Phylogenetic analysis of Antarctic notothenioids illuminates the utility of RADseq for resolving Cenozoic adaptive radiations

Thomas J. Near\textsuperscript{a,b,*}, Daniel J. MacGuigan\textsuperscript{a}, Elyse Parker\textsuperscript{a}, Carl D. Struthers\textsuperscript{a}, Christopher D. Jones\textsuperscript{d}, Alex Dornburg\textsuperscript{e}

\textsuperscript{a} Department of Ecology & Evolutionary Biology, Yale University, P.O. Box 208106, New Haven, CT 06520, USA
\textsuperscript{b} Peabody Museum of Natural History, Yale University, New Haven, CT 06520, USA
\textsuperscript{c} Museum of New Zealand Te Papa Tongarewa, Wellington, New Zealand
\textsuperscript{d} Antarctic Ecosystem Research Division, NOAA Southwest Fisheries Science Center, La Jolla, CA 92037, USA
\textsuperscript{e} North Carolina Museum of Natural Sciences, Raleigh, NC 27601, USA

\textbf{A R T I C L E   I N F O}

Keywords: Notothenioidei, Phylogenomics, Phylogenetic informativeness, Signal noise, Icefishes

\textbf{A B S T R A C T}

Notothenioids are a clade of \textasciitilde{} 120 species of marine fishes distributed in extreme southern hemisphere temperate near-shore habitats and in the Southern Ocean surrounding Antarctica. Over the past 25 years, molecular and morphological approaches have redefined hypotheses of relationships among notothenioid lineages as well as their relationships among major lineages of percomorph teleosts. These phylogenies provide a basis for investigation of mechanisms of evolutionary diversification within the clade and have enhanced our understanding of the notothenioid adaptive radiation. Despite extensive efforts, there remain several questions concerning the phylogeny of notothenioids. In this study, we deploy DNA sequences of \textasciitilde{} 100,000 loci obtained using RADseq to investigate the phylogenetic relationships of notothenioids and to assess the utility of RADseq loci for lineages that exhibit divergence times ranging from the Paleogene to the Quaternary. The notothenioid phylogenies inferred from the RADseq loci provide unparalleled resolution and node support for several long-standing problems including, (1) relationships among species of Trematomus, (2) resolution of Indonotothenia cyanobrancha as the sister lineage of Trematomus, (3) the deep paraphyly of Nototheniidae, (4) the paraphyly of Lepidonotothen s.l., (5) paraphyly of Artedidraco, and 6) the monophyly of the Bathydraconidae. Assessment of site rates demonstrates that RADseq loci are similar to mtDNA protein coding genes and exhibit peak phylogenetic informativeness at the time interval during which the major Antarctic notothenioid lineages originated and diversified. In addition to providing a well-resolved phylogenetic hypothesis for notothenioids, our analyses quantify the predicted utility of RADseq loci for Cenozoic phylogenetic inferences.

1. Introduction

Antarctic notothenioids (Cryonotothenioidea) are one of the most well studied groups of marine fishes and one of few examples of marine teleost adaptive radiation (Eastman, 1993; Clarke and Johnston, 1996; Ingram and Mahler, 2011; Matschiner et al., 2011; Near et al., 2012; Colombo et al., 2015; Dornburg et al., 2017a). In addition to exhibiting an interesting evolutionary history, notothenioids are vital to Antarctic marine ecosystems, as they comprise a substantial component of the biomass, abundance, and species diversity of near shore fishes (Eastman, 1993, 2005). Correspondingly, these species are critical in linking lower level consumers and higher level predators in the Antarctic marine food web (La Mesa et al., 2004), including species of high economic importance for international fisheries interests (Constable et al., 2000; Abrams, 2013). Despite a long history of research, a well-resolved species-level phylogeny of notothenioids is not available and several key phylogenetic questions remain unanswered.

Over the past quarter-century efforts to investigate the phylogenetics of notothenioids have resulted in important discoveries that dramatically altered subsequent taxonomic classifications. For example, early morphological and molecular inferred phylogenies resolve Bovichtidae, historically delimited to include Bovichtus, Cotoperaea, and Pseudaphritis (Eastman, 1993; Nelson, 1994), as paraphyletic and Elephinops maclouvim as the sister lineage of Cryonotothenioidea instead of being closely related to the nototheniid lineage Dissostichus (Balushkin, 1992; Lecointre et al., 1997; Bargelloni and Lecointre, 1998). The most

\textsuperscript{*} Corresponding author at: Department of Ecology & Evolutionary Biology, Yale University, P.O. Box 208106, New Haven, CT 06520, USA.
\textit{E-mail address:} thomas.near@yale.edu (T.J. Near).
recent phylogenetic analyses of notothenioids use DNA sequences sampled from multiple mitochondrial and nuclear genes with a taxon sampling that includes most of the recognized species in the clade (Near et al., 2012; Dornburg et al., 2017a). While these studies provide important insights into the relationships of notothenioids and serve as the basis for comparative analyses investigating the history and mechanisms of notothenioid diversification (Near et al., 2012; Dornburg et al., 2017a), there are several unresolved issues in the phylogenetics of notothenioids: the lack of phylogenetic resolution among the ~14 species of the rapidly diversifying Trematomus (Kuhn and Near, 2009; Janko et al., 2011; Lautredou et al., 2012; Near et al., 2012); the resolution of the neutrally buoyant Pleurogramma antarctica among the major lineages of Cryonotothenioidea (Near and Cheng, 2008; Dettaï et al., 2012; Near et al., 2012); and the lack of support for monophyly of the Antarctic Dragonfishes (Bathydraconidae) in molecular analyses (Bargelloni et al., 2000; Derome et al., 2002; Dettaï et al., 2012; Near et al., 2012). These remaining challenges to the resolution of notothenioid phylogeny inhibit the investigation of important questions, such as the origin of neutral buoyancy (Near et al., 2007, 2012), species relationships within rapidly diversifying lineages (e.g., Trematomus & Arctedirinae; Lecointre et al., 2011; Lautredou et al., 2012), and patterns of hemoglobin evolution in Bathydraconidae that led to the loss of this protein in the Crocodile Icefishes (Channichthyidae) (Bargelloni et al., 1998; Near et al., 2006; Lau et al., 2012).

