Genomic methods take the plunge: recent advances in high-throughput sequencing of marine mammals

Citation for published version:
Cammen, KM, Andrews, KR, Carroll, EL, Foote, AD, Humble, E, Khudyakov, JI, Louis, M, McGowen, MR, Olsen, MT & Van Cise, AM 2016, 'Genomic methods take the plunge: recent advances in high-throughput sequencing of marine mammals' Journal of Heredity, vol. 107, no. 6, pp. 481-495. DOI: 10.1093/jhered/esw044

Digital Object Identifier (DOI):
10.1093/jhered/esw044

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of Heredity

OXFORD UNIVERSITY PRESS | Journal of Heredity

Genomic methods take the plunge: recent advances in high-throughput sequencing of marine mammals

Journal:	*Journal of Heredity*
Manuscript ID	JOH-2016-093.R2
Manuscript Type:	Invited Review
Date Submitted by the Author:	n/a
Complete List of Authors:	Cammen, Kristina; University of Maine, School of Marine Sciences Andrews, Kim; University of Idaho, Department of Fish and Wildlife Sciences Carroll, Emma; University of St Andrews, Scottish Oceans Institute Foote, Andy; University of Bern, Institute of Ecology and Evolution Humble, Emily; University of Bielefeld, Department of Animal Behaviour; British Antarctic Survey Khudyakov, Jane; Sonoma State University, Department of Biology Louis, Marie; University of St Andrews, Scottish Oceans Institute McGowen, Michael; Queen Mary University of London, School of Biological and Chemical Sciences Olsen, Morten; University of Copenhagen, Natural History Museum of Denmark Van Cise, Amy; University of California San Diego, Scripps Institute of Oceanography
Subject Area:	Genomics and gene mapping
Keywords:	RADseq, target sequence capture, whole genome sequencing, RNAseq, SNP array, non-model organisms

Genomic methods take the plunge: recent advances in high-throughput sequencing of marine mammals

KRISTINA M. CAMMEN[1]*, KIMBERLY R. ANDREWS[2], EMMA L. CARROLL[3], ANDREW D. FOOTE[4], EMILY HUMBLE[5,6], JANE I. KHUDYAKOV[7], MARIE LOUIS[3], MICHAEL R. MCGOWEN[8], MORTEN TANGE OLSEN[9], AND AMY M. VAN CISE[10]

[1]School of Marine Sciences, University of Maine, Orono, Maine 04469, USA
[2]Department of Fish and Wildlife Sciences, University of Idaho, 875 Perimeter Drive MS 1136, Moscow, Idaho 83844-1136, USA
[3]Scottish Oceans Institute, University of St Andrews, East Sands, St Andrews, Fife KY16 8LB, UK
[4]Computational and Molecular Population Genetics Lab, Institute of Ecology and Evolution, University of Bern, Bern CH-3012, Switzerland
[5]Department of Animal Behaviour, University of Bielefeld, Postfach 100131, 33501 Bielefeld, Germany
[6]British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 OET, UK
[7]Department of Biology, Sonoma State University, Rohnert Park, California 94928, USA
[8]School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK
[9]Evolutionary Genomics Section, Natural History Museum of Denmark, University of Copenhagen, DK-1353 Copenhagen K, Denmark
[10]Scripps Institution of Oceanography, 8622 Kennel Way, La Jolla, California 92037, USA

*Corresponding author: kristina.cammen@maine.edu

Running title: Marine mammal genomics

Abstract

The dramatic increase in the application of genomic techniques to non-model organisms over the past decade has yielded numerous valuable contributions to evolutionary biology and ecology, many of which would not have been possible with traditional genetic markers. We review this recent progression with a particular focus on genomic studies of marine mammals, a group of taxa that represent key macroevolutionary transitions from terrestrial to marine environments and for which available genomic resources have recently undergone notable rapid growth. Genomic studies of non-model organisms utilize an expanding range of approaches, including whole genome sequencing, restriction site-associated DNA sequencing, array-based sequencing of single nucleotide polymorphisms and target sequence probes (e.g., exomes), and transcriptome sequencing. These approaches generate different types and quantities of data, and many can be applied with limited or no prior genomic resources, thus overcoming one traditional limitation of research on non-model organisms. Within marine mammals, such studies have thus far yielded significant contributions to the fields of phylogenomics and comparative genomics, as well as enabled investigations of fitness, demography, and population structure. Here we review the primary options for generating genomic data, introduce several emerging techniques, and discuss the suitability of each approach for different applications in the study of non-model organisms.

