

California State University, Monterey Bay  
**Digital Commons @ CSUMB**

---

School of Natural Sciences Faculty Publications  
and Presentations

School of Natural Sciences

---

2014

## So, You Want to Use Next-Generation Sequencing in Marine Systems? Insight from the Pan-Pacific Advanced Studies Institute

D A. Willette

F W. Allendorf

P H. Barber

D J. Barshis

K E. Carpenter

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.csUMB.edu/sns\\_fac](https://digitalcommons.csUMB.edu/sns_fac)

---

### Recommended Citation

Willette, D.A., Allendorf, F.W., Barber, P.H., Barshis, D.J., Carpenter, K.E., Crandall, E.D., Cresko, W.A., Fernandez-Silva, I., Matz, M.V., Meyer, E., Seeb, L.W., and J.E. Seeb. 2014. So, you want to use next-generation sequencing in marine systems?: Insight from the Pacific Advanced Studies Institute. *Bulletin of Marine Science* 90(1): 79-122.

This Article is brought to you for free and open access by the School of Natural Sciences at Digital Commons @ CSUMB. It has been accepted for inclusion in School of Natural Sciences Faculty Publications and Presentations by an authorized administrator of Digital Commons @ CSUMB. For more information, please contact [digitalcommons@csUMB.edu](mailto:digitalcommons@csUMB.edu).

---

**Authors**

D A. Willette, F W. Allendorf, P H. Barber, D J. Barshis, K E. Carpenter, E D. Crandall, W A. Cresko, I Fernandez-Silva, M V. Katz, E Meyer, M D. Santos, L W. Seeb, and J E. Seeb



# So, you want to use next-generation sequencing in marine systems? Insight from the Pan-Pacific Advanced Studies Institute

<sup>1</sup> Department of Ecology and Evolutionary Biology, 2141 Terasaki Life Science Building, 610 Charles E Young Drive South, University of California Los Angeles, Los Angeles, California 90095.

<sup>2</sup> Division of Biological Sciences, University of Montana, Missoula, Montana 59812.

<sup>3</sup> Southwest Fisheries Science Center, 110 Shaffer Rd., Santa Cruz, California 95060.

<sup>4</sup> Department of Biological Sciences, Old Dominion University, Norfolk, Virginia 23529.

<sup>5</sup> Institute of Ecology and Evolution, 312 Pacific Hall, University of Oregon, Eugene, Oregon 97403.

<sup>6</sup> Hawai'i Institute of Marine Biology, P.O. Box 1346, Kaneohe, Hawaii 96734.

<sup>7</sup> Section of Integrative Biology, 1 University Station #C0930, University of Texas, Austin, Texas 78712.

<sup>8</sup> Department of Zoology, 3106 Cordley Hall, Oregon State University, Corvallis, Oregon 97331.

<sup>9</sup> National Fisheries Research and Development Institute, 101 Mother Ignacia St. Quezon City, Metro Manila, Philippines, 1101.

<sup>10</sup> School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, Washington 98195.

\* Corresponding author email: <dwillett@odu.edu>.

Date Submitted: 17 January, 2013.

Date Accepted: 1 July, 2013.

Available Online: 10 January, 2014.

**DA Willette**<sup>1\*</sup>

**FW Allendorf**<sup>2</sup>

**PH Barber**<sup>1</sup>

**DJ Barshis**<sup>3</sup>

**KE Carpenter**<sup>4</sup>

**ED Crandall**<sup>3</sup>

**WA Cresko**<sup>5</sup>

**I Fernandez-Silva**<sup>6</sup>

**MV Matz**<sup>7</sup>

**E Meyer**<sup>8</sup>

**MD Santos**<sup>9</sup>

**LW Seeb**<sup>10</sup>

**JE Seeb**<sup>10</sup>

---

**ABSTRACT.**—The emerging field of next-generation sequencing (NGS) is rapidly expanding capabilities for cutting edge genomic research, with applications that can help meet marine conservation challenges of food security, biodiversity loss, and climate change. Navigating the use of these tools, however, is complex at best. Furthermore, applications of marine genomic questions are limited in developing nations where both marine biodiversity and threats to marine biodiversity are most concentrated. This is particularly true in Southeast Asia. The first Pan-Pacific Advanced Studies Institute (PacASI) entitled “Genomic Applications to Marine Science and Resource Management in Southeast Asia” was held in July 2012 in Dumaguete, Philippines, with the intent to draw together leading scientists from both sides of the Pacific Ocean to understand the potential of NGS in helping address the aforementioned challenges. Here we synthesize discussions held during the PacASI to provide perspectives and guidance to help scientists new to NGS choose among the variety of available advanced genomic methodologies specifically for marine science questions.

---

In July 2012, the first Pan-Pacific Advanced Studies Institute (PacASI), funded by the United States National Science Foundation, was held in Dumaguete, Philippines (<http://sci.odu.edu/impac/pacasi/index.html>). Entitled “Genomic Applications to Marine Science and Resource Management in Southeast Asia,” the 2-wk workshop

assembled participants from developing and developed nations to present case studies and methods, and to discuss the emerging field of next-generation sequencing (NGS) in the context of marine molecular ecology. Working groups explored how advanced genomic tools can be used to test hypotheses and to meet marine conservation challenges of food security, biodiversity loss, and climate change (Palumbi et al. 2009, Allendorf et al. 2010, Carpenter et al. 2011, Maralit et al. 2013, Barber et al. 2014). The rapid decline in sequencing costs, increased accessibility to powerful computer clusters, publicly available software, and scripts and tutorials for data management and analysis make NGS applications increasingly viable options. However, navigating the appropriate use of the range of technologies is complicated, especially for the laboratories in developing nations who were the target audience of PacASI. Here we synthesize discussions held during the PacASI to provide a road map to help scientists choose among the variety of advanced genomic methodologies depending on the research questions being asked.

Numerous research questions, pertinent to the marine conservation challenges, were presented by local scientists and resource managers at the PacASI (See Textbox A). However, before addressing any of these with NGS, the first question of any study should be, “What is the central question and what data are needed to answer it?” This question highlights the guiding principle that it is the science that steers research, not the technology. NGS tools are not necessarily the most appropriate approach, and in many instances first-generation sequencing methods may be cheaper, more accessible, easier to use, and provide satisfactory results.

Genomics is the study of complete or large portions of the genome of organisms, and whole genome sequencing is now possible with relatively reasonable time and

Textbox A. Research questions pertinent to marine conservation challenges of food security, biodiversity loss, and climate change.

*Questions aligning with Section I: Population structure and genomics*

- How can the spatial connectivity of populations of commercially important or threatened species be inferred, and can knowledge of spatial connectivity help position Fisheries Management Zones (FMZ) or Marine Protected Areas (MPAs)?
- How can genetics guide better management of multi-species and multi-gear fisheries, and aid the Ecosystem Approach to Fisheries Management (EAFM)?
- How can population boundaries of marine species be assessed; and if populations span national boundaries, how best can shared stocks be managed?
- Can population structure and genetic variability be assessed to a level that can provide guidance in restocking or restoration efforts after the collapse of a local population?
- How can effective population size be estimated in marine organisms?
- How can marine biodiversity be assessed, inclusive of microbes, viruses, and other microscopic organisms?
- How can species be identified en masse, particularly in the absence of discernable morphological features (i.e., post-processed fish products)?

*Questions aligning with Section II: Local adaptation of marine organisms*

- How will corals and other marine organisms respond to environmental stress, particularly those associated with climate change?
- Which populations of corals and other marine organisms have the highest likelihood of surviving increased environmental stresses from climate change?
- Can individuals be screened for favorable traits (i.e., disease-resistance) to identify robust broodstock for the aquaculture industry?

cost investments (see Textbox B). However, many biological questions can be addressed by focusing on a particular region of the genome (i.e., DNA barcodes) or a suite of target loci (see Textbox C). Further, the practical aspects of acquiring and managing genomic data sets may be excessive and overly burdensome for many research laboratories. It is self-evident that, when fiscally and logistically possible, the most appropriate tool should be selected in answering a biological question. However, the most advanced tool does not necessarily equate to the most appropriate tool, and deciding whether or not to use NGS should be properly weighed.

To address the above questions, marine science experts at PacASI shared robust knowledge—often gained by painful first-hand experiences—of available NGS applications, namely transcriptomics, amplicon sequencing, genotyping-by-sequencing, and SNP discovery for high-throughput genotyping. Recent studies have reviewed the wide range of NGS technologies available, including their relative strengths and weaknesses (Harismendy et al. 2009, Davey et al. 2011, Pareek et al. 2011, Boers et al. 2012, Liu et al. 2012, Luo et al. 2012, Quail et al. 2012, Arnold et al. 2013, Gautier et al. 2013), and therefore we limit our technical discussion to some of these NGS tools in textboxes. A streamlined microsatellite discovery method based upon NGS was presented at PacASI (see Fernandez-Silva and Toonen 2013, Fernandez-Silva et al. 2013); however, after consideration of challenges in applying microsatellites and prevailing trends in marine genomic studies, we chose to forego discussing microsatellites here.

The above questions are not novel (though some have only recently become technologically feasible) and examples of scientific papers that address them can be readily found in a search of the literature including specific examples from Southeast Asia (Table 1; also see Begeer et al. 2014, Bowen et al. 2014, Keyse et al. 2014, von der Heyden et al. 2014). Most of these questions fit into two general categories in molecular marine research: (1) population genomics and (2) local adaptation and demographic history of marine organisms. The field of population genomics explores how mutation, selection, gene flow, and genetic drift affect patterns of genetic variation based upon study of large amounts of genomic data from numerous individuals across populations (Allendorf et al. 2010). The study of local adaptation of marine organisms examines how phenotypes are shaped by the relationship between genes and the environment.

Although we focus on these two categories of molecular research, we acknowledge that many others exist beyond the scope of this review. For each category we provide a brief background, explore how NGS applications are used in (or in the case of the Indo-Pacific region, have the potential to be used in) addressing the question, and describe relevant marine research. Complementary are textboxes that provide a short review and technical description of NGS methods discussed at PacASI. These include whole genome sequencing (Textbox B), RAD-sequencing (Textbox C), amplicon sequencing (Textbox D), metagenomics (Textbox E), and RNA-sequencing/gene expression (Textbox F). Lastly, for consideration of those newly considering NGS applications, we include example estimates of cost and time investments (Table 2), a list of genetic problems in marine conservation with potential solutions from genomics (Table 3), and a schematic diagram of interacting factors to consider in genomic studies (Fig. 1). Although we strongly believe the present study will serve as a valuable resource for NGS newcomers, readers should feel free to jump to topics of most interest.

### Textbox B. An “Instant” genome.

Many ecological and evolutionary genetic studies can be performed using next-generation sequencing (NGS) on organisms without a reference genome. Studies of parentage and relatedness, population structure, or even gene expression may be enhanced by a complete genome sequence, but do not necessarily require it. However, other studies require at least the ordering of genetic markers, or ultimately the complete DNA sequence of most of the genome. For example, the identification of genomic regions that are subject to diversifying or stabilizing selection can be more easily identified when co-localized markers reinforce one another. Furthermore, identifying quantitative trait loci (QTL) requires a linkage map, and DNA sequence to identify the causative sequence-level changes.

NGS now provides powerful tools to “genomically enable” non-model organisms. While genome sequencing and assembly remains non-trivial, many molecularly-capable laboratories can do so with a reasonable investment of resources. We suggest that three core datasets are important; a solid genome assembly, a good reference transcriptome for annotation, and a genetic map to scaffold the assemblies. Special laboratory equipment is not required, since sequencing, and even preparation can be outsourced to the growing number of laboratories providing external NGS services. However, access to a high performance computing environment is essential. The assembly of even a moderately complex vertebrate genome requires several weeks of processor time on a computer cluster with dozens of processors, numerous terabytes of storage, and most importantly a large amount of memory (RAM; 500 gigabytes to 1 terabyte).

For the genome assembly NGS sequencing libraries are prepared and sequenced to high depth of coverage ( $>20\times$ ) across the genome. Importantly, libraries of different insert sizes (short overlapping, standard and long mate-pair) appear to increase the quality of the assemblies. The optimal insert size for Illumina sequencing is approximately 500–1000 bp, and this standard library size will form the core assembly data. Short insert libraries allow overlapping paired end reads to “self-correct” the terminal lower quality base calls. Together the standard and short libraries will form the majority of the assembled contiguous segments (contigs). Although very long inserts (5–50 kb) cannot be directly sequenced, several “mate-pair” circularization and shearing techniques have been developed to produce these libraries (van Heesch et al. 2013). These mate pairs are useful for bridging gaps and repetitive regions to scaffold the contigs. Several de Bruijn graph-based assemblers exist for short read assembly, notably Trinity (Grabherr et al. 2011) Velvet (Zerbino and Birney 2008).

The assembly of the reference transcriptome will occur using similar library generation and assembly methods, but require less sequencing because only a proportion of the genome is usually transcribed (approximately 1%–5% in vertebrates). However, several additional considerations are specific to transcriptomes. First RNA from a variety of transcriptionally complex cells and tissues (e.g., embryos) should be sequenced to increase coverage of the transcriptome. The abundance of transcripts can vary widely across cell types, and transcriptome assembly may benefit from normalization. In addition, mRNA of coding genes by itself can be isolated for sequencing (polyA selection), or total RNA without ribosomal RNA (riboMinus), allowing the additional identification of functional noncoding genes such as microRNAs. Similar to genome assembly, de Bruijn graph-based short read assemblers can be used (for example, Trinity (Zerbino and Birney 2008), but gaps need be allowed because of splice variants.

Once the genomic and transcriptomic contigs have been constructed, they can be ordered against a genetic map. While it was once very difficult to produce a dense, high quality genetic map, GBS approaches have made this much simpler (see GBS text box). A panel of backcross or  $F_2$  individuals can be sequenced along with the progenitors of the cross, and these data can be used to create a high density genetic linkage map. Significantly, GBS approaches such as RAD produce a sufficiently large number of genetic markers that an  $F_1$  family produced from a single pair of heterozygous parents can be used to create a genetic map in pseudo-testcross format. Two independent genetic maps based only on the recombination occurring during meiosis in each parent are produced, which are then linked together by the smaller proportion of markers shared between the maps (Catchen et al. 2011, Amores et al. 2011). This approach is very useful for organisms for long lived organisms that produce many offspring (e.g., many trees). Once constructed, the large number of RAD markers (several thousand) can be used to facilitate the ordering of contigs. With even a low level of sequence coverage of an average sized genome, a large number of contigs will contain one or more identifiable RAD tags which can then be used to anchor them to the genetic map via BLAST. Similarly, transcriptomic contigs can be aligned to the contigs directly, or via associated local paired end assemblies (Etter et al. 2011b), resulting in a well annotated genome that covers 80%–95% of the DNA sequence of a new, genomically enabled organism.

SECTION I:  
HOW CAN NEXT-GENERATION SEQUENCING ENABLE STUDIES OF  
PHYLOGEOGRAPHY, POPULATION STRUCTURE, AND INTROGRESSION?

FROM POPULATION GENETICS TO POPULATION GENOMICS.—The field of population genetics comprises a rich theoretical framework and a powerful set of analytical tools for empirical studies. The overall goal of this field is to understand how mutation, selection, gene flow, and genetic drift affect patterns of genetic variation (Fisher 1930, Wright 1978, Allendorf et al. 2010). When performed in an explicitly geographic context this research often focuses on the genetic structure of populations, or the partitioning of genetic variation within and among individuals in different populations. Statistical analyses of samples can be used to estimate core population genetic parameters such as genetic diversity (e.g., nucleotide diversity  $\pi$  or average heterozygosity), patterns of nonrandom mating (Wright's inbreeding coefficient  $F$ ), as well as population parameters such as effective population size ( $N_e$ ) and migration rates ( $m$ ) among populations (Table 3, Fig. 1).

An expanding range of genetic markers has been used to infer population structure; from discrete Mendelian phenotypes to allozymes, microsatellites, and ultimately DNA sequence data. These data have in turn spawned novel theoretical approaches such as the neutral theory of molecular evolution (Kimura 1968) and the coalescent (Kingman 2000). The latter has proven to be particularly useful in population genetics for connecting demographic and evolutionary processes to patterns of genetic variation at loci (Kingman 2000). Coalescent theory focuses retrospectively on the ancestral relatedness of samples of alleles within and among populations, thereby allowing the development of probabilistic models that focus analytical effort on the samples at hand (Wakeley 2009). The basic coalescent model for a single population has been modified to allow inferences on genetic structure, migration, recombination, some kinds of selection, and demographic history (Kaplan et al. 1988, Notohara 1990, Beerli and Felsenstein 1999, Nielsen and Wakeley 2001, Drummond et al. 2005). The kinship of the coalescent with traditional phylogenetics is evident, and coalescent theory can now also describe genealogies across multiple populations and species, providing a way to merge population genetic and phylogenetic methodologies going forward (Edwards 2009).

Until recently the majority of analyses have focused on one or a small number of loci scattered throughout the genome, often in unknown locations. However, evolutionary processes such as natural selection and genetic drift act in concert with genetic factors such as dominance and epistasis, as well as linkage and recombination, to produce the structure of genomic variation observed in natural populations. Although traditional and coalescent population genetics theory has addressed interactions among loci, empirical studies have often been confined to just a handful of traditional genetic markers. Yet, recent breakthroughs in molecular genetic protocols are now allowing detailed population-level genomic studies in ways that were not possible even a few years ago (Asmann et al. 2008, Mardis 2008a,b, Marguerat et al. 2008, Shendure and Ji 2008, Pool et al. 2010, Glenn 2011, McCormack et al. 2011, Fan et al. 2012, Hohenlohe et al. 2012a). The study of large amounts of genomic data from numerous individuals across populations has been labeled population genomics (Beaumont and Balding 2004, Liti et al. 2009, Rockman and Kruglyak 2009).



### Textbox C. RAD-sequencing

The low cost of next-generation sequencing (NGS) technology has made it feasible to profile genetic variation directly by sequencing genomic DNA. Whole-genome shotgun sequencing (WGS) would be the obvious approach, but remains expensive in organisms with larger genomes and provides higher marker densities than required for many studies of ecology and evolution. Instead, recent methods focus sequencing efforts on a dense panel of loci randomly distributed across the genome. These methods genotype thousands of loci at far lower costs than WGS, in high throughput, without requiring prior sequencing information. Broadly, this specificity can be achieved through PCR amplification, sequence capture by hybridization, transcriptome sequencing, or restriction endonucleases (RE; reviewed by Cronn et al. 2012).

Here we focus on RE-based methods, which have been rapidly adopted for use in natural populations and non-model organisms (reviewed in Davey et al. 2011). Several approaches have been developed, including CRoPS (Van Orsouw et al. 2007), RAD-seq (Miller et al. 2007b, Baird et al. 2008, Etter et al. 2011a), RRLs (Van Tassel et al. 2008), GR-RSC (Maughan et al. 2009), and GBS (Elshire et al. 2011). Despite the diversity of names, these methods are united by the use of RE to fragment genomic DNA, after which sequencing libraries are prepared through various combinations of size-selection, adaptor ligation, PCR amplification, and DNA purification. In RAD-seq, for example, restriction fragments are ligated to sample-specific barcoded adaptors. Ligated fragments are then pooled and PCR amplified to enrich for fragments adjacent to restriction sites. RAD-seq enables the sequencing of large regions surrounding essentially all restriction fragments for any RE, regardless of the length of the restriction fragments, by randomly shearing the restriction fragments to a length suitable for the chosen sequencing platform. This approach subsamples the genome at homologous locations, allowing for single nucleotide polymorphisms (SNPs) to be identified and typed for thousands of markers in multiple samples with minimal investment of resources. Recently described modifications (2b-RAD: Wang et al. 2012b; ddRAD: Peterson et al. 2012) have added the ability to customize marker density during library preparation, depending on experimental designs and genome size. These approaches can increase flexibility and reduce preparation time, but with some trade-offs in efficiency and read lengths due to the removal of the shearing step.