Next-generation sequencing through reduced representation methods such as restriction site associated DNA sequencing (RADseq) hold the promise of resolving species-level relationships in notothenioids. By sequencing DNA flanking restriction sites, RADseq captures thousands of single nucleotide polymorphisms (SNPs) across any target genome and have been used to resolve difficult phylogenetic problems in lineages spanning beetles (Craud et al., 2014), plants (Massatti et al., 2016; Wang et al., 2017), corals (Herrera and Shank, 2016), and Lake Victoria cichlids (Wagner et al., 2013). Empirical assessments of RADseq data suggest cost-effective phylogenetic utility across temporal scales spanning recent divergences to those dating back tens of millions of years (Eaton et al., 2017). However, these assessments are based on post-hoc measure of topological support such as bootstrap values that can be positively misled by noise in the dataset, or phylogenetic informativeness profiles (Townsend, 2007), which make no explicit prediction of phylogenetic noise or tree topology (Townsend, 2007; Collins and Hrbek, 2018).

It is important to consider that the length of a given internode can fundamentally alter expectations of phylogenetic utility (Whitfield and Lockhart, 2007; Steel and Leuenberger, 2017; Dornburg et al., 2018). A relationship exists between the timescale of a phylogenetic question (T), the time between branching events (t0), and the rate of character evolution, which allows direct quantification of the predictive utility of a given marker (Townsend et al., 2012; Su et al., 2014). A rapidly evolving marker will provide phylogenetic information for deep divergences when waiting times between branching events are long, while the same marker has a higher probability of being positively misleading at a similar timescale when waiting times between branching events are reduced (Townsend et al., 2012). Although previous research has demonstrated that RADseq successfully resolves phylogenetic relationships dating ~50 million years (Eaton et al., 2017; Collins and Hrbek, 2018), no empirical analyses have analyzed the conditions under which RADseq remains cost-effective versus potentially misleading for phylogenetic inference.

Here we assemble a RADseq dataset for the majority of living notothenioid species, providing the first detailed phylogenomic investigations of this radiation. To evaluate the impact of internode length on the predictive utility of sequenced RADseq loci, we assess patterns of phylogenetic information content across tens of thousands of RADseq loci using a combination of phylogenetic informativeness (PI) that complement those presented in Collins and Hrbek (2018), and approaches to quantifying phylogenetic signal and noise (Townsend, 2007; Townsend et al., 2012). Our results provide a strongly supported phylogenetic hypothesis of notothenioid species-level relationships, the ability to reject several previous phylogenetic hypotheses and refine our perspective on the predicted utility and limits of RADseq data.

2. Methods

2.1. Taxon sampling, sequencing, and data preparation

The taxon sampling included 80 notothenioid species (Supplementary Table 1), which encompasses all of the recognized taxonomic families and is very similar to previous phylogenetic studies using Sanger sequenced legacy markers (Near et al., 2012). For example, the species included in Near et al. (2012) that are not included in this study are Gobionotothen acuta, G. marionensis, Paranothenia angustata, Bathyraco scotiae, Neopagetopsis ionah, and Cryodracoantarcticus. Species included in this study that are not in Near et al. (2012) are Nototheniops cf. nudifrons, Pogonophryne fusca, and Bathyracojoanna. We are confident that the minor differences in taxon sampling do not have a strong effect on the phylogenetic inferences.

DNA was isolated from tissue samples using the Qiagen DNeasy tissue extraction protocols (DNeasy, Qiagen, Valencia, CA). Extractions were gel-quantified in agarose using New England Biolabs 100 bp ladder (NEB, Ipswich, MA) to ensure successful extractions and DNA concentrations were determined using a Qubit v. 3.0 fluorometer (ThermoFisher Scientific, Philadelphia, PA). All DNA extractions were standardized to contain between 17 and 23 ng DNA/µL.

The RADseq protocol was not optimized for notothenioids, but rather is a single enzyme protocol that has been used in lineages of flowering plants (Eaton et al., 2017). Floragenex Inc. (Portland, OR) prepared the RADseq libraries using the SbfI restriction enzyme, a six base pair cutter (5′-CCCTGA-3′) and sample-specific barcodes. The samples were combined into two 95-sample multiplexed libraries. Floragenex created two replicates of each library in order to minimize the influence of PCR duplication bias and other technical errors. We sequenced each library twice on an Illumina HiSeq 2000 using single-end 100 base pair sequencing at the University of Oregon GC3f facility (https://gc3f.uoregon.edu/). We used pyRAD v.3.0.61 to assemble and align the RADseq datasets (Eaton, 2014). Individual reads that contained more than four sites with Phred scores < 20 were excluded. Reads were clustered using vsearch and an 88% similarity threshold, allowing for approximately 11 base differences between reads within a cluster. Analyses did not include clusters with a sequencing depth of less than six reads. We also discarded consensus sequences for each cluster if they contained more than five heterozygous or ambiguous bases. Consensus sequences were clustered as homologous loci across samples with an 88% similarity threshold, excluding loci that were shared by fewer than four of the sampled specimens. Accession numbers for the raw sequence reads deposited in the NCBI BioSample database are given in Supplementary Table 1.

2.2. Phylogenomic inference and testing alternative phylogenetic hypotheses

The RADseq alignment was analyzed using IQ-TREE to determine the optimal molecular evolutionary model and infer a maximum likelihood phylogeny of notothenioids (Nguyen et al., 2015; Kalyaanamoorthy et al., 2017). Node support was assessed using an ultrafast bootstrap analysis with 1000 replicates (Hoang et al., 2018). Based on previous analyses (e.g., Near et al., 2012), we rooted our phylogenetic inferences using the species of Bovichtidae sampled in this study.

In order to accommodate incomplete lineage sorting, we used the species tree approach tetrad v.0.7.19, an implementation of SVDQuartets in the software package iPyrad (Chifman and Kubatko, 2014; http://github.com/dereean/tetrad). We ran tetrad using one randomly selected SNP per locus, for 102,232 SNPs. We inferred all
2,024,785 possible quartets for the 85 sampled specimens representing 80 species. **Tetrad** uses the algorithm implemented by QMCC to join the individual quartet trees into a supertree (Avni et al., 2015). We constructed a 50% majority-rule consensus tree from 100 nonparametric bootstrap replicates.

