Genomics Are Transforming Our Understanding of Responses to Climate Change

JONATHON H. STILLMAN AND ERIC ARMSTRONG

Downloaded from https://academic.oup.com/bioscience/article/65/3/237/236654 by guest on 20 August 2022

Genomic approaches are revolutionizing the biomedical sciences, providing new ways in which to approach the development of therapeutic practices and the understanding of the causal bases of disease. So, too, do genomics approaches have the potential to transform the way in which ecological studies can be conducted by providing powerful tools in which genotype and phenotype data of diverse "wild" populations can be measured with a never-before-possible level of fine-scale resolution. In this article, we examine the potential for genomics-enabled approaches to aid in understanding the responses of populations to climate change and highlight examples in which these tools have been applied to understand physiological adaptation or migration.

Keywords: genomics, ecology, adaptation, migration, population

There are three potential outcomes of population responses to climate change: The population can move (Chen et al. 2011, Pinsky et al. 2013), persist (via local adaptation—Pespeni et al. 2013b—or phenotypic plasticity; Evans and Hofmann 2012), or die (Franks and Hoffmann 2012, Savolainen et al. 2013). A central question in ecological, evolutionary, and organismal biology is which of these outcomes is happening (Kultz et al. 2013) in order to gain predictive understanding of the potential for responses to climate change (Sunday et al. 2014). Although the third outcome in response to climate change—dying—is relatively easy to measure using simple field surveys of abundance, the first two options—moving and adapting—require information about the genotype or phenotype of organisms within study populations (Stillman 2003, Somero 2011, Franks and Hoffmann 2012). DNA sequence data allow scientists to characterize these responses to climate change.

We are in the middle of a revolution in DNA sequencing technologies that are making DNA sequence data available faster and at lower cost than ever before. These technologies are collectively called *next-generation DNA sequencing* (NGS), as is described below. Next-generation sequencing is transforming the way in which biologists generate data to study the movement and adaptation of organisms in natural populations (Evans and Hofmann 2012, Whitehead 2012, Des Marais et al. 2013, Shaffer and Purugganan 2013). In this article, we briefly describe NGS, summarize recent work that has leveraged NGS in making inferences about how populations have responded to climate change, and discuss how additional resources may facilitate this approach more broadly.

What is NGS? *Next-generation sequencing* refers to a number of different technological advancements in DNA sequencing that have been made during the past decade in which enormous numbers of short pieces of DNA are sequenced to determine their order of nucleotides—specifically, adenine, thymine, guanine, or cytosine. Relative to the former generation of Sanger sequencing (Sanger and Coulson 1975), NGS advancements have been made by combining the chemical reactions and detection operations involved with DNA sequencing and by doing this for many DNA molecules simultaneously on a massively parallel scale. There are several leading technologies used to generate millions to billions of sequences that range from 50 nucleotides in length to 1000 nucleotides in length. Here, we describe some of the currently popular sequencing technologies, although these are likely to change with time, because this is a rapidly changing area of biotechnology.

The sequencing platform that currently generates the largest quantity of DNA sequence data is Illumina sequencing (*www.illumina.com*). Illumina generates sequences of nucleotides that are from 36 to150 base pairs (bp) long. A collection of DNA molecules is fragmented into short pieces,

BioScience 65: 237–246. © The Author(s) 2015. Published by Oxford University Press on behalf of the American Institute of Biological Sciences. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com.
doi:10.1093/biosci/biu219 Advance Access publication 21 January 2015

and oligonucleotide (short polymers of nucleotides typically less than 50 bp long) adapters are ligated to the ends. Those oligonucleotides can contain barcode nucleotide sequences that allow DNA molecules from different samples to be mixed together on one sequencing lane of the instrument. The DNA molecules are spread across the lane where they stick to a specific location. Then they are amplified using the polymerase chain reaction so that there are many copies of each DNA molecule at each site on the solid surface. Fluorescently labeled nucleotides and DNA polymerase are flowed across the sequencing lane in cycles, and an image of the sequencing lane is produced after each cycle. Those images allow the last nucleotide added to each sequencing location to be noted, and the DNA sequence is therefore generated by synthesizing the molecule one nucleotide at a time. This type of DNA sequencing is known as *sequencing by synthesis*. Presently, the Illumina sequencers can generate as much as 600 giga-base pairs (Gb) in 11 days, using the HiSeq2500 platform, although if far fewer data are required, the run time can be as short as 4 hours on the Illumina MiSeq platform. The cost per million base pairs is just pennies, although the instruments, themselves, are very expensive and are typically operated by a commercial or an academic genomics core facility.

Other sequencing by synthesis platforms include the Roche 454 Life Sciences method (*www.454.com*), which, like Illumina, uses fluorescently labeled nucleotides but has the ability to generate DNA fragments up to 1000 bp (although only 1 million reads—1 Gb total—are produced per run of about 24 hours), and the Life Technologies Ion Torrent (*www.lifetechnologies.com/us/en/home/brands/ion-torrent.html*), which, unlike Illumina or 454, uses a pH detector that senses the protons released during the addition of nucleotides to each growing DNA molecule and can generate up to 5 million 400-bp sequences (2 Gb total) in less than 8 hours. The greater sequence length produced by the 454 platform is useful for sequencing whole genomes, whereas the Ion Torrent platform dramatically reduces the costs of reagents, because no expensive fluorescently labeled nucleotides are required.

Next-generation sequencing approaches allow a wide range of applications involving analysis of the genome (the DNA sequence), the transcriptome (the DNA sequence of expressed genes), and the epigenome (DNA sequences that have been chemically modified). Next-generation sequencing can be used for sequencing genomes for the first time (*de novo*) to provide the starting reference material for analyses of the transcriptome or the epigenome. Although re-sequencing genomes to identify genomic regions where mutations persist across populations of individual is within the capability of the NGS sequencing technology, often methods are used to reduce the genomic regions being analyzed to those that are specifically targeted. For example, sequencing restriction-site associated DNA markers (RAD-seq) is often used in ecological studies to identify new single-nucleotide polymorphisms (SNPs) useful for population genetics (Miller et al. 2007). Amplicon sequencing can be used to sequence just a few genomic regions across many individuals that are selectively isolated through the polymerase chain reaction. Sequencing the expressed genes (the transcriptome) is used to generate a snapshot of what genes were being activated or repressed at the time the sample was taken (messenger RNA-sequencing or RNA-seq), and has great utility for making physiological inferences. Identifying epigenetic regulation by NGS includes sequencing the genes that are methylated or wrapped around histones that were specifically modified (e.g., acetylated) to activate or repress the expression of those genes (the epigenome). Next-generation sequencing can be used to find out which genes are regulated by specific transcription factors or other proteins through immunoprecipitation of the bound regions of DNA (chromatin immunoprecipitation-sequencing; Johnson et al. 2007). These and other approaches are making NGS technologies increasingly important in ecological studies.