Considering the diverse and growing array of methods available, the optimum choice may not be immediately obvious for researchers entering the field. We suggest that while each method offers advantages and tradeoffs for a particular design, all are broadly applicable. Factors that may influence the choice of methods include access to equipment or procedures each method requires (e.g., random fragmentation or isolation of a precisely defined size fraction out of continuous "smear" of DNA fragments; for overview of considerations see Davey et al. 2011). If development of targeted SNP assays is required, methods that allow for assembly of longer contigs might be favored. In planning these studies, researchers might begin by deciding on the number of SNPs required for their design; this would generally be lower when recombination is limiting (e.g., in experimental mapping populations) and higher when linkage disequilibrium is limiting (e.g., association studies in natural populations). Genome size and SNP frequency can be used to estimate the size of library (combined length of all fragments) required to genotype that number of SNPs. Comparing this target with the library size produced by each method (number of tags times length of flanking sequence included) can inform the selection of a genotyping method for a particular study. *Continues opposite page*

An important conceptual shift with the advent of NGS technologies is that the aforementioned classical statistics, and newer concepts such as the coalescent, can be visualized as continuous variables distributed across a genome (Hohenlohe et al. 2011). A critical aspect of these dense data is that they exhibit correlation among measurements at neighboring genomic regions that result from linkage disequilibrium (Slatkin 2008, Pritchard et al. 2010). The degree to which this autocorrelation itself changes along the genome reflects selection and recombination as well as other evolutionary forces (Charlesworth et al. 2009). Because of this autocorrelation, inferring the evolutionary history of any single locus is complicated by the influence of its genomic neighbors (Nielsen et al. 2005a, Boitard et al. 2009, Pickrell et al. 2009). Conversely, population genomics allows the simultaneous identification of a genome-wide average and outliers for any given statistic. The genome-wide average provides



A number of ecological and evolutionary studies have applied the RAD-seq family of techniques to organisms lacking genomic resources (Barchi et al. 2011, 2012, Baxter et al. 2011, Rowe et al. 2011, Bus et al. 2012, Everett et al. 2012, Houston et al. 2012, Lemmon and Lemmon 2012, Scaglione et al. 2012, Wang et al. 2012a, Yang et al. 2012). Parentage and relatedness, migration and gene flow, population structure, phylogeography, and phylogenetic relationships have been analyzed in aquatic systems such as cichlid species (Keller et al. 2012, Wagner et al. 2013), different lineages of trout (Hohenlohe et al. 2011, Amish et al. 2012, Everett et al. 2012, Hecht et al. 2012a,b, Miller et al. 2012) and freshwater and oceanic threespine stickleback (Hohenlohe et al. 2010, 2012b). RAD-seq can provide data across entire genomes of numerous individuals, allowing the simultaneous identification of a genome-wide average and outliers for any given statistic (Luikart et al. 2003, Nielsen et al. 2005a, Storz 2005, Bowcock 2007, Bonin 2008). Describing this variation improves our understanding of neutral effects of demographic processes such as colonization and range expansion, or the genomically-localized effects of natural selection (Beaumont and Balding 2004, Foll and Gaggiotti 2008, Gaggiotti et al. 2009, Hohenlohe et al. 2010a,b, 2012a). RAD-seq data can be used to link genotype to phenotype either through quantitative trait loci (QTL) analysis of mapping populations or correlation analysis in natural populations (e.g., Barchi et al. 2011, 2012, Chutimanitsakun et al. 2011, Houston et al. 2012, King et al. 2012, Pfender et al. 2011). Genome Wide Association Studies (GWAS) are another promising application for natural populations of marine species (Balding 2006, Rosenberg et al. 2010). For example, RAD-seq has been used to identify loci associated with migration propensity in steelhead salmon as compared to rainbow trout (Hecht et al. 2012a,b).

Analysis of RAD-seq data is conceptually straightforward: genotypes are determined by simply counting alleles in alignments or clusters of reads. A variety of statistical methods may then be employed to evaluate the confidence of genotype calls and estimate genetic parameters. Importantly, this does not require prior sequence information (a complete genome), although including this information is straightforward when available. Computational pipelines have been developed to assemble reads, identify alleles and genotypes, and track those genotypes in a statistically rigorous framework (Lynch 2009, Gompert et al. 2010). One such integrated platform is Stacks, which was initially designed for mapping crosses (Catchen et al. 2011), and has recently been extended to perform population genomic analyses (Catchen et al. 2013; e.g.,  $\pi$ ,  $F_{IS}$  and  $F_{ST}$ ). The output can be directly handled in other analytical packages such as GenePop (Rousset 2008), Arlequin (Excoffier and Lischer 2010), PhyML (Guindon et al. 2010), and Structure (Pritchard et al. 2000a, Falush et al. 2003, Falush et al. 2007). An alternative but less well integrated approach to de novo RAD-tag analysis based on Markov Cluster Learning (MCL) graph clustering algorithm (Peterson et al. 2012) has been implemented in the rtd pipeline (<https://github.com/brantp/rtd>).

As with all genotyping approaches, biases (such as null alleles in microsatellite analyses) can occur in the collection of RAD-seq data. In particular polymorphisms in or near RAD sites (see Davey et al. 2013) can lead to an underestimate of diversity and haplotype distributions. These effects are minor, and because they are linearly dependent upon polymorphism levels, can be quantified (Arnold et al. 2013, Gautier et al. 2013). For example, the coalescent simulations of RAD analyses show that for most reasonable levels of diversity and polymorphism the standard sheared RAD-seq underestimates genetic diversity by only 0.5%–1.7%, while this effect is more significant for ddRAD because of the absence of a shearing step. In addition haplotypes are lost from a small subset of genomic regions, but because of their phylogenetic positions the majority of these do not affect coalescent inferences. In addition to making adjustments based upon these simulations, existing analysis pipelines use data filters to mitigate effects by, for example, selecting only loci that are successfully genotyped in a majority of individuals and in approximate Hardy-Weinberg Equilibrium. Improving corrections for restriction site polymorphisms will be an important area of future research.

a baseline view of both neutral demographic and genetic processes. Outliers from the background may indicate the action on specific loci of evolutionary forces like natural selection, but other demographic processes may produce similar patterns (Luikart et al. 2003, Przeworski et al. 2005, Storz 2005, Butlin 2010). For example, the genetic effects of selection can mimic those of demographic factors, so that effective population size and migration rate are also continuous variables along the genome (Hohenlohe et al. 2010a). Similarly, demographic processes can also affect the

Table 1. Non-exhaustive literature search of existing publications addressing research and management questions posed at the Pan-Pacific Advanced Studies Institute (PacASI). Table includes references to both first-generation and next-generation sequencing methods.

Question	General references	Reference from Southeast Asia
How can the spatial connectivity of populations be obtained and inform management?	Manel et al. 2005, Cowen et al. 2007, Hedgecock et al. 2007, Botsford et al. 2009 (coral), Puritz et al. 2012 (seastar)	Alino et al. 2000, Campos and Alino 2008, Crandall et al. 2008 (seastar), Kockzius and Nuraynto 2008 (clam), Salayo et al. 2008, Shinmura et al. 2012 (mangrove)
How can genetics guide management of multi-species fisheries and EAFM?	Policansky and Magnuson 1998; Leslie and McLeod 2007; Palumbi et al. 2009	Armada et al. 2009, Pomeroy et al. 2010, Taylor et al. 2011
How can population boundaries be assessed and shared stocks be managed?	Begg and Waldman 1991 (fish), Pampoulie et al. 2006 (fish), Riccioni et al. 2010 (fish)	Ablan et al. 2002 (corals), Santos et al. (2010), Dammannagoda et al. 2011 (tuna), Gold et al. 2013 (cobia), Izzo et al. 2012 (sardine)
How can genetic data guide restocking or restoration efforts after a collapse?	Blakenship and Leber 1995 (fish), Hulata 2001 (fish), Bell et al. 2006 (fish), Baums 2008 (coral)	Lipcius et al. 2008, Juinio-Menez et al. 2008 (urchin), Okuzawa et al. 2008
How can effective population size be estimated in marine organisms?	Bazin et al. 2006, Charlesworth 2009, Hare et al. 2011	Arnaud-Haond et al. 2006
How can marine biodiversity be assessed at community level?	Venter et al. 2004, Chariton et al. 2010, Ardura et al. 2011, Caron et al. 2012, Quing Yun and YuHe 2011	–
How do organisms respond to environmental stress?	Ladner and Palumbi 2012, Barshis et al. 2013	–
Can individuals be screened and selected for favorable traits for mariculture?	–	Benzie et al. 2002, Yan et al. 2011, Zhao et al. 2012

variance as well as the average of genome-wide distributions (Teshima et al. 2006, Wares 2010). Therefore, the signatures of all of these neutral and non-neutral processes on both the genome-wide distribution of population genetic statistics, and on specific genomic regions, must be considered simultaneously when making inferences from population genomic data. Fortunately, as more and more annotated genomes become available from NGS methods (see Textbox B), it is becoming possible to identify the function of formerly anonymous outlier loci and experimentally test for an adaptive role (see Section II).

Performing population genomic studies has until recently been difficult to achieve in most organisms because the requisite molecular genetic techniques have been prohibitively expensive for all but a small number of model organisms (Charlesworth et al. 1997, Charlesworth 1998, Stephan et al. 2006, Begun et al. 2007, Bonin 2008, Butlin 2010, Hohenlohe et al. 2010a, Stapley et al. 2010). However, the massive amounts of genetic data generated by genotyping by sequencing (GBS) approaches such as RAD-seq (see Textbox C), and by amplicon approaches (see Textbox D) are changing the scale and nature of ecological and evolutionary genetic studies that can be performed. Numerous ecological and evolutionary studies can now be performed using GBS, and a growing number of studies have successfully employed this family of techniques, even in non-model organisms for which few genomic resources presently exist (Barchi et al. 2011, 2012, Rowe et al. 2011, Bus et al. 2012, Scaglione et al.

#### Textbox D. Amplicon sequencing.

While massively-parallel sequencing technologies can now provide enough sequence data to cover an entire genome, they can also be employed to simultaneously sequence smaller numbers of loci across a large number of individuals. This “amplicon” method is conceptually simple: genetic regions of interest are targeted with specific primers or probes, and enriched via PCR or subtractive hybridization (Hodges et al. 2009, Neiman et al. 2012). Barcodes and adapters are then added with a second PCR or ligation reaction. For PCR-based methods, these two steps may be combined by using “fusion primers” which comprise (from 5’ to 3’) the adapter sequence for the sequencing platform to be used, a barcode sequence, and the normal reverse-complemented recognition sequence for PCR amplification. Multiplexing multiple loci in a single PCR reaction can provide further efficiencies. Following PCR amplification, purification, and quantification, individual libraries can be pooled in equimolar amounts and sequenced. Given the number of reads provided by any given sequencing platform, the method can be applied to any combination of individuals  $\times$  loci, so long as adequate sequencing depth at each locus is allowed. The resultant sequence data for each locus are limited only by the read-length of the sequencing platform; they may contain multiple SNPs, indels, or other informative genetic variation. Because next-generation sequencing (NGS) methods read from single-stranded molecules, the haplotypes of each gene copy are easily inferred without further cloning or statistical haplotype reconstruction.

Until now, amplicon sequencing methods have primarily been used for characterizing microbial diversity in environmental samples (e.g., Sogin et al. 2006) and for clinical detection of mutations (e.g., Kohlmann et al. 2010). However, the applications to molecular-based ecology and evolution studies are clear: Theorists have long called for the characterization of large numbers of loci from a population sample for the estimation of population genetic parameters and phylogenies (Pluzhnikov and Donnelly 1996, Felsenstein 2006). The potential for multiple SNPs means that, like microsatellites, each locus carries more genealogical information for probabilistic coalescent or parentage models than data from SNP genotyping platforms or RAD-tags. However, unlike microsatellites, scoring genotypes from sequencing data is objective and repeatable, and the underlying mutational model is well understood.

Amplicon sequencing protocols are available for most massively parallel sequencing platforms (Glenn 2011), each of which has strengths and weaknesses with regard to this application. 454 and Ion Torrent chemistries can potentially provide longer runs (although this advantage is steadily shrinking). However, the cost per base for this approach is much higher than for Illumina Hi-Seq. While Illumina sequencing is generally cheaper, and read length potential has been increasing, there is currently an issue with sequencing multiplexed “low-complexity” libraries wherein the initial bases of different component libraries are identical and are therefore called as the same sequence cluster (Krueger et al. 2011). This can be overcome by either spiking in higher diversity libraries, (albeit at the cost of up to 50% of sequencing depth), or by including degenerate bases downstream of the sequencing primer binding site in each construct (at the cost of approximately 4 bases per read). Furthermore, Illumina has recently offered new instrument software that aims to deal with this problem. Finally, amplicon reads from Pacific Biosciences technology can be significantly longer than other platforms (up to 1 kb), although they are still relatively expensive and prone to random errors.

As of this writing, five pioneering studies had developed protocols for the identification and targeted sequencing of nuclear loci for use in studies of molecular ecology and evolution (Bybee et al. 2011, Puritz et al. 2012, Lemmon et al. 2012, Zellmer et al. 2012, O’Neill et al. 2013). However, it is still the early days for NGS sequencing of specific amplicons, and undoubtedly, these protocols will continue to evolve.

2012, Wang et al. 2012a, Yang et al. 2012). This emerging availability of genome-wide data is of particular importance for many Indo-Pacific marine species of management or conservation interest that have little to no history of molecular studies, as researchers are able to side-step the cumbersome and expensive process of developing markers, and access the genome via GBS methods. While amplicon sequencing approaches will not cover the genome as densely as other GBS approaches, the typically longer sequences will provide much more information per locus; a desirable

### Textbox E. Metagenomics through next generation sequencing.

Biodiversity studies historically focused on metazoan taxa amenable to morphological identification (e.g., Roberts et al. 2002, Willig et al. 2003). However, the advent of DNA sequencing allowed the characterization of microbial diversity from environmental samples (Pace et al. 1986, Schmidt et al. 1991), ushering in the field, metagenomics (Handelsman et al. 1998). Metagenomic approaches have explored microbial diversity in soils, biofilms, water samples and gut fauna among others (for reviews, see Riesenfeld et al. 2004, Tringe and Rubin 2005), resulting in a major transformation in how we view microbial communities. For example, Rohwer et al. (2001) showed that the majority of microbes associated with the coral *Montastraea franksi* Gregory, 1895 were novel species representing a wide diversity of taxa, but were dominated by cyanobacteria and  $\alpha$ -proteobacteria, a sharp contrast to results from culture studies. Similarly, Venter et al. (2004) identified 148 novel bacterial phylotypes and 1.2 million new genes from 1800 marine microbial taxa, changing our views of the diversity and ecological function of planktonic microbes. The power of metagenomic approaches resulted in its application to diversity studies across all kingdoms of life, including viruses (e.g., Breitbart et al. 2002, Edwards and Rohwer 2005), fungi (Anderson and Carney 2004, Unterseher et al. 2011), planktonic eukaryotes (Ryneron and Palenik 2011, Quing Yun and YuHe 2011), and paleofauna (Noonen et al. 2005, Poinar et al. 2006).

Early metagenomic studies relied on cloning, but cloning large pieces of genomic DNA proved challenging (Handelsman et al. 1998) and cloning PCR products, while easier, often yielded skewed diversity profiles depending on template concentration (PCR bias, Polz and Cavanaugh 1998). Cloning also failed to recover dominant members of microbial communities (e.g., Lindquist et al. 2005), even when other methods confirmed their presence, because the expense of cloning limited the total number of sequences obtained.

Next-generation sequencing (NGS) methods are a major advance for metagenomics with many key advantages. First, NGS methods can be based exclusively on environmental DNA extractions, eliminating PCR bias, providing more accurate diversity estimates. Second, a single NGS run provides many orders of magnitude more sequences than cloning at a fraction of the cost and effort; exceptionally high coverage captures low frequency sequences, resulting in better diversity estimates (Medinger et al. 2010). Lastly, the large numbers of bases sequenced during NGS analysis can produce significant coverage of the entire genomes of organisms with simpler genomes (e.g., microbes and viruses), allowing for the identification of diversity and analyses of novel genes and gene function (see Petrosino et al. 2009 for review), improving our understanding of the ecological function of these communities (Schloss and Handelsman 2003).

NGS data is transforming the field of metagenomics. For example, Turnbaugh et al. (2007) showed that the human microbiome has a “metagenome” that is 100 times larger than the human genome, and that human DNA is only a small portion of the human organism (Haiser and Thurnbaugh 2012). Metcalf et al. (2012) identified novel methane synthesis pathways in marine microbes providing a potential explanation to the “methane paradox” (Kiene 1991), advancing our understanding of biogeochemical cycling. NGS methods have facilitated the comparison of microbiomes in healthy and diseased individuals, improving our understanding of the disease ecology (Cox-Foster et al. 2007).

In the marine realm most NGS metagenomic research effort remains focused on marine microbial diversity (Caron et al. 2012, Gilbert and Dupont 2011, Zinger et al. 2011). However, the potential application of NGS techniques in marine environments is vast. The number of marine metazoans isn't known within an order of magnitude (Sala and Knowlton 2006, Reaka-Kudla 1997) and NGS metagenomic techniques hold great promise for illuminating this hidden diversity (Medinger et al. 2010, Bik et al. 2012). For example, metagenomics approaches have identified the composition of eukaryotic picoplankton communities (Piganeau et al. 2008, Cuvelier et al. 2010), providing new insights to the diversity and function of this community (see Ryneron and Palenik 2011 for review). Similarly, Chariton et al. (2010) explored the diversity of eukaryotic taxa within marine sediment, identifying 640 taxon clusters in 54 unique phyla. NGS-based metagenomic approaches utilizing the DNA barcoding marker COI have recently been applied to rapid assessments of biodiversity in artisanal and commercial finfish fisheries (Ardura et al. 2011). To date, DNA barcoding of marine larvae have been relatively limited in scope, yet have been remarkably effective in identifying cryptic biodiversity (e.g., Barber and Boyce 2006). NGS approaches could be applied to planktonic fish or invertebrate larvae and to marine holoplankton, magnifying our ability to detect and document biodiversity in the plankton, much like Yu et al. (2012) did for terrestrial arthropods.

*Continues opposite page*

NGS methods can also improve our accuracy of biodiversity surveys, providing markedly different estimates of biodiversity from morphological based methods (Grosillier et al. 2006, Fonseca et al. 2010, Pfrender et al. 2010), highlighting sampling and preservation biases (Bik et al. 2012). Additionally, documenting marine biodiversity in taxa ranging from viruses and microbes to larger metazoans will allow the examination of associations among marine taxa, allowing better understanding of the ecological associations and functions of marine communities. This baseline data also opens up new avenues for monitoring biodiversity changes over time, particularly with respect to anthropogenic stressors or climate change (Baird and Hajibabaei 2012).

While there is much promise in the application of NGS metagenomic methods to marine ecosystems, such studies are still in their infancy and as such, challenges still remain. Few biodiversity scientists have the computational skills to manage NGS metagenomics data (Pfrender et al. 2010, Bik et al. 2012). There are also concerns about quality control such as chimeric sequences that can result in inflated biodiversity estimates (Porazinska et al. 2010). However, metagenomic techniques will increasingly provide important insights into marine biodiversity, providing a critical complement to morphological based techniques. As informatics tools and methods are refined, NGS based metagenomic techniques will likely have a transformative effect on marine biodiversity studies, improving our understanding of the evolution and ecological function of these communities.

feature for inference from coalescent and parentage models (Bybee et al. 2011, Puritz et al. 2012, Lemmon et al. 2012, Zellmer et al. 2012).

Of course, more refined genomics studies require a whole genome sequence, and with substantial effort these same NGS approaches may be used to acquire a draft “instant genome” in non-model organisms (see Textbox B). Absent a whole genome, and with somewhat lesser effort, dense meiotic maps may be generated to enable genomics study. Such dense genomic maps as well as association maps based upon GBS are rapidly emerging in non-model organisms (Amores et al. 2011, Baxter et al. 2011, Davey et al. 2011, Everett et al. 2012, Houston et al. 2012; see Textbox B).

