水 研 機 構 研 報, 第 45 号, 21 − 27, 平 成 29 年
Bull. Jap. Fish. Res. Edu. Agen. No. 45, 21 − 27, 2017

Improving Aquaculture Production in *Haliotis* Species through the Development of a Genomic Toolkit

Catherine PURCELL[*1], Andrew SEVERIN[*2], Paula SYLVIA[*3], Nick WEGNER[*1], and John HYDE[*1]

Abstract: Commercial abalone aquaculture has greatly expanded over the past decade, becoming a thriving global industry valued at over $100 million USD. Abalone is one of the few species where culture production dominates the global market as a result of increasing demand and declining natural stocks from overexploitation and disease. U.S. abalone production is also growing due to high market value and demand. Most farms operating in California utilize three native west coast species: red abalone (*Haliotis rufescens*), green abalone (*H. fulgens*), and pink abalone (*H. corrugata*). These species differ in commercially important traits that are key to culture expansion in California and improved production efficiency (e.g., growth rate, disease resistance, thermal tolerance). Next generation sequencing has opened the door for extensive and explorative genetic research on abalone, and several studies were recently published examining transcriptomic data for species including the domestically important red abalone, *H. rufescens*. Recently, researchers working with the Southwest Fisheries Science Center (SWFSC) used restriction site associated DNA sequencing (RAD-Seq) methods to identify genome-wide SNP markers in *H. fulgens* and to examine population structure in wild populations. However, transcriptomic and RAD-mapping analyses were also limited by the paucity of genomic information available for abalone; without knowledge of the genomic structure, it is very difficult to ascertain coverage depth in these studies. We are working to create the first *de novo* abalone genome assembly using sexed *H. rufescens* samples, generate tissue specific transcriptomes for *H. rufescens*, and conduct comparative genomic analyses with other commercially important California abalone. Comparative analyses will include *H. fulgens*, *H. corrugata*, and endangered white (*H. sorenseni*) and black (*H. cracherodii*) abalone. Genomic, transcriptomic, and comparative analyses will improve our understanding of sex-determination, thermal/environmental preference, disease resistance, hybridization outcomes, and local adaptation, especially for commercially important California abalone. This will enable identification of candidate genes of interest and marker development for marker assisted selection to improve aquaculture practices in the U.S. abalone industry and elsewhere. This research will also help guide restoration and wild stock enhancement along the west coast for these species, including for the endangered white and black abalone.

Key words: *Haliotis, Haliotis rufescens, Haliotis sorenseni*, genome assembly, RNA-Seq

Introduction

In the U.S., approximately 200 metric tons of farmed abalone are produced annually (Gordon and Cook, 2013). Most domestic abalone farms operate in California and utilize three native west coast species: red (*Haliotis rufescens*) (Fig. 1), green (*H. fulgens*), and pink abalone (*H. corrugata*) (Allsopp *et al*, 2011) (Fig.2).

2017年 2 月28日受理（Received on February 28, 2017）
*1 National Marine Fisheries Service, 8901 La Jolla Shores Dr., 92037-1508 La Jolla, CA, U. S. A.
*2 Iowa State Genome Informatics Facility, Ames, IA 50011, U. S. A.
*3 Port of San Diego, 3165 Pacific Hwy, San Diego, CA 92101, U. S. A.
E-mail: catherine.purcell@noaa.gov

The dominant culture species, red abalone, grows quickly and reaches a large size in culture; popular in the U.S. market, they are also one of the most valuable species in the mollusk industry globally (Robertson, 2012; Aguilar-Espinoza *et al*., 2014; Brokordt *et al*., 2015b). Of the three species, *red abalone* is the most temperate while the green and pink abalone have more southern distributions and are of greater interest for aquaculture production in Southern California and Mexico, where they can be grown at higher water temperatures (McBride and Conte, 1996).

