**Title of the Study**

Development of a tool to sequence the exome of any organism rapidly and cost-effectively

**Author**

Jacob Michael Green

**Committee**

Dr. Jonathan B. Puritz, Dr. Marta Gomez-Chiarri, Dr. Dina Proestou, Dr. Bryan Dewsbury

**Statement of the Problem**

Next generation sequencing (NGS) techniques, like whole genome sequencing and RNA sequencing, have greatly increased our capacity to explore issues of genetic conservation and adaptation, but are costly and time-intensive to accurately assess regional population diversity and adaptation. Target capture sequencing allow us to reduce cost and selectively enrich only the specific expressed exons, but this design requires a robust and well-annotated genome. We are developing a novel sequencing technique called Expressed Exome Capture Sequencing (EecSeq) to sequence the exome of any organism rapidly and cost-effectively. Our research will transform a proof of concept into a readily available approach by improving and optimizing the molecular laboratory workflow, comparing, and validating EecSeq against other NGS approaches, and developing a reproducible and open source bioinformatic pipeline to analyze EecSeq sequencing data accessible to any evolutionary biologist. This project will provide researchers with the information, justification, and protocols needed to implement EecSeq for their own questions.

**Justification and Significance of the Study**

*“Life is the Universe Developing a Memory” – Dr. Lee Cronin*

Every organism alive today holds in its precious cells the genetic memory of life on earth. The study of genetic memory and how it is passed down (evolutionary biology) is in its infancy, beginning with our understanding of inheritance through breeding to now leveraging the next generation sequencing (NGS) tools that illuminate the deoxyribonucleic acid (DNA) sequences that form our genetic memories. We have gone from tracking pedigrees with Punnett squares to parsing the billions of nucleotides that dictate our biological relationship with the environment. These billions of nucleotides form our genome, and each individual organism has a unique mosaic of nucleotides that tell the evolutionary story of their lineage all the way back to our shared universal common ancestor. New sequencing tools are being developed every year expanding the types of questions we can ask and refining the conversations we create through our research.

NGS tools have transformed our approach to investigating evolution and adaptation. The Human Genome Project is arguably the greatest achievement in evolutionary biology in the 21st century. Through multi-national and multi-institutional collaborations, we declared the Human genome complete in 2003 to the tune of 3 billion dollars and 13 years of development. But we left many parts of the Human genome unexplored primarily due to its complexity and the limitations of our NGS toolkits. It was not until May 2022 that researchers finally declared “We have completed sequencing all of the Human genome!”. Within this 20-year period we have sequenced the genomes and transcriptomes of many other organisms leveraging tools such as whole-genome sequencing (WGS) and ribonucleic acid sequencing (RNA-seq). Organisms such as the fruit fly (*Drosophila melanogaster*), zebra fish (*Danio rerio*), Corn (*Maize*), and the house mouse (*Mus musculus*) have had their model organism status cemented through NGS applications and these resources allow us to deeply investigate the tree of life. While WGS and RNA-seq have been essential to in developing resources for model organisms, these tools can be prohibitively expensive and time-intensive for studying non-model organisms.

Genome reduction methods are needed to apply NGS technologies to non-model or obscure model species. The field of population genomics, understanding how our genes contribute to the health and well-being of groups of organisms, uses multiple tools that significantly reduced the amount of genetic information assessed. This reduction of information reduces costs associated with molecular library preparation, sequencing genomic reads, and processing read data computationally. Restriction-site associated DNA sequencing (RAD-seq) is one of the most widely used genome reduction tools as it uses restriction enzymes that recognize very specific DNA motifs to

My thesis seeks to develop EecSeq as an efficient genome reduction method through taking our initial proof of concept research and moving it towards a grounded method. This project will be completed in three phases. Phase 1 will focus on developing a bioinformatic pipeline to analyze EecSeq data. Phase 2 will optimize the molecular lab protocol to be the most efficient when creating probe pools and capturing exome sequences in the gDNA libraries. Phase 3 will be a comparison with traditional commercial capture probe design and validation of genotyping accuracy with Oysters used in building the Easter Oyster Consortium reference genome. This process will provide a holistic approach from the lab bench to the computational environment for other evolutionary biologists to utilize. All information including lab protocols, bioinformatic pipelines, and data will be made available on the Puritz lab of Marine Evolutionary Ecology [Github](https://github.com/MarineEvoEcoLab).

**Research Design**

Phase 1 of our research design will be developing a bioinformatic pipeline to process, assemble, and analyze EecSeq data. Due to a lack of existing resources for non-model or obscure model organisms we will approach our assembly of Oyster EecSeq reads without using the published reference but will leverage the Eastern Oyster Genome to assess our different assembly approaches. To design a de novo and hybrid pipeline we will use the genetic data generated through the molecular laboratory protocol, specifically the cDNA libraries and the captured gDNA. Data generated from Phases 2 and 3 will be used to test our two EecSeq assembly pipelines, but the development of objective 1a and 1b will utilize previously sequenced gDNA and cDNA reads from CASE and Heat Shock experiments. Our goal for this phase of the project is to publish an open-source bioinformatics pipeline for EecSeq.