**TOWARD POPULATION GENOMICS IN THE INDO-PACIFIC REGION.**—GBS (including amplicon sequencing) data will prove useful to Indo-Pacific researchers for their ability to greatly improve the resolution of population genetic, phylogeographic, and phylogenetic studies. The Indo-Pacific is the largest biogeographic region on Earth, with species ranges that may span up to two-thirds of the globe (Spalding et al. 2007). Because gene flow for most of these species is mediated by pelagic larval dispersal, they will typically large coalescent effective population sizes, which tend to drive up heterozygosity and depress the maximum value of  $F_{ST}$  (Crandall et al. 2012). These large effective population sizes also ensure that genealogical histories and population structure are obfuscated by a longer expected time for lineage sorting to occur. Traditional population genetic and phylogeography studies have typically sampling a small fraction of the genome (often 1 to 20 loci). While this amount of information has certainly been sufficient to answer many important questions about Indo-Pacific population structure (Bowen et al. 2014), GBS approaches will provide many orders of magnitude more markers, which will allow integration across the universe of possible phylogeographic histories created by large marine effective population sizes (Knowles 2009), thus allowing much greater precision in parameter estimates (Rubin et al. 2012).

NGS provides a rich source of data for population and phylogeography studies. For example, an individual’s multilocus genotype provides key information that can be used to infer such important biological parameters as parentage and kinship (Anderson and Garza 2006), the population of origin of a sampled individual



## Textbox F. Gene Expression and RNA-seq.

One application of next-generation sequencing (NGS) technologies is the investigation of which genes and pathways are involved in the response to a given change. The deep sequencing coverage provided by NGS, it is now possible to sequence and quantify the entire transcriptome (the complete set of expressed transcripts). This is analogous to the cDNA microarray (Schena et al. 1995) wherein a broad suite of genes can be examined for changes in expression within the same experimental setup. The advantages of expression profiling via RNA-seq (Nagalakshmi et al. 2008) are threefold: (a) no prior genomic information is required (for a microarray you either need a large clone library or actual sequence for each gene), (b) the NGS approach has a greater dynamic range of detection, and (c) one can examine both changes in expression and also have the direct sequence for every individual in the experiment which makes the data available to a variety of additional analyses not possible with microarrays (population genomics, detection of signatures of selection). RNA-seq can be applied to a variety of biological questions such as:

Which genes respond to a particular environmental stress (e.g., temperature stress, oxidative stress, exposure to toxins, salinity stress)?

Which genes are turned on/off during various stages of development?

Which genes show rhythmic patterns of expression over time (e.g., circadian clock gene discovery)?

The approaches to all of these questions are very similar once samples are obtained from a particular experiment. Common sense experimental guidelines apply [e.g., adequate replication, control samples where appropriate; see for example Auer and Doerge (2010)]. While beyond the scope of this section, approaches such as bulk segregant analysis (e.g., Liu et al. 2012) and genetic mapping of F1 genotypes via parental crosses may be required to fully tease apart the relative influence of genetic background (e.g., Hunter et al. 2013). Lastly, working with RNA is also more challenging than working with DNA because RNA is a more sensitive and unstable molecule.

Here, we describe a case study that employed transcriptome sequencing to profile gene expression changes during heat stress in reef-building corals. The steps outlined here could be applied to any question related to the gene expression response to change (in this case, response to elevated temperature stress). This study was an attempt to examine the potential cellular mechanisms responsible for differences in upper thermal tolerance limits between two populations of corals in American Samoa (see Barshis et al. 2013 for a complete description). We performed a heat stress experiment wherein replicate clonal fragments of 5 individual coral samples from two previously characterized populations (Oliver and Palumbi 2011), termed here the tolerant and susceptible populations, were exposed to control and elevated temperatures.

There are two main types of RNA-seq, “tag-based” or “3Seq”, for species in which a reference transcriptome assembly is available, and “shotgun RNA-seq. Tag-based approaches to RNA-seq (Beck et al. 2010, Meyer et al. 2011) utilize poly-A cDNA construction so that all reads come from a single, narrow window at the 3' end of each transcript. This allows for greater determination of strand specificity, leverages existing transcriptomic resources (instead of *de novo* assembly), and provides a straightforward and cost-effective route to gene expression profiling (approximately \$50 per sample). In shotgun RNA-seq, cDNA libraries are created via random hexamer priming. This approach could introduce unnecessary noise into the data; however, this may only erode the statistical power to detect small changes in expression and would likely not obscure strong differences between treatments. We chose to construct our cDNA library using random hexamer priming, as these sequences were going to be used both for a *de novo* transcriptome assembly (since no species transcriptome reference was yet published) and gene expression profiling. We also chose to perform the more expensive paired-end sequencing on a set of the lanes to increase read length and aid in *de novo* assembly. The remaining lanes were sequenced as single-end runs to avoid the additional cost. Our target coverage was approximately 18–20 million reads per individual sample, which was more than sufficient for the purposes of the present study.

*Continues opposite page*

We assembled the sequences from every individual (single-end and paired-end) into 220,213 contigs. A more detailed description of the assembly, annotation, and data analysis steps can be found in the instructional publication describing the NGS pipeline: Simple Fools Guide to Population Genomics via RNA-seq (SFG; De Wit et al. 2012).

An additional challenge with a symbiotic organism such as coral is the presence of sequences from a variety of different taxa (e.g., coral host, dinoflagellate symbiont, coral-associated fungi, and additional microbial associates). We compared our full assembly to known cnidarian-only sequences from other species (various EST libraries and the first draft coral genome) to extract the subset of contigs that belonged only to the coral/cnidarian host. We ended up with 33,496 putative coral genes, which were used for analysis. We used a similar approach to identify the sequences for fungi (Amend et al. 2012) and *Symbiodinium* dinoflagellates (Ladner and Palumbi 2012). While contamination may be a more pronounced issue in corals due to their association with a wide variety of taxa, for any taxon it is important to consider what additional genomic resources are available that can be used to screen an assembly for sequences that come only from that particular taxon.

After sequencing and assembly, each individual sequence was mapped (i.e., aligned/matched) to the reference for differential expression analyses. We used the *DESeq* package in R (Anders and Huber 2010) to identify gene expression differences between putatively heat-tolerant and heat-susceptible corals in our experiment. Both populations demonstrated a transcriptional response to the heat stress, changing expression significantly between control and heated corals across 159 and 247 contigs (i.e., genes) in the tolerant and susceptible population, respectively. Many of these genes had similar functions and previously characterized roles in the stress response (e.g., molecular chaperones, antioxidants, apoptosis regulators). Interestingly, there were 169 genes that responded significantly to the heat stress in the sensitive corals but did not show significant change in the tolerant corals. Upon further investigation, we found that the tolerant corals not only showed a reduced response (i.e., did not change as much under heat stress) across these 169 genes, but also had higher expression of a number of these genes under control conditions, suggesting that these pathways may already be “frontloaded” or turned-on in the tolerant population. The tolerant corals from our experiment come from a reef section that undergoes extreme daily fluctuations in temperature, pH, and dissolved O<sub>2</sub> (pool 300 from Craig et al. 2001), suggesting that acquired stress tolerance in these corals may be a product of natural environmental exposure to extreme conditions. Future work is needed to demonstrate if this pattern of frontloading is a true cause of the increased tolerance or if some other mechanism may be contributing to the pattern. Additionally, we cannot determine from this first experiment if the increased tolerance of corals from the more extreme areas is genetically based (i.e. adaptive) or a product of acclimatization to the extreme conditions (i.e., phenotypic plasticity) though this represents an area of future research interest (see Barshis et al. 2013 for a full presentation and discussion of the results).

The case study outlined here illustrates a practical application of the RNA-seq approach to investigation of the molecular mechanisms responsible for a particular phenotypic difference (tolerance and sensitivity to heat stress) between two groups of corals. The primary advantages of this approach are: (1) a reference genome or transcriptome is not required, enabling this type of investigation in non-model taxa; (2) the sequence data generated can be used for many other purposes besides gene expression profiling such as SNP discovery or phylogenetic inference (Ladner et al. 2012); and (3) the product is a comprehensive scan across the majority of the transcriptome from the taxon of interest, potentially resulting in the discovery of novel gene targets or pathways. Some of the primary cautions with this approach are: (1) bioinformatics challenges are often underestimated by novice laboratories, and (2) this is only a correlative approach, thus true demonstration of a causative relationship between a particular candidate gene or pathway and a functional phenotype can take years to investigate via gene knockouts/knockdowns, mutant strains, and/or RNA-interference approaches.

(Rannala and Mountain 1997, Novembre et al. 2008), the probability that an individual is the product of recent hybridization among populations or species (Anderson and Thompson 2002), the degree of historical or recent connectivity among populations (Wilson and Rannala 2003, Pickrell and Pritchard 2012) and the effective size and demographic history of a population (Storz and Beaumont 2002, Hare et al. 2011; Table 3). Each of these methods have applications in refining our understanding of marine population structure and evolutionary history in the Indo-Pacific region,



Table 2. Select examples of next-generation sequencing studies using the following tools or methods: (A) SNP discovery, (B) RNA-seq, and (C) metagenomics. Table includes Target species/environment, individuals and populations sampled, sequencing platform, library created, estimated number of raw reads, average read size, if reads were assembled and how, putative markers discovered after filters, and time and cost estimates. For more in-depth and up-to-date review of costs by instrument, see the public database at [http://www.blueseq.com/knowledgebank/#](http://www.molecularecologist.com/next-gen-table-2b-2013).

Species	Individual/ populations	Platform	Library	No. of reads	Average read size	Assembly	Markers discovered after filters	Duration of project	Sample preparation cost	Sequencing and library prep cost	Total project costs	Source
<i>Oncorhynchus mykiss</i> (Walbaum, 1792) and <i>Oncorhynchus clarkii</i> (J. Richardson, 1836)	24/15	Illumina Genome Analyzer	Single-end	40 million	60 bp	Yes, de novo	2,923	N/A	N/A	N/A	N/A	Hohenlohe et al. 2011
<i>Meridiastra calcar</i> (Lamarek, 1816) and <i>Parvulastra exigua</i> (Lamarek, 1816)	640/16	454 GS- FLX	Amplicon	335,153	350 bp	Yes, de novo	N/A	1 yr, 1 grad student	\$12,140	\$12,420	\$24,560	Puritz et al. 2012
<i>Scarus niger</i> (Forsskål, 1775)	96/3	Illumina HiSeq	RAD-tag	233 million	100 bp	No	5,290	3 mo, 2 grad students	\$400	N/A	N/A	Stockwell, Old Dominion Univ, pers comm
<i>Caesio caning</i> (Bloch, 1791)	200/4	454 GS- FLX	genome shotgun	250,000	350 bp	Yes, de novo	N/A	N/A	\$500	\$4,368	\$4,868	Ackiss, Old Dominion Univ, pers comm
<i>Patiria miniata</i> (Brandt, 1835)	80/4	Illumina GAII	RAD-tag	15.6 million	100 bp	No	11,453	3 mo, 1 postdoc	N/A	\$1,200	N/A	Puritz, Univ Hawai'i, pers comm
<i>Cryptasterina lysteri</i> (Dartnall and Byrne, 2003)	10/1	Illumina GAII	RAD-tag	1.6 million	100 bp	Yes, de novo	8,442	3 mo, 1 postdoc	N/A	\$200	N/A	Puritz, pers comm
<i>Cellana talcosa</i> (A.A. Gould, 1846)	48/2	Illumina GAII	RAD-tag	16 million	100 bp	Yes, de novo	21,000	3 mo, 1 postdoc	\$500	\$1,200	\$5,000	Bird, Univ Hawai'i, pers comm

Table 2. Continued.

B. Gene expression using RNA-seq												
Species	Individual/ populations	Platform	Library	No. of reads	Total contigs	Assembly	Total unique significant genes	Duration of project	Sample preparation cost	Sequencing and library prep cost	Total project costs	Source
<i>Acropora hyacinthus</i> (Dana, 1846)	15/2	Illumina Hi-Seq	cDNA, single and paired-end	527 million	33,496	de novo	574	3 yrs, 1 grad student	\$3,500	\$40,000	\$50,000	Barshis et al. 2013
<i>Symbiodinium</i> sp.	15/2	Illumina Hi-Seq	cDNA, single and paired-end	223 million	26,986	de novo	1,695	Part of above project	N/A	N/A	N/A	Barshis, Old Dominion Univ, pers comm
C. Metagenomics studies												
Sample environment		Platform	Library	No. of reads	Assembly	Putative markers discovered	Duration of project	Sample preparation cost	Sequencing and library prep cost	Total project costs	Source	
Marine bacteria	-	454 GS- FLX	Shotgun	55,186	No	N/A	2 yrs, 1 grad student	\$1,500	1 run, 10 samples/ library	\$27,000	Lei, Univ Hawai'i, pers comm	
Marine bacteria- experimental	-	454 GS- FLX	Shotgun	98,620	No	N/A	1.5 yrs, 1 grad student	\$1,500	1 run, 10 samples/ library	\$5,500	Lei, Univ Hawai'i, pers comm	
Coral-algae communities	-	454 GS- FLX	Amplicon	380,000	No	Distinct community structure across interfaces	2 yrs, 2 graduate students	\$5,000	2 runs	\$100,000	Lei, Univ Hawai'i, pers comm	

Table 3. Primary genetic problems in marine conservation and how genomics can contribute to their solutions (after Allendorf et al. 2010).

Primary problem	Possible genomic solution
Estimation of effective population size ( $N_e$ ), migration rate ( $m$ ) and selection coefficient ( $s$ )	Increasing the number of markers, reconstructing pedigrees and using haplotype information will provide greater power to estimate and monitor $N_e$ and $m$ , as well as to identify migrants, estimate the direction of migration and estimate $s$ for individual loci within a population. Studies of juveniles will help to identify patterns of spatially varying selection across life history stages.
Identification of units of conservation: species, distinct population segments, and management units	The incorporation of adaptive genes and gene expression will augment our understanding of conservation units based on neutral genes and improve our ability to conduct phylogeographic studies and evaluation of efforts such as IUCN Red lists. The use of individual-based seascape genetics will help to identify boundaries between conservation units more precisely for design of fisheries management zones (FMX) and/or marine protected areas (MPAs).
Increasing power for mixture and individual assignment	Increasing the number of markers and including non-neutral markers will improve the ability to conduct mixture analyses and individual assignment to population of origin for fisheries management, enforcement, migration, bycatch, and truth-in-labeling studies of post-processed fish products. This information can help guide better management of multi-species fisheries and aid in the Ecosystem Approach to Fisheries Management (EAFM). Mixture analyses can also be used to track stocks with wide-ranging marine migrations that span national boundaries.
Predicting the ability of populations to adapt to climate change and other anthropogenic challenges	Understanding adaptive genetic variation will help to predict the response to harvesting by humans and to evaluate the ability of sensitive marine organisms to adapt to environmental stresses including rapidly changing temperature, acidity, salinity, and sea levels.
Minimizing genetic effects of restoration	Numerous markers throughout the genome could be monitored to detect whether populations used for restoration or restocking are becoming adapted to captivity, help guide the development of robust broodstock for mariculture, and serve as genomic markers for quantitative or disease-related traits.
Predicting the viability of local populations	Understanding the viability of local populations through gene expression profiling, common garden and reciprocal transplant experiments, and estimation of ratios of effective ( $N_e$ ) to census size ( $N$ ) will better inform local conservation planning.

and will benefit from the greater resolution provided by GBS data. A good example of multilocus genotype approaches in a rigorous Bayesian statistical framework is found in the software package Structure (Prichard et al. 2000a). This package allows researchers to analyze different hypotheses of population structure and determine the most likely number of population units, as well as the proportion of each individual's ancestry that is derived from each population. The sensitivity of multilocus genotype tests depends heavily upon the number of markers that have been sampled, and in general a larger number of markers greatly increase the ability to infer the aforementioned parameters. GBS-based approaches provide significantly increased power because of the very large number of markers that can be sampled across the genomes of hundreds or thousands of individuals.

An example of the potential power of GBS approaches for Indo-Pacific phylogeography can be found in a recent study the pitcher plant mosquito *Wyeomyia smithii*

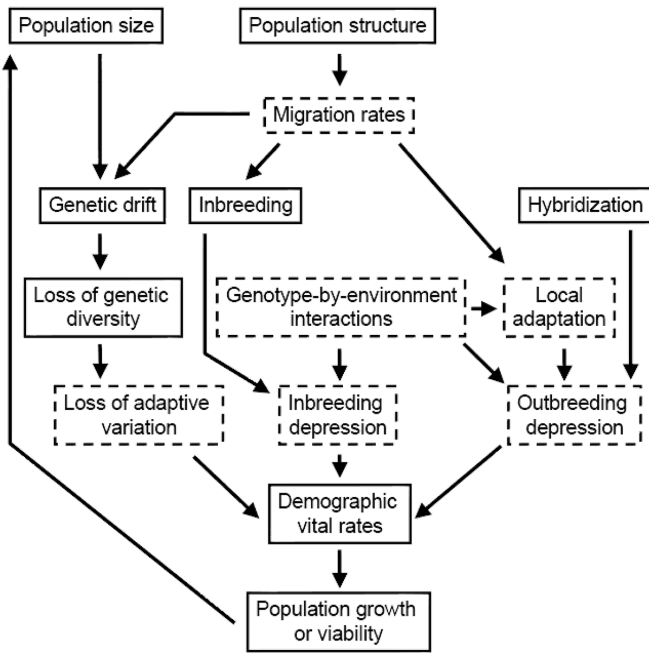


Figure 1. Schematic diagram of interacting factors to consider when considering the study of genomics natural populations (after Allendorf et al. 2010). Traditional population genetics, using small panels of mostly neutral markers, provides direct estimates of some interacting factors (solid boxes). Population genomics using genotyping by sequencing (GBS) can address a wider range of factors (dashed boxes). GBS also promises more precise estimates of neutral processes (solid boxes) and understanding of the specific genetic basis of all of these factors. For example, small panels of markers may be the best tool to estimate overall migration rates or inbreeding coefficients, whereas genomic tools can assess gene flow rates that are specific to adaptive loci or founder-specific inbreeding coefficients.

Coquillett, 1901. The postglacial phylogeography of this species was poorly estimated using mtDNA sequence data but the relationship of 15 populations from the Eastern seaboard of North America was well resolved with only a modest amount of RAD sequence data (Emerson et al. 2010). The resolution of this phylogenetic reconstruction was expanded with the addition of several other populations (Merz 2012). Other such phylogeographic studies using GBS markers have recently been completed in the carnivorous plant *Sarracenia alata* Alph. Wood (Zellmer et al. 2012), cichlid fishes in Lake Victoria (Wagner et al. 2013), ninespine stickleback *Pungitius pungitius* Linnaeus, 1758 in Scandinavia (Bruneaux et al. 2013), and recently diverged species of birds (McCormack et al. 2012). We are not aware of published phylogeographic studies utilizing GBS in the Indo-Pacific region; however, a myriad of studies are underway. These include a pilot study examining three Philippine populations of parrotfishes (B Stockwell, unpubl data), a phylogeographic study comparing two sympatric sea stars with differing life history strategies in Australia using 454 amplicons (Puritz et al. unpubl data), a GBS study of intertidal and subtidal populations of Hawaiian limpets *Cellana talcosa* A.A. Gould, 1846 (CE Bird unpubl data), a phylogenomics approach to the study of species delineation and phenotypic plasticity in Indo-Pacific scleractinian corals (ZH Forsman unpubl data), and a population

genomics study of spinner dolphins (KR Andrews unpubl data). These and other forthcoming studies utilizing GBS in the Indo-Pacific region will broaden our understanding of regional phylogeographic patterns, as well as the geological events that may have shaped them.