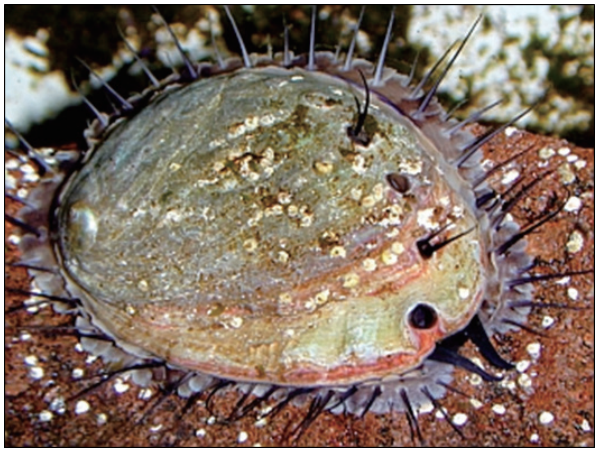

Fig. 1. Red abalone (*Haliotis rufescens*)

Although abalone culture is rapidly growing, it has been hindered by several bottlenecks that limit production capacity and efficiency in this industry. Improvements to the abalone stock in disease resistance and the other economically important traits are imperative to reducing production costs and to accelerating growth of the abalone industry, particularly in the United States (Arai and Okamura, 2013). The ability to overcome these bottlenecks will be significantly improved with better genomic resources for abalone. When correctly applied, these techniques may rapidly improve broodstock selection, characterize variation (both beneficial and detrimental), and provide methods to directly improve the value, efficiency, and production in the target species.

Considerable genetic research has been conducted for these species. Genetic markers have been generated to characterize variation in wild and farmed populations, characterize variation in traits of interest, and identify QTLs for important traits (e.g., Kang *et al*., 2010; Rhode *et al*., 2012; Aguilar-Espinoza *et al*., 2014; De Wit and Palumbi, 2013; Gruenthal *et al*., 2014). SWFSC partners have been developing

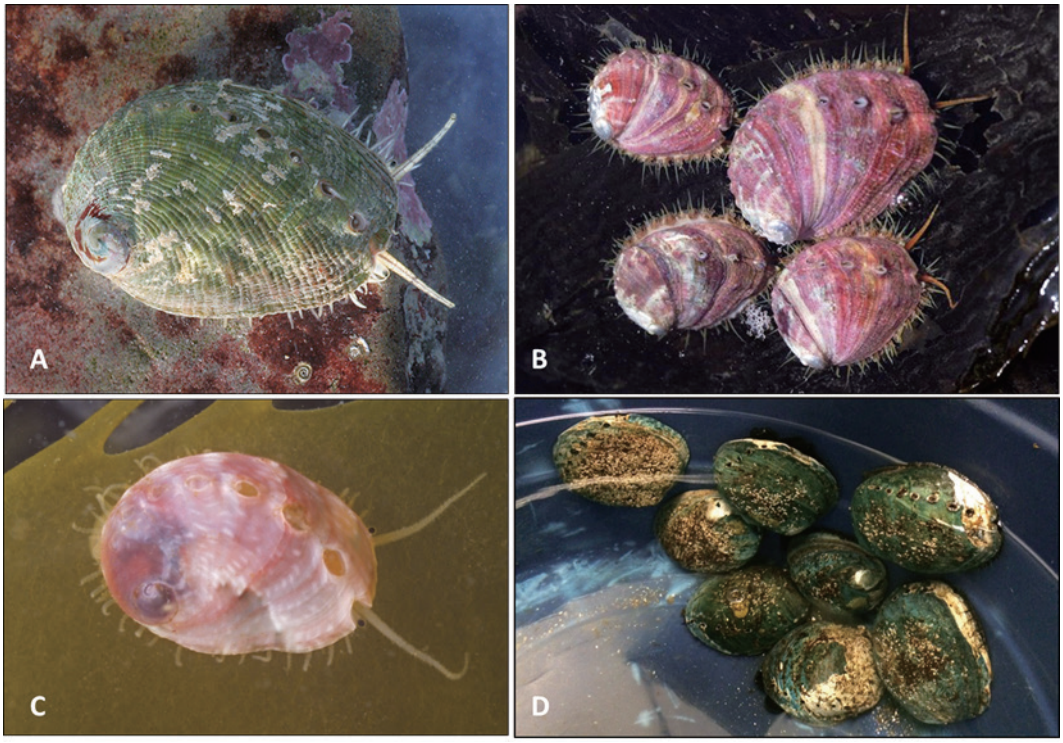

Fig. 2. Green abalone (*H. fulgens*) (a), white abalone (*H. sorenseni*) (b), pink abalone (*H. corrugata*) (c), and black abalone (*H. cracherodii*) (d)