For objective 1a we will use captured gDNA reads to design a de novo assembly pipeline for EecSeq. These reads were sequenced from a simple heat shock exposure of four individuals exposed to 36 C for 1 hour and a control group of 4 individuals at ambient temperatures. Our assembly methods will leverage a diversity of de novo transcriptome assembly workflows and additional read processing steps co-opted from the ddocent pipeline. Transcriptome assembly programs are effective in assembling full length genes and therefore should be able to assemble EecSeq data which is comprised of exon loci. We hypothesize that de novo transcriptome assembly programs will generate high quality exome assemblies that can guide genotyping and SNP calling.

For objective 1b we will be using both the capture gDNA reads and the cDNA EecSeq probes to design a hybrid assembly pipeline. During the laboratory protocol for EecSeq we synthesize high quality cDNA libraries that can be sequenced either concurrently or sequentially in a planned research design. This data was generated from a CASE experiment where we sequenced both the cDNA libraries before probe synthesis and the capture gDNA reads. There are two approaches we will take for cDNA transcript assembly: 1) De novo transcriptome assembly to generate a high-quality reference 2) Merged and clustered cDNA reads that will act as “seeds” to guide Exome assembly. The first approach uses all the same de novo assembly programs in objective 1a which streamlines the overall pipeline architecture and file structure. In this method we are using the de novo transcriptome as a sudo reference to then map the gDNA reads. The second approach leverages a different assembly program that uses cDNA reads as a “seed” from which to assemble the gDNA reads. While it is a guided or reference-based assembly, we are only using data that has been generated within the context of our study. The result of this competitive assembly approach will be integrating only one of these methods into the final published EecSeq bioinformatic pipeline

Phase 2: Optimize molecular lab protocol for probe length, probe diversity, and library hybridization

Data sources

* Objective 2a: Determine the optimal level of probe and capture pool diversity to assess drop out of divergent alleles
* Objective 2b: Optimize capture probe to capture pool insert libraries ratio for efficient hybridization and capture

Phase 3: Compare EecSeq with traditional target capture methods and validate genotyping accuracy with Eastern Oyster Consortium reference individual samples

Data sources

* Objective 3.1: Compare commercial exome capture probes designed from the Eastern Oyster Reference genome to EecSeq probes using identical samples
* Objective 3.2: Validate genotyping accuracy and test ability to target genes of interest by utilizing refence individuals form the Eastern Oyster genome project

**Methods in Development**

**Expressed Exome Capture Sequencing Molecular Protocol**

*Tissue collection and DNA/RNA extraction*

Oysters were dissected live 0 – 2 days after sampling and were stored on ice in a 4 C fridge. Oyster shells were measured with electronic calipers for shell height and length. Each Oyster was weighed before dissection and after dissection to determine tissue weight. Oyster mantle tissue was dissected in triplicate and gill tissue was dissected in quadruplicate and stored in 1.5 ml tubes. Tissue tubes were immediately flash frozen in liquid nitrogen and stored in -80 until extraction.

Oyster tissue DNA/RNA was homogenized by putting ~ 800 uL DNA/RNA shield into a ceramic bead Qiagen Tissuelyser tube. Bead tubes were then placed in Qiagen Tissuelyzer II for 2 minutes at 30 Hz. DNA/RNA extractions used the Zymogen Quick DNA/RNA extractions kits using custom protocols. Tissues were first treated with Proteinase K and incubated for 30 minutes at 25 C and shaking at 600 rpm on an Eppendorf thermomixer. After centrifugation and transfer of supernatant containing DNA/RNA 800 uL of DNA/RNA lysis buffer was added and mixed well via vortex. For DNA purification. 750 uL of supernatant was transferred to DNA spin away column at a time and centrifuged at 16,000 rcf for 1 min. Flow-through was saved in a 5 ml centrifuge tube for RNA purification. After all lysis solution was added to the column it was treated with DNA/RNA prep and wash buffer. We then added 50 uL of heated TRIS-HCl (pH 8.0) to the column and centrifuged at 16,000 rcf for 1 minute. This process was repeated once, and DNA elute was stored in 1.5 ml tubes in -20 C fridge. Flow-through from DNA column was placed in a 5 mL tube and an equal volume of 100% EtOH (~ 1690 uL) was added to the samples and mix well via vortex. For RNA purification, 750 uL of supernatant was added to the RNA spin-away column and the flow through was discarded. Once all supernatant was added to the column, 400 uL of DNA/ RNA wash buffer was added and centrifuged for 16,000 rcf for 1 minute. To remove DNA contamination, 80 uL of DNase/DNA digestion buffer master mix was added directly to the column and incubated for 15 minutes at 32 C in the Eppendorf thermomixer. RNA spin-away column was treated with DNA/RNA pre and wash buffer. We then added 50 uL of heated TRIS-HCl (pH 8.0) to the column and centrifuged at 16,000 rcf for 1 minute. This process was repeated once, and RNA elute was stored in 1.5 ml tubes in -80 C fridge. For an example of this method please see this [protocol](https://github.com/madmolecularman/JMG_Puritz_Lab_Notebook/blob/main/_posts/2021-01-08-Oyster_DNA_RNA_extractions1.md). All extractions were quantified with Biotium BR assay and visualized on the Agilent Tapestation.