Pleistocene sea-level fluctuations have alternately exposed and flooded the massive Sunda and Sahul shelves, creating range expansions and contractions together with intermittent allopatry between populations in the Indian and Pacific oceans (Benzie 1999). This appears to have resulted in frequent introgression events among divergent lineages from each basin (Crandall et al. 2008, Hobbs et al. 2009, Gagnaire et al. 2011). Using data from 858 AFLPs in the catadromous eel *Anguilla marmorata* Quoy and Gaimard, 1824, Gagnaire et al. (2011) were able to detect three distinct populations to the west, east, and north of the Coral Triangle, and apparently rampant hybridization within it. The example below shows how genomic data from GBS approaches will allow even more detailed study of what may be the world's largest suture zone (Carpenter et al. 2011).

In a recent RAD-seq population genomic analysis of oceanic and freshwater stickleback from across the state of Oregon, USA (Catchen et al. 2013), fish were sampled from coastal, Willamette Basin, and central Oregon sites to address specific hypotheses of introgressive hybridization and recent range expansion. RAD-seq SNP data from nearly 1000 individuals were analyzed using Structure. The distribution of Bayesian posterior probabilities of group assignment for each individual with respect to their collecting location exhibited a clear phylogeographic break between coastal and inland populations. In addition, Willamette Basin and central Oregon populations formed a clade of closely related populations, a finding consistent with a recent introduction of stickleback into central Oregon. Interestingly, coastal oceanic and freshwater populations exhibited significant number of individuals with a range of multilocus genotypes from both populations, possibly the result of ongoing hybridization between fish from the two habitats. In contrast to the coastal populations, the Willamette Basin and central Oregon populations exhibited a clear pattern in which the central Oregon specific genotypic combinations were present at low frequency in the Willamette Basin. These data clearly support a recent human introduction of stickleback into habitats east of the Cascade Range, and demonstrate the power of GBS data to produce multilocus genotype data that can elucidate very recent demographic patterns of population connectivity.

Studies examining population structure, recent demographic patterns, phylogenetic relationships, and other aspects within the emerging field of population genomics are benefitting greatly from rich data sets generated by NGS methods. The development of more "user-friendly" software and bioinformatics tools will be the next advance in the field with progress already underway in making these programs broadly accessible. By utilizing large amounts of genomic data that can be sampled from both across the genome and across population, broad-brush perspectives can provide a robust exploration of the demographic, ecological and evolutionary questions central to conservation issues in Southeast Asia. Of the research questions presented by scientists and resource managers at PacASI (Table 1), those targeting the

connectivity and boundary of commercially-important or threatened species, and quantifying genetic variability and population structure for restoration and management are well-suited for the GBS approaches discussed here.

## SECTION II: HOW CAN NEXT-GENERATION SEQUENCING ASSESS LOCAL ADAPTATION IN THE MARINE ENVIRONMENT?

Adaptation to local habitat, being a result of evolutionary divergence in response to spatially variable environment, provides arguably the best experimental platform to elucidate basic mechanisms of evolution (Kawecki and Ebert 2004), which can now be supplemented by molecular insights from genome-wide genetic variation and gene expression profiling based on NGS (Stapley et al. 2010, Radwan and Babik 2012). These studies are of great interest for marine ecology and conservation for four major reasons. First, spatial patterns of adaptation could indicate which types of habitat are more stressful for the studied organisms, which in the marine environment can be far from obvious. Second, understanding the genes and molecular pathways modified during adaptation can identify the mechanistic basis for differences between populations and highlight functional constraints on adaptive responses. Third, availability of genetic markers of adaptation would allow identifying populations at high- and low-risk of extinction with respect to changing selection pressure. Fourth, genetic studies could inform us about the rates of adaptation and the demographic processes involved. A deeper understanding of these different aspects of adaptation would greatly help prioritizing conservation efforts.

We are only beginning to take advantage of the NGS opportunities, and much remains to be learned in the ways of experimental design, data analysis, and especially interpretation and validation. In this section, we outline several questions related to genome-environment interactions in space that can be considerably informed by NGS methods.

**BEFORE NGS: ESTABLISHING THE FACT OF ADAPTIVE DIVERGENCE BETWEEN POPULATIONS.**—A population is said to exhibit local adaptation if its members are more fit in their native habitat than individuals immigrating or transplanted from other locations (Kawecki and Ebert 2004). Two types of experiments are traditionally used to demonstrate local adaptation: reciprocal transplantation (RT) and common garden (CG) (Kawecki and Ebert 2004, Sotka 2005, Sanford and Kelly 2011). RT experiments compare fitness of individuals from different populations when out-planted together to their respective field environments, while CG experiments bring individuals from different populations into the same controlled environment and directly test potential selective factors, looking for population-wise difference in performance. These two approaches are complementary: RTs establish the fact of local adaptation but do not identify the factors driving it, while CGs address the effect of particular factors but may overlook some of the important ones, and therefore cannot guarantee that local adaptation, if present, will always be detected.

Local adaptation has two components: physiological adjustments at the level of an individual (physiological adaptation, or acclimatization), and changes in population-wide frequencies of the alleles affecting organisms' physiology (genetic adaptation). Genetic adaptation in particular is a continuous source of excitement for biologists,

since it directly relates to fundamental evolutionary processes, and has been a subject of excellent reviews recently (Sotka 2005, Sanford and Kelly 2011). The interest in the genetic component of adaptation is such that “local adaptation” as a term is sometimes defined as the genetic component only (Sanford and Kelly 2011). An acceptable way to achieve genetics-only contrast (i.e., free from effects of prior exposure to different environments and from environment-driven maternal effects) is to compare second-generation offspring reared in CG conditions (e.g., Torres-Dowdall et al. 2012), which is not feasible for most marine organisms. In the marine environment, the individuals for RT and CG experiments are typically collected from the field, and therefore there is always a question to which extent the population-wise differences, if observed, are due to genetics vs long-term acclimatization (Sanford and Kelly 2011). Still, much can be learned about the organisms’ ecology by considering genetic and physiological components of adaptation jointly.

Whether the adaptation is due to genetic modifications or long-term acclimatization for which the effects do not dissipate over the length of the experiment [typically, weeks to a year (Sotka 2005, Sanford and Kelly 2011)], the results can still address several fundamental questions about the interaction of the organisms with their environment. RT experiments can determine which environment of the ones tested is more challenging, and whether there are tradeoffs to being able to survive in it, such as diminished fitness in other environments. CG experiments can identify the factors that impose the greatest fitness costs in their natural habitat, and therefore should be the primary focus of attention for marine conservation. A good example of this is the contrast between the degree of coral decline at inshore patch reefs and the offshore reef tract in the Florida Keys. Inshore patch reefs are characterized by elevated turbidity, sedimentation, nutrients, and high temperature variation (Boyer and Briceno 2011), all of which affect coral growth detrimentally in the lab (Stafford-Smith 1993, Ferrier-Pages et al. 2001, Jokiel 2004, Fabricius 2005). The offshore reef tract, on the other hand, is characterized by mild temperature and low turbidity, and generally one would expect that buffering by the Florida Current (a part of the Gulf Stream) and remoteness from sources of pollution on shore would facilitate better coral survival there. Contrary to this expectation, inshore patch reefs in the Florida Keys consistently exhibit higher coral cover and higher coral growth rates than offshore reefs (Causey et al. 2002, Lirman and Fong 2007). A recent review covering coral decline in the Florida Keys suggests that the scientific community should stop debating the causes of reef decline and initiate a blanket strategy to reduce all threats, particularly those of anthropogenic origin (Pandolfi et al. 2005). Yet anthropogenic input does not seem to be responsible for the decline of offshore reefs; while action is needed, how can managers address threats that are undefined? RT and CG experiments can help refine our basic understanding of coral reef stressors, in the Florida Keys, Southeast Asia, and globally, to generate predictive models that match anomalous patterns. Notably, NGS technologies are not necessary for this initial stage of analysis; they come into play later when the presence of adaptive divergence between populations is already established and the focus of inquiry shifts towards its physiological and genetic mechanisms.

**PHYSIOLOGICAL MECHANISMS OF ADAPTATION.**—A major challenge in studies of local adaptation based on RT and CG experiments is to identify its environmental drivers. Guessing from the physical parameters of the environment can be



misleading, and organism-level fitness proxies assessed in these experiments are, by design, not specific enough to reflect the action of any particular factor. Even if CG experiments indicate that populations have different responses to some factor, this does not necessarily mean that this is indeed the major driver of adaptation.

Global gene expression profiling using RNA-sequencing (RNA-seq; Wang et al. 2009, Textbox F) and in particular the cost-efficient tag-based version of it (Meyer et al. 2011), is the NGS-based approach that can help in this case, serving as a powerful tool to generate hypotheses for subsequent testing in CG experiments. The basic idea is to compare gene expression profiles in individuals undergoing an RT experiment, to identify groups of genes that are regulated in response to transplantation and, in the same time, are consistently expressed at different levels in individuals from different populations (see Textbox F for details). Association of these genes with particular organismal functions, such as heat stress response, immunity, detoxification, or various aspects of metabolism, would provide a good lead to the environmental factors driving physiological divergence between populations (Ouborg et al. 2010), whether based on acclimatization or genetics. For example, Barshis et al. (2013) used RNA-seq to evaluate molecular responses to short-term temperature stress in corals from tidal pools in American Samoa that experience different variability in daily temperature regimes. Barshis et al. (2013) found that corals from the hotter pools exhibit constitutively elevated expression of stress response genes such as heat shock proteins, antioxidant enzymes, apoptosis and tumor suppression factors, and innate immune components. They suggested that this mechanism of “frontloading” transcription might facilitate increased thermotolerance in reef-building corals by making them more resistant to the acute stress exposure. Another study (Kenkel et al. 2013) used a cost-efficient tag-based variant of RNA-seq to identify a number of genes whose expression was associated with higher thermotolerance of inshore coral population relative to offshore population. In contrast to the study by Barshis et al. (2013), the main signal in this case was related to adjustments of the core metabolic processes, suggesting divergent strategies to manage energy budgets.

**GENETIC MECHANISMS OF ADAPTATION IN THE SEA.**—To find a gene, or genes, under divergent selection across environments is what nearly every evolutionary biologist interested in adaptation hopes for. The population genetics and molecular function of such gene would shed light on all aspects of adaptation, such as which environmental factors affect the organism most, what are the demographic processes leading to divergence, and which molecular mechanisms are responsible for the phenotypic change (Stapley et al. 2010). In addition, such “adaptation markers” would have a great value for marine conservation since they could be used as indicators of population’s vulnerability in the face of anthropogenic stressors, including climate change.

It has been argued that local adaptation in the sea can result in “phenotype-environment mismatch” (Marshall et al. 2010), which is specialization for a particular kind of environment manifested as reduced fitness in any other environments. For marine organisms that do not move much (or at all) as adults have the potential to disperse very far as larvae, such specialization can have interesting consequences: the degree of genetic connectivity might become more dependent on environmental similarity between locations rather than the time it would take a larva to traverse the distance between them. For example, the pattern of coral decline in the Florida Keys

mentioned above might be partially attributable to the inability of surviving inshore corals to colonize the devastated offshore reefs because of their specialization for inshore habitat.

One particular problem pertaining to local adaptation in the marine environment that merits an in-depth investigation in the near future is the tendency of marine organisms to have broad connectivity ranges relative to the scale of environmental heterogeneity that would require local adaptation. A good example is, once again, the contrast between inshore and offshore reef habitats in barrier reef systems such as the Great Barrier Reef or the Florida Keys. These environments are different in a number of physical and biological parameters (Mieog et al. 2009, Boyer and Briceno 2011), and yet the distance between them is at least an order of magnitude smaller than the dispersal distance of the planktonic larvae of most resident organisms (Cowen and Sponaugle 2009). In the terrestrial environment, this would present a serious problem since one of the key prerequisites to achieve local genetic adaptation is restricted migration between populations (Felsenstein 1976). Otherwise, immigration of maladapted genotypes could swamp the adaptive gene pool, which in theory may even result in the extinction of the population, or “migrational melt-down” (Ronce and Kirkpatrick 2001). In the sea, however, the migratory stages in the vast majority of cases are larvae rather than reproductively capable adults, so that the migration does not immediately result in gene flow. Upon their recruitment to a particular habitat, selection has considerable time to deplete the unfit individuals before they reach maturity, so that the resulting adult population would show a much higher proportion of individuals possessing genetic variants and traits matching the local environment than the original population of recruits. In most terrestrial animals, the migrating individuals are adults ready to reproduce, thus their misadaptation must be much more pronounced to achieve the same degree of exclusion from the reproductive pool. In contrast, in the sea the time lag between recruitment and reproduction would allow selection to be very efficient, so that local genetic adaptation would be achievable within each generation despite high migration, as long as the adaptive polymorphisms are segregating in the metapopulation (Levene 1953, Hedgecock 1986, Slatkin 1987). Some authors choose to call this mechanism “spatially varying selection” (Gagnaire et al. 2012a), “balanced polymorphism” (Sanford and Kelly 2011) or “polymorphism-based local selection” (Somero 2010) rather than “local adaptation,” to emphasize differences in underlying demographic processes. Importantly for the GBS approaches to detect such adaptive genetic variants (see below), this mode of adaptation might be very advantageous, since the rest of the genome outside adaptive loci can be expected to show little or no genetic differentiation because of pervasive gene flow (unless selection is extremely strong and eliminates all or nearly all the individuals bearing unfavorable alleles). An excellent recent example of how selection can shape locally adapted populations from a common pool of larvae is the story of the panmictic American eels, which all spawn in the Sargasso sea but then end up in freshwater basins all along the Atlantic coast of North America, from Florida to Labrador, exhibiting strong genetic differentiation only at some, potentially adaptive, genetic loci (Gagnaire et al. 2012b). In marine invertebrates, the evidence for this adaptation mechanism still amounts to just a few studies (Schmidt and Ran 2001, Pespeni et al. 2012), neither in the Indo-West Pacific region.

**IDENTIFYING SIGNATURE OF GENETIC SELECTION FROM POPULATION GENOMICS VIA GBS.**—Because of the massive amounts of genomic data, GBS approaches provide not only global estimates of population structure and gene flow (see Section I above), but allow determination of the variance across the genome of the effects of evolutionary and ecological processes on the structuring of genetic diversity. In particular, natural selection shapes patterns of genetic variation among individuals, populations, and species, and it does so differentially across genomes. Tests for selection in population genomics are diverse (for an overview, see Hohenlohe et al. 2010b). These assays of genetic variation focus on a plethora of genetic patterns, from nucleotide diversity and allele frequency spectra within and among populations, to haplotype structure and linkage disequilibrium, to fixed DNA sequence divergence among related taxa. The influence of selection on single loci in natural populations can be tentatively inferred from summary statistics describing population differentiation (e.g.,  $F_{ST}$ ; Beaumont 2005), allele frequency spectrum (Tajima's  $D$ ; Tajima 1989), or genetic sequence divergence between species ( $dN/dS$ ; Yang and Nelsen 1998) at specific loci; however, this inference must still be confirmed in cross-validation studies (see below).

One promise of a GBS-based population genomics approach is the simultaneous identification of both a genome-wide average and outliers for any given statistic, whether traditional measures of allele frequencies or aspects of coalescent genealogy. The genome-wide average is taken to provide a baseline examination of neutral processes, both demographic (e.g., population size, migration rate) and genetic (e.g., mutation rate, recombination). Outliers from this background can indicate the action of natural selection or other evolutionary forces on specific loci (Luikart et al. 2003; but see Hohenlohe et al. 2010b for potential difficulties). The GBS family of methods is well suited to identify genomic regions under natural selection in differentiated populations because GBS markers occur at high density throughout the genome and are therefore likely to be found in close proximity of the loci under selection, at a physical distance less than the extend of linkage disequilibrium. Genetic adaptation due to divergent selection at certain loci would be manifested at increased  $F_{ST}$  accompanied by a decrease in genetic diversity within either or both of the diverging populations. Other forms of selection can also be inferred: for example, genomic regions under balancing selection, either because of over-dominance or frequency dependent selection, would be detectable as significant increases in genetic diversity and heterozygosity over that seen in the wider genome. RAD-seq approaches have already been used in many such studies on organisms from nematode worms (Andersen et al. 2012), to butterflies (Nadeau et al. 2013), plants (Stolting et al. 2013, Andrew et al. 2013), but especially fishes (Hohenlohe et al. 2010b, 2011, 2012b, Amish et al. 2012, Everett et al. 2012, Hecht et al. 2012a,b, Keller et al. 2012, Miller et al. 2012a, Wagner et al. 2013). It must be noted, however, that GBS methods, with the only exception of full genome resequencing, do not guarantee that markers would be found at or near every locus under selection, and hence there is always a possibility that some loci under selection would escape detection.

Perhaps the most powerful inference about selection affecting specific loci in the genome can be made from analyzing multiple pairs of populations that diverged in parallel in response to the same environmental gradient. While  $F_{ST}$  outliers in a single pair of populations can be indicative of adaptive evolution due to selection, other factors such as hidden population structure or stochastic processes can result

in false positives (e.g., Excoffier et al. 2009). However, finding  $F_{ST}$  outliers localized to the same genetic region in additional, independently diverging pairs of populations would constitute much stronger evidence for the adaptive significance of that region with respect to the environmental gradient in question. For example, Hohenlohe et al. (2010a, 2011) used a high-throughput sequencing technique to genotype thousands of single nucleotide polymorphisms (SNPs) in two oceanic and three independently derived freshwater populations of Alaskan threespine sticklebacks, *Gasterosteus aculeatus* Linnaeus, 1758. Each of the three freshwater populations exhibited elevated genetic differentiation from the oceanic ancestor at many of the same genomic regions, strongly suggesting adaptive divergence due to spatially variable selection. It is important to note, however, that selection may have acted on just one or two genes in each of these regions, but the extent of elevated population differentiation—the effect of selection—may cover dozens of genes in each region due to extensive linkage disequilibrium (Hohenlohe et al. 2010b).

**NEED FOR CROSS-VALIDATION.**—Despite the promise of GBS and RNA-seq, each of these approaches in isolation is not much more than a hypothesis-generating tool since each of these methods is prone to artifacts both of technical and analytical nature. The most convincing inference therefore can potentially be achieved by combining the two approaches in a cross-validating study involving both genotyping and gene expression profiling. Since the majority of adaptive mutations are hypothesized to be cis-regulatory (Wray 2007, Jones et al. 2012), a protein-coding gene under divergent selection between populations is expected to show both the signature of genetic divergence revealed by GBS and evidence of differential expression in CG and RT experiments revealed by RNA-seq. The third approach that can be added to these two is mapping of quantitative trait loci (QTL) for phenotypic traits that diverge between populations. This approach is particularly feasible in broadcast-spawning marine organisms such as reef-building corals from which abundant outbred F1 progeny can be obtained. RAD and related methods (Textbox C) can be used for associating genotypic variation in the F1 progeny with trait variation. We hope that in the future additional adaptation-related studies will be designed to take advantage of these cross-validation opportunities.

## CONCLUSIONS

As marine conservation challenges of food security, biodiversity loss, and climate change move to the forefront of national priorities and are the focus of region-wide conservation initiatives in Southeast Asia (George and Hussin 2010, Fidelman et al. 2012), genetics and genomics becomes an increasingly important tool in marine research and management to the region. Addressing these challenges will benefit from greater investments in molecular research infrastructure, a step recognized by Southeast Asian nations with the establishment of national genomic centers including the Thailand Genome Institute (<http://gi.biotec.or.th/GI>) and the Philippines Genomic Center (<http://pgcbioinformatics.blogspot.com/>). The importance of genetics and genomics is further recognized in national development plans, such as the Comprehensive National Fisheries Industry Development Plan (CNFIDP) of the Philippines.