Alternative phylogenetic hypotheses were compared to the maximum likelihood tree inferred from the IQ-TREE analysis using the approximately unbiased (AU) test based on the resampling of estimated log-likelihoods (RELL) method (Kishino et al., 1990; Shimodaira, 2002). The alternative phylogenetic hypotheses tested include (1) the monophyly of Nototheniidae as delimited in standard references for fish and notothenioid taxonomy (DeWitt et al., 1990; Eastman and Eakin, 2000; Nelson et al., 2016: 465–466), (2) the monophyly of Pleuranagrammatinae, containing *Pleurogramma antarctica*, *Aethotaxis mitopteryx*, *Dissostichus eleginoides*, and *D. mawsoni* (Balushkin, 2000; Near et al., 2007; Near and Cheng, 2008), (3) the monophyly of *Lepidonot then* s.l. (DeWitt et al., 1990), (4) the monophyly of *Arctedidraco* (Lecointre et al., 2011), (5) the monophyly of *Indonotothenia cyanobrancha* and sampled species of Notothenia (DeWitt et al., 1990; Balushkin, 2000), (6) *Pogotethina borchevichini* as not phylogenetically nested in *Trematomus*, (7) *Trematomus nevensis* and *T. borchevichini* are sister taxa and all other species of *Trematomus* form a monophyletic group (Balushkin, 2000: Fig. 17), which Balushkin classified in *Pseu dotrematomus* (Balushkin, 1982), and (8) the phylogeny of *Trematomus* presented in Lautredou et al. (2012: Fig. 2). The trees with the highest likelihoods consistent with the alternative phylogenies were estimated using the constraint tree search option in IQ-TREE (Nguyen et al., 2015).

### 2.3. Quantifying predicted phylogenetic utility

We compared RADseq loci to a dataset of legacy markers used in previous investigations of notothenioid relationships with very similar taxon sampling (Near et al., 2012; Dornburg et al., 2017a) that consisted of two mitochondrial genes (16S and ND2), one intron (57), and five exons (*Rag1*, *Tbr1*, *S3EPK3*, *gylt*, and *zic1*; DOI: https://doi.org/10.5281/zenodo.801836). We used the program HyPhy (Pond et al., 2005) in the PhyDef web interface (López-Giraldez and Townsend, 2011) to quantify site-specific rates of substitution for each legacy marker as well as the newly generated RADseq data using a publicly available notothenioid chronogram (Near et al., 2012; Dornburg et al., 2017a) downloaded from zenodo (DOI: https://doi.org/10.5281/zenodo.801836) that was pruned to mirror the taxon sampling of our alignment. The R package vioplot was used to generate violin plots of rate distributions (http://wosuppenkiste.wiso.uni-goettingen.de/~dadar). By combining a rotated kernel density plot with a box plot of data quantiles, violin plots allow for simultaneous visualization of both the quartiles and the underlying probability distribution of the site rates (Hintze and Nelson, 1998). Using the equations presented in Townsend (2007), phylogenetic informativeness (PI) was quantified for each locus in the R package PhyInformR (Dornburg et al., 2016). Visual detection of declines in PI over time have been considered a signature of homoplasy (Townsend and Leuenberger, 2011) and linked to phylogenetic estimation error (Dornburg et al., 2017c). Given that a RADseq dataset consists of thousands of loci, PI profiles were visualized using hexagon binning to assess overall trends in PI between loci using the hexbinplot function package hexbin (http://github.com/edzer/hexbin). The hexbin plots were compared with PI profiles of the sampled legacy markers and mapped to the 95% highest probability density interval of notothenioid divergence times estimated from previous relaxed molecular clock analyses (Near et al., 2012; Dornburg et al., 2017a).

Although the shape of the PI profiles provides an indication of overall trends of phylogenetic information content, these visualizations make no explicit quantification of how convergence in character state not reflecting evolutionary history (noise) will impact topological resolution (Townsend and Leuenberger, 2011; Dornburg et al., 2017c). Townsend et al. (2012) proposed theory that allows the utility of a locus to be quantified by comparing the predicted phylogenetic signal supporting a correct resolution (R) versus the amount of phylogenetic noise supporting resolution based on homoplasy (H) of a hypothetical phylogenetic quartet. We quantified the difference between R and H to explicitly quantify and visualize trends of the predicted utility of RADseq data for phylogenetic resolution. Specifically, we simultaneously assessed the predicted impact of alignment length and temporal depth on RADseq based inferences of short internodes. We quantified R and H for resolving a short (0.25 million year) internode (t5 in Townsend et al., 2012) using increasing quartet depths of 5 or 10 million years beginning 5 million years ago (7 in Townsend et al., 2012). For each hypothetical quartet, we increasingly added 500 bp of variable sites from the concatenated RADseq alignment to determine changes in R and H as a function of alignment length. Given that RADseq alignments are often comprised of over one million variable sites, this approach allows us to assess at what point we would expect the contribution of R to mitigate any potential impacts of H. All quantifications were conducted in the R package PhyInformR (Dornburg et al., 2016) and results were visualized as horizon plots in the LatticeExtra Package (Sarkar, 2008).

### 3. Results

#### 3.1. RADseq data

We collected RAD data for 80 notothenioid species, representing all major lineages (Supplementary Table 1). After initial quality filtering, we retained an average of $1.7 \times 10^6$ reads per sample, subsequently reduced to an average of 52,000 clusters per sample. Each of these clusters has a minimum sequencing depth of 6×, with an average depth of 23× across all sampled specimens. After additional filtering to remove clusters with excess heterozygosity, we retained an average of 51,000 consensus sequences per sample. The average estimated heterozygosity in these clusters is $3.6 \times 10^{-3}$, with an average estimated base calling error rate of 6.0 × 10^{-4}. After clustering consensus sequences into homologous loci, excluding loci shared by fewer than four taxa, and filtering identified paralogs, the final dataset contains 104,709 loci. The average number of loci per specimen is 25,881 (SD = 12110). The proportion of missing data in the final concatenated alignment of all loci is 76.4%.