The analysis of NGS data involves overcoming computational challenges involved in the correct analysis of the sequence data. For example, the billions of sequences generated in an Illumina sequencing run must be aligned with each other; a process that is challenging at the level of algorithm development and computational power. Incorrect alignments can lead to errors in estimating gene expression, genetic diversity, and poorly finished genomes, so improving sequencing alignment is an important aspect of bioinformatics tool development. Sequence homology—comparing the DNA sequence of one organism with that of another, whether it be from the same species or a different species—is important for the identification of what genes are represented by the DNA sequence. Functional interpretation of those genes represents a third challenge: Many expressed gene sequences encode proteins that have no known function or for which functional roles may be diverse across organisms. Despite these challenges, great advancements in ecology and functional biology are being made using NGS data.

Generating samples for sequencing is possible in just about any standard molecular biology laboratory using starting materials that can include fresh, frozen, ethanol preserved, and archival museum specimens (Bi et al. 2013). Next-generation sequencing has been used to analyze ancient DNA from specimens found at least partially preserved, such as wooly mammoths buried in the ice (Miller et al. Nature 2008) or Neanderthals from caves (Mednikova et al. 2011). Next-generation sequencing approaches can make use of the latter archived or ancient samples, in which DNA is partially to wholly degraded because DNA fragmentation is required in the short read technologies, whereas these samples are far less useful for conventional DNA approaches (e.g., cloning into bacterial plasmid DNA), in which longer DNA molecules are typically required.

Although NGS remains costly, the price per base sequenced is far less than that of any traditional approach.

Downloaded from https://academic.oup.com/bioscience/article/65/3/237/236654 by guest on 20 August 2022

More important, the time required for researchers to generate enormous quantities of data is dramatically reduced. The savings realized by lowered cost per base pair and salary per data point have made it more appealing for many scientists to adopt an NGS approach than a traditional cloning approach, even if they are only interested in the sequence of one or a few genes; the unneeded data generated in the process can be saved for other questions.

A present challenge for making biological inferences from NGS data centers around the informatics strategies for processing the data to link the short sequences together into larger contiguous sequences—so called *contigs* (Staden 1979). The ultimate goal is to have those contigs represent the entire gene sequence in the original genome. That is achieved by linking contigs together into a piece large enough (a so called *scaffold*) to represent a whole chromosome, which are the units of DNA in cells. Doing so has required new computational approaches in a field known as *bioinformatics*, and the development of computational resources capable of processing that information in a reasonable time frame. Solutions to these informatics challenges are presently opening the doors to allow NGS approaches to be taken using any organism, although there are still greater resources for model organisms for which genome sequences are available (e.g., human, mouse, nematode worm, fruit fly).

Next-generation sequencing is transforming all fields of biology. In particular, the biomedical fields are using NGS to understand diseases and disorders and to develop personalized medical approaches by which therapeutic strategies can be tailored to patients in a way that will be most effective. Next-generation sequencing is transforming our ability to make evolutionary inferences of the diversity of life and the origins of specific types of tissues or cellular features (e.g., nervous systems, mitochondria). Ecological studies of how climate change is affecting populations are also taking advantage of NGS approaches, enabling a finer-scale resolution in shifts of population structure (Savolainen et al. 2013) and physiology (De La Torre et al. 2014) than ever before. In doing so, ecological genomics has great potential to enhance our understanding of local adaptation in response to climate change (Savolainen et al. 2013). In this article, we examine specific examples of how NGS has changed ecological studies of climate change responses.

NGS data help scientists measure what cannot be seen or studied using conventional methods. Microbial diversity in environmental samples can be of paramount importance in determining how nutrients are cycled in the oceans (Swan et al. 2013), in lakes and rivers (Song et al. 2014), or in soils (Delmont et al. 2012). For example, the characterization of microbial diversity led to the discovery of the formerly unknown anaerobic ammonia oxidation (*anammox*) portion of the nitrogen cycle in marine environments (Strous et al. 1999). Understanding that diversity through traditional methods of culture and isolation of microbes was dramatically underestimating diversity because many of the microbes remain unculturable (Epstein 2013, Swan et al. 2013). The application of genomics approaches in environmental samples has revolutionized our ability to understand the genetic and functional diversity of microbial communities (Rappe and Giovannoni 2003, Amir et al. 2013, Swan et al. 2013) and has led to important new inferences in understanding ecosystem responses to climate change (Delmont et al. 2012, Sharp and Ritchie 2012).

Next-generation sequencing allows researchers to characterize the populations that they are studying to a higher resolution than ever before possible (De Wit et al. 2012). Although genetic markers, specific loci that have variable DNA sequences within and among populations, have long been used to characterize the genetic differentiation of individuals within and between locations, those markers have become more specific and numerous with NGS approaches. For example, genomic scale data can identify thousands of SNPs in conserved regions of the genome. Single-nucleotide polymorphisms are specific locations in the genome (loci) at which one nucleotide is substituted for a different nucleotide (e.g., adenine to cytosine). For example, a study of two species of spruce tree—*Picea glauca* and *Picea engelmannii*, as well as their hybrid crosses—in a hybrid zone used hundreds to thousands of SNPs identified in genes related to putative selective pressures (e.g., cold hardiness, insect herbivory resistance, growth and phenology) to demonstrate the physiological adaptation to climate variation among individuals of either species or hybrids (De La Torre et al. 2014). Similarly, techniques classically used to screen for genetic differences among individuals within a population using restriction enzymes (e.g., amplified fragment length polymorphisms) that were traditionally performed on just a handful of genes can now be performed on a genome-wide basis (e.g., RAD-seq; see Savolainen et al. 2013 for a review).

Genomic approaches can also help us to understand the mechanisms by which adaptation to climate change occurs. For example, through inherited gene regulation differences by epigenetic mechanisms such as DNA methylation or histone modification (Bossdorf et al. 2008, Franks and Hoffmann 2012, Brautigam et al. 2013, Yakovlev et al. 2014) or through the fixation of specific alleles during adaptive shifts (Hohenlohe et al. 2010).

How can NGS help to understand how populations move in response to climate change?

One possible consequence of climate change is that individuals move into locations from which they were previously excluded because of physical factors, such as cold temperatures, or because the effects of those physical factors decreased the abundance of a competitive dominant species. Range shifts of this type have been demonstrated in many ecosystems using phylogeographic methods. For example, in the Indo-Pacific Coral Triangle, mitochondrial DNA cytochrome oxidase I haploid genotype mapping in three species of giant clams (genus *Tridacna*) has provided evidence of multiple range expansions after the most recent Pleistocene

Downloaded from https://academic.oup.com/bioscience/article/65/3/237/236654 by guest on 20 August 2022

low stand in sea level (Kochzius and Nuryanto 2008, DeBoer et al. 2014). Concordant phylogeographic patterns have been exhibited in a wide variety of other marine species in this region, including the benthic false clownfish (*Amphiprion ocellaris*; Timm and Kochzius 2008), pelagic tuna and mackerel (Jackson et al. 2014), and a host of benthic invertebrates (Carpenter et al. 2011). The similarities between such ecologically diverse taxa provide strong evidence that rising sea level after the maximum Pleistocene low stand resulted in ecosystem-wide population movements and played a strong role in structuring genetic diversity throughout the Coral Triangle (DeBoer et al. 2014). Climate-induced range shifts have also been demonstrated in a host of Australasian bird lineages (Dolman and Joseph 2012); Caribbean and Hawaiian reef communities (Rocha et al. 2007); and numerous terrestrial species, with salient examples being the postglacial northward expansion of many temperature European species (ranging from angiosperms to insects to mammals) from Mediterranean refugia (Hewitt 2000).