*Custom adapter and primers*

Custom oligonucleotide adapters were designed specifically for the EecSeq protocol. The RNA adapters have a similar design to the Illumina TruSeq adapters but include the SaII-HF restriction site at the 3’ end of the Universal adapter and the 5’ end of the indexed adapter. The location of the SaII restriction site allows the Illumina universal adapter and primers to be removed during probe synthesis to prevent interference during hybridization between probe and gDNA pools. RNA adapters are first diluted to 200 uM with a solution of TRIS-HCl (pH 8.0), NaCl, and EDTA. Then SaIIa and SaIIb adapters are annealed at 40 uM by adding 20 uL of each 200 uM adapter to 10 uL of 10 X annealing solution and 50 uL of RNase free H20. This solution is then heated to 97.5 C for 2.5 min and cooled at a rate of 3 C/min until solution reaches a temperature of 21 C. The annealed working solution can be used to create dilutions for different library preparations depending on the input RNA concentration. EecSeq DNA adapters are designed as custom TruSeq style barcoded adapters that facilitate multiplexing various gDNA libraries on one sequencing lane and subsequent bioinformatic demultiplexing. These DNA adapters do not need to be annealed in equimolar concentrations but should be diluted to allow for an adapter to gDNA library insert ratio of 50:1. For an adapter walkthrough please see this [walkthrough](https://github.com/madmolecularman/JMG_Puritz_Lab_Notebook/blob/main/protocols/EecSeq_adapter_design_walkthrough_v1.md). All adapter sequences can be found in Table 1.

*mRNA library preparation and normalization*

After selecting individuals that comprise the probe set, RNA is pooled into individual tubes to ~ 4000 ng/uL. To prepare mRNA libraries we used the mRNA KAPA Hyperprep kit with the following modifications: Custom RNA Sall adapters were used, adapters were used at 7 uM concentration for input adapter ligation, half volume reactions were used for all steps except for the volume of EtOH used in bead cleanups, custom i5 and i7 illumina primers were used for library amplification PCR, 12-14 cycles of PCR were used for library enrichment. For a full walkthrough please see this [protocol](https://github.com/madmolecularman/JMG_Puritz_Lab_Notebook/blob/main/_posts/2022-04-22-mRNA-library-preparations.md). Completed libraries were quantified with Biotium BR assay and visualized on the Agilent Tapestation with the D1000 kit.