#### 3.2. Phylogenomic inference

Maximum likelihood IQ-TREE (MLiq) and SVDquartets analyses of the RADseq dataset result in phylogenies that are highly congruent with one another (Figs. 1 and 2) and very similar to previous inferences using Sanger sequenced legacy markers (Near et al., 2012; Dornburg et al., 2017a). The phylogenies were rooted with the three sampled species of Bovichtidae and relationships among the non-Antarctic lineages *Pseudaphritis urvilli* and *Eleginos maclovinus* are identical to previous analyses (Near et al., 2012, 2015). These lineages are pruned out of the trees shown in Figs. 1 and 2 to allow focus on the relationships among the Cryonotothenioidea. The MLiq and SVDquartets phylogenies differ at five nodes: the resolution of *Pleurogramma antarctica*, and in four additional apical relationships that are not strongly supported in either analysis (Figs. 1 and 2).

The MLiq phylogeny resolves two major cryonotothenioid clades, (1) a clade containing *Trematomus*, *Indonotothenia cyanobrancha*, *Dissostichus*, *Aethotaxis mitopteryx*, *Pleurogramma antarctica*, *Lepidonotothen squamifrons*, *Nototheniops*, and *Patagonotothen* and (2) a clade containing *Gobionotothen*, *Notothenia*, *Harpagifer*, *Artedidraconidae*, *Bathydraconidae*, and *Channichthyidae*. The Pleuragrammatinae, composed of *P. antarctica*, *A. mitopteryx*, and the two species of *Dissostichus*, is resolved as paraphyletic in both analyses;
Fig. 1. Maximum likelihood phylogeny of Cryonotothenioidea inferred from RADseq dataset using IQ-TREE. Numbers at nodes are bootstrap values for those with less than 100% support. Photographs of nototheniid specimens by P. Marriott, P. McMillan, R. McPhee, T. J. Near, and C. Struthers and are deposited at the Museum of New Zealand Te Papa Tongarewa and Peabody Museum of Natural History, Yale University.
Fig. 2. Species tree inferred using SVDquartets. Numbers at nodes are bootstrap support values. Nodes that differ from the maximum likelihood IQ-TREE (Fig. 1) are marked with a filled black circle. Photographs of notothenioid specimens by P. Marriott, P. McMillan, R. McPhee, T. J. Near, and C. Struthers and are deposited at the Museum of New Zealand Te Papa Tongarewa and Peabody Museum of Natural History, Yale University.
however, the subclade containing *A. mitopteryx* and *Dissostichus* is strongly supported (Fig. 1). The Trematomiinae, comprising *Patagonotothen*, *Nototheniops*, *L. squamifrons*, *I. cyanobrancha*, and *Trematomus*, form a clade. *Indonotothenia cyanobrancha* is resolved as the sister lineage of *Trematomus* and there is strong node support for the relationships among species of *Trematomus* (Fig. 1). *Nototheniops*, *Patagonotothen*, and *Lepidonotothen* form a clade, but *Lepidonotothen* as traditionally delimited to contain *L. squamifrons* and the species of *Nototheniops* is resolved as non-monophyletic as *L. squamifrons* is the sister lineage of *Patagonotothen* (Figs. 1 and 2).

### Table 1
Comparison of phylogenetic hypotheses of Notothenioidei using the approximately unbiased (AU) test based on the resampling of estimated log-likelihoods (RELL) method.

<table>
<thead>
<tr>
<th>Phylogenetic hypothesis</th>
<th>logLn</th>
<th>ΔlogLn</th>
<th>bp-RELL</th>
<th>p-AU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal ML tree (Fig. 1)</td>
<td>−7765781.407</td>
<td>0.000</td>
<td>0.713</td>
<td>0.774</td>
</tr>
<tr>
<td>Nototheniidae monophyletic</td>
<td>−7766949.513</td>
<td>1168.106</td>
<td>0.000</td>
<td>0.010</td>
</tr>
<tr>
<td>Pleurogrammatinae monophyletic</td>
<td>−7765888.625</td>
<td>107.218</td>
<td>0.283</td>
<td>0.250</td>
</tr>
<tr>
<td>Lepidonotothen monophyletic</td>
<td>−7768175.483</td>
<td>2394.076</td>
<td>0.000</td>
<td>0.007</td>
</tr>
<tr>
<td>Arctocephalina monophyletic</td>
<td>−7767136.315</td>
<td>1354.908</td>
<td>0.000</td>
<td>0.006</td>
</tr>
<tr>
<td><em>Indonotothenia</em> and <em>Notothenia</em> sister taxa</td>
<td>−7767942.406</td>
<td>2160.999</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td><em>Patagonotothenia</em> not nested in <em>Trematomus</em></td>
<td>−7768068.765</td>
<td>2287.358</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td><em>Pseudotrematomus</em> (Balushkin 1982, 2000)</td>
<td>−7772505.186</td>
<td>6723.779</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>Phylogeny of <em>Trematomus</em> in Lautredou et al. (2012)</td>
<td>−7771618.300</td>
<td>5836.893</td>
<td>0.000</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Fig. 3. Predictions of phylogenetic utility for RADseq and legacy DNA sequence datasets. A. Violin plot comparing the distribution of site rates for each dataset, with the size of each dataset in base pairs (bp) indicated above each plot. An asterisk marks the truncation of the upper tail of the site rate distribution for graphical purposes. B. Hexbin plot of the relative phylogenetic informativeness (PI) over time of each RADseq locus. Colors correspond to the number of loci with a measured value of informativeness. Curved lines represent PI profiles of each legacy DNA sequence dataset. C. Highest 95% posterior density interval of cryonototheniod divergence times taken from Dornburg et al. (2017a). Dashed vertical lines corresponding to the previously estimated most recent common ancestor of cryonotothenioids and the onset of rapid lineage diversification hypothesized in Near et al. (2012). Photograph of *Histiodraco velifer* by A. Stewart and is deposited at the Museum of New Zealand Te Papa Tongarewa.
**Gobionotothen** is strongly supported as the sister lineage of an inclusive clade containing *Notothenia*, *Harpagifer*, Artedidraconidae, Bathydraconidae, and Channichthyidae (Figs. 1 and 2). The 14 sampled species of Bathydraconidae are a monophyletic group and are resolved as sister lineage of the Channichthyidae (Figs. 1 and 2). Relationships within the bathydraconids and channichthyids are well supported in the MLiq phylogeny, but relationships of the Antarctic Dragnishes *Racovitza glacialis*, Vomeridens *infuscipinnis*, and species of *Bathydraco* are only moderately supported (Fig. 1). *Harpagifer* and Artedidraconidae are strongly supported as a clade, but within Artedidraconidae *Artedidraconis* is paraphyletic with *A. skottsbergii* resolved as the sister lineage of all other artedidraconids and all other sampled species of *Artedidraconis* (*A. arianae*, *A. mirus*, *A. glareobarbasus*, and *A. shackletoni*) form a clade that is the sister lineage of Dolloidraco *longedorsalis* (Fig. 1). *Histidraco velifer* and the sampled species of *Pogonophryne* are resolved as sister lineages (Figs. 1 and 2).