A benefit of using high-throughput genotyping techniques is that they can allow for sequencing of previously poor-quality historical samples, which greatly expands the scope of data available to researchers (Morin and McCarthy 2007, Bi et al. 2013). For example, exome capture, a process by which the sequences of coding sequences are specifically isolated from genomic DNA extracts (Choi et al. 2009), was used to sequence approximately 11,000 genes from early twentieth-century museum skins to test for changes in genomic diversity accompanying a climate-related range retraction in alpine chipmunks (*Tamias alpinus*) in the high Sierra Nevada mountains of California (Bi et al. 2013). Although NGS methods made it possible to use the museum skin archives, a primary benefit of NGS approaches in addressing questions of biogeographical range shifts is their ability to greatly increase the number of loci that can be sampled (Helyar et al. 2011), as well as improving throughput for increasing sample size. The difference in sampling scale between traditional multilocus and genomics-based phylogeography can be immense, with the former using tens of loci as compared to tens of thousands in the latter (Brito and Edwards 2009). However, the most practical application of NGS approaches in phylogeography may be when there is discordance in distribution patterns revealed from studies using single-loci approaches, and the use of NGS as a standard approach may not be needed unless discordance is observed in single-locus studies (Bowen et al. 2014). However, the ease of generating NGS data may make the effort involved in single-locus studies greater if discordant data require subsequent NGS application, and is free of the challenges associated with choosing the appropriate divergence within the single-locus marker chosen. Because NGS-based sampling guarantees the best chance of revealing a set of markers that have the most appropriate divergence time scales for the phylogeographic problem, adopting an NGS-sampling scheme may be ultimately beneficial, even if a single locus is sampled from the NGS data for final analyses. In other words, NGS data are the most likely to contain the right locus for the question being asked, whereas in a single locus approach, there is no guarantee that the a priori information regarding the locus of choice holds true for the next set of samples.

Next-generation sequencing approaches can also provide information on which loci within a population are likely to be under selective pressure as range boundaries shift (e.g., comparative genomics, quantitative trait loci [QTL] mapping). Most phylogeographic studies to date have used putative selectively neutral markers (e.g., mitochondrial DNA; for a debate regarding neutrality, see Galtier et al. 2009), which do not provide insight into the distribution of adaptive sequence variation within populations (Stinchcombe and Hoekstra 2008, Lexer et al. 2013). Next-generation sequencing technologies offer researchers the ability to recognize loci that are subject to selection and therefore address questions of adaptive differentiation between locally adapted populations (Lexer et al. 2013). For example, a retrospective population genomics analysis of Atlantic cod (*Gadus morhua*) population structure off the coast of Greenland used more than 900 SNPs in 847 individuals to search for signs of divergent selection over a 78-year period (Therkildsen et al. 2013). These researchers identified four genetically distinct populations that exhibited increased sequence divergence in several genomic regions that were consistent with differential local adaption between populations (Therkildsen et al. 2013). These results highlight the potential of SNP-based NGS approaches for identifying genomic regions that are likely to be under selective pressure as a result of climate change.

How can NGS help to understand how populations persist in response to climate change?

There are two potential ways in which populations persist in response to climate change: local adaptation, whereby specific genotypes are favored as climate conditions change, or phenotypic plasticity, whereby existing genetic diversity can produce new phenotypes that are equally fit in the changed environment. Next-generation sequencing approaches have great utility in demonstrating persistence of populations during climate change and in distinguishing between these two potential modes of persistence. For example, in local adaptation, we expect to see cryptic genetic diversity accompany phenotypic diversity, whereas in phenotypic plasticity, a single genotype gives rise to observed phenotypic diversity.

Local adaptation. Adaptation to climate change resulting in shifted reaction norms to environmental drivers—for example, to temperature—has traditionally been assessed using fitness traits (e.g., fruiting; Agren et al. 2013), or with traits that are correlated to reproductive success (e.g., cold hardiness—Hamilton et al. 2013, De La Torre et al. 2014—or bud performance; De La Torre et al. 2014). In many organisms, these traits are difficult or potentially impossible to measure, and alternative performance traits that derive from

Downloaded from https://academic.oup.com/bioscience/article/65/3/237/236654 by guest on 20 August 2022

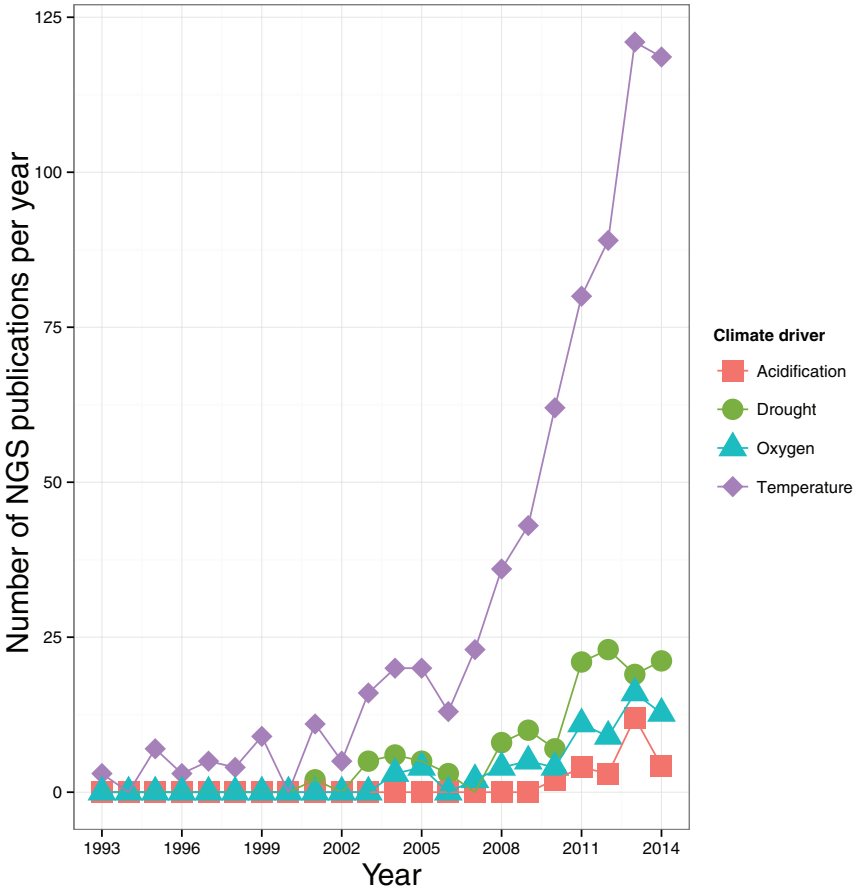

Figure 1. The number of published research articles that use genomics enabled approaches to understand biotic responses to four different climate drivers: temperature, acidification, drought, oxygen. The data were obtained from the Web of Science. The data for 2014 include only the period of January–June and have been normalized to that time period. Note the increase of studies after 2010, when next-generation sequencing (NGS) approaches became more widely available.

