular mechanisms to loss of algae during cnidarian bleaching. *PLoS One* 11, e0152693
21. Ricaurte, M. *et al.* (2016) Proteomic analysis of bleached and unbleached *Acropora palmata*, a threatened coral species of the Caribbean. *Mar. Pollut. Bull.* 107, 224–232
22. Chen, Z. *et al.* (2017) Epigenetic regulation: a new frontier for biomedical engineers. *Annu. Rev. Biomed. Eng.* 19, 195–219
23. Flores, K.B. *et al.* (2013) The role of methylation of DNA in environmental adaptation. *Integr. Comp. Biol.* 53, 359–372
24. Eirin-Lopez, J.M. and Putnam, H.M. (2019) Marine environmental epigenetics. *Annu. Rev. Mar. Sci.* 11, 335–368
25. Durante, M.K. *et al.* (2019) What drives phenotypic divergence among coral clonemates of *Acropora palmata*? *Mol. Ecol.* 28, 3208–3224
26. Xu, X. and Qi, L.S. (2019) A CRISPR–dCas toolbox for genetic engineering and synthetic biology. *J. Mol. Biol.* 431, 34–47
27. Putnam, H.M. *et al.* (2016) Ocean acidification influences host DNA methylation and phenotypic plasticity in environmentally susceptible corals. *Evol. Appl.* 9, 1165–1178
28. Liew, Y.J. *et al.* (2018) Intergenerational epigenetic inheritance in reef-building corals. *bioRxiv*. .Published online February 21, 2018. https://doi.org/10.1101/269076
29. Putnam, H.M. and Gates, R.D. (2015) Preconditioning in the reef-building coral *Pocillopora damicornis* and the potential for trans-generational acclimatization in coral larvae under future climate change conditions. *J. Exp. Biol.* 218, 2365–2372
30. Putnam, H.M. *et al.* (2018) Nurtured by nature: considering the role of environmental and parental legacies in coral ecological performance. *bioRxiv*. .Published online May 8, 2018. https://doi.org/10.1101/317453
31. Torda, G. *et al.* (2017) Rapid adaptive responses to climate change in corals. *Nat. Clim. Chang.* 7, 627–636
32. Liu, Q. *et al.* (2019) Detection of DNA base modifications by deep recurrent neural network on Oxford Nanopore sequencing data. *Nat. Commun.* 10, 2449
33. Cziesielski, M.J. *et al.* (2019) The past, present, and future of coral heat stress studies. *Ecol. Evol.* 9, 10055–10066
34. Seneca, F.O. and Palumbi, S.R. (2015) The role of transcriptome resilience in resistance of corals to bleaching. *Mol. Ecol.* 24, 1467–1484
35. Barshis, D.J. *et al.* (2013) Genomic basis for coral resilience to climate change. *Proc. Natl Acad. Sci. U. S. A.* 110, 1387–1392
36. Hrdlickova, R. *et al.* (2017) RNA-seq methods for transcriptome analysis. *Wiley Interdiscip. Rev. RNA* 8, e1374
37. Zhang, Y. *et al.* (2019) Development of a transcriptomic database for 14 species of scleractinian corals. *BMC Genomics* 20, 387
38. Traylor-Knowles, N. *et al.* (2017) Early transcriptional responses during heat stress in the coral *Acropora hyacinthus*. *Biol. Bull.* 232, 91–100
39. Oakley, C. and Davy, S. (2018) Cell biology of coral bleaching. In *Coral Bleaching* (van Oppen, M.J.H. and Lough, J.M. eds), pp. 189–211, Springer
40. Harel, A. *et al.* (2015) Deciphering primordial cyanobacterial genome functions from protein network analysis. *Curr. Biol.* 25, 628–634
41. Harel, A. *et al.* (2014) Evolutionary history of redox metal-binding domains across the tree of life. *Proc. Natl Acad. Sci. U. S. A.* 111, 7042–7047
42. Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559
43. Fabina, N.S. *et al.* (2013) Symbiotic specificity, association patterns, and function determine community responses to global changes: defining critical research areas for coral–*Symbiodinium* symbioses. *Glob. Change Biol.* 19, 3306–3316
44. Sorrells, T.R. and Johnson, A.D. (2015) Making sense of transcription networks. *Cell* 161, 714–723