The AU test rejects seven of the eight alternative hypotheses when compared to the optimal tree inferred from the IQ-TREE analysis (Table 1). The only hypothesis that is not rejected is the monophyly of Pleuragrammatinae (Table 1), which is sampled with *Pleuragramma antarctica*, *Aethotaxis mitopteryx*, *Dissostichus eleginoides*, and *D. mawsonii* (Figs. 1 and 2).

### 3.3. Predicted phylogenetic information content

A violin plot based comparison of the probability density of calculated site rates for RADseq loci and other classes of Sanger sequenced legacy markers reveal that RADseq loci possess a similar distribution of rates as observed in the previously sampled mtDNA genes (Fig. 3A). Visualizing the density of PI profiles through time for the RADseq loci as a hexagonal heatmap (Fig. 3B), shows a high-predicted level of informativeness through the Miocene (~23.0 to 5.3 Ma) (Fig. 3B), with particularly high densities of informative loci corresponding to the geologic intervals hypothesized as the periods of both the origin and diversification of major cryonotothenid lineages (Fig. 3B & C). Mapping the PI profiles of all other classes of legacy markers used to investigate notothenioid phylogeny onto this heat map demonstrates little decline in PI during the interval corresponding to the hypothesized geologic time period and estimated 95% highest posterior density intervals of molecular ages for the origin and diversification in the clade (Fig. 3B & C). Quantification of R, the predicted phylogenetic signal supporting a correct resolution, and H, the amount of phylogenetic noise supporting resolution based on homoplaspy, also reveals high probabilities of R for short internodes through the Late Miocene (~20 Ma; Fig. 4). In all cases, R rapidly maximized with the addition of data, predicting that RADseq loci contain high levels of phylogenetic information for resolving short internodes. However, our quantifications reveal that the predicted information content declines at internodes dating to early periods of the Cenozoic (Fig. 4). By the Cretaceous-Paleogene boundary (~66 Ma), there is very little phylogenetic information remaining, even when the dataset contains 10⁶ variable characters (Fig. 4).

### 4. Discussion

#### 4.1. Resolving the phylogenetic history of the notothenioid adaptive radiation

Over the past quarter century, the phylogeny of notothenioids has come increasingly into focus (Balushkin, 1992; Bargelloni et al., 1994, 2000; Lecointre et al., 1997; Balushkin, 2000; Near and Cheng, 2008; Dettai et al., 2012; Near et al., 2012, 2015; Dornburg et al., 2017a); however, there remain considerable areas of uncertainty. The trees inferred using the RADseq dataset have an unprecedented degree of resolution and node support for relationships within cryonototheniid lineages that exhibit high rates of diversification (Fig. 1). The phylogeny inferred in this study provides important insight regarding relationships among species of *Trematomus*, the phylogeny of species of Artedidraconidae and a clarification of the paraphyly of *Artedidraconis*, the deep paraphyly of Nototheniidae, consistent non-monophyly of *Lepidonotothen squamifrons* and the species of *Nototheniops*, the relationships of Bathydraconidae, and continued uncertainty in the phylogenetic placement of *Pleuragramma antarctica*. The well-supported resolution of relationships throughout the notothenioid phylogeny, especially among the radiations of closely related species (e.g., *Trematomus*, Artedidraconidae, Channichthyidae, and Bathydraconidae), provides...
opportunities to investigate mechanisms of speciation and evolutionary diversification in one of the most iconic and threatened marine ecosystems of our planet, an avenue of research identified as one of the six priorities for Antarctic science (Kennicutt et al., 2014).

Species of Trematomus comprise an important element of the nototheniid adaptive radiation (Janko et al., 2011; Lautredou et al., 2012; Near et al., 2012) as well as a dominant component of the nearshore nototheniid fish fauna (Kock, 1992). However, no prior study provides strong support for the resolution of the inter-relationships of these species. The RADseq data reject all of the alternative phylogenetic relationships involving Trematomus that includes a phylogeny inferred from a set of legacy Sanger sequenced mitochondrial and nuclear genes, the proposal that Trematomus is limited to T. newnesi, and all other species are classified as Pseudotrematomus (Balushkin, 1982), and that the Bald Nototen, Pagothenia borchgrevinki, is not phylogenetically nested in Trematomus (Table 1). Previous molecular studies have consistently resolved T. scotti as the sister lineage to other species of Trematomus (Ritchie et al., 1996; Sanchez et al., 2007; Kuhn and Near, 2009; Janko et al., 2011; Lautredou et al., 2012; Near et al., 2012), a result also strongly supported in the RADseq inferred phylogenies (Figs. 1 and 2). The phylogeny of Trematomus resolves strongly supported monophyletic groups containing epibenthic species T. loennbergi, T. lepidorhinus, and T. eudiplodus as a clade, as well as a clade containing the demersal species T. bernacchii and T. hansoni (Figs. 1 and 2). These relationships, initially suggested from analysis of sequence data of mtDNA rRNA genes, are consistent with the expectations of habitat use driven adaptive radiation within this clade (Ritchie et al., 1996).