NGS data (e.g., the magnitude of gene expression change) could allow the detection of shifted reaction norms. Next-generation sequencing allows us to detect when genetically distinct populations may be locally adapted across an environmental gradient by changing the magnitude of expression of the same set of genes or when they express a different set of genes by activation of specific allelic variants of multicopy genes (more than one gene for a particular protein in the genome; Franks and Hoffmann 2012) that produce proteins with kinetic properties tuned to the local environment (Hochachka and Somero 2002). Next-generation sequencing approaches have made it possible to obtain these molecular characteristics for any target taxa at unprecedented levels of high speed and low cost.

One important question is just how much genomic information is required as evidence to detect local adaptation—or, in other words, how much of the genome shifts in local adaptation. An NGS study of inbred lines of *Arabidopsis* established from source populations that were locally adapted to differing climates established that selection in response to climate change can involve relatively few genetic markers; only 10–15 QTLs were required to explain shifts in population reaction norms (Agren et al. 2013). Genes from those QTLs regulate flowering time and freezing tolerance, suggesting that regulation of phenology and survival to extreme cold temperature are two of the most important aspects of organismal responses to climate change (Agren et al. 2013).

Alteration in gene expression across populations in response to shifts in climatic conditions has been demonstrated using NGS approaches (figure 1). For example, NGS approaches have been used to understand the mechanistic bases of fixed differences in the heat stress temperatures that induce bleaching between two populations of the scleractinian coral (*Acropora hyacinthus*) from tidal pools of varying degrees of low-tide associated heating (Barshis et al. 2013). Genes that differed in expression under control (non–heat stressed) conditions, representing prestress variation among locally adapted corals, indicate that regulation of the extracellular matrix (cell adhesion, potentially regulation of endosymbiotic zooxanthellae algae) was upregulated in heat-adapted corals, whereas genes typically associated with the cellular stress responses (CSR; e.g., heat shock protein 70, tumor necrosis factor, peroxidases and proteinases) were downregulated in the heat adapted corals (Barshis et al. 2013). The downregulation of CSR genes is consistent with evidence suggesting that there is an evolutionary cost in maintaining elevated levels of expression in those genes. For example, genes involved in regulation of actin and the cytoskeleton, oxidative stress responses, were regulated to a much greater extent in heat tolerant corals (Barshis et al. 2013), a finding that has been shown in other studies of heat stress responses among organisms acclimatized to differing climatic regimes (Stillman and Tagmount 2009).

Local adaptation to temperature has been demonstrated using genomic-scale approaches for the brown trout (*Salmo trutta*), following several generations in common garden captive conditions (Meier et al. 2014). Roughly two percent of the expressed genes were differentially expressed among populations, with a significant fraction of those representing variants of hemoglobin, the main oxygen transport and delivery respiratory protein (Meier et al. 2014). Given the

Downloaded from https://academic.oup.com/bioscience/article/65/3/237/236654 by guest on 20 August 2022

close relationship between environmental temperature and oxygen availability and the narrow thermal optima of many salmonid fishes, shifts in respiratory proteins may represent an adaptive response to climate variation.

Temperature is not the only climate variable for which local adaptation has been shown. Sea urchins (*Strongylocentrotus purpuratus*), living across an upwelling-driven gradient in pH variability along the California to Oregon Pacific coast have allelic variation at loci that are functionally important in response to acidification (Pespeni et al. 2013b). Allelic shifts in genes for ion homeostasis were correlated with differences in larval sea urchin performance under the present and future conditions of acidification (Pespeni et al. 2013b).

Using microarrays made from genome scale transcriptomics data (Grasso et al. 2008), *Acropora millepora* corals from different locations were shown to have fixed gene expression differences, potentially indicating local adaptation (Granados-Cifuentes et al. 2013). However, there was little correlation between the fixed common-garden differences in gene expression and the genotype of the host coral, suggesting that other factors that influence how expression of an organism's genome is regulated (e.g., symbiotic interactions) could play a large role in shaping patterns of local adaptation (Granados-Cifuentes et al. 2013).

Phenotypic plasticity. Genomics-enabled studies of phenotypic plasticity, or flexibility in response to changes in important climate-change associated drivers (e.g., temperature, pH, salinity), can provide information regarding loci of critical importance, the potential for populations to persist under shifting local conditions, as well as to provide transcripts that are regulated with high sensitivity across ecologically relevant environmental conditions (i.e., high-resolution biomarkers) useful for assessment of physiological state (e.g., energetics, stress) across ecological or spatial gradients (Evans and Hofmann 2012). In many cases, transcriptomics projects have been used to generate complementary DNA (cDNA) microarrays used in assessing variation in gene expression profiles across populations or organisms within a population exposed to different conditions. RNA sequencing is quickly replacing cDNA microarrays as the tool of choice for genome-scale inferences regarding physiological plasticity, although there are fewer RNA-seq-based studies than the number of microarray studies to date.

For example, genome-scale studies have been undertaken to understand the capacity of coastal marine organisms to respond to the elevated acidity associated with ocean acidification (Evans and Hofmann 2012, Evans et al. 2013, Pespeni et al. 2013a, 2013b). Sea urchins from pH-variable habitats along the California coast show life stage (Todgham and Hofmann 2009, Evans et al. 2013) and population dependent responses to ocean acidification (Pespeni et al. 2013b). Although plasticity is certainly of great importance, in cases in which local adaptation has been demonstrated, the role of plasticity may be muted; differences in gene expression of locally adapted *S. purpuratus* larvae was limited to 32 genes across the entire genome (Pespeni et al. 2013b). A microarray study of gene expression changes that correlated with shifts in larval morphology showed concordance between larval traits (arm length) and genes involved with cellular structure (e.g., tubulin, actin) and skeletal formation (e.g., spicule matrix proteins) (Padilla-Gamino et al. 2013), providing important targets for adaptive plasticity.