suitable parentage and kinship microsatellite panels, genotyping broodstock animals, and applying these markers to monitor survival in outplanted green abalone. Genetic linkage maps have also been constructed for several abalone species using various genetic methods (e.g., amplified fragment length polymorphisms (AFLPs), microsatellites, and single nucleotide polymorphisms (SNPs)) (Liu *et al*., 2006; Sekino and Hara, 2007; Vervalle *et al*., 2013). While linkage-mapping and other genetic studies have started to improve culture practices for abalone, the usefulness of these studies is restricted by the low marker density that limits an understanding of mechanisms controlling traits of interest (Jones *et al*., 2013).

Next generation sequencing has opened the door for extensive and explorative genetic research on abalone, and several studies were recently published examining transcriptomic data for species including the domestically important red abalone (Franchini *et al*., 2011; Bester-Van der Merwe *et al*., 2013; De Wit and Palumbi, 2013). Recently, researchers working with the SWFSC used restriction site associated DNA sequencing (RAD-Seq) methods to identify genome-wide SNP markers in green abalone and examine population structure in wild populations (Gruenthal *et al*., 2014). However, transcriptomic and RAD-mapping analyses were also limited by the paucity of genomic information available for abalone; without knowledge of the genomic structure, it is very difficult to ascertain coverage depth in these studies (Franchini *et al*., 2011; Arai and Okamura, 2013). Additionally, abalone species are poorly represented in public genomic resource databases compared with other aquaculture species (Franchini *et al*., 2011), further limiting usefulness of the available data.

To develop the genomic resources that have benefited other breeding programs, we are working to generate a *de novo* genome assembly for the commercially valuable red abalone, generate tissue specific transcriptomes in red abalone, and conduct comparative genomic analyses through resequencing the pink, green, black (*H. cracherodii*) and white (*H. sorenseni*) abalone. Genomic resource development in red abalone will help identify and characterize economically and/or biologically relevant variation among the abalone species. This information will be an important step toward understanding of the genetic basis of traits that currently limit the growth of domestic abalone culture (i.e., infectious disease, thermal tolerance, slow growth, and the difficulty in determining sex). This research will greatly advance the state of knowledge of abalone culture and improve productivity in the U.S. abalone industry. The work will also provide insight for recovery efforts for the decimated endangered white and black abalone populations.

Methods

Mature and sexed live red abalone specimens will be sampled for both the genomic and transcriptomic sequencing. Sampling for the transcriptomic work will involve collection of several tissue types (e.g. foot muscle, gonad, digestive gland, epipodium). Sexed (male and female) pink, green, white and black abalone specimens were collected from the abalone culture tanks at the SWFSC and from the California Department of Fish and Wildlife.

Genomic DNA samples were stored in vials containing 95% ethanol; RNA samples submerged in RNAlater (Qiagen) and stored at -20 °C until use. Genomic DNA and RNA were extracted using DNeasy, Genomic Tip, and RNeasy kits (Qiagen), respectively.

Following preparation of the abalone specimens and quality assessment, the DNA and RNA samples were sent to the DNA facility at ISU (Ames, IA) for Illumina and PacBio library preparation and sequencing (Fig. 3). Sequencing will be conducted with the Illumina Hiseq 2500 on two lanes of 150 bp paired-end read data and four lanes of Illumina mate-paired end libraries, with large insert sizes of approximately 12 kb, for at least 100X genomic coverage. Twenty smart cells of PacBio data will be collected for approximately 5x coverage of the genome with expected average read lengths of 6 kb. The resequencing of the other abalone species will involve four lanes of 100bp paired-end Illumina sequencing. For all sequencing data, half of the lanes/smart cells will be male and half will be female.