Keywords: RADseq, SNP array, target sequence capture, whole genome sequencing, RNAseq, non-model organisms

Introduction

Recent advances in sequencing technologies, coincident with dramatic declines in cost, have increasingly enabled the application of genomic sequencing in non-model systems (Ekblom and Galindo 2011; Ellegren 2014). These advances in molecular technologies have in many ways begun to blur the distinction between model and non-model organisms (Armengaud et al. 2014). Non-model organisms (NMOs) have traditionally been defined as those for which whole-organism experimental manipulation is rarely, if ever, possible due to logistical and/or ethical constraints (Ankeny and Leonelli 2011). Further, NMOs have typically been characterized by limited genomic resources, but this is becoming increasingly less so as the number of NMO reference genomes grows rapidly, for example through efforts like the Genome 10K Project (Koepfli et al. 2015). In fact, in some taxonomic orders, we are approaching the point at which all species have at least one representative reference genome available for a closely related species (Fig 1).

Despite the limitations of working with NMOs, including potentially small sample sizes, low DNA quantity, and limited information on gene function, genetic and genomic investigations of NMOs have yielded numerous valuable contributions to understanding their evolutionary biology and ecology. For the past several decades, traditional genetic markers such as microsatellites and short fragments of mitochondrial DNA (e.g., the control region) have been extensively used in molecular ecology. These markers, which typically evolve under neutral expectations, have proven useful for identifying population structure and reconstructing population demographic history (Hedrick 2000). However, the power of such studies is limited by the number of markers that can feasibly be evaluated using traditional approaches. The advent of low-cost high-throughput sequencing has led to dramatic increases in the number of neutral markers that can be evaluated, in many cases improving our power to resolve fine-scale or cryptic population structure in species with high dispersal capability (e.g., Corander et al. 2013) and improving the accuracy of estimating some (though not all) demographic parameters (Li and Jakobsson 2012; Shafer et al. 2015). Importantly, high-throughput sequencing has also further enabled genomic studies of non-neutral processes in NMOs, for example, characterizing both deleterious and adaptive variation within and across species (Stinchcombe and Hoekstra 2008;

Künstner et al. 2010). It is increasingly evident that genomic analyses of NMOs can and have provided important insights that could not be identified with traditional genetic markers.

Many molecular ecologists now face the challenge of deciding which of the broad range of genomic approaches to apply to their study systems. Here we review the primary options for generating genomic data and their relative suitability for different applications in the study of NMOs. We focus on marine mammals, which represent several mammalian clades with notably rapid growth in available genomic resources in recent years. This growth is clearly evident in both publication rate (Fig 2) and the rise in number and size of genomic sequences deposited in public resources (Fig 3). We comprehensively review the literature on marine mammal genomics, highlighting recent trends in methodology and applications, and then describe in detail the molecular approaches that are most commonly applied to studies of NMO genomics. Our hope is that this review will highlight the promise of genomics for NMOs and offer guidance to researchers considering the application of genomic techniques in their non-model study system of choice.

Why study marine mammal genomics?

Marine mammals represent key macroevolutionary transitions from terrestrial to marine environments (McGowen et al. 2014) and accordingly are an exemplary system for investigating the evolution of several morphological and physiological adaptations (Foote et al. 2015) associated with locomotion (Shen et al. 2012), sight (Meredith et al. 2013), echolocation (Parker et al. 2013; Zou and Zhang 2015), deep diving (Mirceta et al. 2013), osmoregulation (Ruan et al. 2015), and cognition (McGowen et al. 2012). Furthermore, studies of marine mammal evolution to date have characterized several unique aspects of their genome evolution that merit further investigation, including low genomic diversity and a relatively slow molecular clock, especially in cetaceans (Jackson et al. 2009; McGowen et al. 2012; Zhou et al. 2013). As many cetacean species are highly mobile with no obvious physical geographic barriers to dispersal, they provide a unique opportunity to study the role of behavior and culture in shaping population structure and genetic diversity (Riesch et al. 2012; Carroll et al. 2015; Alexander et al. 2016). Though highly mobile, many marine mammals exhibit evidence of local adaptation; for example, several species show parallel divergent morphological and behavioral adaptations to coastal and pelagic environments (Moura et al. 2013; Louis et al. 2014; Viricel and Rosel 2014). These species may be studied across ocean basins as emerging examples of ecological adaptation and speciation (Morin et al. 2010a).

Beyond their value as systems of evolutionary study, many marine mammals are also of broader interest relating to their historical and present conservation status. Many marine mammal populations share histories of dramatic decline due to hunting and other human impacts. Genomics provides a promising tool with which to expand our insights into these historical population changes, which so far primarily have been derived from archival review and traditional genetic approaches (Ruegg et al. 2013; Sremba et al. 2015). More recently, since the implementation of national and international protections, many marine mammal populations have partially or fully recovered (Magera et al. 2013), yet the conservation status of certain marine mammal populations remains of concern. Such vulnerable populations could benefit greatly from an improved understanding of their genetic diversity and evolution, especially in ways that can inform predictions of adaptive capacity to anthropogenic pressures and expand the toolkit for conservation policy (Garner et al. 2016; Taylor and Gemmell 2016).

