Why not do phylogeography on every chthamalid barnacle? The case of *Jehlius cirratus*

John Wares, University of Georgia; jpwares@uga.edu

06 November, 2014

Abstract

Here I evaluate the population genetic structure of the barnacle *Jehlius cirratus* across a broad portion of its geographic distribution using data from the mitochondrial cytochrome oxidase I (COI) gene region. Despite sampling diversity from close to 2500km of the linear range of this barnacle, there is little evidence for population structure. Hudson's Snn value across all sites is not significantly different from null expectation, and no other phylogeographic structure is evident. Unbalanced sample sizes and their effect on such population genetic and phylogeographic analyses are discussed, but the general conclusion of this study is that *J. cirratus* can be considered panmictic along the Chilean coast.

Introduction

There is an expanding interest in exploration of genetic diversity within and among regional populations of intertidal species along the coast of Chile (see Haye et al. 2014 for a recent synthesis). Such data are being used to explore the underlying causes of biogeographic transition (Zakas et al. 2009), to inform management and aquacultural concerns (Núñez-Acuña et al. 2012; Haye and Munoz-Herrera 2013), and better understand how the dynamics of a coastal ocean influence local diversity.

As an example, it is known that the Humboldt current creates a shift in oceanographic conditions at ~ 30-32°S latitude off the coast of Chile (Hormazabal et al. 2004; Rutllant et al. 2004). This shift includes a change in the strength and seasonality of upwelling, which influences population structure and connectivity in many coastal species (Navarrete et al. 2005; Haye et al. 2014). An example among coastal barnacles, which all have high propensity for larval dispersal, is the high intertidal chthamalid barnacle *Notochthamalus scabrosus*; Zakas et al. (2009) first showed that there is a significant transition in mitochondrial diversity in this species associated with the biogeographic transition in central Chile. Such transitions may have independent origins but are readily maintained by strong coastal oceanographic structure, and can have significant implications for coastal biodiversity (Wares et al. 2001; Haye et al. 2014).

However, the same study (Zakas et al. 2009) did not find such evidence of structure in *Jehlius cirratus* (see Figure 1). Though there was considerable genetic diversity noted in *Jehlius*, the regional samples in that study only covered ~800km of the Chilean coast; analysis of those data indicated only a marginal pattern of structure over that range, perhaps evidence of broad-scale isolation by distance. Here, we add sequences from subsequent samples within this range, as well as samples from the edges of the distribution (Häussermann and Försterra 2009), now spanning close to 2500km of the species' range, to explore more fully the potential for spatial diversity in intertidal taxa from this coast. As chthamalid barnacles have a propensity to harbor cryptic genetic diversity (Dando and Southward 1981; Zardus and Hadfield 2005; Tsang et al. 2008; Wares et al. 2009; Meyers et al. 2013), we specifically look for any phylogeographic structure that may add to our understanding of coastal biodiversity in Chile.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.596v1 | CC-BY 4.0 Open Access | rec: 7 Nov 2014, publ: 7 Nov 2014

Figure 1. Magnified image of *Jehlius cirratus* (larger, lower left; actual size ~1cm diameter) next to confamilial *Notochthamalus scabrosus* in intertidal (photo by JPW, location near Cucao on island of Chiloé). Opercular margins between terga and scuta are primary field identifiers.

Methods

Sequence data from the mitochondrial cytochrome oxidase I locus (COI) for *J. cirratus* from Zakas et al. (2009) were accessed from GenBank. Additional data, obtained from sequencing efforts across the range of *J. cirratus*, were added to the data matrix. Sample locations are indicated in Figure 2.

We isolated DNA from all tissues using a standard PureGene protocol and sequenced the mitochondrial COI gene region, as in Zakas et al. (2009). In most cases, sequence data were obtained from assembly of forward and reverse reads of the same PCR amplicon; nucleotides with quality scores <20 were called as ambiguities.

These data were aligned using the MUSCLE algorithm (default parameters) implemented in the R package **ape** (Paradis et al. 2004). A neighbor-joining analysis (with 1000 bootstrap data replicates) was performed to identify any clades supported in 70% or more of replicates, and Rosenberg's (2007) monophyly test (applied to the output neighbor-joining tree in Geneious R7) was implemented for any visibly apparent clades.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.596v1 | CC-BY 4.0 Open Access | rec: 7 Nov 2014, publ: 7 Nov 2014

Data were transformed for class GENOME in R; using the **popgenome** package (Pfeifer et al. 2014), sequences were sorted to geographic location and Hudson's Snn (Hudson 2000) statistic was calculated (with 1000 randomizations of the data matrix to assess statistical significance) across the geographic range of *J. cirratus* with local collections grouped into Northern (north of 25°S), North Central (25-32°S), South Central (32-35°S), and Southern (south of 35°); groupings reflect the consideration of the ~32°S biogeographic break as well as evaluating the more extreme locations separately (see Figure 2).