*Probe synthesis*

*gDNA library synthesis*

*Hybridization*

**Bioinformatic Assembly of Exome Capture Reads**

*Exome read processing*

*De novo exome assembly*

*Hybrid exome assembly*

*Assessment of exome assemblies*

*Exome read mapping*

*Genotyping and SNP analysis*

**Figure 1:**

**Table 1:**

**Literature Cited**

1. Abrouk M, Athiyannan N, Müller T, Pailles Y, Stritt C, Roulin AC, Chu C, Liu S, Morita T, Handa H, et al. 2021. Population genomics and haplotype analysis in spelt and bread wheat identifies a gene regulating glume color. Commun. Biol. 4:1–11. doi:10.1038/s42003-021-01908-6.
2. Alex Buerkle C, Gompert Z. 2013. Population genomics based on low coverage sequencing: How low should we go? Mol. Ecol. 22:3028–3035. doi:10.1111/mec.12105.
3. Alexander MA, Scott JD, Friedland KD, Mills KE, Nye JA, Pershing AJ, Thomas AC. 2018. Projected sea surface temperatures over the 21st century: Changes in the mean, variability and extremes for large marine ecosystem regions of Northern Oceans. Elem Sci Anth 6:9. doi:10.1525/elementa.191.
4. Ali OA, O’Rourke SM, Amish SJ, Meek MH, Luikart G, Jeffres C, Miller MR. 2016. Rad capture (Rapture): Flexible and efficient sequence-based genotyping. Genetics 202:389–400. doi:10.1534/genetics.115.183665.
5. Allendorf FW, Hohenlohe PA, Luikart G. 2010. Genomics and the future of conservation genetics. Nat. Rev. Genet. 11:697–709. doi:10.1038/nrg2844.
6. Almany GR, Planes S, Thorrold SR, Berumen ML, Bode M, Saenz-Agudelo P, Bonin MC, Frisch AJ, Harrison HB, Messmer V, et al. 2017. Larval fish dispersal in a coral-reef seascape. Nat. Ecol. Evol. 1:1–7. doi:10.1038/s41559-017-0148.
7. Anand S, Mangano E, Barizzone N, Bordoni R, Sorosina M, Clarelli F, Corrado L, Boneschi FM, D’Alfonso S, De Bellis G. 2016. Next generation sequencing of pooled samples: Guideline for variants’ filtering. Sci. Rep. 6:1–9. doi:10.1038/srep33735.
8. 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 3:1–7. doi:10.1371/journal.pone.0003376.
9. Baruzzo G, Hayer KE, Kim EJ, DI Camillo B, Fitzgerald GA, Grant GR. 2017. Simulation-based comprehensive benchmarking of RNA-seq aligners. Nat. Methods 14:135–139. doi:10.1038/nmeth.4106.
10. Benestan L, Quinn BK, Maaroufi H, Laporte M, Clark FK, Greenwood SJ, Rochette R, Bernatchez L. 2016. Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (Homarus americanus). Mol. Ecol. 25:5073–5092. doi:10.1111/mec.13811.
11. Bernatchez S, Xuereb A, Laporte M, Benestan L, Steeves R, Laflamme M, Bernatchez L, Mallet MA. 2019. Seascape genomics of eastern oyster (Crassostrea virginica) along the Atlantic coast of Canada. Evol. Appl. 12:587–609. doi:10.1111/eva.12741.
12. Bernatchez S, Xuereb A, Laporte M, Benestan L, Steeves R, Laflamme M, Bernatchez L, Mallet MA. 2019. Seascape genomics of eastern oyster (Crassostrea virginica) along the Atlantic coast of Canada. Evol. Appl. 12:587–609. doi:10.1111/eva.12741.
13. Bi K, Vanderpool D, Singhal S, Linderoth T, Moritz C, Good JM. 2012. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. BMC Genomics 13. doi:10.1186/1471-2164-13-403.
14. Booker TR, Yeaman S, Whitlock MC. 2020. Variation in recombination rate affects detection of outliers in genome scans under neutrality. Mol. Ecol. 29:4274–4279. doi:10.1111/mec.15501.
15. Booker TR, Yeaman S, Whitlock MC. 2020. Variation in recombination rate affects detection of outliers in genome scans under neutrality. Mol. Ecol. 29:4274–4279. doi:10.1111/mec.15501.
16. Brennan RS, Garrett AD, Huber KE, Hargarten H, Pespeni MH. 2019. Rare genetic variation and balanced polymorphisms are important for survival in global change conditions. Proc. R. Soc. B Biol. Sci. 286. doi:10.1098/rspb.2019.0943.
17. Brind’Amour A, Mahévas S, Legendre P, Bellanger L. 2018. Application of Moran Eigenvector Maps (MEM) to irregular sampling designs. Spat. Stat. 26:56–68. doi:10.1016/j.spasta.2018.05.004.
18. Capblancq T, Forester BR. 2021. Redundancy analysis: A Swiss Army Knife for landscape genomics. Methods Ecol. Evol. 12:2298–2309. doi:10.1111/2041-210X.13722.
19. Caron DA, Alexander H, Allen AE, Archibald JM, Armbrust EV, Bachy C, Bell CJ, Bharti A, Dyhrman ST, Guida SM, et al. 2017. Probing the evolution, ecology and physiology of marine protists using transcriptomics. Nat. Rev. Microbiol. 15:6–20. doi:10.1038/nrmicro.2016.160.
20. Chapman RW, Mancia A, Beal M, Veloso A, Rathburn C, Blair A, Holland AF, Warr GW, Didinato G, Sokolova IM, et al. 2011. The transcriptomic responses of the eastern oyster, Crassostrea virginica, to environmental conditions. Mol. Ecol. 20:1431–1449. doi:10.1111/j.1365-294X.2011.05018.x.
21. Chapman RW, Mancia A, Beal M, Veloso A, Rathburn C, Blair A, Sanger D, Holland AF, Warr GW, Didonato G. 2009. A transcriptomic analysis of land-use impacts on the oyster, Crassostrea virginica, in the South Atlantic bight. Mol. Ecol. 18:2415–2425. doi:10.1111/j.1365-294X.2009.04194.x.