Recent trends in marine mammal genomics

We conducted a meta-analysis of the peer-reviewed marine mammal genomics literature to evaluate trends in publication rates across research methodologies and aims. A search of the Web of Science database using the term “genom*” and one of the following terms indicating study species - “marine mammal”, “pinniped”, “seal”, “sea lion”, “sea otter”, “whale”, “dolphin”, “polar bear”, “manatee” - identified 825 records on December 11, 2015. We excluded 77% of the search results that were not directly related to genomic studies in marine mammal systems. The remaining 101 articles that were relevant to marine mammal genomics were further categorized by primary research methodology and general research aim. A subset of these articles is described briefly in Supplemental Table 1.

From the early 1990s through 2015, published literature in the field shifted from an early focus on mitogenome sequencing to more sequence-intensive approaches, such as transcriptome and whole genome sequencing (Figs 2 and 4). This trajectory closely follows trends in sequencing technologies, from Sanger sequencing of short- and long-range PCR products for mitogenome sequencing (Arnason et al. 1991) and SNP discovery (Olsen et al. 2011), to high-throughput sequencing of reduced-representation genomic libraries (RRLs) that consist of selected subsets of the genome (e.g., Viricel et al. 2014), to high-throughput sequencing of whole genomes with varying levels of depth, coverage, and contiguity. Today, high-throughput sequencing can be used both to generate high-quality reference genome assemblies (Yim et al. 2014; Foote et al. 2015; Humble et al. 2016) and to re-sequence whole genomes at a population scale (Liu et al. 2014a; Foote et al. 2016). Similarly, the scale of gene expression studies has increased from quantitative real-time PCR of candidate genes (Tabuchi et al. 2006) to microarrays containing hundreds to thousands of genes (Mancia et al. 2007) and high-throughput RNAseq that evaluates hundreds of thousands of contigs across the genome (Khudyakov et al. 2015b). As the cost of high-throughput sequencing continues to decline, we anticipate an increase in studies that sequence RRLs, whole genomes, and transcriptomes in NMOs at a population scale.

Marine mammal genomic studies thus far have primarily contributed to the fields of phylogenomics and comparative genomics (Fig 2, Table S1). Several of these comparative genomics studies have aimed to improve our understanding of the mammalian transition to an aquatic lifestyle and describe the evolutionary relationships within and among marine mammals and their terrestrial relatives (McGowen et al. 2014; Foote et al. 2015). Whereas such studies require only a single representative genome per species, an emerging class of studies applying genomic techniques at a population scale enables further investigations of fitness, demography, and population structure within species (Table S1). However, expanding the scale of genomic studies requires careful selection of an appropriate method for data generation and analysis from a growing number of approaches that are becoming available to non-model systems.

Data generation

Our review of marine mammal genomics highlights an increasing number of options for the generation and analysis of genomic data. Choosing which of these sequencing strategies to apply is a key step in any genomics study. Here we describe approaches that have been used successfully in order to help guide future studies of ecological, physiological, and evolutionary genomics in NMOs. Across data generation methods, we highlight approaches that can be used with limited or no prior genomic resources, overcoming one traditional challenge of genomic studies of NMOs (the need for a reference genome to which sequencing reads can be mapped). These methods produce a range in quantity and type of data output, from hundreds of SNPs to whole genome sequences, and from single individuals to population samples, reflecting the trade-off between number of samples and amount of data generated per sample.

Sample collection, storage and extraction

Prior to starting a genomic study, researchers must recognize that many recent methods for high-throughput sequencing require genetic material of much higher quality and quantity than techniques used to characterize traditional genetic markers. These more stringent sample requirements necessitate new standards for tissue sampling, storage, and DNA/RNA extraction. Ideally, samples should be collected from live or newly deceased individuals and stored at -80°C, or when this is not possible at -20°C in RNAlater, Trizol, ethanol, salt-saturated DMSO, or dry, depending on the intended application. Given the sensitivity of new sequencing methods, great care should be taken to minimize cross-contamination during sampling, as even minute amounts of genetic material from another individual can bias downstream analyses, for example variant genotyping and gene expression profiles. Choice of extraction method varies with sample type and study aim, but typically genomic methods require cleanup and treatment with RNase to yield pure extracts, whereas RNAseq methods require rigorous DNase treatment to remove genomic contamination that can bias expression results. Depending on the genomic methodology, target quantities for a final sample may range from as low as 50 ng of DNA for some RRL sequencing methods (Andrews et al. 2016) up to ~1 mg for sequencing the full set of libraries (of different insert sizes) necessary for high-quality genome assemblies (Ekblom and Wolf 2014). Most commercial RNAseq library preparation services require at least 500-1,000 ng of pure total RNA that shows minimal degradation as measured by capillary gel electrophoresis (RNA Integrity Number (RIN) ≥ 8). Samples should ideally consist of high molecular weight genetic material (with little shearing), though continuing molecular advances enable genomic sequencing even of low quantity or poor quality starting material. Extreme examples of the latter include successfully sequenced whole genomes from ancient material (e.g., Rasmussen et al. 2010; Meyer et al. 2012; Allentoft et al. 2015), including a more than 500,000-year-old horse (Orlando et al. 2013).