Raw sequencing reads will be assembled with AllpathsLG (Gnerre *et al*., 2011) and MaSuRCA (Neale *et al*., 2014). The PacBio reads will either be error

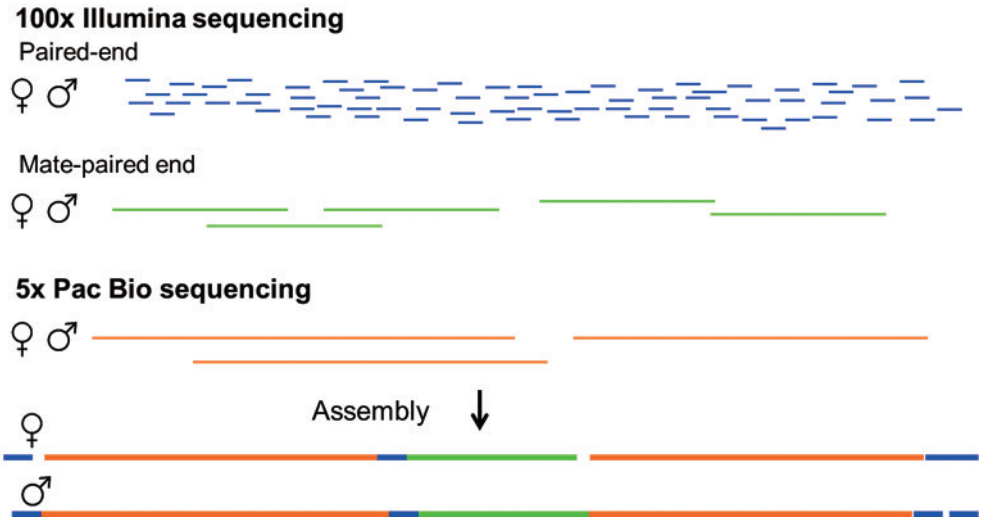

Fig. 3. The sequencing strategy using Illumina and PacBio technologies to generate the red abalone (*H. rufescens*) genome

corrected and included in the genome assembly as long "Illumina" reads, or used in super scaffolding after assembly. Transcripts will be assembled *de novo* from the available transcriptome data using Trinity (Haas *et al*., 2013). These transcripts will be aligned to the draft genome produced in the first step. Transcripts that map uniquely to the ends of two scaffolds will be used as evidence to join the two scaffolds. If the transcript alignment spans multiple small scaffolds (one or two exons per scaffold), multiple small scaffolds could be joined. We will use L_RNA_scaffolder (Xue et al. 2013) to perform this step. Gene models will be generated using MAKER (Holt and Yandell, 2011) genome annotation pipeline. The final set of genes will be predicted combining the results from all the predictors, EST evidence, protein alignment and BLAST similarity to other closely related genomes.

Raw reads from the other four species will be aligned to the red abalone genome using GSNAP (Wu and Nacu, 2010). Single Nucleotide Polymorphisms (SNPs) and Insertions/Deletions (InDels) will be called using GATK (McKenna *et al*., 2010). A pseudogenome of each of the four species will be generated using the aligned read file and the SNPs file. The pseudo genome assemblies will be directly comparable to each other and have annotations that correspond to annotations in red abalone, which will provide straightforward cross-referencing among red, green, pink, white, and black abalone gene models.

Results and discussion

Abalone samples are currently being prepared for sequencing and therefore no sequencing results are available at this time. However, it is expected that the *de novo* assemblies of the resequenced species will be explored for gene loss or gain of candidate genes that are known or suspected to be involved in sex, disease resistance, thermal tolerance, and growth rates (Brokordt *et al*., 2015a; Choi *et al*., 2015; Liang *et al*., 2014; Zippay and Hofmann, 2010; Klingbunga *et al*., 2009; Liu *et al*., 2006). Comparison of synonymous and nonsynonymous nucleotide substitution rates in homologous genes will be performed to measure the evolutionary distance among the five abalone species and detect genomic regions under selective pressure. Genes undergoing selective pressure will be interesting from an aquaculture perspective as they may relate to local adaptation and relevant genetic variation among the species (e.g. thermal tolerance, disease resistance).