Next-generation sequencing approaches have been used to assess the mechanistic responses to ocean acidification in the first phase of benthic life and calcification in corals, in what is hypothesized to be a highly sensitive life stage (Moya et al. 2012). RNA-seq was used to assess gene expression differences among *A. millepora* coral during exposure to acidified conditions on days 2–4 after settlement (Moya et al. 2012). Genes involved in metabolism and extracellular matrix formation were repressed under acidified conditions. Unexpectedly, expression of genes putatively thought to be of widespread importance in the calcification and ion transport processes were not strongly differentially expressed with changes in seawater pH (Moya et al. 2012). A similar lack of correspondence between genes identified as integral to the calcification process and the expression of those genes under acidified conditions has also been demonstrated in calcifying phytoplankton, the coccolithophore *Emiliania huxleyi* (Benner et al. 2013). Repression of metabolic processes at the transcriptomic level was also observed in the coral *Pocillopora damicornis* under acidified conditions, though in this case genes putatively involved with the calcification process were also induced under acidification (Vidal-Dupiol et al. 2013), unlike in other studies.

Next-generation sequencing-derived biomarkers for important physiological processes in corals provide coral reef ecologists with new high-resolution tools that can be used to assess the health of reefs. Changes in coral fitness-relevant physiological processes, such as calcification or bleaching, can be characterized by transcriptomic shifts under different environmental conditions (Desalvo et al. 2008, Moya et al. 2012). Monitoring expression, mutation, or other changes in those specific transcript biomarkers can provide coral reef ecologists a powerful tool for assessing coral health *in situ*. Similar biomarker approaches may be generated in marine organisms that occur along a gradient of climate variables in temperature (e.g., crab, Stillman and Tagmount 2009; fish, Jeffries et al. 2014), heat stress (e.g., scallop, Fu et al. 2014; a gradient of pH—e.g., sea urchin, Pespeni et al. 2013b). For example, processes involved in enhancing messenger RNA (mRNA) stability are regulated following thermal stress in a range of organisms (Fu et al. 2014).

Not all increases in the amount of mRNA for a gene (the transcript) are reflected in the production of protein encoded by that gene; in other words, the processes of transcription and translation are not necessarily coupled. Roughly 81% of the differentially expressed transcripts of *Arabidopsis* following thermal stress were associated with polyribosomal mRNA (an mRNA molecule that is bound to multiple ribosomes), suggesting that those transcripts were actively

Downloaded from https://academic.oup.com/bioscience/article/65/3/237/236654 by guest on 20 August 2022

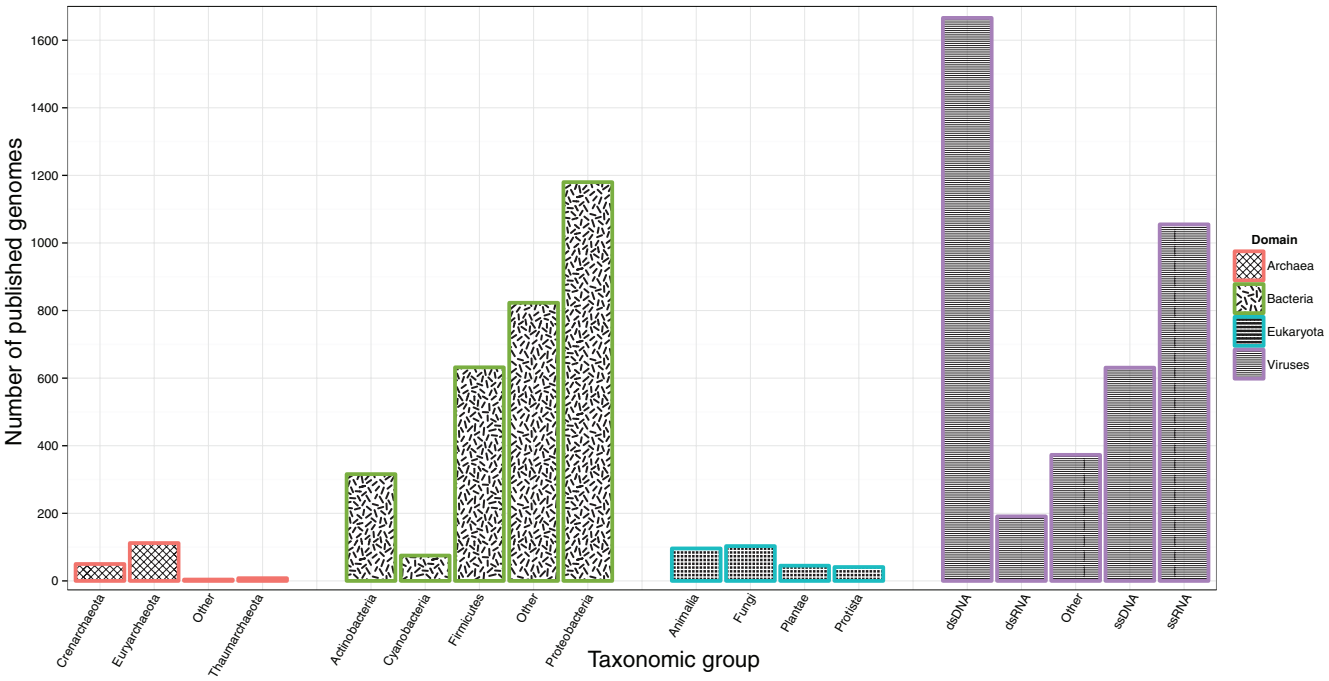

Figure 2. The number of available reference genomes and their distribution across the domains of life. The data were obtained on 30 September 2014 from the National Center for Biotechnology Information's Genome Browser (*www.ncbi.nlm.nih.gov/genome/browse*) and the Kyoto Encyclopedia of Genes and Genomes' genome catalog (*http://genome.jp/kegg/catalog/org_list.html*).

being translated (Yangueez et al. 2013). Among the actively translated differentially expressed transcripts, 95% of the genes were repressed in thermally stressed as compared with control plants (Yangueez et al. 2013). However, of the 19% of the differentially expressed genes that were not likely to be translated (i.e., not associated with polyribosomal mRNA), only 60% were repressed, suggesting that there was a greater equivalence of induction and repression (Yangueez et al. 2013). That difference in repression between translated and nontranslated transcripts could indicate variation in gene regulatory mechanisms in genes whose products are important for the heat stress response.

What resources are needed to make NGS approaches to understanding responses to climate change more easily applied across a broad range of taxa?

Genome projects. Genomes of additional nonmodel organisms are needed to fully enable climate change related studies using genomics-based approaches such as RNA-seq transcriptomics, or RAD-seq and DNA-seq phylogenomics, the study of relatedness using genomics data. Although generating genome-scale data at the level of DNA or RNA is getting easier and easier, there are still large challenges in assembling those data into an annotated representation of the chromosomal structure of organisms, or the genome. Two genome-sequenced arthropods, the fruit fly *Drosophila* and the water flea *Daphnia*, have great ecological breadth that makes them attractive targets for climate change related ecological studies, and *Daphnia* has principally been hailed as the newest genomics-enabled organism for environmental biology (Colbourne et al. 2011). Because of their small genome size, there are relatively more completed genome projects for microbes (figure 2), some of which are of critical importance in understanding the relationships between climate change and planetary biogeochemical cycling (e.g., the calcifying marine phytoplankton *Emiliania huxleyi*; Read et al. 2013).