<u>Reduced-representation genome sequencing</u>

i. RADseq

Reduced-representation sequencing methods evaluate only a small portion of the genome, allowing researchers to sequence samples from a larger number of individuals within a given budget in comparison to sequencing whole genomes. Restriction site-associated DNA sequencing (RADseq) is currently the most widely used RRL sequencing method for NMOs (Davey et al. 2011; Narum et al. 2013; Andrews et al. 2016). RADseq generates sequence data from short regions adjacent to restriction cut sites and therefore targets markers that are distributed relatively randomly across the genome and occur primarily in non-coding regions. This method allows simultaneous discovery and genotyping of thousands of genetic markers for virtually any species, regardless of availability of prior genomic resources. Of greatest interest are variable markers, characterized either as single SNPs or phased alleles that can be resolved from the identification of several variants within a single locus.

The large number of markers generated by RADseq dramatically increases genomic resolution and statistical power for addressing many ecological and evolutionary questions when compared to studies using traditional markers (Table S1). For example, heterozygosity fitness correlations in harbor seals (*Phoca vitulina*) were nearly fivefold higher when using 14,585 RADseq SNPs than when using 27 microsatellite loci (Hoffman et al. 2014). A recent study on the Atlantic walrus (*Odobenus rosmarus rosmarus*) using 4,854 RADseq SNPs to model demographic changes in connectivity and effective population size associated with the Last Glacial Maximum (Shafer et al. 2015) both supported and extended inferences from previous studies using traditional markers (Shafer et al. 2010; Shafer et al. 2014).

Furthermore, RADseq can provide sufficient numbers of markers across the genome to identify genomic regions influenced by natural selection. These analyses require large numbers (thousands to tens of thousands) of markers to ensure that some markers will be in linkage disequilibrium with genomic regions under selection and to minimize false positives, particularly under non-equilibrium demographic scenarios (Narum and Hess 2011; De Mita et al. 2013; Lotterhos and Whitlock 2014). Extreme demographic shifts, as experienced by many marine mammal populations (e.g., killer whales, Foote et al. 2016), can drive shifts in allele frequencies that confound the distinction of drift and selection and make it difficult to detect genomic signatures of selection (Poh et al. 2014). Proof of concept of the application of RADseq for identifying genomic signatures of selection in wild populations was demonstrated in three-spined sticklebacks (*Gasterosteus aculeatus*), for which analyses of over 45,000 SNPs (Hohenlohe et al. 2010) identified genomic regions of known evolutionary importance associated with differences between marine and freshwater forms (Colosimo et al. 2005; Barrett et al. 2008). RADseq studies with similar aims in marine mammals have resulted in comparatively sparser sampling of SNPs (<10,000), likely due to both methodological differences and generally low genetic diversity particularly among cetaceans. Nonetheless, genomic regions associated with resistance to harmful algal blooms in common bottlenose dolphins (*Tursiops truncatus*) were identified across multiple pairwise comparisons using 7,431 RADseq SNPs (Cammen et al. 2015), and genomic regions associated with habitat use and resource specialization in killer whales (*Orcinus orca*) were identified using 3,281 RADseq SNPs (Moura et al. 2014a). Some of these RADseq SNPs associated with diet in killer whales were later also confirmed as occurring in genomic regions of high differentiation and reduced diversity consistent with a signature of selection identified in a study utilizing whole genome re-sequencing (Foote et al. 2016). It will remain important for further studies of genomic signatures of selection in NMOs to carefully consider which approach will generate a sufficiently large number of SNPs to accurately identify the range of putatively neutral F_{ST} values (and thus outliers) given the demographic history of the population (Lotterhos and Whitlock 2014).

Numerous laboratory methods have been developed for generating RADseq data (reviewed in Andrews et al. 2016), with the most popular library preparation methods currently being the original RAD (Miller et al. 2007; Baird et al. 2008), Genotyping by Sequencing (GBS, Elshire et al. 2011; Poland et al. 2012), and double digest RAD (ddRAD, Peterson et al. 2012). All RADseq methods share the common goal of sequencing regions adjacent to restriction cut sites across the genome, but differ in technical details, such as the number and type of restriction enzymes used, the mechanisms for reducing genomic DNA fragment sizes, and the strategies for attaching sequencing adapters to the target DNA fragments. For example, both the original RAD method and GBS use a single enzyme digest, but the original RAD method uses a rare-cutting enzyme and mechanical shearing to reduce DNA fragment size (Baird et al. 2008), whereas GBS uses a more frequent-cutting enzyme and relies on preferential PCR amplification of shorter fragments for indirect size selection (Elshire et al. 2011). These modifications lead to differences across methods in the time and cost of library preparation, the number and lengths of loci produced, and the types of error and bias present in the resulting data. Different RADseq methods will be better suited to different research questions, study species, and research budgets, and therefore researchers embarking on a RADseq study should carefully consider the suitability of each method for their individual projects. Further details on the advantages and disadvantages of each method are described in Andrews et al. (2016).