22. Cherry KM, Qian L. 2018. Scaling up molecular pattern recognition with DNA-based winner-take-all neural networks. Nature 559:370–388. doi:10.1038/s41586-018-0289-6.
23. Christmas MJ, Biffin E, Breed MF, Lowe AJ. 2016. Finding needles in a genomic haystack: targeted capture identifies clear signatures of selection in a nonmodel plant species. Mol. Ecol. 25:4216–4233. doi:10.1111/mec.13750.
24. Combs M, Puckett EE, Richardson J, Mims D, Munshi-South J. 2018. Spatial population genomics of the brown rat (Rattus norvegicus) in New York City. Mol. Ecol. 27:83–98. doi:10.1111/mec.14437.
25. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szcześniak MW, Gaffney DJ, Elo LL, Zhang X, et al. 2016. A survey of best practices for RNA-seq data analysis. Genome Biol. 17:1–19. doi:10.1186/s13059-016-0881-8.
26. Coutelier M, Hammer MB, Stevanin G, Monin ML, Davoine CS, Mochel F, Labauge P, Ewenczyk C, Ding J, Gibbs JR, et al. 2018. Efficacy of exome-targeted capture sequencing to detect mutations in known cerebellar ataxia genes. JAMA Neurol. 75:591–599. doi:10.1001/jamaneurol.2017.5121.
27. Cowley MJ, Liu YC, Oliver KL, Carvill G, Myers CT, Gayevskiy V, Delatycki M, Vlaskamp DRM, Zhu Y, Mefford H, et al. 2019. Reanalysis and optimisation of bioinformatic pipelines is critical for mutation detection. Hum. Mutat. 40:374–379. doi:10.1002/humu.23699.
28. Davey JL, Blaxter MW. 2010. RADseq: Next-generation population genetics. Brief. Funct. Genomics 9:416–423. doi:10.1093/bfgp/elq031.
29. 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. doi:10.1111/mec.12084.
30. De Mita S, Thuillet AC, Gay L, Ahmadi N, Manel S, Ronfort J, Vigouroux Y. 2013. Detecting selection along environmental gradients: Analysis of eight methods and their effectiveness for outbreeding and selfing populations. Mol. Ecol. 22:1383–1399. doi:10.1111/mec.12182.
31. De Oliveira AL, Wollesen T, Kristof A, Scherholz M, Redl E, Todt C, Bleidorn C, Wanninger A. 2016. Comparative transcriptomics enlarges the toolkit of known developmental genes in mollusks. BMC Genomics 17. doi:10.1186/s12864-016-3080-9.
32. De Villemereuil P, Frichot É, Bazin É, François O, Gaggiotti OE. 2014. Genome scan methods against more complex models: When and how much should we trust them? Mol. Ecol. 23:2006–2019. doi:10.1111/mec.12705.
33. Dolby GA, Ellingson RA, Findley LT, Jacobs DK. 2018. How sea level change mediates genetic divergence in coastal species across regions with varying tectonic and sediment processes. Mol. Ecol. 27:994–1011. doi:10.1111/mec.14487.
34. Dong C, Zhang L, Chen Z, Xia C, Gu Y, Wang J, Li D, Xie Z, Zhang Q, Zhang X, et al. 2020. Combining a New Exome Capture Panel With an Effective varBScore Algorithm Accelerates BSA-Based Gene Cloning in Wheat. Front. Plant Sci. 11:1–12. doi:10.3389/fpls.2020.01249.
35. Dumbauld BR, Ruesink JL, Rumrill SS. 2009. The ecological role of bivalve shellfish aquaculture in the estuarine environment: A review with application to oyster and clam culture in West Coast (USA) estuaries. Aquaculture 290:196–223. doi:10.1016/j.aquaculture.2009.02.033.
36. Epelboin Y, Quintric L, Guévélou E, Boudry P, Pichereau V, Corporeau C. 2016. The kinome of Pacific oyster Crassostrea gigas, its expression during development and in response to environmental factors. PLoS One 11:1–24. doi:10.1371/journal.pone.0155435.
37. Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61:717–726. doi:10.1093/sysbio/sys004.
38. Flanagan SP, Jones AG. 2018. Substantial differences in bias between single-digest and double-digest RAD-seq libraries: A case study. Mol. Ecol. Resour. 18:264–280. doi:10.1111/1755-0998.12734.
39. Gasc C, Peyretaillade E, Peyret P. 2016. Sequence capture by hybridization to explore modern and ancient genomic diversity in model and nonmodel organisms. Nucleic Acids Res. 44:4504–4518. doi:10.1093/nar/gkw309.
40. Geniza M, Jaiswal P. 2017. Tools for building de novo transcriptome assembly. Curr. Plant Biol. 11–12:41–45. doi:10.1016/j.cpb.2017.12.004.
41. Gibert P, Debat V, Ghalambor CK. 2019. Phenotypic plasticity, global change, and the speed of adaptive evolution. Curr. Opin. Insect Sci. 35:34–40. doi:10.1016/j.cois.2019.06.007.
42. Glover N, Dessimoz C, Ebersberger I, Forslund SK, Gabaldón T, Huerta-Cepas J, Martin MJ, Muffato M, Patricio M, Pereira C, et al. 2019. Advances and Applications in the Quest for Orthologs. Mol. Biol. Evol. 36:2157–2164. doi:10.1093/molbev/msz150.
43. Golestan Hashemi FS, Razi Ismail M, Rafii Yusop M, Golestan Hashemi MS, Nadimi Shahraki MH, Rastegari H, Miah G, Aslani F. 2018. Intelligent mining of large-scale bio-data: Bioinformatics applications. Biotechnol. Biotechnol. Equip. 32:10–29. doi:10.1080/13102818.2017.1364977.
44. Grover CE, Salmon A, Wendel JF. 2012. Targeted sequence capture as a powerful tool for evolutionary analysis. Am. J. Bot. 99:312–319. doi:10.3732/ajb.1100323.
45. Hart MW, Marko PB. 2010. It’s about time: Divergence, demography, and the evolution of developmental modes in marine invertebrates. Integr. Comp. Biol. 50:643–661. doi:10.1093/icb/icq068.
46. Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, Poss ML, Reed LK, Storfer A, Whitlock MC. 2016. Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. Am. Nat. 188:379–397. doi:10.1086/688018.
47. Hodel RGJ, Chen S, Payton AC, McDaniel SF, Soltis P, Soltis DE. 2017. Adding loci improves phylogeographic resolution in red mangroves despite increased missing data: Comparing microsatellites and RAD-Seq and investigating loci filtering. Sci. Rep. 7:1–14. doi:10.1038/s41598-017-16810-7.
48. Hoffberg SL, Kieran TJ, Catchen JM, Devault A, Faircloth BC, Mauricio R, Glenn TC. 2016. RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. Mol. Ecol. Resour. 16:1264–1278. doi:10.1111/1755-0998.12566.
49. Hölzer M, Marz M. 2019. De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. Gigascience 8:1–16. doi:10.1093/gigascience/giz039.
50. Honaas LA, Wafula EK, Wickett NJ, Der JP, Zhang Y, Edger PP, Altman NS, Chris Pires J, Leebens-Mack JH, DePamphilis CW. 2016. Selecting superior de novo transcriptome assemblies: Lessons learned by leveraging the best plant genome. PLoS One 11:1–42. doi:10.1371/journal.pone.0146062.
51. Jones MR, Good JM. 2016. Targeted capture in evolutionary and ecological genomics. Mol. Ecol. 25:185–202. doi:10.1111/mec.13304.
52. Kersey PJ, Allen JE, Allot A, Barba M, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Grabmueller C, et al. 2018. Ensembl Genomes 2018: An integrated omics infrastructure for non-vertebrate species. Nucleic Acids Res. 46:D802–D808. doi:10.1093/nar/gkx1011.
53. LaCava MEF, Aikens EO, Megna LC, Randolph G, Hubbard C, Buerkle CA. 2020. Accuracy of de novo assembly of DNA sequences from double-digest libraries varies substantially among software. Mol. Ecol. Resour. 20:360–370. doi:10.1111/1755-0998.13108.
54. Lal MM, Southgate PC, Jerry DR, Bosserelle C, Zenger KR. 2017. Swept away: Ocean currents and seascape features influence genetic structure across the 18,000 Km Indo-Pacific distribution of a marine invertebrate, the black-lip pearl oyster Pinctada margaritifera. BMC Genomics 18:1–21. doi:10.1186/s12864-016-3410-y.
55. Langa J, Estonba A, Conklin D. 2020. EXFI: Exon and splice graph prediction without a reference genome. Ecol. Evol. 10:8880–8893. doi:10.1002/ece3.6587.
56. Lelieveld SH, Veltman JA, Gilissen C. 2016. Novel bioinformatic developments for exome sequencing. Hum. Genet. 135:603–614. doi:10.1007/s00439-016-1658-6.
57. Li B, Fillmore N, Bai Y, Collins M, Thomson JA, Stewart R, Dewey CN. 2014. Evaluation of de novo transcriptome assemblies from RNA-Seq data. Genome Biol. 15:1–21. doi:10.1186/s13059-014-0553-5.
58. Liao X, Li M, Zou Y, Wu FX, Yi-Pan, Wang J. 2019. Current challenges and solutions of de novo assembly. Quant. Biol. 7:90–109. doi:10.1007/s40484-019-0166-9.
59. Lim HC, Braun MJ. 2016. High-throughput SNP genotyping of historical and modern samples of five bird species via sequence capture of ultraconserved elements. Mol. Ecol. Resour. 16:1204–1223. doi:10.1111/1755-0998.12568.
60. Linck EB, Hanna ZR, Sellas A, Dumbacher JP. 2017. Evaluating hybridization capture with RAD probes as a tool for museum genomics with historical bird specimens. Ecol. Evol. 7:4755–4767. doi:10.1002/ece3.3065.
61. Lindegren M, Holt BG, MacKenzie BR, Rahbek C. 2018. A global mismatch in the protection of multiple marine biodiversity components and ecosystem services. Sci. Rep. 8:1–8. doi:10.1038/s41598-018-22419-1.
62. Liu J, Li G, Chang Z, Yu T, Liu B, McMullen R, Chen P, Huang X. 2016. BinPacker: Packing-Based De Novo Transcriptome Assembly from RNA-seq Data. PLoS Comput. Biol. 12:1–15. doi:10.1371/journal.pcbi.1004772.
63. Lotterhos KE. 2019. The effect of neutral recombination variation on genome scans for selection. G3 Genes, Genomes, Genet. 9:1851–1867. doi:10.1534/g3.119.400088.
64. Lotterhos KE, Card DC, Schaal SM, Wang L, Collins C, Verity B. 2017. Composite measures of selection can improve the signal-to-noise ratio in genome scans. Methods Ecol. Evol. 8:717–727. doi:10.1111/2041-210X.12774.
65. Lotterhos KE, Whitlock MC. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol. Ecol. 24:1031–1046. doi:10.1111/mec.13100.
66. Lotterhos KE, Whitlock MC. 2014. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. Mol. Ecol. 23:2178–2192. doi:10.1111/mec.12725.
67. Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, Storfer A. 2017. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Mol. Ecol. Resour. 17:142–152. doi:10.1111/1755-0998.12635.
68. Maisano Delser P, Corrigan S, Hale M, Li C, Veuille M, Planes S, Naylor G, Mona S. 2016. Population genomics of C. melanopterus using target gene capture data: Demographic inferences and conservation perspectives. Sci. Rep. 6:1–12. doi:10.1038/srep33753.
69. Marko PB, Hart MW. 2011. The complex analytical landscape of gene flow inference. Trends Ecol. Evol. 26:448–456. doi:10.1016/j.tree.2011.05.007.
70. Matz M V. 2018. Fantastic Beasts and How To Sequence Them: Ecological Genomics for Obscure Model Organisms. Trends Genet. 34:121–132. doi:10.1016/j.tig.2017.11.002.