Comparing differential gene expression among populations or species across environmental gradients related to climate change requires the ability to be sure that the genes being compared are representing the same genomic regions—or, in other words, are homologous. Inferring gene homology within a species (paralogy; gene duplication leads to multiple gene copies) and across species (orthology; genes descended from the same ancestral sequence) is difficult when mapping to *de novo* transcriptomes instead of genomes, because *de novo* transcriptomes are likely to have multiple gene models for a particular locus. In addition, multiple loci in a genome may have coding sequences similar enough so that those loci cannot be distinguished at the level of the mRNA transcript. Inferences from RNA-seq studies without a genome reference at best can be considered hypotheses for further study rather than provide definitive answers regarding genomic changes that are

Downloaded from https://academic.oup.com/bioscience/article/65/3/237/236654 by guest on 20 August 2022

occurring during population responses to climate change. In addition, the use of SNP data in phylogenomic studies is limited without a reference genome because there is no easy way in which to assess linkage among the SNPs. Because phylogenetic models assume that the genetic markers used are independent, if SNPs are actually closely linked on a chromosome, they would be in violation of those models. Incorrectly assuming independence of loci that are closely linked within the genome could potentially bias estimations of divergence time because such assumptions would not take into account reductions in sequence divergence within a subpopulation.

Computational resources. Next-generation sequencing analyses are best performed on high performance, multithreaded, parallel computing platforms. Genomic-scale sequence alignment is an intrinsically parallel process involving millions to billions of reads being simultaneously compared and is an immense computing operation. Although computationally intensive analyses of NGS data (i.e., sequence alignment) could be conducted on local desktop computers, the time to complete those analyses could be unacceptably long. In recognition of the demanding computer infrastructure required to perform some steps of the NGS analysis pipeline, institutionally shared computing resources may be the most powerful approach. Although companies involved in biomedicine and agribusiness may have the financial leverage to obtain their own high-performance computers, most academic researchers working in environmental biology would not have that capital. Luckily, there are other ways for environmental biologists interested in using NGS approaches to access needed computational resources. For example, XSEDE (for *Extreme Science and Engineering Discovery Environment*; *www.xsede.org*) provides a portal for partner institutions who host supercomputers best suited for specific computational tasks and make awards to researchers, allowing them to access supercomputers needed for their research. In addition, distributed cloud computing resources (e.g., Amazon, Google) provide an alternative to using a single supercomputer or purchasing a local high-performance computing cluster.

Data storage is a challenge, but the decreased costs of data storage devices have made the archival storage of terabytes of data far less of an issue than it once was. Nevertheless, NGS approaches generate too many data for the current data infrastructure (e.g., GenBank) to archive. Because the data are costly to generate and have terrific potential and utility for future studies, it is imperative that they be archived and made available to other researchers. How best to store and serve those data? Should research communities invest in generating taxon-specific databases to support researchers across institutions and international borders (e.g., a crustacean database that would house NGS data for all crustacean studies)? Would subsets of those data be housed in biological problem–inspired databases in which data of a particular type are stored and served together (e.g., a Pacific Rim phylogenomics database that would house all NGS data from samples in a certain region—e.g., the Pacific Rim)? Should there be institutional databases that store and serve all data generated at their sequencing facilities (e.g., the University of California Berkeley QB3 database)? Should nations invest in a new model for a data service that can handle the enormous mass of data generated (e.g., a GenBank-II), perhaps based on a cloud computing distributed database structure? These are challenges that research communities spanning disciplines and international boundaries must meet in the near future.

Accessing the computational tools needed for NGS data analysis takes some expertise. Although commercial software packages are available for analysis of some types of NGS data (e.g., genome assembly, transcriptomics, but not RAD-seq), the research community in NGS bioinformatics is primarily noncommercial academic users who freely distribute and frequently update or modify computing code that requires programming expertise in order to use it. Relatively few biologists decide to study environmental biology because of their love and passion for learning how to script in Python or Perl (although R is quite popular). In any case, additional training in computer programming and the Unix operating system would seem absolutely essential for the next generation of genome-enabled scientists.

New computational tools to work with NGS data are continuously being developed and are helping to better refine transcriptome assemblies, sequence read mapping, statistical analyses, and generating biological inferences from the results of those studies. The reader is directed toward the Bioconductor portal for updated software resources available for use in the open-source environment R (Gentleman et al. 2004). Tools for network analyses of differentially expressed genes is an area for continued focus as network visualizations are complex, and it can be difficult to infer biological functional significance from them (e.g., Fu et al. 2014).

Funding. Genomics tools, as well as any cutting edge tools in the biological sciences, are always likely to be expensive. The biomedical and agribusiness industries that are driving the development of new technologies in biology are exceptionally well funded relative to environmental biologists studying biotic responses to climate change. Broad scale adoption of NGS approaches by environmental biologists will require new thinking about how to fund researchers or research consortia who are attempting to use the best available tools to tackle what may be one of the largest but most poorly funded problems facing society today—understanding and mitigating the impacts of climate change.

Conclusions

Next-generation sequencing approaches are fundamentally changing the way in which environmental scientists undertake studies to understand how organisms are responding or may respond in the future to climate change. Never before

Downloaded from https://academic.oup.com/bioscience/article/65/3/237/236654 by guest on 20 August 2022

have scientists been able to generate so many data about the genomes of an organism in such little time. We are just at the dawn of the genomics era in environmental biology, but we have already learned much that would not have been possible using traditional molecular biology approaches. In the relatively few examples to date, NGS approaches have resulted in finer scale resolution of population structure, have identified local adaptation, and have characterized the loci most likely to be adaptive in response to climate change. In order to leverage fully the available NGS tools for the climate change research of tomorrow and beyond, we need more training, more computing, and more money.

Acknowledgments

We thank three anonymous reviewers for their helpful comments in improving this manuscript. This material is based on work supported by National Science Foundation grant no. 1041225 to JHS. EJA was supported by the Department of Defense through the National Defense Science and Engineering Graduate Fellowship Program.

References cited

Agren J, Oakley CG, McKay JK, Lovell JT, Schemske DW. 2013. Genetic mapping of adaptation reveals fitness tradeoffs in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the United States of America 110: 21077–21082.

Amir A, Zeisel A, Zuk O, Elgart M, Stern S, Shamir O, Turnbaugh PJ, Soen Y, Shental N. 2013. High-resolution microbial community reconstruction by integrating short reads from multiple 16S rRNA regions. Nucleic Acids Research 41 (art. e205).