However, many researchers and scientists have yet to fully understand the nuances of NGS to maximize its potential applications for marine research and conservation. Since food security is a primary priority in many countries in Southeast Asia, research using molecular techniques is focused on aquaculture and commercially important species and typically utilizes a limited number of DNA markers. To date, the few publications from the region utilizing NGS methods are preliminary investigations and remain focused on aquaculture and marine biotechnology (Yan et al. 2011, Arockiaraj et al. 2012, Lluisma et al. 2012, Zhao et al. 2012, Maralit and Santos unpubl data).

As described throughout this paper, knowledge on the ecology and evolution of marine organisms can also be gained from NGS studies (Table 3). Not only will this allow for greater understanding of the target species, but more importantly in the region, it can help improve marine and resource management, which in turn will promote sustainable harvest and conservation to meet food security objectives. Research and education partnerships such as ctPIRE (for detailed review, see Barber et al. 2014) and PacASI, which are geared towards initiating capacity building of Southeast Asian researchers and students marine molecular genomics, are very good catalysts for Southeast Asian nations to fully embrace and apply genetics and genomics. Further, these partnerships open new opportunities to foreign scientists to expand their research efforts and build new collaborations to uncover answers to marine conservation challenges.

Increased accessibility to NGS tools is revolutionizing marine research and generating excitement and anxiety for marine scientists who are eager to add these methods to their research toolbox. Here, we highlight a few key issues discussed in the PacASI for NGS neophytes to consider as they explore the utility of NGS in their research.

**SCIENCE, NOT THE TECHNOLOGY, STEERS THE RESEARCH.**—Given its importance, we reiterate the value of scientists first deciding if NGS is the most appropriate tool for obtaining the data set needed to answer their research question. NGS tools may be at a disadvantage to first-generation sequencing methods that may be cheaper, more time efficient, and benefit from a well-developed data analysis infrastructure.

**DATA MANAGEMENT PLANNING.**—In most cases, obtaining data from NGS methods is the easier part of a study. The larger challenges are data storage, management, analysis and interpretation. Further, with whole genome sequencing on the horizon, marine scientists must figure out how the information can and should be used. Planning for the bioinformatics component of any NGS study is not trivial and ample time and resources must be allocated to this process (see Table 2 for examples). Although collaboration with trained bioinformaticians can be productive at this stage, we emphasize that biologists should not plan to outsource all details of the analysis, just as they would not rely entirely on professional statisticians for basic analyses such as ANOVA. Appreciating the bioinformatics and data management requirements beforehand will reduce frustration at a later stage, after the data are obtained.

**UTILIZING EXISTING DATA SETS.**—NGS projects can be greatly aided by existing transcriptome and genome data, particularly if those data have already been used to create a reference assembly. Generally, this can save the investigator the time

and additional costs associated with de novo assemblies for new taxa of interest. However, caution should be taken when using a reference assembly from a separate study, as sequence divergence between the reference individual/population and the study individual/population may create additional mismatches that could prevent successful mapping of the new data to the old assembly. This can usually be corrected in the specific mismatch thresholds, used during the mapping process. Many studies have successfully used previous transcriptome/genome assemblies to investigate genome-wide divergence patterns among phenotypes (Hohenlohe et al. 2011, Jones et al. 2012) or multiple lineages (e.g., Amemiya et al. 2013, Axelsson et al. 2013) in a combinatorial approach, in which all or parts of the genome are re-sequenced from multiple new samples of interest and existing assemblies are utilized to map and compare sequence diversity and divergence among samples. Lastly, a properly annotated genome or transcriptome can allow for functional investigation of any interesting patterns of divergence. For instance, after identifying RAD-sequences or transcripts that show signs of selection, one could interrogate neighboring sequence regions for specific genes and regulatory pathways upon which selection might be acting (e.g., starch metabolism in domesticated dogs; Axelsson et al. 2013).

**DATA AND RESOURCE SHARING.**—Lastly, we strongly encourage marine scientists to include a strategy of how to share and disseminate their discoveries when applying NGS tools to marine research questions. This can be accomplished in the traditional fashion of submitting genomic data to public domain databases and publishing resource notes, as well as in the very rewarding ways of mentoring young scientists to utilize these tools in their research, organizing forums for the exchange of tools and ideas, and by growing research collaborations between laboratories in developing and developed nations.

The success of the PacASI served as a first step in bringing together scientists from both sides of the Pacific in realizing the potential of pan-Pacific collaborations in using NGS tools to help address challenges of global climate change and to preserving imperiled food stocks and centers of biodiversity. The next steps will come as scientists from across the Indo-Pacific region work to address these questions and decide if they indeed want to use next-generation sequencing in marine systems.

#### AUTHOR CONTRIBUTIONS

The PacASI meeting was organized by DA Willette and KE Carpenter, and this manuscript was conceived and developed by DA Willette and JE Seeb. WA Cresko, ED Crandall, and E Meyer wrote the population genomics section; and MV Matz wrote the local adaptation section. The Instant genome textbox was written by WA Cresko; RAD-seq textbox by WA Cresko, ED Crandall, and E Meyer; Amplicon sequencing by ED Crandall; Metagenomics by P Barber; and Gene expression and RNA-seq by DJ Barshis. Tables and figures were created by DA Willette, LW Seeb, JE Seeb, and FW Allendorf. I Fernandez-Silva, MD Santos, and all above authors contributed in writing and editing the manuscript.

#### ACKNOWLEDGMENTS

The authors would like to thank H Calumpang, Silliman University, and Silliman University Institute of Environmental and Marine Science for co-hosting the PacASI meeting where this paper was originally proposed. Also, the authors extend thanks to the participants



of the PacASI that made generous contributions to robust discussions during the meeting. Lastly, we thank three anonymous reviewers and the guest editor for their constructive comments that improved the manuscript. The PacASI was supported by NSF grant OISE 1206614 to KE Carpenter, and while writing this manuscript, D Willette was supported by NSF grant OISE 0730256 to KE Carpenter and P Barber.

#### LITERATURE CITED

- Ablan M, McManus JW, Chen CA, Shao KT, Bell J, Cabanaban AS, Tuanh VS, Arthana JW. 2002. Meso-scale transboundary units for the management of coral reefs in the South China Sea area. *WorldFish Center Quarterly*. 25:4–9.
- Alino PM, Palomar NE, Arceo HO, Uychiaoco AT. 2000. Challenges and opportunities for marine protected area (MPA) in the Philippines. *Proc 9th Intl Coral Reef Symp*. 9:1–6.
- Allendorf FW, Hohenlohe PA, Luikart G. 2010. Genomics and the future of conservation genetics. *Nat Rev Genet*. 11(10):697–709. PMID:20847747. <http://dx.doi.org/10.1038/nrg2844>
- Amemiya CT, Alfoldi J, Lee AP, Fan S, Philippe H, MacCallum I, Braasch I, Manousaki I, Schneider I, Rohner N, et al. 2013. The African coelacanth genome provides insight into tetrapod evolution. *Nature*. 496:311–316. PMID:23598338. <http://dx.doi.org/10.1038/nature12027>
- Amend A, Barshis DJ, Oliver TA. 2012. Coral-associated marine fungi form novel lineages and heterogeneous assemblages. *ISME J*. 6:1291–1301. PMID:22189500. PMCid:PMC3379630. <http://dx.doi.org/10.1038/ismej.2011.193>
- Amish SJ, Hohenlohe PA, Painter S, Leary R, Muhlfeld C, Allendorf FW, Luikart G. 2012. RAD sequencing yields a high success rate for westslope cutthroat and rainbow trout species-diagnostic SNP assays. *Mol Ecol Res*. 12:653–660. PMID:22672623. <http://dx.doi.org/10.1111/j.1755-0998.2012.03157.x>
- Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait J. 2011. Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics*. 188(4):799–808. PMID:21828280. PMCid:PMC3176089. <http://dx.doi.org/10.1534/genetics.111.127324>
- Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol*. 11:R106. PMID:20979621. PMCid:PMC3218662. <http://dx.doi.org/10.1186/gb-2010-11-10-r106>
- Anderson E, Thompson E. 2002. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*. 160(3):1217–1229. PMID:11901135. PMCid:PMC1462008.
- Anderson IC, Cairney JW. 2004. Diversity and ecology of soil fungal communities: increased understanding through the application of molecular techniques. *Environ Microbiol*. 6(8):769–779. PMID:15250879. <http://dx.doi.org/10.1111/j.1462-2920.2004.00675.x>
- Anderson E, Garza J. 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics*. 172(4):2567–2582. PMID:16387880. PMCid:PMC1456362. <http://dx.doi.org/10.1534/genetics.105.048074>
- Anderson J, Rodriguez Mari A, Braasch I, Amores A, Hohenlohe P, Batzel P, Postlethwait J. 2012. Multiple sex-associated regions and a putative sex chromosome in zebrafish revealed by RAD mapping and population genomics. *PLoS One*. 7(7):e40701. PMID:22792396. PMCid:PMC3392230. <http://dx.doi.org/10.1371/journal.pone.0040701>
- Andrew R, Kane N, Bauté G, Grassa C, Rieseberg L. 2013. Recent nonhybrid origin of sunflower ecotypes in a novel habitat. *Mol Ecol*. 22(3):799–813. PMID:23072494. <http://dx.doi.org/10.1111/mec.12038>
- Ardura A, Planes S, Garcia-Vazquez E. 2011. Beyond biodiversity: fish metagenomes. *PLoS One*. 6(8). PMID:21829636. PMCid:PMC3150381. <http://dx.doi.org/10.1371/journal.pone.0022592>



- Armada N, White A, Christie P. 2009. Managing fisheries resources in Danajon Bank, Bohol, Philippines: an ecosystem-based approach. *Coast Manage.* 37(3):308–330. <http://dx.doi.org/10.1080/08920750902851609>
- Arnold B, Corbett-Detig RB, Hartl D, Bomblies K. 2013. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotypes sampling. *Mol Ecol.* 22:3179–3190. PMID:23551379. <http://dx.doi.org/10.1111/mec.12276>
- Arockiaraj J, Easwaran S, Vanaraja P, Singh A, Othman RY, Bhassu S. 2012. Effect of infectious hypodermal and haematopoietic necrosis virus (IHHNV) infection on caspase 3c expression and activity in freshwater prawn *Macrobrachium rosenbergii*. *Fish Shell Immunol.* 32:161–169. PMID:22119573. <http://dx.doi.org/10.1016/j.fsi.2011.11.006>
- Asmann Y, Wallace M, Thompson E. 2008. Transcriptome profiling using next-generation sequencing. *Gastroenterol.* 135(5):1466–1468. PMID:18848555. <http://dx.doi.org/10.1053/j.gastro.2008.09.042>
- Auer P, Doerge R. 2010. Statistical design and analysis of RNA sequencing data. *Genetics.* 185:405–416. PMID:20439781. PMCid:PMC2881125. <http://dx.doi.org/10.1534/genetics.110.114983>
- Axelsson E, Ratnakumar A, Arendt MJ, Maqbool K, Webster MT, Perloski M, Liberg O, Arnemo JM, Hedhammar A, Lindblad-Toh K. 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature.* 495:360–364. PMID:23354050. <http://dx.doi.org/10.1038/nature11837>
- Baird DJ, Hajibabaei M. 2012. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Mol Ecol.* 21(8):2039–2044. <http://dx.doi.org/10.1111/j.1365-294X.2012.05519.x>
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One.* 3(10):e3376. PMID:18852878. PMCid:PMC2557064. <http://dx.doi.org/10.1371/journal.pone.0003376>
- Balding DJ. 2006. A tutorial on statistical methods for population association studies. *Nature Rev Genet.* 7:781–791. PMID:16983374. <http://dx.doi.org/10.1038/nrg1916>
- Barber PH, Ablan M, Ambariyanto, Berlinck R, Cahyani D, Crandall E, Gotanco R, Junio-Menez A, Mahardika N, Shankar K, et al. 2014. Advancing biodiversity research in developing countries: the need for a new paradigm. *Bull Mar Sci.* 90:187–210. <http://dx.doi.org/10.5343/bms.2012.1108>
- Barber P, Boyce SL. 2006. Estimating diversity of Indo-Pacific coral reef stomatopods through DNA barcoding of stomatopod larvae. *Proc R Soc Biol Sci.* 273(1597):2053–2061. PMID:16846913. PMCid:PMC1635474. <http://dx.doi.org/10.1098/rspb.2006.3540>
- Barchi L, Lanteri S, Portis E, Acquadro A, Vale G, Toppino L, Rotino G. 2011. Identification of SNP and SSR markers in eggplant using RAD tag sequencing. *BMC Genomics.* 12:304. PMID:21663628. PMCid:PMC3128069. <http://dx.doi.org/10.1186/1471-2164-12-304>
- Barchi L, Lanteri S, Portis E, Vale G, Volante A, Pulcini L, Ciriaci T, Acciarri N, Barbierato V, Toppino L, et al. 2012. A RAD tag derived marker based eggplant linkage map and the location of QTLs determining anthocyanin pigmentation. *PLoS One.* 7(8):e43740. PMID:22912903. PMCid:PMC3422253. <http://dx.doi.org/10.1371/journal.pone.0043740>
- Barshis DJ, Ladner JT, Oliver TA, Seneca F, Traylor-Knowles N, Palumbi SR. 2013. Genomic basis for coral resilience to climate change. *Proc Natl Acad Sci.* 110:1387–1392. PMID:23297204. PMCid:PMC3557039. <http://dx.doi.org/10.1073/pnas.1210224110>
- Baums IB. 2008. A synopsis of coral restoration genetics. *In: Leewis RJ, Janse M, editors. Advances in coral husbandry in public aquariums. Public Aquarium Husbandry Series Vol 2.* p. 335–338.
- Baxter S, Davey J, Johnston J, Shelton A, Heckel D, Jiggins C, Blaxter M. 2011. Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PLoS One.* 6(4):e19315. PMID:21541297. PMCid:PMC3082572. <http://dx.doi.org/10.1371/journal.pone.0019315>

- Bazin E, Glemin S, Galtier. 2006. Population size does not influence mitochondrial genetic diversity in animals. *Science*. 312:570–572. PMID:16645093. <http://dx.doi.org/10.1126/science.1122033>
- Beaumont MA, Balding DJ. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol*. 13(4):969–980. <http://dx.doi.org/10.1111/j.1365-294X.2004.02125.x>
- Beaumont M. 2005. Adaptation and speciation: what can  $F_{st}$  tell us? *Trends Ecol Evol*. 20(8):435–440. PMID:16701414. <http://dx.doi.org/10.1016/j.tree.2005.05.017>
- Beck A, Weng Z, Witten D, Zhu S, Foley J, Lacroute P, Smith C, Tibshirani R, van de Rijn M, Sidow A, et al. 2010. 3'-End sequencing for expression quantification (3SEQ) from archival tumor samples. *PLoS One*. 5(1-e8768):1–11.
- Beerli P, Felsenstein J. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*. 152(2):763–773. PMID:10353916. PMCid:PMC1460627.
- Begg GA, Waldman JR. 1999. An holistic approach to fish stock identification. *Fish Res*. 43:35–44. [http://dx.doi.org/10.1016/S0165-7836\(99\)00065-X](http://dx.doi.org/10.1016/S0165-7836(99)00065-X)
- Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh Y-P, Hahn MW, Nista PM, Jones CD, Kern AD, Dewey CN, et al. 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol*. 5(11):e310. PMID:17988176. PMCid:PMC2062478. <http://dx.doi.org/10.1371/journal.pbio.0050310>
- Benzie JAH. 1999. Major genetic differences between crown-of-thorn starfish (*Acanthaster planci*) populations in the Indian and Pacific oceans. *Evolution*. 53:1782–1795. <http://dx.doi.org/10.2307/2640440>
- Beger M, Selkoe KA, Trembl EA, Barber PH, von der Heyden S, Crandall ED, Toonen RJ, Riginos CR. 2014. Evolving coral reef conservation with genetic information. *Bull Mar Sci*. 90:187–210. <http://dx.doi.org/10.5343/bms.2012.1108>
- Bell JD, Bartley DM, Lorenzen K, Loneragan NR. 2006. Restocking and stock enhancement of coastal fisheries: potential, problems and progress. *Fisher Res*. 80(1):1–8. <http://dx.doi.org/10.1016/j.fishres.2006.03.008>
- Benzie JAH, Ballment SE, Forbes AT, Demetriades T, Sugama K, Haryanti, Moria S. 2002. Mitochondrial DNA variation in Indo-Pacific populations of the giant tiger prawn, *Penaeus monodon*. *Mol Ecol*. 11:2553–2569. PMID:12453239. <http://dx.doi.org/10.1046/j.1365-294X.2002.01638.x>
- Bik HM, Porazinska DL, Creer S, Caporaso JG, Knight R, Thomas WK. 2012. Sequencing our way towards understanding global eukaryotic biodiversity. *Trend Ecol Evol*. 27(4):233–243. PMID:22244672. PMCid:PMC3311718. <http://dx.doi.org/10.1016/j.tree.2011.11.010>
- Blakenship HL, Leber KM. 1995. A responsible approach to marine stock enhancement. *Am Fisher Soc Symp*. 15:167–175
- Boers SA, van der Reijden WA, Jansen R. 2012. High-throughput multilocus sequence typing: bringing molecular typing to the next level. *PLoS One*. 7(7):e39630. <http://dx.doi.org/10.1371/journal.pone.0039630>
- Boitard S, Schlötterer C, Futschik A. 2009. Detecting selective sweeps: a new approach based on hidden markov models. *Genetics*. 181(4):1567–1578. PMID:19204373. PMCid:PMC2666521. <http://dx.doi.org/10.1534/genetics.108.100032>
- Bonin A. 2008. Population genomics: a new generation of genome scans to bridge the gap with functional genomics. *Mol Ecol*. 17(16):3583–3584. PMID:18662224. <http://dx.doi.org/10.1111/j.1365-294X.2008.03854.x>
- Botsford LW, White JW, Coffroth MA, Paris CB, Planes S, Shearer TL, Thorrold SR, Jones GP. (2009) Connectivity and resilience of coral reef metapopulations in marine protected areas: matching empirical efforts to predictive needs. *Coral Reefs*. 28:327–337. PMID:22833699. PMCid:PMC3402229. <http://dx.doi.org/10.1007/s00338-009-0466-z>
- Bowcock AM. 2007. Genomics: guilt by association. *Nature*. 447:645–646. PMID:17554292. <http://dx.doi.org/10.1038/447645a>