Identification of sex determining regions that are specific to one sex will also be explored using the coverage depth of aligned reads to the red abalone genome assembly. Since one male and one female will be sequenced for each species, regions of the genome that show reduced coverage (typically one-half of the coverage for each sex, or one-fourth of the coverage of the combined male/female raw reads)

compared to the rest of the genome are prime candidates for a sex-determining region (Vicoso *et al*., 2013). Markers in these regions will be designed and tested by SWFSC to determine if a sex specific marker can be identified. This will be done for all five species where male and female individuals can be identified for sequencing.

To achieve the greatest impact on sustainable abalone aquaculture, we will create a genomic toolbox for abalone species that facilitates the integration of very large sequencing data sets, molecular markers, QTL data and genetic maps into an easy-to-use web interface. The website will include a Genome Browser utilizing JBrowse (Skinner *et al*., 2009) for visualization of the red abalone genome assembly and tracks that display the raw read and *de novo* assembled alignments of the other species. The genome assembly for red abalone and the pseudo genome assemblies for the other abalone species will be available for download for each species and deposited in the appropriate databases (e.g. NCBI-GenBank).

These tools will enable abalone aquaculture to move past the primary bottlenecks in production by moving toward a marker assisted selection model with breeding and/or hybridization programs that exploit natural diversity in abalone species to improve production capacity and efficiency. By selecting individual abalone or developing hybrids with desirable traits (e.g., disease resistance, thermal tolerance, and growth rate) the value of the aquaculture product will be greatly improved. In support of conservation and restoration efforts, the developed genetic toolkit can be used to monitor the health and viability of wild populations and guide breeding programs for stock replenishment.

References

Aguilar-Espinoza A., Valderrama-Aravena N., Farlora R., Lafarga-De la Cruz L., and Cristian Gallardo-Escàrate C., 2014: Development of novel polymorphic EST-SSR markers in Californian abalone *Haliotis rufescens* and genetic analysis in wild and hatchery-bred populations. *Aquac. Res.* **45**, 1942-1952.

Allsopp M., Lafarga-De la Cruz F., Flores-Aguilar R., and Watts E., 2011: Abalone culture: In Recent advances and new species in aquaculture (eds. Fotedar R. K. and Phillips B. F.), Wiley-Blackwell, Oxford,. 416pp.

Arai K. and Okumura S. I., 2013: Aquaculture-oriented genetic researches in abalone: Current status and future perspective. *African J. Biotech.* **12**, 4044-4052.

Bester-Van Der Merwe A., Blaauw S., Du Plessis J., and Roodt-Wilding R., 2013: Transcriptome-wide single nucleotide polymorphisms (SNPs) for Abalone (*Haliotis midae*): validation and application using GoldenGate medium-throughput genotyping assays. *Int. J. Mol. Sci.* **14**, 19341-19360.

Brokordt K., González R., Farías W., and Winkler F., 2015a: Potential response to selection of HSP70 as a component of innate immunity in the abalone *Haliotis rufescens*. *PloS One* **10**, e0141959.

Brokordt K. B., Winkler F. M., Farìas W. J., González R. C., Castaño F., Fullsack P., and Herbinger C. M., 2015b: Changes of heritability and genetic correlations in production traits over time in red abalone (*Haliotis rufescens*) under culture. *Aquac. Res.* **46**, 2248-2259.

Choi M., Kim D., Kim J. and Lim H., 2015: Differentially-expressed genes associated with faster growth of the Pacific abalone, *Haliotis discus hannai*. *Int. J. Mol. Sci.* **16**, 27520-27534.

De Wit P. and Palumbi S., 2013: Transcriptome-wide polymorphisms of red abalone *Haliotis rufescens*) reveal patterns of gene flow and local adaptation. *Mol. Ecol.* **22**, 2884-2897.

Franchini P., van der Merwe M., and Roodt-Wilding R., 2011: Transcriptome characterization of the South African abalone *Haliotis midae* using sequencing-by-synthesis. *BMC Res. Notes* **4**, 59.

Gnerre S., MacCallum I., Przybylski D., Ribeiro F. J., Burton J. N., Bruce J. Walker B. J., Sharpe T., Giles Hall G., Shea T. P., Sean Sykes S., Berlin A. M., Daniel Aird D., Maura Costello M., Daza R., Williams L., Nicol R., Gnirke A., Nusbaum C., Lander E. S., and Jaffe D. B., 2011: High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad Sci.* **108**, 1513–1518.