Barshis DJ, Ladner JT, Oliver TA, Seneca FO, Traylor-Knowles N, Palumbi SR. 2013. Genomic basis for coral resilience to climate change. Proceedings of the National Academy of Sciences of the United States of America 110: 1387–1392.

Benner I, Diner RE, Lefebvre SC, Li D, Komada T, Carpenter EJ, Stillman JH. 2013. *Emiliania huxleyi* increases calcification but not expression of calcification-related genes in long-term exposure to elevated temperature and pCO_2. Philosophical Transactions of the Royal Society B 368: 20130049.

Bi K, Linderoth T, Vanderpool D, Good JM, Nielsen R, Moritz C. 2013. Unlocking the vault: next-generation museum population genomics. Molecular Ecology 22: 6018–6032.

Bossdorf O, Richards CL, Pigliucci M. 2008. Epigenetics for ecologists. Ecology Letters 11: 106–115.

Bowen BW, et al. 2014. Phylogeography unplugged: comparative surveys in the genomic era. Bulletin of Marine Science 90: 13–46.

Brautigam K, et al. 2013. Epigenetic regulation of adaptive responses of forest tree species to the environment. Ecology and Evolution 3: 399–415.

Brito PH, Edwards SV. 2009. Multilocus phylogeography and phylogenetics using sequence-based markers. Genetica 135: 429–455.

Carpenter KE, et al. 2011. Comparative phylogeography of the coral triangle and implications for marine management. Journal of Marine Biology 2011: Article ID 396982.

Chen I-C, Hill JK, Ohlemüller R, Roy DB, Thomas CD. 2011. Rapid range shifts of species associated with high levels of climate warming. Science 333: 1024–1026.

Choi M, et al. 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proceedings of the National Academy of Sciences 106: 19096–19101.

Colbourne JK, et al. 2011. The ecoresponsive genome of *Daphnia pulex*. Science 331: 555–561.

DeBoer TS, Naguit MRA, Erdmann MV, Ablan-Lagman MCA, Ambariyanto, Carpenter KE, Toha AHA, Barber PH. 2014. Concordance between phylogeographic and biogeographic boundaries in the Coral Triangle: Conservation implications based on comparative analyses of multiple giant clam species. Bulletin of Marine Science 90: 277–300.

De La Torre AR, Wang TL, Jaquish B, Aitken SN. 2014. Adaptation and exogenous selection in a *Picea glauca* x *Picea engelmannii* hybrid zone: Implications for forest management under climate change. New Phytologist 201: 687–699.

Delmont TO, et al. 2012. Structure, fluctuation and magnitude of a natural grassland soil metagenome. Isme Journal 6: 1677–1687.

Des Marais DL, Hernandez KM, Juenger TE. 2013. Genotype-by-environment interaction and plasticity: Exploring genomic responses of plants to the abiotic environment. Annual Review of Ecology Evolution and Systematics 44: 5–29.

Desalvo MK, Voolstra CR, Sunagawa S, Schwarz JA, Stillman JH, Coffroth MA, Szmant AM, Medina M. 2008. Differential gene expression during thermal stress and bleaching in the Caribbean coral *Montastraea faveolata*. Molecular Ecology 17: 3952–3971.

De Wit P, Pespeni MH, Ladner JT, Barshis DJ, Seneca F, Jaris H, Therkildsen NO, Morikawa M, Palumbi SR. 2012. The simple fool's guide to population genomics via RNA-Seq: An introduction to high-throughput sequencing data analysis. Molecular Ecology Resources 12: 1058–1067.

Dolman G, Joseph L. 2012. A species assemblage approach to comparative phylogeography of birds in southern Australia. Ecology and Evolution 2: 354–369.

Epstein SS. 2013. The phenomenon of microbial uncultivability. Current Opinion in Microbiology 16: 636–642.

Evans TG, Chan F, Menge BA, Hofmann GE. 2013. Transcriptomic responses to ocean acidification in larval sea urchins from a naturally variable pH environment. Molecular Ecology 22: 1609–1625.

Evans TG, Hofmann GE. 2012. Defining the limits of physiological plasticity: How gene expression can assess and predict the consequences of ocean change. Philosophical Transactions of the Royal Society B 367: 1733–1745.

Franks SJ, Hoffmann AA. 2012. Genetics of climate change adaptation. Annual Review of Genetics 46: 185–208.

Fu X, et al. 2014. Sequencing-based gene network analysis provides a core set of gene resource for understanding thermal adaptation in Zhikong scallop *Chlamys farreri*. Molecular Ecology Resources 14: 184–198.

Galtier N, Nabholz B, Glemin S, Hurst GDD. 2009. Mitochondrial DNA as a marker of molecular diversity: A reappraisal. Molecular Ecology 18: 4541–4550.

Gentleman RC, et al. 2004. Bioconductor: Open software development for computational biology and bioinformatics. Genome Biology 5 (art. R80).

Granados-Cifuentes C, Bellantuono AJ, Ridgway T, Hoegh-Guldberg O, Rodriguez-Lanetty M. 2013. High natural gene expression variation in the reef-building coral *Acropora millepora*: Potential for acclimative and adaptive plasticity. BMC Genomics 14 (art. 228).

Grasso LC, Maindonald J, Rudd S, Hayward DC, Saint R, Miller DJ, Ball EE. 2008. Microarray analysis identifies candidate genes for key roles in coral development. BMC Genomics 9 (art. 540).

Hamilton JA, Lexer C, Aitken SN. 2013. Genomic and phenotypic architecture of a spruce hybrid zone (*Picea sitchensis P. glauca*). Molecular Ecology 22: 827–841.

Helyar SJ, et al. 2011. Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. Molecular Ecology Resources 11: 123–136.

Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. Nature 405: 907–913.

Hochachka PW, Somero GN. 2002. Biochemical Adaptation. Oxford University Press.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLOS Genetics 6 (art. e1000862).

Jackson AM, Ambariyanto, Erdmann MV, Toha AHA, Stevens LA, Barber PH. 2014. Phylogeography of commercial tuna and mackerel in the Indonesian archipelago. Bulletin of Marine Science 90: 471–429.

Downloaded from https://academic.oup.com/bioscience/article/65/3/237/236654 by guest on 20 August 2022

Jeffries KM, Hinch SG, Sierocinski T, Pavlidis P, Miller KM. 2014. Transcriptomic responses to high water temperature in two species of Pacific salmon. Evolutionary Applications 7: 286–300.

Johnson DS, Mortazavi A, Myers RM, Wold B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. Science 316: 1497–1502.

Kochzius M, Nuryanto A. 2008. Strong genetic population structure in the boring giant clam, *Tridacna crocea*, across the Indo-Malay Archipelago: implications related to evolutionary processes and connectivity. Molecular Ecology 17: 3775–3787.

Kultz D, et al. 2013. New frontiers for organismal biology. Bioscience 63: 464–471.