- Bowen BW, Shanker K, Yasuda N, Malay MCD, von der Heyden S, Paulay G, Rocha LA, Selkoe KA, Barber PH, Williams ST, et al. 2014. Phylogeography unplugged: comparative geographic surveys in the genomic era. *Bull Mar Sci.* 90:13–46. <http://dx.doi.org/10.5343/bms.2013.1007>
- Boyer JN, Briceno HO. 2011. 2010 annual report of the water quality monitoring project for the water quality protection program of the Florida Keys National Marine Sanctuary. *In: Southeast Environmental Research Center, Florida International University, Miami.*
- Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, Azam F, Rohwer F. 2002. Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci USA.* 99(22):14250–14255. PMID:12384570. PMCid:PMC137870. <http://dx.doi.org/10.1073/pnas.202488399>
- Bruneaux M, Johnson SE, Herczeg G, Merila J, Primmer CR, Vasemagi A. 2013. Molecular evolutionary and population genomic analysis of the nine-spined stickleback using a modified restriction-site-associated DNA tag approach. *Mol Ecol.* 22:565–582. PMID:22943747. <http://dx.doi.org/10.1111/j.1365-294X.2012.05749.x>
- Bus A, Hecht J, Huettel B, Reinhardt R, Stich B. 2012. High-throughput polymorphism detection and genotyping in *Brassica napus* using next-generation RAD sequencing. *BMC Genomics.* 13:281. PMID:22726880. PMCid:PMC3442993. <http://dx.doi.org/10.1186/1471-2164-13-281>
- Butlin R. 2010. Population genomics and speciation. *Genetica.* 138(4):409–418. PMID:18777099. <http://dx.doi.org/10.1007/s10709-008-9321-3>
- Bybee S, Bracken-Grissom H, Haynes B, Hermansen R, Byers R, Clement M, Udall J, Wilcox E, Crandall K. 2011. Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. *Genome Biol Evol.* 3:1312–1323. PMID:22002916. PMCid:PMC3236605. <http://dx.doi.org/10.1093/gbe/evr106>
- Campos WL, Alino PM. 2008. Recent advances in the management of Marine Protected Areas in the Philippines. *Kuroshio Sci.* 2:29–34.
- Caron DA, Countway PD, Jones AC, Kim DY, Schnetzer A. 2012. Marine protistan diversity. *In: Carlson CA, Giovannoni SJ, editors. Ann Rev Mar Sci.* 4:467–493. <http://dx.doi.org/10.1146/annurev-marine-120709-142802>
- Carpenter KE, Barber PH, Crandall ED, Ablan-Lagman MCA, Ambariyanto, Mahardika GN, Manhazi-Matsumoto BM, Juinio-Menez MA, Santos MD, Starger CJ, et al. 2011. Comparative phylogeography of the Coral Triangle and implications for marine management. *J Mar Biol.* 2011:1–14. <http://dx.doi.org/10.1155/2011/396982>
- Catchen J, Amores A, Hohenlohe P, Cresko W, Postlethwait J. 2011. Stacks: building and genotyping Loci de novo from short-read sequences. *G3 (Bethesda).* 1(3):171–182.
- Catchen J, Bassham S, Wilson T, Currey M, O'Brian C, Yeates Q, Cresko WA. 2013. The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Mol Ecol.* 22:2864–2883. PMID:23718143. <http://dx.doi.org/10.1111/mec.12330>
- Casey B, Delaney J, Diaz E, Dodge D, Garcia J, Higgins J, Keller B, Kelty R, Jaap WC, Matos C, et al. 2002. Status of coral reefs in the US Caribbean and Gulf of Mexico: Florida, Texas, Puerto Rico, US Virgin Islands, Navassa. *In: Wilkinson C, editor. Status of Coral Reefs of the World.* Australian Institute of Marine Science, Townsville. p. 251–276.
- Chariton AA, Court LN, Hartley DM, Colloff MJ, Hardy CM. 2010. Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. *Front Ecol Environ* 8(5):233–238. <http://dx.doi.org/10.1890/090115>
- Charlesworth B, Nordborg M, Charlesworth D. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet Res.* 70(2):155–174. PMID:9449192. <http://dx.doi.org/10.1017/S0016672397002954>

- Charlesworth B. 1998. Measures of divergence between populations and the effect of forces that reduce variability. *Mol Biol Evol.* 15(5):538–543. PMID:9580982. <http://dx.doi.org/10.1093/oxfordjournals.molbev.a025953>
- Charlesworth B. 2009. Effective population size and patterns of molecular evolution and variation. *Nature Rev Genet.* 10(3):195–205. PMID:19204717. <http://dx.doi.org/10.1038/nrg2526>
- Chutimanitsakun Y, Nipper R, Cuesta-Marcos A, Cistue L, Corey A, Filichkina T, Johnson E, Hayes P. 2011. Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley. *BMC Genomics.* 12:4. PMID:21205322. PMID:PMC3023751. <http://dx.doi.org/10.1186/1471-2164-12-4>
- Cowen RK, Gawarkiewicz G, Pineda J, Thorrold SR, Werner FE. 2007. Population connectivity in marine systems: an overview. *Oceanography.* 20(3)14–21. <http://dx.doi.org/10.5670/oceanog.2007.26>
- Cowen RK, Sponaugle S. 2009. Larval dispersal and marine population connectivity. *Annu Rev Mar Sci.* 1:443–466. <http://dx.doi.org/10.1146/annurev.marine.010908.163757>
- Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA, Quan P-L, Briese T, Hornig M, Geiser DM, et al. 2007. A metagenomic survey of microbes in honey bee colony collapse disorder. *Science.* 318(5848):283–287. PMID:17823314. <http://dx.doi.org/10.1126/science.1146498>
- Craig P, Birkeland C, Belliveau S. 2001. High temperatures tolerated by a diverse assemblage of shallow-water corals in American Samoa. *Coral Reefs.* 20(2):185–189. <http://dx.doi.org/10.1007/s003380100159>
- Crandall ED, Jones ME, Munoz MM, Akinronbi B, Erdmann MV, Barber PH. 2008. Comparative phylogeography of two seastars and their ectosymbionts within the Coral Triangle. *Mol Ecol.* 17:5276–5290. PMID:19067797. <http://dx.doi.org/10.1111/j.1365-294X.2008.03995.x>
- Crandall ED, Trembl EA, Barber PH. 2012. Coalescent and biophysical models of stepping-stone gene flow in neritid snails. *Mol Ecol.* 21:5579–5598. PMID:23050562. <http://dx.doi.org/10.1111/mec.12031>
- Cronn R, Knaus BJ, Litson A, Maughan PJ, Parks M, Syring JV, Udall J. 2012. Targeted enrichment strategies for next-generation plant biology. *Am J Bot.* 99:291–311. PMID:22312117. <http://dx.doi.org/10.3732/ajb.1100356>
- Cuvellier ML, Allen AE, Monier A, McCrow JP, Messier M, Tringe SG, Woyke T, Welsh RM, Ishoye T, Lee J-H, et al. 2010. Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proc Natl Acad Sci USA.* 107(33):14,679–14,684. PMID:20668244. PMID:PMC2930470. <http://dx.doi.org/10.1073/pnas.1001665107>
- Dammannagoda ST, Hurwood D, Mather P. 2011. Genetic analysis reveals two stocks of skipjack tuna (*Katsuwonus pelamis*) in the north western Indian Ocean. *Canad J Fisher Aquat Sci.* 68(2):210–223. <http://dx.doi.org/10.1139/F10-136>
- Davey J, Blaxter M. 2010. RADSeq: next-generation population genetics. *Brief Funct Genomics.* 9(5–6):416–423. PMID:21266344. PMID:PMC3080771. <http://dx.doi.org/10.1093/bfpg/elq031>
- Davey J, Hohenlohe P, Etter P, Boone J, Catchen J, Blaxter M. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet.* 12(7):499–510. PMID:21681211. <http://dx.doi.org/10.1038/nrg3012>
- Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML. 2013. Special features of RAD Sequencing data: implications for genotyping. *Mol Ecol.* 22:3151–3164. PMID:23110438. PMID:PMC3712469. <http://dx.doi.org/10.1111/mec.12084>
- 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. *Mol Ecol Res.* 12:1058–1067. PMID:22931062. <http://dx.doi.org/10.1111/1755-0998.12003>
- Drummond R, Pinheiro A, Rocha C, Menossi M. 2005. ISER: selection of differentially expressed genes from DNA array data by non-linear data transformations and

- local fitting. *Bioinformatics*. 21(24):4427–4429. PMID:16249264. <http://dx.doi.org/10.1093/bioinformatics/bti729>
- Edwards RA, Rohwer F. 2005. Viral metagenomics. *Nat Rev Microbiol*. 3(6):504–510. PMID:15886693. <http://dx.doi.org/10.1038/nrmicro1163>
- Edwards SV. 2009. Is a new and general theory of molecular systematics emerging? *Evolution*. 63:1–19. PMID:19146594. <http://dx.doi.org/10.1111/j.1558-5646.2008.00549.x>
- Elshire KJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*. 6(5):e19379. <http://dx.doi.org/10.1371/journal.pone.0019379>
- Emerson KJ, Merz CR, Catchen JM, Hohenlohe PA, Cresko WA, Bradshaw WE, Holzapfel CM. 2010. Resolving postglacial phylogeography using high-throughput sequencing. *Proc Natl Acad Sci USA*. 107(37):16,196–16,200. PMID:20798348. PMCid:PMC2941283. <http://dx.doi.org/10.1073/pnas.1006538107>
- Etter P, Bassham S, Hohenlohe P, Johnson E, Cresko W. 2011a. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Meth Mol Biol*. 772:157–178. PMID:22065437. PMCid:PMC3658458. [http://dx.doi.org/10.1007/978-1-61779-228-1\\_9](http://dx.doi.org/10.1007/978-1-61779-228-1_9)
- Etter P, Preston J, Bassham S, Cresko W, Johnson E. 2011b. Local de novo assembly of RAD paired-end contigs using short sequencing reads. *PLoS One*. 6(4):e18561. PMID:21541009. PMCid:PMC3076424. <http://dx.doi.org/10.1371/journal.pone.0018561>
- Everett MV, Miller MR, Seeb JE. 2012. Meiotic maps of sockeye salmon derived from massively parallel DNA sequencing. *BMC Genomics*. 13(1):521. PMID:23031582. PMCid:PMC3563581. <http://dx.doi.org/10.1186/1471-2164-13-521>
- Excoffier L, Hofer T, Foll M. 2009. Detecting loci under selection in a hierarchically structured population. *Heredity*. 103(4):285–298. PMID:19623208. <http://dx.doi.org/10.1038/hdy.2009.74>
- Excoffier L, Lischer H. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res*. 10(3):564–567. PMID:21565059. <http://dx.doi.org/10.1111/j.1755-0998.2010.02847.x>
- Fabricius KE. 2005. Effects of terrestrial runoff on the ecology of corals and coral reefs: review and synthesis. *Mar Pollut Bull*. 50:125–146. PMID:15737355. <http://dx.doi.org/10.1016/j.marpolbul.2004.11.028>
- Fan HC, Gu W, Wang J, Blumenfeld YJ, El-Sayed YY, Quake SR. 2012. Non-invasive prenatal measurements of the fetal genome. *Nat Comm*. 487:320–324. PMID:22763444. PMCid:PMC3561905. <http://dx.doi.org/10.1038/nature11251>
- Falush D, Stephens M, Pritchard J. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*. 164(4):1567–1587. PMID:12930761. PMCid:PMC1462648.
- Falush D, Stephens M, Pritchard J. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes*. 7(4):574–578. PMID:18784791. PMCid:PMC1974779. <http://dx.doi.org/10.1111/j.1471-8286.2007.01758.x>
- Felsenstein J. 1976. The theoretical population genetics of variable selection and migration. *Ann Rev Genet*. 10:253–280. PMID:797310. <http://dx.doi.org/10.1146/annurev.ge.10.120176.001345>
- Felsenstein J. 2006. Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? *Mol Biol Evol*. 23:691–700. PMID:16364968. <http://dx.doi.org/10.1093/molbev/msj079>
- Fernandez-Silva I, Toonen RJ. 2013. Microsatellite loci from 454 pyrosequencing via post-sequencing bioinformatic analyses. *Microsatellites: Meth Mol Biol*. 1006:101–120. PMID:23546786. [http://dx.doi.org/10.1007/978-1-62703-389-3\\_7](http://dx.doi.org/10.1007/978-1-62703-389-3_7)
- Fernandez-Silva I, Whitney J, Wainwright B, Andrews KR, Ylitalo-Ward H, Bowen BW, Toonen RJ, Goetze E, Karl SA. 2013. Microsatellites for next-generation ecologists: a post-sequencing bioinformatics pipeline. *PLoS One*. 8:e55990. <http://dx.doi.org/10.1371/journal.pone.0055990>



- Ferrier-Pages C, Schoelzke V, Jaubert J, Muscatine L, Hoegh-Guldberg O. 2001. Response of a scleractinian coral, *Stylophora pistillata*, to iron and nitrate enrichment. *J Exper Mar Biol Ecol.* 259:249–261. [http://dx.doi.org/10.1016/S0022-0981\(01\)00241-6](http://dx.doi.org/10.1016/S0022-0981(01)00241-6)
- Fidelman, P, Evans L, Fabinyi M, Foale S, Cinner J, Rosen F. 2012. Governing large-scale marine commons: contextual challenges in the Coral Triangle. *Mar Pol.* 36:42–53. <http://dx.doi.org/10.1016/j.marpol.2011.03.007>
- Fisher RA. 1930. The genetic theory of natural selection. Oxford, England. p. 272.
- Fisher RA. 1958. The genetical theory of natural selection. Dover, New York.
- Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics.* 180:977–993. PMID:18780740. PMID:PMC2567396. <http://dx.doi.org/10.1534/genetics.108.092221>
- Fonseca VG, Carvalho GR, Sung W, Johnson HF, Power DM, Neill SP, Packer M, Blaxter ML, Lamshead PJD, Thomas WK, et al. 2010. Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature Comm.* 1:98. <http://dx.doi.org/10.1038/ncomms1095>
- Gaggiotti OE, Bekkevold D, Jorgensen HB, Foll M, Carvalho GR, Andre C, Ruzzante DE. 2009. Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic herring as a case study. *Evolution.* 63:2939–2951. PMID:19624724. <http://dx.doi.org/10.1111/j.1558-5646.2009.00779.x>
- Gagnaire P-A, Minegishi Y, Zenboudji S, Valade P, Aoyama J, Berrebi P. 2011. Within-population structure highlighted by differential introgression across semipermeable barriers to gene flow in *Anguilla marmorata*. *Evolution.* 65:3413–3427. PMID:22133215. <http://dx.doi.org/10.1111/j.1558-5646.2011.01404.x>
- Gagnaire P-A, Normandeau E, Pavey SA, Bernatchez L. 2012. Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the Lake Whitefish (*Coregonus clupeaformis*). *Mol Ecol.* 22:3036–3048. PMID:23181719. <http://dx.doi.org/10.1111/mec.12127>
- Gagnaire PA, Normandeau E, Cote C, Hansen MM, Bernatchez L. 2012. The genetic consequences of spatially varying selection in the panmictic American eel (*Anguilla rostrata*). *Genetics.* 190(2):725–703. PMID:22135355. PMID:PMC3276646. <http://dx.doi.org/10.1534/genetics.111.134825>
- Gautier M, Gharbi K, Cezard T, Foucaud J, Kerdelhue C, Pudlo P, Cornuet JM, Estoup. 2013. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Mol Ecol.* 22:3165–3178. PMID:23110526. <http://dx.doi.org/10.1111/mec.12089>
- George M, Hussin A. 2010. Current legal developments south east Asia: the Coral Triangle initiative on coral reefs, fisheries and food security. *Intl J Mar Coast Law.* 25:443–454. <http://dx.doi.org/10.1163/157180810X520336>
- Gilbert JA, Dupont CL (2011) Microbial metagenomics: beyond the genome. *In:* Carlson CA, Giovannoni SJ, editors. *Ann Rev Mar Sci.* 3:347–371.
- Glenn T. 2011. Field guide to next-generation DNA sequencers. *Mol Ecol Resour.* 11(5):759–769. PMID:21592312. <http://dx.doi.org/10.1111/j.1755-0998.2011.03024.x>
- Gold JR, Giresi MM, Renshaw MA, Gwo JC. 2013. Population genetic comparisons among Cobia from the Northern Gulf of Mexico, US western Atlantic, and Southeast Asia. *N Am J Aquacult.* 75:57–63. <http://dx.doi.org/10.1080/15222055.2012.713899>
- Gompert Z, Lucas L, Fordyce J, Forister M, Nice C. 2010. Secondary contact between *Lycaeides idas* and *L. melissa* in the Rocky Mountains: extensive admixture and a patchy hybrid zone. *Mol Ecol.* 19(15):3171–3192. PMID:20618903. <http://dx.doi.org/10.1111/j.1365-294X.2010.04727.x>
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotech.* 29(7):644–652. PMID:21572440. PMID:PMC3571712. <http://dx.doi.org/10.1038/nbt.1883>



- Groissillier A, Massana R, Valentin K, Vaulot D, Guillou L. 2006. Genetic diversity and habitats of two enigmatic marine alveolate lineages. *Aquat Microb Ecol.* 42(3):277–291. <http://dx.doi.org/10.3354/ame042277>
- Guindon S, Dufayard J, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59(3):307–321. PMID:20525638. <http://dx.doi.org/10.1093/sysbio/syq010>
- Haiser HJ, Turnbaugh PJ. 2012. Is it time for a metagenomic basis of therapeutics? *Science.* 336(6086):1253–1255. PMID:22674325. <http://dx.doi.org/10.1126/science.1224396>
- Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol.* 5(10):R245–R249. [http://dx.doi.org/10.1016/S1074-5521\(98\)90108-9](http://dx.doi.org/10.1016/S1074-5521(98)90108-9)
- Hare MP, Nunney L, Schwartz MK, Ruzzante DE, Burford M, Waples RS, Ruegg K, Palstra F. 2011. Understanding and estimating effective population size for practical application in marine species management. *Conserv Biol.* 25:438–449. PMID:21284731. <http://dx.doi.org/10.1111/j.1523-1739.2010.01637.x>
- Harismendy O, Ng PC, Strausberg RL, Wang XY, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S, et al. 2009. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol.* <http://dx.doi.org/10.1186/gb-2009-10-3-r32>
- Hecht B, Campbell N, Holecek D, Narum S. 2012a. Genome-wide association reveals genetic basis for the propensity to migrate in wild populations of rainbow and steelhead trout. *Mol Ecol.* 22:3061–3076. PMID:23106605. <http://dx.doi.org/10.1111/mec.12082>
- Hecht B, Thrower F, Hale M, Miller M, Nichols K. 2012b. Genetic architecture of migration-related traits in rainbow and steelhead trout, *Oncorhynchus mykiss*. *G3: Genes | Genomes | Genetics.* 2(9):1113–1127.
- Hedgecock D. 1986. Is gene flow from pelagic larval dispersal important in the adaptation and evolution of marine-invertebrates? *Bull Mar Sci.* 39:550–564.
- Hedgecock D, Barber PH, Edmonds S. 2007. Genetic approaches to measuring connectivity. *Oceanography.* 20(3):70–79. <http://dx.doi.org/10.5670/oceanog.2007.30>
- Hobbs J-PA, Frisch AJ, Allen GR, van Herwerden L. 2009. Marine hybrid hotspot at Indo-Pacific biogeographic border. *Biol Lett.* 5:258–261. PMID:19126528. PMID:PMC2665801. <http://dx.doi.org/10.1098/rsbl.2008.0561>
- Hodges E, Smith AD, Kendall J, Xuan Z, Ravi K, Rooks M, Zhang MQ, Ye K, Bhattacharjee A, Brizuela L, et al. 2009. *Genome Res.* 19:1593–1605. PMID:19581485. PMID:PMC2752124. <http://dx.doi.org/10.1101/gr.095190.109>
- Hohenlohe PA, Amish S, Catchen J, Allendorf FW, Luikart G. 2011. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Mol Ecol Res.* 1(11 Suppl):117–122. PMID:21429168. <http://dx.doi.org/10.1111/j.1755-0998.2010.02967.x>
- Hohenlohe P, Bassham S, Currey M, Cresko W. 2012a. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philos Trans R Soc Lond B Biol Sci.* 367(1587):395–408. PMID:22201169. PMID:PMC3233713. <http://dx.doi.org/10.1098/rstb.2011.0245>
- Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010a. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.* 6(2):e1000862. PMID:20195501. PMID:PMC2829049. <http://dx.doi.org/10.1371/journal.pgen.1000862>
- Hohenlohe P, Catchen J, Cresko W. 2012b. Population genomic analysis of model and non-model organisms using sequenced RAD tags. *Meth Mol Biol.* 888:235–260. PMID:22665285. [http://dx.doi.org/10.1007/978-1-61779-870-2\\_14](http://dx.doi.org/10.1007/978-1-61779-870-2_14)
- Hohenlohe P, Phillips P, Cresko W. 2010b. Using Population genomics to detect selection in natural populations: key concepts and methodological considerations. *Int J*