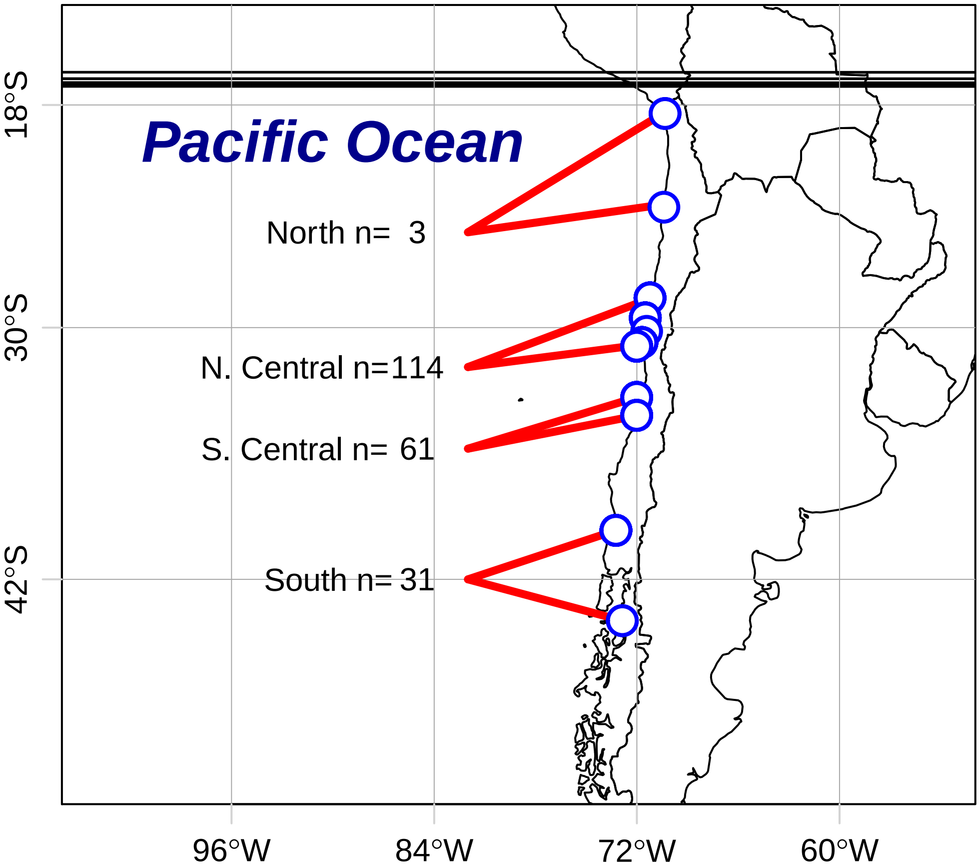

Figure 2. Sample locations (blue dots) and regional sample sizes for *Jehlius cirratus* sequence diversity at the mitochondrial COI locus. Horizontal lines near top are plot error that is difficult to fix in R(oce), please ignore.

Results

A total of 209 sequences of the COI gene region are represented in the data matrix, with 55 new sequences to report (GenBank accession numbers pending, **Bankit 1772410**, submitted via Geneious). The sequence data for *J. cirratus* were 710 nt in length, with 56 variable positions. Alignment of these data was unambiguous and suggested no pattern of pseudogene or NUMT behavior. Resultant analysis indicated similar coalescent divergence as in Zakas et al. (2009), with no clades strongly supported either with bootstrap support or Rosenberg's test for monophyly (probability of observed clades being randomly distinct 0.56), and no apparent geographic frequency variation across subclades of the tree; thus, the output trees are not shown but are available from the author.

Nucleotide diversity for all data is 0.0017 (π per nucleotide). Across all sequences, Tajima's D statistic is -2.5964, a typical value for a population of cirripedes (see Ewers and Wares 2012). Spatial variation in these data suggest little population genetic structure, with Snn across all locations of 0.415 (p=0.086 with 95% distribution of permuted Snn values from 0.386 and 0.4187).