ii. SNP arrays

An alternative high-throughput reduced-representation genotyping approach involves the use of custom arrays designed to capture and sequence targeted regions of the genome. Such array-based approaches may provide certain advantages over RADseq, including the ability to easily estimate genotyping error rates, scalability to thousands of samples, lower requirements for DNA quantity/quality and technical effort, greater comparability of markers across studies, and the ability to genotype SNPs within candidate genomic regions. However, unlike RADseq, array-based techniques require prior knowledge of the study system's genome or the genome of a closely related species, which remains unavailable for some NMOs. Furthermore, SNP arrays must take into account the potential for ascertainment bias (e.g., Malenfant et al. 2015), whereas RADseq avoids ascertainment bias by simultaneously discovering and genotyping markers.

To identify SNPs for NMO array development, researchers must rely on existing genomic resources or generate new reference sequences, in the form of whole or reduced-representation genomes or transcriptomes (Hoffman et al. 2012; Malenfant et al. 2015). When a whole genome reference assembly is available for the target species or a related species, multiplex shotgun sequencing can facilitate the rapid discovery of hundreds of thousands of SNPs for array development. This SNP discovery approach involves high-throughput sequencing of sheared genomic DNA that can be sequenced at a low depth of coverage (i.e., low mean read depth across the genome) if suitable genotype likelihood-based methods (O'Rawe et al. 2015) are used to identify polymorphic sites. Thus, this approach is less restrictive in terms of DNA quality. For example, shotgun sequencing of 33 Northeast Atlantic common bottlenose dolphins, which included degraded DNA collected from stranded specimens, on one Illumina HiSeq2000 lane of 100 bp single-end sequencing identified 440,718 high-quality SNPs (M. Louis unpublished data). Such dense sampling of SNPs is essential for studies of population genomics that require a large number of markers, such as for inferences of demographic history (Gutenkunst et al. 2009; Excoffier et al. 2013; Liu and Fun 2015) and selective sweeps (Chen et al. 2010). Once a set of putative markers has been identified, hybridization probes can be designed from their flanking sequences and printed onto a SNP array. The two principal SNP genotyping platforms supporting thousands to millions of SNPs are the Illumina Infinium iSelect® and Affymetrix Axiom® arrays.

The use of SNP arrays in NMOs has thus far been somewhat limited, potentially due to low SNP validation rates (Chancerel et al. 2011; Helyar et al. 2011), issues of ascertainment bias (Albrechtsen et al. 2010; McTavish and Hillis 2015), and cost of SNP discovery. However, using both SNP data and whole genome sequence from the Antarctic fur seal (*Arctocephalus gazella*), Humble et al. (2016) recently demonstrated that careful filtering based on SNP genomic context prior to array development has the potential to substantially increase assay success rates. Further, ascertainment bias can be reduced by selecting samples for SNP discovery that span the geographic range of populations that will be target sequenced (Morin et al. 2004). By accounting for ascertainment bias, Malenfant et al. (2015) were able to demonstrate population structure in Canadian polar bears (*Ursus maritimus*) more clearly using a 9K SNP array than 24 microsatellite markers.

iii. Target sequence capture

Target sequence capture (TSC, also called target enrichment, direct selection, or Hyb-seq) has many of the same advantages and disadvantages as the array-based SNP approaches described above, but differs in library preparation, sequencing platform, and resulting sequence data. While SNP arrays genotype single variable positions, TSC can be used to sequence selected short fragments. With TSC, researchers can amplify and sequence up to a million target probes on solid-state arrays, and even more if in-solution arrays are used. This gives the user the ability to choose to sequence many samples in parallel (Cummings et al. 2010), as many as 100-150 per

Illumina HiSeq lane, or to sequence many regions per individual. Recent advances in target enrichment, such as genotyping in thousands (Campbell et al. 2015), anchored hybrid enrichment (Lemmon et al. 2012), and target capture of ultraconserved elements (UCEs, Faircloth et al. 2012; McCormack et al. 2012), have further increased the number of regions and individuals that can be sampled in a single lane. In addition, UCEs overcome the need for a reference genome, enabling their wide application across many NMOs (though designing custom probe sets from closely related species will remain preferable in many cases (Hancock-Hanser et al. 2013)). Although a number of methodological variants have been developed and optimized (Bashiardes et al. 2005; Noonan et al. 2006; Hodges et al. 2009; Cummings et al. 2010; Mamanova et al. 2010; Hancock-Hanser et al. 2013), TSC generally relies on hybridization and amplification of specially prepared libraries consisting of fragmented genomic DNA. Many companies offer kits for TSC, such as Agilent (SureSelect) and MYcroarray (MYbaits), with MYcroarray specifically marketing their kits for use with NMOs.