The RADseq phylogenies resolve Indonotothenia cyanobrancha as the sister lineage of Trematomus (Figs. 1 and 2), a result consistent with analyses of mtDNA genes and a combination of mtDNA and nuclear genes (Bargelloni and Lecointre, 1998; Bargelloni et al., 2000; Dornburg et al., 2017a). Indonotothenia cyanobrancha was long classified in the genus Notothina (DeWitt et al., 1990), but Balushkin (1984: 13) described the monotypic genus Indonotothenia to classify this species. Based on morphological analyses, Indonotothenia is resolved as the sister lineage of a clade containing Notothina and Paranotothenia (Voskoboinikova, 1993: Fig. 8; Balushkin, 2000: Fig. 15). However, assessment of external morphological characters argues for a closer relationship with Trematomus (Hureau, 1970: 225, Table 14). The results of the phylogenetic analyses of the RADseq data strongly reject a close relationship of L. cyanobrancha with other species of Notothina and instead suggest that it is a species of Trematomus (Figs. 1 and 2; Table 1).

Similar to Trematomus, previous molecular phylogenetic analyses do not provide strong node support for relationships inferred among species of Arctedidracini. Arctedidracoins occupy benthic habitats of the Antarctic continental shelf and exhibit a high diversification rate relative to other nototheniid lineages (Eakin, 1990; Near et al., 2012). Consistent with previous molecular analyses (Derome et al., 2002; Lecointre et al., 2011; Near et al., 2012), the phylogeny inferred from the RADseq dataset resolves Arctedidraco as paraphyletic, with A. skottsbergi as the sister lineage of all other arctedidracoins (Figs. 1 and 2). The AU test rejects the best alternative phylogeny that resolves Arctedidraco as a monophyletic group (Fig. 1). The remaining species of Arctedidraco sampled in this study, including the type species A. mirus (Lönnberg, 1905: 39), comprise a strongly supported clade in the MLIQ tree that is sister to Dollopoidraco longidorsalis (Fig. 1), a relationship not hypothesized in previous molecular phylogenetic analyses (Derome et al., 2002; Lecointre et al., 2011; Near et al., 2012). The deep paraphyly of Arctedidraco s.l. and the lack of an available genus group name to accommodate A. skottsbergi necessitates a taxonomic revision that reflects the phylogeny and consistent resolution of Arctedidraco s.l. as a paraphyletic group.

Previous molecular analyses and the RADseq inferred phylogenies strongly support monophyly of the species-rich arctedidracoine lineage Pogonophryne (Figs. 1 and 2; Near and Cheng, 2008; Eakin et al., 2009; Lecointre et al., 2011; Near et al., 2012); however, the taxonomy of this group warrants reevaluation. Species of Pogonophryne are currently classified into five species groups on the basis of external morphological traits (Balushkin and Eakin, 1998). Although previous molecular analyses support these groupings, the relationships among them remain unresolved (Eakin et al., 2009). Analysis of the RADseq dataset results in a strongly supported phylogenetic hypothesis for the relationships among the species groups of Pogonophryne; however, the P. mentella group, which includes the sampled species P. macropogon, P. cerebropon, P. eakinii, and P. fusca, is not monophyletic as P. fusca is the sister lineages of P. barsukovi (Figs. 1 and 2). These results suggest a need for a reexamination of the classification of Pogonophryne species diversity using a dataset with greater taxonomic sampling. Given that species of Pogonophryne exhibit substantial overlap in bathymetric distribution (Eakin, 1990; Eastman, 2017) and diet (Wyanski and Targett, 1981; Eakin, 1990; Lombarte et al., 2003), and are characterized by very little morphological disparity among the species (Eakin, 1977; Lombarte et al., 2003), increased inter- and intraspecific sampling is needed to assess the delimitation of species and investigate the mechanisms driving diversification in this clade.

In addition to resolving relationships among the most closely related species, the RADseq inferred phylogenies also provide much-needed resolution to several higher-level relationships within Cryoontothenioidae. Since the first molecular analyses of nototheniid phylogeny, the inclusive family Nototheniidae is consistently resolved as a paraphyletic group (Bargelloni et al., 1994; Sanchez et al., 2007; Dettai et al., 2012; Near et al., 2012; Dornburg et al., 2017a). The concept of Nototheniidae communicated in taxonomic references (e.g., Eastman and Eakin, 2000; Nelson et al., 2016) has an origin with the pioneering work of Regan (1913: 249–251) and Norman (1938: 7–10) who provided important taxonomic revisions of nototheniid classification. Specifically, Regan (1913) classified species of Harpagifer and Arctedidraco in Nototheniidae along with Notothina s.l., Trematomus, Pleuragramma, Dissostichus, and Eleginops. Norman (1938) classified Arctedidraco, Dollopoidraco, Pogonophryne, and Harpagifer in Harpagiferidae and limited Nototheniidae to Notothina s.l., Trematomus, Pleuragramma, Dissostichus, and Eleginops. Subsequent discoveries of new species added Aethotaxis, Cryoontothenia, and Gvozdzars to Nototheniidae (DeWitt, 1962; Daniels, 1981; Balushkin, 1989). Taxonomic revisions by Balushkin (1976, 1992) led to the removal of Eleginops maclovini from Nototheniidae to the monotypic Eleginopsidae and the description of several new genera: Gobionotothen, Lepidonotothen, Notothina, Patagonotothen, and Paranotothenia all of which contain species that were previously classified as Notothina (Norman, 1938; Andriashev, 1965). The classification of these disparate lineages as Notothina at the time of the origin of modern classifications for notothenioids contributed to the long-held idea that Nototheniidae is a natural group (e.g., Norman, 1938). Molecular phylogenies, including those inferred from the RADseq data (Figs. 1 and 2), provide a strong inference that Nototheniidae as traditionally delimited is not monophyletic. The alternative phylogeny with the highest likelihood that depicts Nototheniidae, as traditionally delimited, is rejected in the AU test (Table 1). While this study did not sample species of Paranotothenia, previous molecular phylogenetic analyses consistently resolve Notothina and Paranotothenia as a clade (Cheng et al., 2003; Sanchez et al., 2007; Near and Cheng, 2006; Near et al., 2012; Dornburg et al., 2017a). We recommend that Nototheniidae is limited to Notothina and Paranotothenia.