Discussion

Using sequence data from individuals collected across a range of ~2500km, there is no evident population genetic structure in the chthamalid barnacle *J. cirratus*. Although the analysis shown here is relatively simple - using Hudson's (2000) Snn statistic to evaluate whether panmixia can be rejected for mitochondrial sequence data - it is the most statistically powerful analysis for situations in which sample sizes may be small and diversity at the analyzed locus is high (Hudson 2000). With little evidence for rejecting panmixia based on Snn, and the results of Zakas et al. (2009) already showing no sign of population structure (including isolation by distance) across a portion of this geographic range, there seems to be little reason to proceed with more complicated consideration of the data.

One result that is of interest is the upward bias of Snn when unbalanced sample sizes are considered across sample regions. Both because of the density of available samples - including the effort from a previous study (Zakas et al. 2009) - and the spatial pooling of proximal sites to address very small samples in the northernmost locations, there are huge differences in sample sizes among spatial collections. Ordinarily, the null (no population structure) expectation for Hudson's Snn is thought to be $1/n$, where n is the number of locations evaluated; for example, for two locations the null expectation would be 0.5, and for four locations the null expectation would be 0.25. In a multi-species analysis of mitochondrial diversity along the Atlantic coast of North America, Díaz-Ferguson et al. (2009) also grouped locations into four regions for analysis with Snn, and sample sizes for each location in each species were close to 20. In this case, the "null" results were as expected, close to 0.25, and the "significant" results exceeded that value and the subsequent permutational test distribution.

However, it is easy to consider the effect of highly unbalanced sample sizes on the Snn statistic. As the statistic is a weighted average of a sum over all tips (sequences), in which a 'nearest-neighbor' sequence collected in the same spatial population adds a 1, and a 'nearest-neighbor' sequence from a distinct location adds a 0, a very large sample from one location in an unstructured system like *J. cirratus* will be likely to have closely-related sequences from within that sampled location, an effect that will inflate the Snn statistic (the small populations will be very unlikely to have a 'nearest-neighbor' within the population, and will reduce the statistic, but only proportional to the sample from that area). At the extremes of unbalanced sampling, the statistic is necessarily biased towards 1 regardless of the actual population structure (see Figure 3). Certainly, the effect of unbalanced samples from evolutionarily distinct populations on population genetic statistics can be significant as well (Ewers and Wares 2012). The context for the Snn statistic is not a simple algebraic expectation, however, but a nonparametric resampling of data to generate a null distribution for the statistic for a given set of population data (Hudson 2000). It is worth considering the limits on the power of this statistic when there are unbalanced sample sizes as this will clearly compress the possible range of the statistic. Randomized data subsets of equivalent sizes may be necessary in some cases (see Wares and Cunningham 2005 for example).

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.596v1 | CC-BY 4.0 Open Access | rec: 7 Nov 2014, publ: 7 Nov 2014