The most common use of TSC has been the capture of whole exomes in model organisms, including humans (Ng et al. 2009). However, as costs have plummeted, TSC is increasingly being used in investigations of NMOs. TSC is particularly useful in sequencing ancient DNA, where it can enrich the sample for endogenous DNA content relative to exogenous DNA (i.e., contamination) and thereby increase the relative DNA yield (Ávila-Arcos et al. 2011; Enk et al. 2014). For example, TSC has been used to generate mitogenome sequences from subfossil killer whale specimens originating from the mid-Holocene for comparison with modern lineages (Foote et al. 2013). TSC was also recently utilized to compare >30 kb of exonic sequence from museum specimens of the extinct Steller's sea cow (*Hydrodamalis gigas*) and a modern dugong (*Dugong dugon*) specimen to investigate evolution within Sirenia (Springer et al. 2015). Springer et al. (2016) further used TSC to examine gene evolution related to dentition across edentulous mammals, including mysticetes. Finally, TSC of both exonic and intronic regions has been used to assess genetic divergence across cetacean species (Hancock-Hanser et al. 2013; Morin et al. 2015). These studies show the potential use of TSC across evolutionary timescales for population genomics, phylogenomics, and studies of selection and gene loss across divergent lineages (Table S1).

Whole genome sequencing

Beyond advances enabled by the reduced-representation methods presented above, our power and resolution to elucidate evolutionary processes, including selection and demographic shifts, can be further increased by sequencing whole genomes.

i. Reference genome sequencing

At the time of publication, there are 12 publicly available[1] whole (or near-whole) marine mammal genomes of varying quality representing 10 families, including 7 cetaceans (Fig 1A), 3 pinnipeds (Fig 1B), the West Indian manatee (*Trichechus manatus*), and the polar bear. The first sequenced marine mammal genome was that of the common bottlenose dolphin, which was originally sequenced to ~2.5x depth of coverage using Sanger sequencing (Lindblad-Toh et al. 2011). This genome was later improved upon by adding both 454 and Illumina HiSeq data (Foote et al. 2015). Other subsequent marine mammal genomes were produced solely using Illumina sequencing and mate-paired or paired-end libraries with varied insert sizes (Miller et al. 2012; Zhou et al. 2013; Yim et al. 2014; Foote et al. 2015; Keane et al. 2015; Kishida et al. 2015; Humble et al. 2016).

Whole genome sequencing has been used to address many issues in marine mammal genome evolution, usually by comparison with other existing mammalian genomes. Biological insights discussed in the genome papers listed above include the evolution of transposons and repeat elements, gene evolution and positive selection, predicted population structure through time, SNP validation, molecular clock rates, and convergent molecular evolution (Table S1). For example, analyses of the Yangtze river dolphin (*Lipotes vexillifer*) genome confirmed that a bottleneck occurred in this species during the last period of deglaciation (Zhou et al. 2013). In addition, following upon earlier smaller-scale studies (e.g., Deméré et al. 2008; McGowen et al. 2008; Hayden et al. 2010), genomic analyses have confirmed the widespread decay of gene families involved in olfaction, gustation, enamelogenesis, and hair growth in some cetaceans (Yim et al. 2014; Kishida et al. 2015). Perhaps the most widespread use of whole genome studies

[1] These genomes are available on NCBI's online genome database or Dryad, but they have not all been published. As agreed upon in the Fort Lauderdale Convention, the community standard regarding such unpublished genomic resources is to respect the data generators' right to publish with these data first.

has been the use of models of selection to detect protein-coding genes that show evidence of natural selection in specific lineages. A recent study by Foote et al. (2015) extended this approach to investigate convergent positive selection among cetaceans, pinnipeds, and sirenians. This study exemplifies a trend in recent genomic studies that sequence multiple genomes to address a predetermined evolutionary question, in this case, the molecular signature of aquatic adaptation.

In addition to these evolutionary insights that typically stem from a comparative genomics approach, the development of high-quality reference genome assemblies provide an important resource that facilitates mapping of reduced-representation genomic data (see previous section) as well as short-read sequencing data with relatively low depth of coverage (see following section). These data types can be generated at relatively low cost on larger sample sizes enabling population-scale genomic studies. In many cases, genome assemblies from closely related species are sufficient for use as a reference. Particularly among marine mammals, given their generally slow rate of nucleotide divergence, it is therefore likely unnecessary to sequence a high-quality reference genome assembly for every species. Instead, resources could be allocated toward population-scale studies, including genome re-sequencing efforts.