Similar to the traditional delimitation of Nototheniidae, the RADseq tree and previous molecular phylogenetic analyses consistently fail to resolve Lepidonotothen squamifrons and species of Notothina as a monophyletic group (Bargelloni et al., 2000; Near and Cheng, 2008; Dettai et al., 2012; Near et al., 2012). Specifically, L. squamifrons is resolved as the sister lineage of Patagonotothen (Figs. 1 and 2). In his revision of Notothina, Balushkin (1976, 1979) described the genera...
Lepidonotothen (containing L. squamifrons and the two synonyms L. kempi and L. macrophthalma), Nototheniops (containing N. larmani, and synonyms N. loeha, N. rybelini, and N. tchizh), and Lindbergichthys (containing N. mizops, N. nudifrons, and the undescribed N. cf. nudifrons). The most recent taxonomic revision of these lineages treats Nototheniops and Lindbergichthys as subgenera of Lepidonotothen without identifying any morphological evidence to support the hypothesis that they share common ancestry (DeWitt et al., 1990; 294–295). Phylogenetic analysis of morphological characters does not resolve Lepidono- tothen, Nototheniops, and Lindbergichthys as a monophyletic group (Balushkin, 2000: Fig. 15) and this phylogeny is rejected in the AU test (Table 1). We recommend that Nototheniops (Balushkin, 1976) is the appropriate genus group name for N. larmani, N. mizops, N. nudifrons, and N. cf. nudifrons. The monophyly of Nototheniops, as delimited here (Figs. 1 and 2), is supported with several synapomorphic morphological characters that include an upper lateral line with perforated scales and a supraorbital canal with four pores (Andersen, 1984: 24).

The evolutionary loss of red blood cells, hemoglobin, and the variable loss of myoglobin expression in Channichthyidae are unique among vertebrates (Ruud, 1954; Sidell et al., 1997; Sidell and O’Brien, 2006). The genetics and physiological consequences of these highly unusual traits are well studied (Egginton and Rankin, 1998; Egginton et al., 2002; Near et al., 2006; Beers and Sidell, 2009; Beers et al., 2010; Beers and Sidell, 2011; Lewis et al., 2015; Xu et al., 2015; Kuhn et al., 2016); however increased resolution of the phylogenetic relationships of Channichthyidae has potential to provide insights into the evolution of hemoglobin loss (Bargelloni et al., 1998; Near et al., 2006; Lau et al., 2012). While morphological and molecular phylogenetic analyses consistently resolve Channichthyidae and Bathydraconidae as a clade (Iwami, 1985: Fig. 174; Balushkin, 1992: Fig. 6; 2000: Fig. 11; Bargelloni et al., 2000; Near and Cheng, 2008), most molecular analyses result in phylogenies where the channichthyids are nested in a paraphyletic Bathydraconidae. In these previous molecular analyses strong support for the resolution of the sister lineage of Channichthyidae is lacking (e.g., Derome et al., 2002; Dettaï et al., 2012; Near et al., 2012). The RADseq dataset resolves Bathydraconidae as a monophyletic group and the sister lineage of Channichthyidae with strong bootstrap support in the MLiq inferred phylogeny, but with lower support in the SVDquartet phylogeny (Figs. 1 and 2). The common ancestor of Channichthyidae and Bathydraconidae likely exhibited decreased hematocrit, lower concentrations of hemoglobin, and reduced globin chain multiplicity (D’Avino and Di Prisco, 1988; di Prisco, 1998; Verde et al., 2007; Wujick et al., 2007). The phylogenetic resolution of the channichthyid sister lineage facilitates contextualizing the genomic pathways that have given rise to these highly unusual phenotypes.

Despite the well-supported phylogenetic resolution of notothenioid relationships resulting from analysis of the RADseq dataset, the placement of the Antarctic Silverfish, Pleuragramma antarctica, remains unresolved. This monotropic pelagic lineage is a radical departure in phenotype from other notothenioids (Eastman, 1997; Voskoboinikova et al., 2017) and is of key importance to Antarctic food webs (Zane et al., 2006; Mintenbeck and Torres, 2017). Balushkin (2000: S101, Fig. 14) delimited the clade Pleuragrammatinae to contain P. antarctica, the two species of Dissostichus, Aethotaxis mitopteryx, and Gvozdarus svetovidovi based on pleural ribs originating from the 4th or 5th vertebrae and the reduction or absence of a basihenipon in the skull. Among notothenioids, Pleuragramma antarctica, Dissostichus mawsoni, and Aethotaxis mitopteryx are neutrally buoyant as adults (Eastman and DeVries, 1982; Near et al., 2003, 2007), and common ancestry of these lineages would indicate a single origin of neutral buoyancy in notothenioids (Near et al., 2007). However, previous molecular analyses do not confidently support monophyly of Pleuragrammatinae or result in different resolutions for P. antarctica. Our analyses do not resolve Pleuragrammatinae as a clade, but one poorly supported node separates P. antarctica from a monophyletic group containing A. mitopteryx and the two species of Dissostichus (Figs. 1 and 2). Gvozdarus svetovidovi was not included in our analysis. Investigation of the uncertainty regarding phylogenetic relationships of P. antarctica requires additional work as evidenced by the inability of the RADseq dataset to reject the best alternative phylogeny that depicts Pleuragrammatinae as a monophyletic group (Table 1).