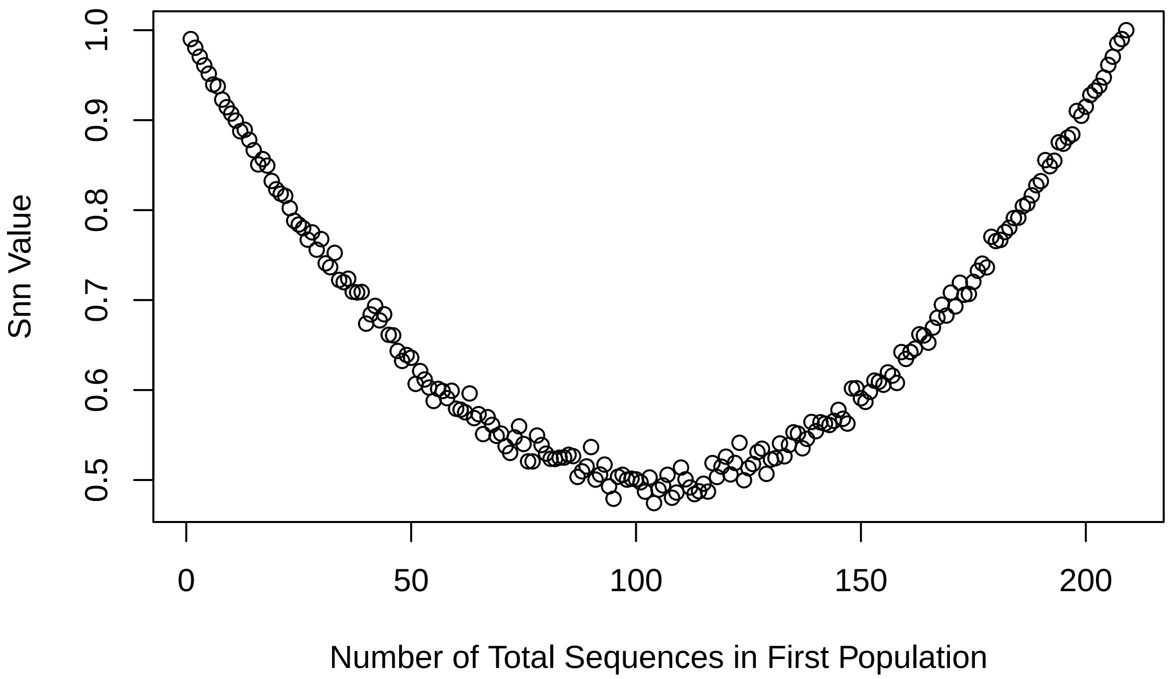

Figure 3. Distribution of Snn for sequence data from *J. cirratus* with random samples for 2 populations ranging from [1,(n-1)] to [(n-1),1] to illustrate the upwards bias of asymmetric sample sizes on the distribution of Snn.

A remaining question is why particular environmental boundaries appear to influence the connectivity among locations for some species but not others. As with the comparison between *N. scabrosus*, which exhibits strong population structure along the Chilean coast (Zakas et al. 2009; Laughlin et al. 2012), and *J. cirratus*, which does not, we can also consider the structure of *Balanus glandula* (Sotka et al. 2004) (not a chthamalid barnacle, however) against the lack of structure in the codistributed *Chthamalus dalli* (Wares and Castañeda 2005), the structure in *C. montagui* across a broad geographic range (Dando and Southward 1981) and lack of structure across a similar range in *C. stellatus* (O'Riordan et al. 2010), along with other examples (Zardus and Hadfield 2005; Tsang et al. 2008; Wares et al. 2009; Meyers et al. 2013). Are the differences in structure based on larval behavior or developmental time (Burton and Feldman 1982)? Are the differences in population structure associated with distinct settlement, ecology, or history (Marko 2004)? A systematic analysis of the chthamalid barnacles is warranted to better understand how to predict the analytical outcome of the overall historical ecology of species in this group.

Acknowledgments

The author would like to thank Sergio Navarrete, John Binford, Christina Zakas, Leah Besch, Jenna Shinen, Arica colleague, Daniel Saucedo, and Christine Ewers-Saucedo for assistance in collecting specimens and sequence data. I also thank colleagues at ROpenSci and RStudio for assisting in fully-repeatable RMarkdown manuscript, with extensive technical help by Katie Bockrath and Christine Ewers-Saucedo on the writing and analysis. Funding for the project came from the National Science Foundation Biological Oceanography panel (NSF-OCE-1029526) to JPW. Data and .Rmd code are provided as supplemental information.

Literature Cited

Burton, R. S., and M. W. Feldman. 1982. Population genetics of coastal and estuarine invertebrates: does larval behavior influence population structure?

Dando, P. R., and A. J. Southward. 1981. Existence of Atlantic and Mediterranean forms of Chthamalus montagui (Crustacea, Cirripedia) in the Western Mediterranean. Marine Biology Letters 2:239–248.

Díaz-Ferguson, E., J. D. Robinson, B. R. Silliman, and J. P. Wares. 2009. Comparative phylogeography of east coast American salt marsh communities. Estuaries & Coasts DOI 10.1007/s12237-009-9220-6.

Ewers, C., and J. P. Wares. 2012. Examining an outlier: Molecular diversity in the Cirripedia. Integrative and Comparative Biology 1–8.

Haye, P. A., and N. C. Munoz-Herrera. 2013. Isolation with differentiation followed by expansion with admixture in the tunicate Pyura chilensis. BMC Evol Biol 13:252.

Haye, P. A., N. I. Segovia, N. C. Munoz-Herrera, F. E. Galvez, A. Martinez, A. Meynard, M. C. Pardo-Gandarillas, E. Poulin, and S. Faugeron. 2014. Phylogeographic structure in benthic marine invertebrates of the southeast Pacific coast of Chile with differing dispersal potential. PLoS One 9:e88613.