- Plant Sci. 171(9):1059–1071. PMID:21218185. PMCID:PMC3016716. <http://dx.doi.org/10.1086/656306>
- Houston R, Davey J, Bishop S, Lowe N, Mota-Velasco J, Hamilton A, Guy D, Tinch A, Thomson M, Blaxter M, et al. 2012. Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC Genomics*. 13:244. PMID:22702806. PMCID:PMC3520118. <http://dx.doi.org/10.1186/1471-2164-13-244>
- Hulata G. 2001. Genetic manipulations in aquaculture: a review of stock improvement by classical and modern technologies. *Genetica*. 111:155–173. PMID:11841164. <http://dx.doi.org/10.1023/A:1013776931796>
- Hunter B, Wright KM, Bomblies K. 2013. Short read sequencing in studies of natural variation and adaptation. *Curr Opin Plant Biol*. 16:85–91. PMID:23177206. <http://dx.doi.org/10.1016/j.pbi.2012.10.003>
- Izzo C, Gillanders BM, Ward TM. 2012. Movement patterns and stock structure of Australian sardine (*Sardinops sagax*) off South Australia and the east coast: implications for future stock assessment and management. *SARDI Research Report Series 611*, p. 102.
- Jokiel PL (2004) Temperature stress and coral bleaching. *In*: Rosenberg E, Loya Y, editors. *Coral health and disease*. Springer, Berlin, Germany. [http://dx.doi.org/10.1007/978-3-662-06414-6\\_23](http://dx.doi.org/10.1007/978-3-662-06414-6_23)
- Jones FC, Grabherr MG, Chan YE, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*. 484(7392):55–61. PMID:22481358. PMCID:PMC3322419. <http://dx.doi.org/10.1038/nature10944>
- Juinio-Menez MA, Bangi HG, Malay MC, Pastor D. 2008. Enhancing the recovery of depleted *Tripneustes gratilla* stocks through grow-out culture and restocking. *Rev Fish Sci*. 16:35–43. <http://dx.doi.org/10.1080/10641260701678116>
- Kaplan N, Darden T, Hudson R. 1988. The coalescent process in models with selection. *Genetics*. 120(3):819–829. PMID:3066685. PMCID:PMC1203559.
- Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. *Ecol Lett*. 7(12):1225–1241. <http://dx.doi.org/10.1111/j.1461-0248.2004.00684.x>
- Keller I, Wagner C, Greuter L, Mwaiko S, Selz O, Sivasundar A, Wittwer S, Seehausen O. 2012. Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Mol Ecol*. 22:2848–2863. PMID:23121191. <http://dx.doi.org/10.1111/mec.12083>
- Kenkel CD, Meyer E, Matz MV. 2013. Gene expression under chronic heat stress in populations of the mustard hill coral (*Porites asteroides*) from different thermal environments. *Mol Ecol*. 22:4322–4334. PMID:23899402. <http://dx.doi.org/10.1111/mec.12390>
- Keyse J, Crandall ED, Toonen RJ, Meyer CP, Trembl EA, Riginos CR. 2014. The scope of published population genetic data for Indo-Pacific marine fauna, and future research opportunities. *Bull Mar Sci*. 90:47–78. <http://dx.doi.org/10.5343/bms.2012.1107>
- Kiene RP. 1991. Production and consumption of methane in aquatic systems. *In*: Rogers JE, Whitman WB, editors. *Microbial production and consumption of greenhouse gases, methane, nitrogen oxides, and halomethanes*. Am Soc Microbiol, Washington DC. p. 111–146.
- Kimura M. 1968. Evolutionary rate at the molecular level. *Nature*. 217:624–626. PMID:5637732. <http://dx.doi.org/10.1038/217624a0>
- King E, Macdonald S, Long A. 2012. Properties and power of the *Drosophila* Synthetic Population Resource for the routine dissection of complex traits. *Genetics*. 191(3):935–949. PMID:22505626. PMCID:PMC3389985. <http://dx.doi.org/10.1534/genetics.112.138537>
- Kingman J. 2000. Origins of the coalescent. 1974–1982. *Genetics*. 156(4):1461–1463. PMID:11102348. PMCID:PMC1461350.
- Knowles LL. 2009. Statistical phylogeography. *Ann Rev Ecol Evol Syst*. 40:593–612. <http://dx.doi.org/10.1146/annurev.ecolsys.38.091206.095702>
- 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. *Mol Ecol.* 17:3775–3787. PMID:18662232. <http://dx.doi.org/10.1111/j.1365-294X.2008.03803.x>
- Kohlmann A, Grossmann V, Klein H-U, Schindela S, Weiss T, Kazak B, Dicker F, Schnittger S, Dugas M, Kern W, et al. 2010. Next-generation sequencing technology reveals a characteristic pattern of molecular mutations in 72.8% of chronic myelomonocytic leukemia by detecting frequent alterations in TET2, CBL, RAS, and RUNX1. *J Clin Oncol.* 28:3858–3865. PMID:20644105. <http://dx.doi.org/10.1200/JCO.2009.27.1361>
- Krueger F, Andrews SR, Osborne CS. 2011. Large scale loss of data in low-diversity Illumina sequencing libraries can be recovered by deferred cluster calling. *PLoS ONE.* 6:e16607. PMID:21305042. PMCid:PMC3030592. <http://dx.doi.org/10.1371/journal.pone.0016607>
- Ladner JT, Barshis DJ, Palumbi SR. 2012. Amino acid sequence evolution in four *Symbiodinium* clades: an exploration into the genetic basis of thermal tolerance in *Symbiodinium* clade D. *BMC Evol Biol.* 12:217. PMID:23145489. PMCid:PMC3740780. <http://dx.doi.org/10.1186/1471-2148-12-217>
- Ladner JT, Palumbi SR. 2012. Extensive sympatry, cryptic diversity and introgression throughout the geographic distribution of two coral species complexes. *Mol Ecol.* 21(9):2224–2238. PMID:22439812. <http://dx.doi.org/10.1111/j.1365-294X.2012.05528.x>
- Lemmon A, Emme S, Lemmon E. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst Biol.* 61(5):727–744. PMID:22605266. <http://dx.doi.org/10.1093/sysbio/sys049>
- Lemmon A, Lemmon E. 2012. High-throughput identification of informative nuclear loci for shallow-scale phylogenetics and phylogeography. *Syst Biol.* 61(5):745–761. PMID:22610088. <http://dx.doi.org/10.1093/sysbio/sys051>
- Leslie HM, McLeod KL. 2007. Confronting the challenges of implementing marine ecosystem-based management. *Front Ecol Environ.* 5:540–548. <http://dx.doi.org/10.1890/060093>
- Levene H. 1953. Genetic equilibrium when more than one ecological niche is available. *Am Nat.* 87:331–333. <http://dx.doi.org/10.1086/281792>
- Lindquist N, Barber PH, Weisz JB. 2005. Episymbiotic microbes as food and defence for marine isopods: unique symbioses in a hostile environment. *Proc R Soc Biol Sci.* 272(1569):1209–1216. PMID:16024384. PMCid:PMC1564109.
- Lipcius RN, Eggleston DB, Schreiber SJ, Seitz RD, Shen J, Sisson M, Stockhausen WT, Wang HV. 2008. Importance of metapopulation connectivity to restocking and restoration of marine species. *Rev Fisher Sci.* 16(1):101–110. <http://dx.doi.org/10.1080/10641260701812574>
- Lirman D, Fong P. 2007. Is proximity to land-based sources of coral stressors an appropriate measure of risk to coral reefs? An example from the Florida Reef Tract. *Mar Pollut Bull.* 54:779–791. PMID:17303183. <http://dx.doi.org/10.1016/j.marpolbul.2006.12.014>
- Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V, et al. 2009. Population genomics of domestic and wild yeast. *Nature.* 458:337–341. PMID:19212322. PMCid:PMC2659681. <http://dx.doi.org/10.1038/nature07743>
- Liu L, Li YH, Li SL, Hu N, He YM, Pong R, Lin DN, Lu LH, Law M. 2012. Comparison of Next-Generation Sequencing Systems. *J Biomed Biotech.* Article ID 251364, 11 p. <http://dx.doi.org/10.1155/2012/251364>
- Liu S, Yeh C-T, Tang HM, Nettleton D, Schmale PS. 2012. Gene mapping via bulked segregant RNA-seq (BSR-Seq). *PLoS One.* 7:e36406. PMID:22586469. PMCid:PMC3346754. <http://dx.doi.org/10.1371/journal.pone.0036406>
- Lluisma AO, Milash BA, Moore B, Olivera BM, Bandyopadhyay. 2012. Novel venom peptides from the cone snail *Conus pulicarius* discovered through next-generation sequencing of its venom duct transcriptome. *Mar Genom.* 5:43–51. PMID:22325721. <http://dx.doi.org/10.1016/j.margen.2011.09.002>
- Luikart G, England P, Tallmon D, Jordan S, Taberlet P. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet.* 4(12):981–994. PMID:14631358. <http://dx.doi.org/10.1038/nrg1226>

- Luo CW, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT. 2012. Direct comparisons of Illumina vs Roche 454 sequencing technologies on the Same Microbial Community DNA Sample. *PLoS One*. 7(2):e30087. <http://dx.doi.org/10.1371/journal.pone.0030087>
- Lynch M. 2009. Estimation of allele frequencies from high-coverage genome-sequencing projects. *Genetics*. 182(1):295–301. PMID:19293142. PMCid:PMC2674824. <http://dx.doi.org/10.1534/genetics.109.100479>
- Manel S, Gaggiotti OE, Waples RS. 2005. Assignment methods: matching biological questions with appropriate techniques. *Trends Ecol Evol*. 20(3):136–142. PMID:16701357. <http://dx.doi.org/10.1016/j.tree.2004.12.004>
- Maralit BA, Aguila RD, Ventolero MFH, Perez SKL, Willette DA, Santos MD. 2013. Detection of mislabeled commercial fishery products in the Philippines using DNA barcodes and its implication to food traceability and safety, *J Food Contr*. 33:119–125. <http://dx.doi.org/10.1016/j.foodcont.2013.02.018>
- Mardis E. 2008a. The impact of next-generation sequencing technology on genetics. *Trends Genet*. 24(3):133–141. PMID:18262675. PMCid:PMC2680276. <http://dx.doi.org/10.1016/j.tig.2007.12.007>
- Mardis E. 2008b. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet*. 9:387–402. PMID:18576944. <http://dx.doi.org/10.1146/annurev.genom.9.081307.164359>
- Marguerat S, Wilhelm B, Bahler J. 2008. Next-generation sequencing: applications beyond genomes. *Biochem Soc Trans*. 36:1091–1096. PMID:18793195. PMCid:PMC2563889. <http://dx.doi.org/10.1042/BST0361091>
- Marshall DJ, Monro K, Bode M, Keough MJ, Swearer S. 2010. Phenotype-environment mismatches reduce connectivity in the sea. *Ecol Lett*. 13(1):128–140. PMID:19968695. <http://dx.doi.org/10.1111/j.1461-0248.2009.01408.x>
- Maughan PJ, Yourstone SM, Jellen EN, Udall JA. 2009. SNP discovery via genomic reduction, barcoding, and 454-pyrosequencing in Amaranth. *Plant Genome*. 2:260–270. <http://dx.doi.org/10.3835/plantgenome2009.08.0022>
- McCormack J, Hird S, Zellmer A, Carstens B, Brumfield R. 2011. Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol Phylogenet Evol*. 66:526–538. PMID:22197804. <http://dx.doi.org/10.1016/j.ympev.2011.12.007>
- McCormack JE, Maley JM, Hird SM, Derryberry EP, Graves GP, Brumfield RT. 2012. Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. *Mol Phylogenet Evol*. 62:397–406. PMID:22063264. <http://dx.doi.org/10.1016/j.ympev.2011.10.012>
- Medinger R, Nolte V, Pandey RV, Jost S, Ottenwaelder B, Schloetterer C, Boenigk J. 2010. Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. *Mol Ecol*. 19:32–40. PMID:20331768. PMCid:PMC2953707. <http://dx.doi.org/10.1111/j.1365-294X.2009.04478.x>
- Merz C. 2012. Independent replication of phylogeographies: how repeatable are they? Dissertation. University of Oregon, Eugene.
- Metcalfe WW, Griffin BM, Cicchillo RM, Gao J, Janga SC, Cooke HA, Circello BT, Evans BS, Martens-Habbena W, Stahl DA, et al. 2012. Synthesis of methylphosphonic acid by marine microbes: a source for methane in the aerobic Ocean. *Science*. 337(6098):1104–1107. PMID:22936780. PMCid:PMC3466329. <http://dx.doi.org/10.1126/science.1219875>
- Meyer E, Aglyamova GV, Matz MV. 2011. Profiling gene expression responses of coral larvae (*Acropora millepora*) to elevated temperature and settlement inducers using a novel RNA-seq procedure. *Mol Ecol*. 20:3599–3616. PMID:21801258.
- Mieog JC, Van Oppen MJH, Berelmans K, Stam WT, Olsen JL. 2009. Quantification of algal endosymbionts (*Symbiodinium*) in coral tissue using real-time PCR. *Mol Ecol Resour* 9:74–82. PMID:21564569. <http://dx.doi.org/10.1111/j.1755-0998.2008.02222.x>
- Miller M, Atwood T, Eames B, Eberhart J, Yan Y, Postlethwait J, Johnson E. 2007a. RAD marker microarrays enable rapid mapping of zebrafish mutations. *Genome Biol*. 8(6):R105. PMID:17553171. PMCid:PMC2394753. <http://dx.doi.org/10.1186/gb-2007-8-6-r105>

- Miller MR, Brunelli JP, Wheeler PA, Liu SX, Rexroad CE, Palti Y, Doe CQ, Thorgaard GH. 2012. A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Mol Ecol.* 21(2):237–249. PMID:21988725. PMCID:PMC3664428. <http://dx.doi.org/10.1111/j.1365-294X.2011.05305.x>
- Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. 2007b. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res.* 17(2):240–248. PMID:17189378. PMCID:PMC1781356. <http://dx.doi.org/10.1101/gr.5681207>
- Nadeau N, Martin S, Kozak K, Salazar C, Dasmahapatra K, Davey J, Baxter S, Blaxter M, Mallet J, Jiggins C. 2013. Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Mol Ecol.* 22(3):814–826. PMID:22924870. <http://dx.doi.org/10.1111/j.1365-294X.2012.05730.x>
- Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science.* 320(5881):1344–1349. PMID:18451266. PMCID:PMC2951732. <http://dx.doi.org/10.1126/science.1158441>
- Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, et al. 2005a. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* 3(6):e170. PMID:15869325. PMCID:PMC1088278. <http://dx.doi.org/10.1371/journal.pbio.0030170>
- Nielsen R, Wakeley J. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics.* 158(2):885–896. PMID:11404349. PMCID:PMC1461674.
- Neiman M, Sundling S, Gronberg H, Hall P, Czene K, Lindberg J, Klevebring D. 2012. Library preparation and multiplex capture for massive parallel sequencing applications made efficient and easy. *PLoS One.* 7:e48616. PMID:23139805. PMCID:PMC3489721. <http://dx.doi.org/10.1371/journal.pone.0048616>
- Noonan JP, Hofreiter M, Smith D, Priest JR, Rohland N, Rabeder G, Krause J, Dettler JC, Paabo S, Rubin EM. 2005. Genomic sequencing of Pleistocene cave bears. *Science.* 309(5734):597–600. PMID:15933159. <http://dx.doi.org/10.1126/science.1113485>
- Notohara M. 1990. The coalescent and the genealogical process in geographically structured population. *J Math Biol.* 29(1):59–75. PMID:2277236. <http://dx.doi.org/10.1007/BF00173909>
- Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, et al. 2008. Genes mirror geography within Europe. *Nature.* 456:98–101. PMID:18758442. PMCID:PMC2735096. <http://dx.doi.org/10.1038/nature07331>
- O'Neill EM, Schwartz R, Bullock CT, Williams JS, Shaffer HB, Aguilar-Miguel X, Parra-Olea G, Weisrock DW. 2013. Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (*Ambystoma tigrinum*) species complex. *Mol Ecol.* 22:111–129. PMID:23062080. <http://dx.doi.org/10.1111/mec.12049>
- Okuzawa K, Maliao RJ, Quintio ET, Buen-Ursua SMA, Lebata MJHL, Gallardo WG, Garcia LMB, Primavera JH. 2008. Stock enhancement of threatened species in Southeast Asia. *Rev Fisher Sci.* 16(1):394–402. <http://dx.doi.org/10.1080/10641260701678496>
- Oliver TA, Palumbi SR. 2011. Do fluctuating temperature environments elevate coral thermal tolerance? *Coral Reefs.* 30:429–440. <http://dx.doi.org/10.1007/s00338-011-0721-y>
- Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma R, Hedrick PW. 2010. Conservation genetics in transition to conservation genomics. *Trend Genet.* 26(4):177–187. PMID:20227782. <http://dx.doi.org/10.1016/j.tig.2010.01.001>
- Pace NR, Stahl DA, Lane DJ, Olsen GJ. 1986. The analysis of natural microbial-populations by ribosomal-RNA sequences. *Adv Microbial Ecol.* 9:1–55. [http://dx.doi.org/10.1007/978-1-4757-0611-6\\_1](http://dx.doi.org/10.1007/978-1-4757-0611-6_1)
- Palumbi SR, Sandifer PA, Allan JD, Beck MW, Fautin DG, Fogarty MJ, Halpern BS, Incze LS, Leong JA, et al. 2009. Managing for ocean biodiversity to sustain marine ecosystem services. *Front Ecol Environ.* 7(4):204–211. <http://dx.doi.org/10.1890/070135>



- Palumbi SR, Sandifer PA, Allan JD, Beck MW, Gautin DG, Fogarty MJ, Halpern BS, Incze LS, Leong JA, Norse E, et al. 2009. Managing for ocean biodiversity to sustain marine ecosystem services. *Front Ecol Environ*. 7:204–211. <http://dx.doi.org/10.1890/070135>
- Pampoulie C, Ruzzante DE, Chosson V, Jorundsdottir ID, Taylor L, Thosteinnsson V, Danielsdottir AK, Marteinsdottir G. 2006. The genetic structure of Atlantic cod (*Gadus morhua*) around Iceland: insight from microsatellites, the Pan I locus, and tagging experiments. *Can J Fish Aquat Sci*. 63:2660–2674. <http://dx.doi.org/10.1139/f06-150>
- Pandolfi JM, Jackson JBC, Baron N, Bradbury RH, Guzman HM, Hughes TP, Kappel CV, Micheli F, Ogden JC, Possingham HP, et al. 2005. Ecology—are US coral reefs on the slippery slope to slime? *Science*. 307(5716):1725–1726. PMID:15774744. <http://dx.doi.org/10.1126/science.1104258>
- Pareek CS, Smoczynski R, Tretyn A. 2011. Sequencing technologies and genome sequencing. *J App Genet*. 52(4):413–435. PMID:21698376. PMCid:PMC3189340. <http://dx.doi.org/10.1007/s13353-011-0057-x>
- Pespeni MH, Garfield DA, Manier MK, Palumbi SR. 2012. Genome-wide polymorphisms show unexpected targets of natural selection. *Proc Roy Soc Biol Sci*. 279(1732):1412–1420.
- Peterson B, Weber J, Kay E, Fisher H, Hoekstra H. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*. 7(5):e37135. PMID:22675423. PMCid:PMC3365034. <http://dx.doi.org/10.1371/journal.pone.0037135>
- Petrosino JE, Highlander S, Luna RA, Gibbs RA, Versalovic J. 2009. Metagenomic pyrosequencing and microbial identification. *Clinic Chem*. 55(5):856–866. PMID:19264858. PMCid:PMC2892905. <http://dx.doi.org/10.1373/clinchem.2008.107565>
- Pfrender ME, Hawkins CP, Bagley M, Courtney GW, Creutzburg BR, Epler JH, Fend S, Ferrington LC Jr, Hartzell PL, Jackson S, et al. 2010. Assessing macroinvertebrate biodiversity in freshwater ecosystems: advances and challenges in DNA-based approaches. *Quart Rev Biol*. 85(3):319–340. PMID:20919633. <http://dx.doi.org/10.1086/655118>
- Pfender W, Saha M, Johnson E, Slabaugh M. 2011. Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust resistance in *Lolium perenne*. *Theor Appl Genet*. 122(8):1467–1480. PMID:21344184. <http://dx.doi.org/10.1007/s00122-011-1546-3>
- Pickrell J, Coop G, Novembre J, Kudaravalli S, Li J, Absher D, Srinivasan B, Barsh G, Myers R, Feldman M, et al. 2009. Signals of recent positive selection in a worldwide sample of human populations. *Genome Res*. 19(5):826–837. PMID:19307593. PMCid:PMC2675971. <http://dx.doi.org/10.1101/gr.087577.108>
- Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet*. 8:e1002967. PMID:23166502. PMCid:PMC3499260. <http://dx.doi.org/10.1371/journal.pgen.1002967>
- Piganeau G, Desdevises Y, Derelle E, Moreau H. 2008. Picoeukaryotic sequences in the Sargasso Sea metagenome. *Genome Biol*. 9:R5. <http://dx.doi.org/10.1186/gb-2008-9-1-r5>
- Pluzhnikov A, Donnelly P. 1996. Optimal sequencing strategies for surveying molecular genetic diversity. *Genetics*. 144:1247–1262. PMID:8913765. PMCid:PMC1207616.
- Poinar HN, Schwarz C, Qi J, Shapiro B, MacPhee RDE, Buigues B, Tikhonov A, Huson DH, Tomsho LP, Auch A, et al. 2006. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science*. 311(5759):392–394. PMID:16368896. <http://dx.doi.org/10.1126/science.1123360>
- Policansky D, Magnuson JJ. 1998. Genetics, metapopulations, and ecosystem management of fisheries. *Ecol Appl*. 8(1):S119–S123. <http://dx.doi.org/10.2307/2641369>
- Polz ME, Cavanaugh CM. 1998. Bias in template-to-product ratios in multitemplate PCR. *App Environ Microbiol*. 64(10):3724–3730. PMID:9758791. PMCid:PMC106531.
- Pomeroy R, Garces L, Pido M, Silvestre G. 2010. Ecosystem-based fisheries management in small-scale tropical marine fisheries: emerging models of governance arrangements in the Philippines. *Mar Pol*. 34:298–308. <http://dx.doi.org/10.1016/j.marpol.2009.07.008>