4.2. RADseq and Cenozoic radiations

Rapidly radiating clades characterize some of the most investigated portions of the Tree of Life (Venditti et al., 2010; Leache et al., 2016; Brennan and Oliver, 2017). Regardless of timescale, the short internodes separating species divergences that characterize rapid radiations also present some of the most challenging problems in phylogenetics (Sharma et al., 2014; Eytan et al., 2015). As such, assessing the phylogenetic utility of different classes of genomic markers is important for cost-effective phylogenomic experimental design. Our investigation of RADseq loci in notothenioids reveals high levels of predicted information for resolving short internodes that date from the Late Oligocene through the Pleistocene (Figs. 3 & 4). This predicted utility is reflected in high levels of support for most nodes in the notothenioid phylogeny that correspond to these geologic ages (Fig. 1). However, for radiations dating to the Early Cenozoic (~65 to 50 mya), our results suggest the massive amount of data offered by RADseq provides diminishing returns (Fig. 4). These findings provide quantitative support substantiating claims that RADseq loci are useful for resolving interspecific phylogenetic problems (Cariou et al., 2013; Massatti et al., 2016; Eaton et al., 2017), while also setting new expectations of the temporal limits for this class of data. Our results suggest RADseq loci contain similar levels of phylogenetic information as the flanking regions of ultra-conserved elements (UCEs), which contain high levels of phylogenetic information for divergences dating to the Miocene (~20 mya) (Gilbert et al., 2015). Both the RADseq loci sampled across the diversity of notothenioids and UCE-flanking regions in teleost fishes exhibit a rapid decay in phylogenetic information for divergences arising prior to the K-Pg boundary, 66 mya (Fig. 3; Gilbert et al., 2015). While our study suggests RADseq loci exhibit high utility for resolving difficult Late Cenozoic phylogenetic problems, it is important to consider that predictions of utility are not guarantees of successful phylogenetic resolution (Townsend and Leuenberger, 2011).

Neither PI profiles nor quartet internode resolution probabilities make explicit statements on the predicted levels of node support (Townsend and Leuenberger, 2011). These approaches merely indicate whether there is a probability of phylogenetic information given the theoretical expectations of phylogenetic experimental design. This does not imply that phylogenetic information sufficient for strong node support is present in the dataset. It is possible that limited phylogenetic information explains the lack of confidence in the phylogenetic resolution of Pleuragramma antarctica and the inability for the RADseq dataset to reject the monophyly of the Pleuragrammatinae (Figs. 1 and 2; Table 1). Alternatively, the nature of RADseq data imposes several challenges to predictions of utility that could also explain our lack of confidence in the resolution of these nodes.

The low number of variable sites per locus limits the power of experimental design approaches for finely dissecting information content by locus. While it is possible to generate PI profiles from few or even single sites, the dependency of the interaction of locus length on experimental design approaches for

276
number of methods (e.g., Xia et al., 2003; Goremykin et al., 2009, 2010) will lead to amplification of an erroneous “signal” in the data leading to confidence for an erroneous phylogenetic resolution (Dornburg et al., 2017c). Given that we found base frequencies to be near stationarity (A = 0.265; C = 0.239; G = 0.242; T = 0.253) nucleotide bias is not likely a major axis of error in the RADseq dataset. However, even a small number of loci dominated by this or other forms of homoplasy can impact the topological resolution (Shen et al., 2017). Additionally, RADseq datasets do contain high levels of missing data, so it is possible that information rich sites for this node were simply not captured during sequencing. As such, determining the factors limiting the confident resolution of Pleurogramma antarctica remains an open question. However, in the case of the other inferred notothenioid relationships, congruence in phylogenies inferred from different sets of genes (e.g., Dettai et al., 2012; Near et al., 2012) and expectations of utility based on analysis of phylogenetic informativeness (Fig. 3) provide confidence in our resolution for the notothenioid phylogeny (Figs. 1 and 2).

Understanding the drivers of diversification in rapidly radiating clades is a primary area of research in evolutionary biology (e.g., Gavriltsev and Losos, 2009). Given that many of the iconic vertebrate adaptive radiations such as anoles (Poe et al., 2017) and cichlids (Friedman et al., 2013) are Cenozoic in origin, our findings coupled with the effectiveness of capturing large amounts of data for non-model organisms underscore the utility of RADseq for providing phylogenetic resolution to recent radiations and species flocks (e.g., Wagner et al., 2013). However, our study also demonstrates the heterogeneity of phylogenetic information within this class of genomic markers, offering insights into phylogenetic informativeness as it related to divergence time (Fig. 3), which is consistent with results presented by Collins and Hrbek (2018). While RADseq is of tremendous utility for late Cenozoic radiations, returns increasingly diminish for radiations moving deeper in time towards the beginning of the Cenozoic or earlier (Figs. 3 and 4). Our results provide an important context for the application of RADseq data to resolving interspecific phylogenetics and complement similar studies conducted on other types of next-generation sequence data such as UCE or loci captured by anchored hybrid enrichment (Gilbert et al., 2015; Prum et al., 2015; Dornburg et al., 2017b; Reddy et al., 2017; Collins and Hrbek, 2018). Future studies comparing the relative performance of multiple classes of markers targeted by next-generation sequencing techniques will contribute to the optimization of phylogenetic experimental design and lead to an efficient and cost effective resolution of the Genomic Tree of Life.

Acknowledgements

Field and laboratory support was provided by W. Brooks, H. W. Detrich, J. Kendrick, K. H. Kuhn, K. Lamb, J. A. Moore, and J. Pennington, G. Watkins-Colwell, K. Zapfe. P. Brickle (South Atlantic Environmental Research Institute, Falkland Islands), A. L. DeVries and C.-H. C. Cheng (University of Illinois, USA), D. A. Fernández (Centro Austral de Investigaciones Científicas, Argentina), and A. Dettai and G. Lecontinent (Muséum National d’Histoire Naturelle, France) provided critical specimens. Fieldwork was facilitated through the United States Antarctic Marine Living Resources Program and the officers and crew of the RV Yuzhmorgeologiya, and the 2004 ICEFISH cruise aboard the RVIB Nathaniel B. Palmer. Fresh color images of notothenioids used in Figs. 1 and 2 were taken by P. Marriott, P. McMillan (NIWA), and R. McPhee (Te Papa). Specimens and data collected by and made available through the New Zealand National Institute of Water and Atmospheric Research Ltd Core Funded Coasts & Oceans Programme 2: Biological Resources subcontract for fundamental knowledge of marine fish biodiversity with the Museum of New Zealand Te Papa Tongarewa. This research was supported by the Bingham Oceanographic fund of the Peabody Museum of Natural History, Yale University and the United States National Science Foundation ANT-0839007. Two reviewers provided helpful comments on the manuscript. All data and analysis files have been archived on Zenodo (DOI 10.5281/zenodo.1406314).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ympev.2018.09.001.

References