Gordon H. R. and Cook P. A., 2013: World abalone

supply, markets, and pricing: 2011 update. *J. Shell. Res.* **32**, 5-7.

Gruenthal K., Witting D., Ford T., Neuman M., Williams J., Pondella D., Bird A., Caruso N., Hyde J., Seeb L., and Larson W., 2014: Development and application of genomic tools to the restoration of green abalone in southern California. *Conservat. Genet.* **15**, 109-121.

Haas B. J., Papanicolaou A., Yassour M., Grabherr M., Blood P. D., Bowden J., Couger M. B., Eccles D., Li B., Lieber M., MacManes M. D., Ott M., Orvis J., Pochet N., Strozzi F., Weeks N., Westerman R., William T., Dewey C. N., Robert Henschel R., LeDuc R. D., Friedman N., and Regev A., 2013: De novo transcript sequence reconstruction from RNA-Seq using the Trinity platform for reference generation and analysis. *Nat. Prot.* **8**, 1494–1512.

Holt C. and Yandell M., 2011: MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinform.* **12**, 491.

Jones D. B., Jerry D. R., Khatkar M. S., Raadsma H. W., and Zenger K. R., 2013: A high-density SNP genetic linkage map for the silver-lipped pearl oyster, Pinctada maxima: a valuable resource for gene localisation and marker-assisted selection. *BMC Gen.* **14(1)**, 810.

Kang J.-H., Appleyard S. A., Elliott N. G., Jee Y.-J., Lee J.-B., Kang S.-W., Baek M. K., Han Y. S., Choi T.-J., and Lee T.-S., 2011: Development of genetic markers in abalone through construction of a SNP database. *Anim. Genet.* **42**, 309-315.

Klinbunga S., Amparyup P., Khamnamtong B., Hirono,I., Aoki T. and Jarayabhand P., 2009: Isolation and characterization of testis-specific DMRT1 in the tropical abalone (*Haliotis asinina*). *Biochem. Genet.* **47**, 66-79.

Liang S., Luo X., You W., Luo L. and Ke C., 2014: The role of hybridization in improving the immune response and thermal tolerance of abalone. *Fish & Shellfish Immunol.* **39**, 69-77.

Liu X., Liu X., Guo X., Gao Q., Zhao H., Zhang G., 2006: A preliminary genetic linkage map of the Pacific abalone *Haliotis discus hannai Ino. Mar. Biotech.* **8**, 386–397.

McBride S. and Conte F. S., 1996: California abalone culture: In California Culture. University of California, Davis, Dept. of Animal Science (ASAQ-A10), 5pp.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., and DePristo M. A., 2010: The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Gen. Res.* **20**, 1297–1303.

Neale D. B., Wegrzyn J. L., Stevens K. A., Zimin A. V., Daniela Puiu D., Crepeau M. W., Cardeno C., Koriabine M., Holtz-Morris A. E., Liechty J. D., Martínez-García P. J., Vasquez-Gross H. A., Lin B. Y., Zieve J. J., Dougherty W. M., Fuentes-Soriano S., Wu L. S., Gilbert D., Marçais G., Roberts M., Holt C., Yandell M., Davis J. M., Smith K. E., Dean J. F. D., Lorenz W. W., Whetten R. W., Sederoff R., Wheeler N., McGuire P. E., Main D., Loopstra C. A., Mockaitis K., deJong P. J., Yorke J. A., Salzberg S. L., and Langley C. H., 2014: Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Gen. Biol.* **15**, R59.

Rhode C., Hepple J. A., Jansen S., Davis T., Vervalle J., Bester-van der Merwe A. E., and Roodt-Wilding R., 2012: A population genetic analysis of abalone domestication events in South Africa: Implications for the management of the abalone resource. *Aquaculture* **356**, 235-242.

Robertson A., 2012: Abalone (*Haliotis* species): International - contained production systems. Seafood Watch Report, Monterey Bay Aquarium, 34pp.

Sekino M. and Hara M., 2007: Linkage maps for the Pacific abalone (genus *Haliotis*) based on microsatellite DNA markers. *Genetics* **175**, 945-958.