Lexer C, Mangili S, Bossolini E, Forest F, Stolting KN, Pearman PB, Zimmermann NE, Salamin N. 2013. "Next generation" biogeography: Towards understanding the drivers of species diversification and persistence. Journal of Biogeography 40: 1013–1022.

Meier K, Hansen MM, Normandeau E, Mensberg K-LD, Frydenberg J, Larsen PF, Bekkevold D, Bernatchez L. 2014. Local adaptation at the transcriptome level in brown trout: Evidence from early life history temperature genomic reaction norms. PLOS ONE 9 (art. e85171).

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Research 17: 240–248.

Morin PA, McCarthy M. 2007. Highly accurate SNP genotyping from historical and low-quality samples. Molecular Ecology Notes 7: 937–946.

Moya A, Huisman L, Ball EE, Hayward DC, Grasso LC, Chua CM, Woo HN, Gattuso JP, Foret S, Miller DJ. 2012. Whole transcriptome analysis of the coral *Acropora millepora* reveals complex responses to CO_2-driven acidification during the initiation of calcification. Molecular Ecology 21: 2440–2454.

Padilla-Gamino JL, Kelly MW, Evans TG, Hofmann GE. 2013. Temperature and CO_2 additively regulate physiology, morphology and genomic responses of larval sea urchins, *Strongylocentrotus purpuratus*. Proceedings of the Royal Society B 280 (art. 20130155).

Pespeni MH, Barney BT, Palumbi SR. 2013a. Differences in the regulation of growth and biomineralization genes revealed through long-term common-garden acclimation and experimental genomics in the purple sea urchin. Evolution 67: 1901–1914.

Pespeni MH, et al. 2013b. Evolutionary change during experimental ocean acidification. Proceedings of the National Academy of Sciences 110: 6937–6942.

Pinsky ML, Worm B, Fogarty MJ, Sarmiento JL, Levin SA. 2013. Marine taxa track local climate velocities. Science 341: 1239–1242.

Rappe MS, Giovannoni SJ. 2003. The uncultured microbial majority. Annual Review of Microbiology 57: 369–394.

Read BA, et al. 2013. Pan genome of the phytoplankton *Emiliania* underpins its global distribution. Nature 499: 209–213.

Rocha LA, Craig MT, Bowen BW. 2007. Phylogeography and the conservation of coral reef fishes. Coral Reefs 26: 513–513.

Sanger F, Coulson AR. 1975. Rapid method for determining sequences in DNA by primed synthesis with DNA-polymerase. Journal of Molecular Biology 94: 441–448.

Savolainen O, Lascoux M, Merila J. 2013. Ecological genomics of local adaptation. Nature Reviews Genetics 14: 807–820.

Shaffer BH, Purugganan MD. 2013. Introduction to theme "genomics in ecology, evolution, and systematics." Annual Review of Ecology Evolution and Systematics 44: 1–4.

Sharp KH, Ritchie KB. 2012. Multi-partner interactions in corals in the face of climate change. Biological Bulletin 223: 66–77.

Somero GN. 2011. Comparative physiology: A "crystal ball" for predicting consequences of global change. American Journal of Physiology: Regulatory Integrative and Comparative Physiology 301: R1–R14.

Song C, Ballantyne F, Smith VH. 2014. Enhanced dissolved organic carbon production in aquatic ecosystems in response to elevated atmospheric CO_2. Biogeochemistry 118: 49–60.

Staden R. 1979. Strategy of DNA sequencing employing computer-programs. Nucleic Acids Research 6: 2601–2610.

Stillman JH. 2003. Acclimation capacity underlies susceptibility to climate change. Science 301: 65–65.

Stillman JH, Tagmount A. 2009. Seasonal and latitudinal acclimatization of cardiac transcriptome responses to thermal stress in porcelain crabs, *Petrolisthes cinctipes*. Molecular Ecology 18: 4206–4226.

Stinchcombe J, Hoekstra H. 2008. Combining population genomics and quantiative genetics: Finding the genes underlying ecologically important traits. Heredity 100: 158–170.

Strous M, Fuerst JA, Kramer EHM, Logemann S, Muyzer G, van de Pas-Schoonen KT, Webb R, Kuenen JG, Jetten MSM. 1999. Missing lithotroph identified as new planctomycete. Nature 400: 446–449.

Sunday JM, Calosi P, Dupont S, Munday PL, Stillman JH, Reusch TBH. 2014. Evolution in an acidifying ocean. Trends in Ecology and Evolution 29: 117–125.

Swan BK, et al. 2013. Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. Proceedings of the National Academy of Sciences 110: 11463–11468.

Therkildsen NO, Hemmer-Hansen J, Hedeholm RB, Wisz MS, Pampoulie C, Meldrup D, Bonanomi S, Retzel A, Olsen SM, Nielsen EE. 2013. Spatiotemporal SNP analysis reveals pronounced biocomplexity at the northern range margin of Atlantic cod *Gadus morhua*. Evolutionary Applications 6: 690–705.

Timm J, Kochzius M. 2008. Geological history and oceanography of the Indo-Malay Archipelago shape the genetic populations structure in the false clownfish (*Amphiprion ocellaris*). Molecular Ecology 17: 3999–4014.

Todgham AE, Hofmann GE. 2009. Transcriptomic response of sea urchin larvae *Strongylocentrotus purpuratus* to CO_2-driven seawater acidification. Journal of Experimental Biology 212: 2579–2594.

Vidal-Dupiol J, Zoccola D, Tambutte E, Grunau C, Cosseau C, Smith KM, Freitag M, Dheilly NM, Allemand D, Tambutte S. 2013. Genes related to ion-transport and energy production are upregulated in response to CO_2-driven pH decrease in corals: New insights from transcriptome analysis. PLOS ONE 8 (art. e58652).

Whitehead A. 2012. Comparative genomics in ecological physiology: Toward a more nuanced understanding of acclimation and adaptation. Journal of Experimental Biology 215: 884–891.

Yakovlev IA, Lee Y, Rotter B, Olsen JE, Skroppa T, Johnsen O, Fossdal CG. 2014. Temperature-dependent differential transcriptomes during formation of an epigenetic memory in Norway spruce embryogenesis. Tree Genetics and Genomes 10: 355–366.

Yangueez E, Castro-Sanz AB, Fernandez-Bautista N, Oliveros JC, Castellano MM. 2013. Analysis of genome-wide changes in the translatome of *Arabidopsis* seedlings subjected to heat stress. PLOS ONE 8 (art. e71425).

Jonathon H. Stillman (stillmaj@sfsu.edu) and Eric Armstrong are affiliated with the Romberg Tiburon Center at San Francisco State University, in San Francisco, California, and with the Department of Integrative Biology at the University of California Berkeley. JHS is also affiliated with the Department of Biology at San Francisco State University.

Downloaded from https://academic.oup.com/bioscience/article/65/3/237/236654 by guest on 20 August 2022