- Pool J, Hellmann I, Jensen J, Nielsen R. 2010. Population genetic inference from genomic sequence variation. *Genome Res.* 20(3):291–300. PMID:20067940. PMCid:PMC2840988. <http://dx.doi.org/10.1101/gr.079509.108>
- Porazinska DL, Sung W, Giblin-Davis RM, Thomas WK. 2010. Reproducibility of read numbers in high-throughput sequencing analysis of nematode community composition and structure. *Mol Ecol Res.* 10(4):666–676. PMID:21565071. <http://dx.doi.org/10.1111/j.1755-0998.2009.02819.x>
- Pritchard J, Pickrell J, Coop G. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol.* 20(4):R208–R215. PMID:20178769. PMCid:PMC2994553. <http://dx.doi.org/10.1016/j.cub.2009.11.055>
- Pritchard J, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics.* 155(2):945–959. PMID:10835412. PMCid:PMC1461096.
- Przeworski M, Coop G, Wall J. 2005. The signature of positive selection on standing genetic variation. *Evolution.* 59(11):2312–2323. PMID:16396172. <http://dx.doi.org/10.1554/05-273.1>
- Puritz J, Addison J, Toonen R. 2012. Next-generation phylogeography: a targeted approach for multilocus sequencing of non-model organisms. *PLoS One.* 7(3):e34241. PMID:22470543. PMCid:PMC3314618. <http://dx.doi.org/10.1371/journal.pone.0034241>
- Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics.* 13:341. <http://dx.doi.org/10.1186/1471-2164-13-341>
- Quing Yun Y, YuHe Y. 2011. Metagenome-based analysis: A promising direction for plankton ecological studies. *Sci China Life Sci.* 54:75–81. PMID:21104154. <http://dx.doi.org/10.1007/s11427-010-4103-4>
- Radwan J, Babik W. 2012. The genomics of adaptation Introduction. *Proc Roy Soc Biol Sci.* 279(1749):5024–5028.
- Rannala B, Mountain J. 1997. Detecting immigration by using multilocus genotypes. *Proc Natl Acad Sci USA.* 94(17):9197–9201. PMID:9256459. PMCid:PMC23111. <http://dx.doi.org/10.1073/pnas.94.17.9197>
- Reaka-Kudla ML. 1997. The global biodiversity of coral reefs: a comparison with rain forests. *In: Reaka-Kudla ML, Wilson DE, Wilson EO, editors. Biodiversity II: understanding and protecting our biological resources.* Joseph Henry Press, Washington, DC. p. 83–108.
- Riccioni G, Landi M, Ferrara G, Milano I, Cariani A, Zane L, Sella M, Barbujani, Tinti F. 2009. Spatio-temporal population structuring and genetic diversity retention in depleted Atlantic bluefin tuna of the Mediterranean Sea. *Proc Nat Acad Sci USA.* 107:2102–2107. PMID:20080643. PMCid:PMC2836650. <http://dx.doi.org/10.1073/pnas.0908281107>
- Riesenfeld CS, Schloss PD, Handelsman J. 2004. Metagenomics: genomic analysis of microbial communities. *Ann Rev Genet.* 38:525–552. PMID:15568985. <http://dx.doi.org/10.1146/annurev.genet.38.072902.091216>
- Roberts CM, McClean CJ, Veron JEN, Hawkins JP, Allen GR, McAllister DE, Mittermeier CG, Schueler FW, Spalding M, Wells F, et al. 2002. Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science.* 295(5558):1280–1284. PMID:11847338. <http://dx.doi.org/10.1126/science.1067728>
- Rockman M, Kruglyak L. 2009. Recombinational landscape and population genomics of *Caenorhabditis elegans*. *PLoS Genet.* 5(3):e1000419. PMID:19283065. PMCid:PMC2652117. <http://dx.doi.org/10.1371/journal.pgen.1000419>
- Rohwer F, Breitbart M, Jara J, Azam F, Knowlton N. 2001. Diversity of bacteria associated with the Caribbean coral *Montastraea franksi*. *Coral Reefs.* 20(1):85–91. <http://dx.doi.org/10.1007/s003380100138>
- Ronce O, Kirkpatrick M. 2001. When sources become sinks: migrational meltdown in heterogeneous habitats. *Evolution.* 55(8):1520–1531. PMID:11580012.

- Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. 2010. Genome-wide associated studies in diverse populations. *Nature Rev Genet.* 11:356–366. PMID:20395969. PMCid:PMC3079573. <http://dx.doi.org/10.1038/nrg2760>
- Rousset F. 2008. Dispersal estimation: demystifying Moran's I. *Heredity.* 100(3):231–232. PMID:17895903. <http://dx.doi.org/10.1038/sj.hdy.6801065>
- Rowe H, Renaut S, Guggisberg A. 2011. RAD in the realm of next-generation sequencing technologies. *Mol Ecol.* 20(17):3499–3502. PMID:21991593.
- Rubin B, Ree R, Moreau C. 2012. Inferring phylogenies from RAD sequence data. *PLoS One.* 7(4):e33394. PMID:22493668. PMCid:PMC3320897. <http://dx.doi.org/10.1371/journal.pone.0033394>
- Rynerston TA, Palenik B. 2011. Learning to read the oceans: genomics of marine phytoplankton. In: Lesser M, editor. *Adv Mar Biol.* 60:1–39.
- Sala E, Knowlton N. 2006. Global marine biodiversity trends. *Annu Rev Environ Res.* 31:93–122. <http://dx.doi.org/10.1146/annurev.energy.31.020105.100235>
- Salayo N, Garcés L, Pido M, Viswanathan K, Pomeroy R, Ahmed M, Siason I, Seng K, Masae. 2008. Managing excess capacity in small-scale fisheries: perspectives from stakeholders in three Southeast Asian countries. *Mar Pol.* 32:692–700. <http://dx.doi.org/10.1016/j.marpol.2007.12.001>
- Sanford E, Kelly MW. 2011. Local adaptation in marine invertebrates. *Ann Rev Mar Sci.* 3:509–535. <http://dx.doi.org/10.1146/annurev-marine-120709-142756>
- Santos MD, Lopez GV, Barut NC. 2010. A pilot study on the genetic variation of Eastern little tuna (*Euthynnus affinis*) in Southeast Asia. *Philippine J Sci.* 139(1):43–50.
- Scaglione D, Acquadro A, Portis E, Tirone M, Knapp S, Lanteri S. 2012. RAD tag sequencing as a source of SNP markers in *Cynara cardunculus* L. *BMC Genomics.* 13:3. PMID:22214349. PMCid:PMC3269995. <http://dx.doi.org/10.1186/1471-2164-13-3>
- Schena M, Shalon D, Davis RW, Brown PO. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA Microarray. *Science.* 270(5235):467–470. PMID:7569999. <http://dx.doi.org/10.1126/science.270.5235.467>
- Schloss PD, Handelsman J. 2003. Biotechnological prospects from metagenomics. *Curr Opin Biotech.* 14(3):303–310. [http://dx.doi.org/10.1016/S0958-1669\(03\)00067-3](http://dx.doi.org/10.1016/S0958-1669(03)00067-3)
- Schmidt TM, Delong EF, Pace NR. 1991. Analysis of a marine picoplankton community by 16S ribosomal-RNA gene cloning and sequencing. *J Bacteriol.* 173(14):4371–4378. PMID:2066334. PMCid:PMC208098.
- Schmidt PS, Rand DM. 2001. Adaptive maintenance of genetic polymorphism in an intertidal barnacle: habitat- and life-stage-specific survivorship of *Mpi* genotypes. *Evolution.* 55(7):1336–1344. PMID:11525458.
- Shendure J, Ji H. 2008. Next-generation DNA sequencing. *Nat Biotech.* 26(10):1135–1145. PMID:18846087. <http://dx.doi.org/10.1038/nbt1486>
- Shinmura Y, Wee AKS, Takayama K, Asakawa T, Yllano OB, Salmo SG III, Ardli ER, Tung NX, Malekal NB, et al. 2012. Development and characterization of 15 polymorphic microsatellite loci in *Sonneratia alba* (Lythraceae) using next-generation sequencing. *Conserv Genet Res.* 4:811–814. <http://dx.doi.org/10.1007/s12686-012-9650-5>
- Slatkin M. 1987. Gene flow and the geographic structure of natural-populations. *Science.* 236(4803):787–792. PMID:3576198. <http://dx.doi.org/10.1126/science.3576198>
- Slatkin M. 2008. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nat Rev Genet.* 9:477–485. PMID:18427557. <http://dx.doi.org/10.1038/nrg2361>
- Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ. 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc Nat Acad Sci USA.* 103:12,115–12,120. PMID:16880384. PMCid:PMC1524930. <http://dx.doi.org/10.1073/pnas.0605127103>

- Somero GN. 2010. The physiology of climate change: how potentials for acclimatization and genetic adaptation will determine 'winners' and 'losers'. *J Exper Biol.* 213(6):912–920. PMID:20190116. <http://dx.doi.org/10.1242/jeb.037473>
- Sotka EE. 2005. Local adaptation in host use among marine invertebrates. *Ecol Lett.* 8(4):448–459. <http://dx.doi.org/10.1111/j.1461-0248.2004.00719.x>
- Spalding MD, Fox HE, Halpern BS, McManus MA, Molnar J, Allen GR, Davidson N, Jorge ZA, Lombana AL, Lourie SA, et al. 2007. Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *Bioscience.* 57:573–583. <http://dx.doi.org/10.1641/B570707>
- Stafford-Smith MG. 1993. Sediment-rejection efficiency of 22 species of Australian scleractinian corals. *Mar Biol.* 115:229–243. <http://dx.doi.org/10.1007/BF00346340>
- Stapley J, Reger J, Feulner PGD, Smadja C, Galindo J, Ekblom R, Bennison C, Ball AD, Beckerman AP, Slate J. 2010. Adaptation genomics: the next generation. *Trend Ecol Evol.* 25(12):705–712. PMID:20952088. <http://dx.doi.org/10.1016/j.tree.2010.09.002>
- Stephan W, Song YS, Langley CH. 2006. The hitchhiking effect on linkage disequilibrium between linked neutral loci. *Genetics.* 172(4):2647–2663. PMID:16452153. PMCid:PMC1456384. <http://dx.doi.org/10.1534/genetics.105.050179>
- Stolting KN, Nipper R, Lindtke D, Caseys C, Waeber S, Castiglione S, Lexer C. 2013. Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Mol Ecol.* 22:842–855. PMID:22967258. <http://dx.doi.org/10.1111/mec.12011>
- Storz J, Beaumont MA. 2002. Testing for genetic evidence of population expansion and contraction: an empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evolution.* 56:154–166. PMID:11913661.
- Storz JF. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol.* 14:671–688. PMID:15723660. <http://dx.doi.org/10.1111/j.1365-294X.2005.02437.x>
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* 123:585–595. PMID:2513255. PMCid:PMC1203831.
- Taylor NG, McAllister MK, Lawson GL, Carruthers T, Block BA. 2011. Atlantic bluefin tuna: a novel multistock spatial model for assessing population biomass. *PLoS One.* 6(12):e27693. <http://dx.doi.org/10.1371/journal.pone.0027693>
- Teshima K, Coop G, Przeworski M. 2006. How reliable are empirical genomic scans for selective sweeps? *Genome Res.* 16:702–712. PMID:16687733. PMCid:PMC1473181. <http://dx.doi.org/10.1101/gr.5105206>
- Torres-Dowdall J, Handelsman CA, Reznick DN, Ghalambor CK. 2012. Local adaptation and the evolution of phenotypic plasticity in Trinidadian guppies (*Poecilia reticulata*). *Evolution.* 66(11):3432–3443. PMID:23106708. <http://dx.doi.org/10.1111/j.1558-5646.2012.01694.x>
- Tringe SG, Rubin EM. 2005. Metagenomics: DNA sequencing of environmental samples. *Nat Rev Genet.* 6(11):805–814. PMID:16304596. <http://dx.doi.org/10.1038/nrg1709>
- Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. 2007. The Human Microbiome Project. *Nature.* 449(7164):804–810. PMID:17943116. PMCid:PMC3709439. <http://dx.doi.org/10.1038/nature06244>
- Unterseher M, Jumpponen A, Oepik M, Tedersoo L, Moora M, Dormann CF, Schnittler M. 2011. Species abundance distributions and richness estimations in fungal metagenomics—lessons learned from community ecology. *Mol Ecol.* 20(2):275–285. PMID:21155911. <http://dx.doi.org/10.1111/j.1365-294X.2010.04948.x>
- van Heesch S, Kloosterman WP, Lansu N, Ruzius FP, Levandowsky E, Lee CC, Zhou S, Gouldstein S, Schwartz DC, Harkins TT, et al. 2013. Improving mammalian genome scaffolding using large insert mate-pair next generation sequencing. *BMC Genomics.* 14:257. <http://dx.doi.org/10.1186/1471-2164-14-257>
- Van Orsouw NJ, Hogers RCJ, Janssen A, Yalcin E, Snoeijers S, Verstege E, Scheiders H, van der Poel H, van Oeveren J, Verstegen H, et al. 2007. Complexity reduction of polymorphic

- sequences (CRoPS): a novel approach for large-scale polymorphism discovery in complex genomes. *PLoS ONE*. 2:e1172. <http://dx.doi.org/10.1371/journal.pone.0001172>
- van Tassell CP, Smith TPL, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CJ, Haudenschild CD, Moore SS, Warren WC, Sonstegard TS. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Meth*. 5(3):247–252. PMID:18297082. <http://dx.doi.org/10.1038/nmeth.1185>
- Venter JC, Remington K, Heidelberg JE, Halpern AL, Rusch D, Eisen JA, Wu DY, Paulsen I, Nelson KE, Nelson W, et al. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science*. 304(5667):66–74. PMID:15001713. <http://dx.doi.org/10.1126/science.1093857>
- von der Heyden S, Beger M, Toonen RJ, van Herwerden L, Juinio-Meñez MA, Ravago-Gotanco R, Fauvelot C, Bernardi G. 2014. The application of genetics to marine management and conservation: examples from the Indo-Pacific. *Bull Mar Sci*. 90:123–158. <http://dx.doi.org/10.5343/bms.2012.1079>
- Wagner C, Keller I, Wittwer S, Selz O, Mwaiko S, Greuter L, Sivasundar A, Seehausen O. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Mol Ecol*. 22(3):787–798. PMID:23057853. <http://dx.doi.org/10.1111/mec.12023>
- Wakeley J. 2009. *Coalescent theory: an introduction*. Harvard University Press, Massachusetts. p. 352.
- Wang N, Thomson M, Bodles W, Crawford R, Hunt H, Featherstone A, Pellicer J, Buggs R. 2012a. Genome sequence of dwarf birch (*Betula nana*) and cross-species RAD markers. *Mol Ecol*. 22:3098–3111. PMID:23167599. <http://dx.doi.org/10.1111/mec.12131>
- Wang S, Meyer E, McKay J, Matz M. 2012b. 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nat Meth*. 9(8):808–810. PMID:22609625. <http://dx.doi.org/10.1038/nmeth.2023>
- Wang Z, Gerstein M, Snyder M. 2009. RNA-seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 10(1):57–63. PMID:19015660. PMID:PMC2949280. <http://dx.doi.org/10.1038/nrg2484>
- Wares J. 2010. Natural distributions of mitochondrial sequence diversity support new null hypotheses. *Evolution*. 64(4):1136–1142. PMID:19863588. <http://dx.doi.org/10.1111/j.1558-5646.2009.00870.x>
- Willig MR, Kaufman DM, Stevens RD. 2003. Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annual Ann Rev Ecol Evol Syst*. 34:273–309. <http://dx.doi.org/10.1146/annurev.ecolsys.34.012103.144032>
- Wilson G, Rannala B. 2003. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*. 163(3):1177–1191. PMID:12663554. PMID:PMC1462502.
- Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet*. 8(3):206–216. PMID:17304246. <http://dx.doi.org/10.1038/nrg2063>
- Wright S. 1978. *Evolution and the genetics of populations*. University of Chicago Press, Chicago.
- Yan Y, Cui H, Jiang S, Huang Y, Huang X, Wei S, Xu W, Qin Q. 2011. Identification of a novel marine fish virus, Singapore grouper iridovirus-encoded microRNAs expressed in grouper cells by solexa sequencing. *PLoS One*. 6:e19148. PMID:21559453. PMID:PMC3084752. <http://dx.doi.org/10.1371/journal.pone.0019148>
- Yang H, Tao Y, Zheng Z, Li C, Sweetingham M, Howieson J. 2012. Application of next-generation sequencing for rapid marker development in molecular plant breeding: a case study on anthracnose disease resistance in *Lupinus angustifolius* L. *BMC Genomics*. 13:318. PMID:22805587. PMID:PMC3430595. <http://dx.doi.org/10.1186/1471-2164-13-318>
- Yang Z, Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J Mol Evol*. 46:409–418. PMID:9541535. <http://dx.doi.org/10.1007/PL00006320>
- Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, Ding Z. 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Meth Ecol Evol*. 3(4):613–623. <http://dx.doi.org/10.1111/j.2041-210X.2012.00198.x>

- Zellmer A, Hanes M, Hird S, Carstens B. 2012. Deep phylogeographic structure and environmental differentiation in the carnivorous plant *Sarracenia alata*. *Syst Biol.* 61(5):763–777. PMID:22556200. <http://dx.doi.org/10.1093/sysbio/sys048>
- Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>
- Zhao X, Yu H, Kong L, Li Q. 2012. Transcriptomic responses to salinity stress in the Pacific Oyster *Crassostrea gigas*. *PLoS One.* 7:e46244. PMID:23029449. PMCid:PMC3459877. <http://dx.doi.org/10.1371/journal.pone.0046244>
- Zinger L, Amaral-Zettler LA, Fuhrman JA, Horner-Devine MC, Huse SM, Welch DBM, Martiny JBH, Sogin M, Boetius A, Ramette A. 2011. Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *PLoS One.* 6:e24570. <http://dx.doi.org/10.1371/journal.pone.0024570>