ii. Population-level genome re-sequencing

In contrast to reference genome sequencing that today often exceeds 100x mean read depth and typically combines long- and short-insert libraries to generate high-quality assemblies for one to a few individuals, genome re-sequencing studies aim to achieve only ≥2x mean read depth on tens to hundreds of individuals from short-insert libraries whose reads are anchored to existing reference assemblies. Despite the inherent trade-offs between cost, read depth, coverage, and sample size, genome re-sequencing of large numbers of individuals for population-level inference can be conducted at a relatively low cost. In the past five years, several influential studies have used genome re-sequencing to advance our understanding of the genomic underpinnings of different biological questions in model systems. For example, population genomics of *Heliconius* butterflies highlighted the exchange of genes between species that exhibit convergent wing patterns (The *Heliconius* Genome Consortium 2012); whole genome re-sequencing of three-spined sticklebacks highlighted the re-use of alleles in replicated divergences associated with ecological speciation and local adaptation (Jones et al. 2012); and combined population genomics and phylogenomics have identified regions of the genome associated with variation in beak shape and size in Darwin's finches (Lamichhaney et al. 2015).

To date only two marine mammal population genomics studies using whole genome re-sequencing have been published. These studies involved re-sequencing the genomes of 79 individuals from three populations of polar bears (Liu et al. 2014a) and 48 individuals from five evolutionarily divergent ecotypes of killer whale (Foote et al. 2016). The findings of Foote et al. (2016) confirmed results of population differentiation that had previously been established using traditional genetic markers (Morin et al. 2010a). However, the study also provided new insights into the demographic history, patterns of selection associated with ecological niche, and evidence of episodic ancestral admixture that could not have been obtained using traditional markers.

Several new resources have made such population genomic studies economically possible for a greater number of NMOs, including the availability of reference genome assemblies (see section above), relatively low-cost high-throughput sequencing (further increases in throughput expected with the new Illumina HiSeq X Ten (van Dijk et al. 2014)), and crucially, the development of likelihood-based methods that allow estimation of population genetic metrics from re-sequencing data (Fumagalli et al. 2013; O'Rawe et al. 2015). One last consideration is the ease of laboratory methods necessary to generate whole genome re-sequencing data when compared to other methods such as RADseq or TSC. DNA simply needs to be extracted from the samples and, using proprietary kits, built into individually index-amplified libraries that are equimolarly pooled and submitted for sequencing.

Many population genomic analyses are based on the coalescent model that gains most information from the number of independent genetic markers, not the number of individuals sampled. Sample sizes of ~10 individuals are usually considered sufficient (Robinson et al. 2014) and have been standard in many genome-wide studies in the eco-evolutionary sciences (Ellegren et al. 2012; Jones et al. 2012). Thus, sampling fewer individuals by whole genome re-sequencing is a salient approach that allows us to consider many more gene trees, whilst continuing to provide robust estimates of per-site genetic metrics (e.g., F_{ST}). The robustness of inference from data with low mean read depth across the genome was recently confirmed using a comparison of per-site F_{ST} estimates for the same sites from high-depth (≥20x) RADseq data and low-depth (≈2x) whole genome re-sequencing data in pairwise comparisons between the same two killer whale ecotypes (Foote et al. 2016).

Beyond the increased power afforded by sequencing more polymorphic sites, whole genome re-sequencing also allows inference of demographic history from the genome of even just a single individual by identifying Identical By Descent (IBD) segments and runs of homozygosity (Li and Durbin 2011; Harris and Nielsen 2013). For example, Liu et al. (2014a) found evidence for ongoing gene flow from polar bears into brown bears after the two species initially diverged. Genome re-sequencing of sufficient numbers of individuals also facilitates haplotype phasing, which has many applications, including the detection of ongoing selective sweeps (Ferrer-Admetlla et al. 2014) and the inference of demographic history of multiple populations based on coalescence of pairs of haplotypes in different individuals (Schiffels and Durbin 2014). However, haplotype phasing typically requires genomic data with higher mean read depth (~20x) from tens of individuals (though recent advances in genotype imputation suggest success with data of lower mean read depth (VanRaden et al. 2015)). Thus far, phasing has been restricted to relatively few NMO studies, and no marine mammal studies to the best of our knowledge.

Transcriptome sequencing

In comparison with the DNA-based genomic approaches described above, RNA-based genomic approaches are a relatively new and emerging application in NMOs such as marine mammals. Transcriptomics by RNA sequencing (RNAseq) can rapidly generate vast amounts of information regarding genes and gene expression without any prior genomic resources. This approach can resolve differences in global gene expression patterns between populations, individuals, tissues, cells, and physiological or environmental conditions, and can yield insights into the molecular basis of environmental adaptation and speciation in wild animals (Wolf 2013; Alvarez et al. 2015). Furthermore, RNAseq is a valuable tool for resource development, for example as a precursor to designing SNP and TSC arrays (e.g., Hoffman et al. 2012). However, applying RNAseq to NMOs requires several unique considerations in comparison to the DNA-based methods described above. Most importantly, the labile nature of gene transcription and high detection sensitivity of RNAseq have the potential to amplify transcriptional "noise" and are thus extremely sensitive to experimental design.