71. McKinney GJ, Larson WA, Seeb LW, Seeb JE. 2017. RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al. (2016). Mol. Ecol. Resour. 17:356–361. doi:10.1111/1755-0998.12649.
72. Micheletti SJ, Matala AR, Matala AP, Narum SR. 2018. Landscape features along migratory routes influence adaptive genomic variation in anadromous steelhead (Oncorhynchus mykiss). Mol. Ecol. 27:128–145. doi:10.1111/mec.14407.
73. Milbury CA, Lee JC, Cannone JJ, Gaffney PM, Gutell RR. 2010. Fragmentation of the large subunit ribosomal RNA gene in oyster mitochondrial genomes. BMC Genomics 11:1–17. doi:10.1186/1471-2164-11-485.
74. Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. 17:240–248. doi:10.1101/gr.5681207.
75. O’Leary SJ, Puritz JB, Willis SC, Hollenbeck CM, Portnoy DS. 2018. These aren’t the loci you’e looking for: Principles of effective SNP filtering for molecular ecologists. Mol. Ecol. 27:3193–3206. doi:10.1111/mec.14792.
76. O’Leary SJ, Puritz JB, Willis SC, Hollenbeck CM, Portnoy DS. 2018. These aren’t the loci you’e looking for: Principles of effective SNP filtering for molecular ecologists. Mol. Ecol. 27:3193–3206. doi:10.1111/mec.14792.
77. Patnaik BB, Wang TH, Kang SW, Hwang HJ, Park SY, Park EB, Chung JM, Song DK, Kim C, Kim S, et al. 2016. Sequencing, de novo assembly, and annotation of the transcriptome of the endangered freshwater pearl bivalve, Cristaria plicata, provides novel insights into functional genes and marker discovery. PLoS One 11:1–28. doi:10.1371/journal.pone.0148622.
78. Payne A, Holmes N, Clarke T, Munro R, Debebe B, Loose M. 2020. Nanopore adaptive sequencing for mixed samples, whole exome capture and targeted panels. bioRxiv:1–28.
79. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat. Protoc. 11:1650–1667. doi:10.1038/nprot.2016.095.
80. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7. doi:10.1371/journal.pone.0037135.
81. Portnoy DS, Puritz JB, Hollenbeck CM, Gelsleichter J, Chapman D, Gold JR. 2015. Selection and sex-biased dispersal in a coastal shark: The influence of philopatry on adaptive variation. Mol. Ecol. 24:5877–5885. doi:10.1111/mec.13441.
82. Puritz JB, Gold JR, Portnoy DS. 2016. Fine-scale partitioning of genomic variation among recruits in an exploited fishery: Causes and consequences. Sci. Rep. 6:1–6. doi:10.1038/srep36095.
83. Puritz JB, Keever CC, Addison JA, Barbosa SS, Byrne M, Hart MW, Grosberg RK, Toonen RJ. 2017. Life-history predicts past and present population connectivity in two sympatric sea stars. Ecol. Evol. 7:3916–3930. doi:10.1002/ece3.2938.
84. Puritz JB, Lotterhos KE. 2018. Expressed exome capture sequencing: A method for cost-effective exome sequencing for all organisms. Mol. Ecol. Resour. 18:1209–1222. doi:10.1111/1755-0998.12905.
85. Puritz JB, Matz M V., Toonen RJ, Weber JN, Bolnick DI, Bird CE. 2014. Demystifying the RAD fad. Mol. Ecol. 23:5937–5942. doi:10.1111/mec.12965.
86. Puritz JB, Toonen RJ. 2011. Coastal pollution limits pelagic larval dispersal. Nat. Commun. 2. doi:10.1038/ncomms1238.
87. Quattrini AM, Faircloth BC, Dueñas LF, Bridge TCL, Brugler MR, Calixto-Botía IF, DeLeo DM, Forêt S, Herrera S, Lee SMY, et al. 2018. Universal target-enrichment baits for anthozoan (Cnidaria) phylogenomics: New approaches to long-standing problems. Mol. Ecol. Resour. 18:281–295. doi:10.1111/1755-0998.12736.
88. Rellstab C, Dauphin B, Zoller S, Brodbeck S, Gugerli F. 2019. Using transcriptome sequencing and pooled exome capture to study local adaptation in the giga-genome of Pinus cembra. Mol. Ecol. Resour. 19:536–551. doi:10.1111/1755-0998.12986.
89. RI Coastal Resources Management Council. 2014. Shellfish management plan. RI Gen. Laws.Richmond JQ, Backlin AR, Galst-Cavalcante C, O’Brien JW, Fisher RN. 2018. Loss of dendritic connectivity in southern California’s urban riverscape facilitates decline of an endemic freshwater fish. Mol. Ecol. 27:369–386. doi:10.1111/mec.14445.
90. Russell PH, Johnson RL, Ananthan S, Harnke B, Carlson NE. 2018. A large-scale analysis of bioinformatics code on GitHub. PLoS One 13:1–35. doi:10.1371/journal.pone.0205898.
91. Schiebelhut LM, Puritz JB, Dawson MN. 2018. Decimation by sea star wasting disease and rapid genetic change in a keystone species, Pisaster ochraceus. Proc. Natl. Acad. Sci. U. S. A. 115:7069–7074. doi:10.1073/pnas.1800285115.
92. Schlötterer C, Tobler R, Kofler R, Nolte V. 2014. Sequencing pools of individuals-mining genome-wide polymorphism data without big funding. Nat. Rev. Genet. 15:749–763. doi:10.1038/nrg3803.
93. Schmid S, Genevest R, Gobet E, Suchan T, Sperisen C, Tinner W, Alvarez N. 2017. HyRAD-X, a versatile method combining exome capture and RAD sequencing to extract genomic information from ancient DNA. Methods Ecol. Evol. 8:1374–1388. doi:10.1111/2041-210X.12785.