Skinner M., Uzilov A., Stein L., Mungall C., and Holmes I., 2009: JBrowse: a next-generation genome browser. *Genome Res.* **19**, 1630-1638.

Vervalle J., Hepple J. A., Jansen S., Du Plessis J., Wang P., Rhode C., and Roodt-Wilding R., 2013: Integrated linkage map of *Haliotis midae* Linnaeus based on microsatellite and SNP markers. *J. Shell. Res.* **32**, 89-103.

Vicoso B., Emerson J., Zektser Y., Mahajan S., and

Bachtrog D., 2013: Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. *PLoS Biol* **11**, e1001643.

Wu T.-D. and Nacu S., 2010: Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* (Oxford, England) **26**, 873–881.

Xue W., Li J.-T., Zhu Y.-P., Hou G.-Y., Kong X.-F., Kuang Y.-Y., and Sun X.-W., 2013: L_RNA_ scaffolder: scaffolding genomes with transcripts. *BMC Gen.* **14**, 604.

Zippay M. and Hofmann G., 2010: Effect of pH on gene expression and thermal tolerance of early life history stages of red abalone (*Haliotis rufescens*). *J. Shell. Res.* **29**, 429-439.

Annotated bibliography

(1) De Wit P. and Palumbi S., 2013: Transcriptome-wide polymorphisms of red abalone (*Haliotis rufescens*) reveal patterns of gene flow and local adaptation. *Mol. Ecol.* 22, 2884-2897.

The authors of this study tested whether signals of environmental selection could be detected in samples of red abalone (*Haliotis rufescens*) collected from three locations in California: Monterey Bay, Sonoma, and north of Cape Mendocino. These particular areas are especially distinct in terms of their temperature, aragonite saturation, exposure to hypoxia stress, and disease pressure; as such, the authors hypothesized that genes related to shell biomineralization, resistance to hypoxia, temperature tolerance, and resistance to pathogens would show the strongest signals of local adaptation. The authors tested this by conducting RNA-Seq analyses on mantle tissue of 39 red abalone individuals from the above locations. A total of 21,579 SNPs were genotyped for each individual, and out of these 691 showed significant differentiation. From this set of 691, 163 loci could be identified through BLAST annotation; many of these genes had functions related to biomineralization, energy metabolism, heat-, and disease- or hypoxia-tolerance. These genes are now candidates for further studies to look for signals of local adaption.

(2) Franchini P., van der Merwe M., and Roodt-Wilding R., 2011: Transcriptome characterization of the South African abalone Haliotis midae using sequencing-by-synthesis. BMC Res. Notes 4, 59.

This study describes the use of applying next generation sequencing technology to develop molecular tools for a South African abalone species, Haliotis midae. They use the Illumina Genome Analyzer II to generate 25 million sequences. Using the transcriptome sequences, they did a de novo assembly that resulted in 27,761 contigs with an average length of 260 bp. Importantly, although abalone have a relatively poor representation in genome databases likely due to their large genome size), a good number of the contigs had BLAST matches to known annotated genes in Genbank; with a stringent e-value set, 16.8% of the contigs had a homologous BLAST match against Genbank. These sequences were assigned to functional categories using GO and KOG databases. The authors were also able to use this data to identify thousands of SNPs, and out of those, they developed 420 primer sets.

(3) Gruenthal K., Witting D., Ford T., Neuman M., Williams J., Pondell, D., Bird A., Caruso N., Hyde J., Seeb L., and Larson W., 2014: Development and application of genomic tools to the restoration of green abalone in southern California. *Conservat. Genet.* 15, 109-121.

The authors describe the use of next generation sequencing technology to develop an extensive set of markers to test population structure and effective population size in green abalone in Southern California. In this study, RADSeq restriction site associated DNA sequencing) was used to generate millions of short sequences, from which, many thousands of SNPs may be identified that span a greater proportion of the genome compared to previous types of marker development. A total of 1209 polymorphic SNPs were developed from this sequencing. While the extensive set of markers did not detect population structure in green abalone in the range that was sampled, they were able to estimate an effective population size Ne) of 1,100-3,600 individuals. Importantly, this work generated valuable genomic resources that can be used to further build the set of tools available to study Haliotis species.