45. Rose, N.H. *et al.* (2015) Gene networks in the wild: identifying transcriptional modules that mediate coral resistance to experimental heat stress. *Genome Biol. Evol.* 8, 243–252
46. Ruiz-Jones, L.J. and Palumbi, S.R. (2017) Tidal heat pulses on a reef trigger a fine-tuned transcriptional response in corals to maintain homeostasis. *Sci. Adv.* 3, e1601298
47. Bay, R.A. and Palumbi, S.R. (2017) Transcriptome predictors of coral survival and growth in a highly variable environment. *Ecol. Evol.* 7, 4794–4803
48. Kenkel, C.D. and Matz, M.V. (2016) Gene expression plasticity as a mechanism of coral adaptation to a variable environment. *Nat. Ecol. Evol.* 1, 14
49. King, D.G. and Kashi, Y. (2007) Mutation rate variation in eukaryotes: evolutionary implications of site-specific mechanisms. *Nat. Rev. Genet.* 8, 902
50. Wienert, B. *et al.* (2015) Editing the genome to introduce a beneficial naturally occurring mutation associated with increased fetal globin. *Nat. Commun.* 6, 7085
51. Moore, J.C. *et al.* (2018) "Site and mutation"-specific predictions enable minimal directed evolution libraries. *ACS Synth. Biol.* 7, 1730–1741
52. Barroso, G.V. *et al.* (2018) The evolution of gene-specific transcriptional noise is driven by selection at the pathway level. *Genetics* 20, 173–189

Please cite this article in press as: Cleves et al., Unknown to Known: Advancing Knowledge of Coral Gene Function, Trends in Genetics (2019), https://doi.org/10.1016/j.tig.2019.11.001

53. Thattai, M. and van Oudenaarden, A. (2004) Stochastic gene expression in fluctuating environments. *Genetics* 167, 523–530
54. Yasuoka, Y. *et al.* (2016) The mesoderm-forming gene *brachyury* regulates ectoderm–endoderm demarcation in the coral *Acropora digitifera*. *Curr. Biol.* 26, 2885–2892
55. Blum, M. *et al.* (2015) Morpholinos: antisense and sensibility. *Dev. Cell* 35, 145–149
56. Doudna, J.A. and Charpentier, E. (2014) The new frontier of genome engineering with CRISPR–Cas9. *Science* 346, 1258096
57. Bhaya, D. *et al.* (2011) CRISPR–Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.* 45, 273–297
58. Wang, H. *et al.* (2016) CRISPR/Cas9 in genome editing and beyond. *Annu. Rev. Biochem.* 85, 227–264
59. Jinek, M. *et al.* (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821
60. Jinek, M. *et al.* (2013) RNA-programmed genome editing in human cells. *eLife* 2, e00471
61. Cleves, P.A. *et al.* (2018) CRISPR/Cas9-mediated genome editing in a reef-building coral. *Proc. Natl Acad. Sci. U. S. A.* 115, 5235–5240
62. van Oppen, M.J. *et al.* (2015) Building coral reef resilience through assisted evolution. *Proc. Natl Acad. Sci. U. S. A.* 112, 2307–2313
63. van Oppen, M.J.H. *et al.* (2017) Shifting paradigms in restoration of the world's coral reefs. *Glob. Change Biol.* 23, 3437–3448
64. Anthony, K. *et al.* (2017) New interventions are needed to save coral reefs. *Nat. Ecol. Evol.* 1, 1420–1422
65. Nguyen, L.T. *et al.* (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274
66. Crooks, G.E. *et al.* (2004) WebLogo: a sequence logo generator. *Genome Res* 14, 1188–1190
67. Huang, S. *et al.* (2017) More is better: recent progress in multi-omics data integration methods. *Front. Genet.* 8, 84
68. Kim, M. and Tagkopoulos, I. (2018) Data integration and predictive modeling methods for multi-omics datasets. *Mol. Omics* 14, 8–25
69. Li, H. *et al.* (2018) An integrated systems genetics and omics toolkit to probe gene function. *Cell Syst.* 6, 90–102
70. Vilanova, C. and Porcar, M. (2016) Are multi-omics enough? *Nat. Microbiol.* 1, 16101
71. Hu, Y. *et al.* (2016) Simultaneous profiling of transcriptome and DNA methylome from a single cell. *Genome Biol.* 17, 88
72. Horvath, S. *et al.* (2012) Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol.* 13, R97