94. Schmitz-Abe K, Li Q, Rosen SM, Nori N, Madden JA, Genetti CA, Wojcik MH, Ponnaluri S, Gubbels CS, Picker JD, et al. 2019. Unique bioinformatic approach and comprehensive reanalysis improve diagnostic yield of clinical exomes. Eur. J. Hum. Genet. 27:1398–1405. doi:10.1038/s41431-019-0401-x.
95. Schweizer RM, VonHoldt BM, Harrigan R, Knowles JC, Musiani M, Coltman D, Novembre J, Wayne RK. 2016. Genetic subdivision and candidate genes under selection in North American grey wolves. Mol. Ecol. 25:380–402. doi:10.1111/mec.13364.
96. Selkoe KA, D’Aloia CC, Crandall ED, Iacchei M, Liggins L, Puritz JB, Von Der Heyden S, Toonen RJ. 2016. A decade of seascape genetics: Contributions to basic and applied marine connectivity. Mar. Ecol. Prog. Ser. 554:1–19. doi:10.3354/meps11792.
97. Shafer ABA, Peart CR, Tusso S, Maayan I, Brelsford A, Wheat CW, Wolf JBW. 2017. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods Ecol. Evol. 8:907–917. doi:10.1111/2041-210X.12700.
98. Shih BB, Angus T, Barnett MMW, Chen S, Summers KKM, Klein K, Faulkner GJG, Saini HKH, Watson M, Dongen S Van, et al. 2018. Visualisation and analysis of RNA-Seq assembly graphs. bioRxiv:1–14. doi:10.1101/409573.
99. Silliman K. 2019. Population structure, genetic connectivity, and adaptation in the Olympia oyster (Ostrea lurida) along the west coast of North America. Evol. Appl. 12:923–939. doi:10.1111/eva.12766.
100. Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT. 2014. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Syst. Biol. 63:83–95. doi:10.1093/sysbio/syt061.
101. Smukowski CS, Noor MAF. 2011. Recombination rate variation in closely related species. Heredity (Edinb). 107:496–508. doi:10.1038/hdy.2011.44.Sohn J Il, Nam JW. 2018. The present and future of de novo whole-genome assembly. Brief. Bioinform. 19:23–40. doi:10.1093/bib/bbw096.
102. Suchan T, Pitteloud C, Gerasimova NS, Kostikova A, Schmid S, Arrigo N, Pajkovic M, Ronikier M, Alvarez N. 2016. Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. PLoS One 11:1–22. doi:10.1371/journal.pone.0151651.
103. Syring J V., Tennessen JA, Jennings TN, Wegrzyn J, Scelfo-Dalbey C, Cronn R. 2016. Targeted capture sequencing in whitebark pine reveals range-wide demographic and adaptive patterns despite challenges of a large, repetitive genome. Front. Plant Sci. 7:1–15. doi:10.3389/fpls.2016.00484.
104. Therkildsen NO, Palumbi SR. 2017. Practical low-coverage genome wide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. Mol. Ecol. Resour. 17:194–208. doi:10.1111/1755-0998.12593.
105. Tian S, Yan H, Klee EW, Kalmbach M, Slager SL. 2018. Comparative analysis of de novo assemblers for variation discovery in personal genomes. Brief. Bioinform. 19:893–904. doi:10.1093/bib/bbx037.
106. Verity R, Collins C, Card DC, Schaal SM, Wang L, Lotterhos KE. 2017. minotaur: A platform for the analysis and visualization of multivariate results from genome scans with R Shiny. Mol. Ecol. Resour. 17:33–43. doi:10.1111/1755-0998.12579.
107. Via S, Lande R. 2016. Genotype-Environment Interaction and the Evolution of Phenotypic Plasticity Author ( s ): Sara Via and Russell Lande Published by : Society for the Study of Evolution Stable URL : http://www.jstor.org/stable/2408649 Accessed : 01-03-2016 03 : 35 UTC Your . Evolution (N. Y). 39:505–522.
108. Wang J, Chen K, Ren Q, Zhang Y, Liu J, Wang G, Liu A, Li Y, Liu G, Luo J, et al. 2021. Systematic Comparison of the Performances of De Novo Genome Assemblers for Oxford Nanopore Technology Reads From Piroplasm. Front. Cell. Infect. Microbiol. 11:1–9. doi:10.3389/fcimb.2021.696669.
109. Wang S, Meyer E, Mckay JK, Matz M V. 2012. 2b-RAD: A simple and flexible method for genome-wide genotyping. Nat. Methods 9:808–810. doi:10.1038/nmeth.2023.
110. Wang S, Gribskov M. 2017. Comprehensive evaluation of de novo transcriptome assembly programs and their effects on differential gene expression analysis. Bioinformatics 33:327–333. doi:10.1093/bioinformatics/btw625.
111. Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva E V., Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol. Biol. Evol. 35:543–548. doi:10.1093/molbev/msx319.
112. Willis SC, Hollenbeck CM, Puritz JB, Portnoy DS. 2022. Genetic recruitment patterns are patchy and spatiotemporally unpredictable in a deep-water snapper (Lutjanus vivanus) sampled in fished and protected areas of western Puerto Rico. Conserv. Genet. doi:10.1007/s10592-021-01426-2.
113. Xuereb A, Benestan L, Normandeau É, Daigle RM, Curtis JMR, Bernatchez L, Fortin MJ. 2018. Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RADseq, in a highly dispersive marine invertebrate (Parastichopus californicus). Mol. Ecol. 27:2347–2364. doi:10.1111/mec.14589.
114. Xuereb A, D’Aloia CC, Andrello M, Bernatchez L, Fortin MJ. 2021. Incorporating putatively neutral and adaptive genomic data into marine conservation planning. Conserv. Biol. 35:909–920. doi:10.1111/cobi.13609.
115. Yahav T, Privman E. 2019. A comparative analysis of methods for de novo assembly of hymenopteran genomes using either haploid or diploid samples. Sci. Rep. 9:1–10. doi:10.1038/s41598-019-42795-6.