Häussermann, V., and G. Försterra. 2009. Marine Benthic Fauna of Chilean Patagonia. 1st ed. Nature in Focus.

Hormazabal, S., G. Shaffer, and O. Leth. 2004. Coastal transition zone off Chile. Journal of Geophysical Research-Oceans 109.

Hudson, R. R. 2000. A new statistic for detecting genetic differentiation. Genetics 155:2011–2014.

Laughlin, K., C. Ewers, and J. P. Wares. 2012. Mitochondrial lineages in Notochthamalus scabrosus as indicators of coastal recruitment and interactions. Ecology and Evolution 2:1584–1592.

Marko, P. B. 2004. 'What's larvae got to do with it?' disparate patterns of post-glacial population structure in two benthic marine gastropods with identical dispersal potential. Molecular Ecology 13:597–611.

Meyers, M., M. S. Pankey, and J. P. Wares. 2013. Genealogical approaches to the temporal origins of the Central American Gap: Speciation and divergence in Pacific Chthamalus. Revista Biologia Tropical 61:75–88.

Navarrete, S. A., E. A. Wieters, B. R. Broitman, and J. C. Castilla. 2005. Scales of benthic-pelagic coupling and the intensity of species interactions: From recruitment limitation to top-down control. PNAS 102:18046–18051.

Núñez-Acuña, G., F. J. Tapia, P. A. Haye, and C. Gallardo-Escárate. 2012. Gene expression analysis in Mytilus chilensis populations reveals local patterns associated with ocean environmental conditions. J. exp. mar. Biol. Ecol. 420-421:56–64.

O'Riordan, R. M., A. M. Power, and A. A. Myers. 2010. Factors, at different scales, affecting the distribution of species of the genus Chthamalus Ranzani (Cirripedia, Balanomorpha, Chthamaloidea). Journal of Experimental Marine Biology and Ecology 392:46–64.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.

Pfeifer, B., U. Wittelsbuerger, S. E. Ramos-Onsins, and M. J. Lercher. 2014. PopGenome: An efficient swiss army knife for population genomic analyses in r. Molecular Biology and Evolution 31:1929–1936.

Rosenberg, N. A. 2007. Statistical tests for taxonomic distinctiveness from observations of monophyly. Evolution 61:317–323.

Rutllant, J. A., B. Rosenbluth, and S. Hormazabal. 2004. Intraseasonal variability of wind-forced coastal upwelling off central Chile (30 degrees S). Continental Shelf Research 24:789–804.

Sotka, E. E., J. P. Wares, J. A. Barth, R. K. Grosberg, and S. R. Palumbi. 2004. Strong genetic clines and geographical variation in gene flow in the rocky intertidal barnacle Balanus glandula. Molecular Ecology 13:2143–2156.

Tsang, L. M., B. K. K. Chan, T. H. Wu, W. C. Ng, T. Chatterjee, G. A. Williams, and K. H. Chu. 2008. Population differentiation in the barnacle Chthamalus malayensis: postglacial colonization and recent connectivity across the Pacific and Indian oceans. Mar. Ecol. Prog. Ser. 364:107–118.

Wares, J. P., and A. E. Castañeda. 2005. Geographic range in Chthamalus along the west coast of North America. J. Mar. biol. Ass. U.K. 85:327–331.

Wares, J. P., and C. W. Cunningham. 2005. Diversification predates glaciation in Balanus glandula. The Biological Bulletin 208:60–68.

Wares, J. P., S. D. Gaines, and C. W. Cunningham. 2001. A comparative study of asymmetric migration events across a marine biogeographic boundary. Evolution 55:295–306.

Wares, J. P., M. S. Pankey, F. B. Pitombo, L. E. Gómez Daglio, and Y. Achituv. 2009. A "shallow phylogeny" of shallow barnacles (Chthamalus). PLoS One 4:e5567.

Zakas, C., J. Binford, S. A. Navarrete, and J. P. Wares. 2009. Restricted gene flow in Chilean barnacles reflects an oceanographic and biogeographic transition zone. Marine Ecology Progress Series 394:165–177.

Zardus, J. D., and M. G. Hadfield. 2005. Multiple origins and incursions of the Atlantic barnacle Chthamalus proteus in the Pacific. Molecular Ecology 14:3719–3733.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.596v1 | CC-BY 4.0 Open Access | rec: 7 Nov 2014, publ: 7 Nov 2014