If the experimental goal is to capture a comprehensive transcriptome profile for a study organism, multiple tissues from individuals of varied life history stages should be sampled. However, if the aim is to characterize transcriptional responses to physiological or environmental stimuli, efforts should focus on minimizing variability in individuals and sampling conditions (Wolf 2013). For differential expression analyses, pairwise comparisons should be made within the same individual if at all possible (e.g., before and after treatment, between two developmental stages). As RNAseq only captures a 'snapshot' of gene expression in time, repeated sampling or time-course studies are necessary to obtain a more complete picture of cellular responses to the condition(s) in question (Spies and Ciaudo 2015). Sampling and sequencing depth requirements will depend on the study design. Simulation studies have shown that a minimum of 5-6 biological replicates sequenced at a depth of 10-20 million reads per sample is necessary for differential expression analysis (Liu et al. 2014b; Schurch et al. 2015). RNAseq can also be used for biomarker development to expand molecular toolkits for NMOs without sequenced genomes (Hoffman et al. 2013). In this case, higher sequencing depths of 30-60 million reads per sample are recommended for SNP discovery and genotyping (De Wit et al. 2015).

Following sequence generation, transcript annotation remains a challenge for NMOs without reference transcriptomes or genomes. *De novo* transcriptomes can be annotated through detection of assembled orthologs of highly conserved proteins, but these analyses remain limited by the quality of reference databases. As a result, NMO transcriptomes are biased in favor of highly conserved terrestrial mammal genes and therefore provide an incomplete understanding of animal adaptations to natural environments (Evans 2015). For example, while 70.0% of northern elephant seal (*Mirounga angustirostris*) skeletal muscle transcripts had BLASTx hits to mouse genes, only 54.1% of blubber transcripts could be annotated due to poor representation of this tissue in terrestrial mammal reference proteomes (Khudyakov et al. 2015b).

To date, RNAseq has been used for gene discovery and phylogenomics analyses in Antarctic fur seal (Hoffman 2011; Hoffman et al. 2013), polar bear (Miller et al. 2012), Indo-Pacific humpback dolphin (*Sousa chinensis* (Gui et al. 2013)), spotted seal (*Phoca largha* (Gao et al. 2013)), bowhead whale (*Balaena mysticetus* (Seim et al. 2014)), narrow-ridged finless porpoise (*Neophocaena asiaeorientalis* (Ruan et al. 2015)), and humpback whale (*Megaptera novaeangliae* (Tsagkogeorga et al. 2015)) (Table S1). Due to the challenges of repeated sampling of wild marine mammals, few studies have examined cetacean or pinniped transcriptome responses to environmental or experimental stimuli. The majority of such functional gene expression studies have used microarrays (Mancia et al. 2008; Mancia et al. 2012; Mancia et al. 2015); however, RNAseq has been employed to profile sperm whale (*Physeter macrocephalus*) skin cell response to hexavalent chromium (Pabuwal et al. 2013) and free-ranging northern elephant seal skeletal muscle response to an acute stress challenge (Khudyakov et al. 2015a; Khudyakov et al. 2015b). With decreasing sequencing costs and improvements in bioinformatics tools, RNAseq has the potential to accelerate molecular discoveries in marine mammal study systems and supplement existing functional genomics approaches.

Emerging techniques

In addition to the relatively proven NMO genomic data generation techniques described above, a suite of emerging techniques is entering the field, with exciting promise for exploration of existing and new research areas. For example, high-throughput shotgun sequencing is increasingly being used to identify genetic material from multiple species in a single sample (metagenomics and metatranscriptomics), rather than focus on characterizing variation in a single target individual. These multi-species approaches can be used, for example, to characterize diet from fecal samples (Deagle et al. 2009) and to investigate microbiomes (Nelson et al. 2015), objectives with implications for improving our understanding of both basic ecology and health in natural populations of NMOs. Furthermore, high-throughput sequencing of environmental DNA dramatically increases the throughput of NMO detection in environmental (e.g., seawater) samples (Thomsen et al. 2012), using degenerate primers for multi-species detection rather than requiring the design and implementation of numerous single-species protocols (Foote et al. 2012).

A second broad area of emerging interest moves beyond the study of variation at the DNA and RNA levels to examine epigenetic effects of histone modification on gene regulation and evolution. Epigenomic studies often examine changes in DNA methylation in association with processes such as cancer and ageing. Such approaches, from targeted gene to genome-wide, have only very recently and not yet frequently been applied in NMOs. Polanowski et al. (2014) used a targeted gene approach to examine changes in DNA methylation in age-associated genes, previously identified in humans and mice, in humpback whales of known age. The most informative markers were able to estimate humpback whale ages with standard deviations of approximately 3-5 years, demonstrating the potential transferability of these approaches from model to non-model organism. Villar et al. (2015) utilized a genome-wide approach – chromatin immunoprecipitation followed by high-throughput sequencing (ChIPseq) – to examine gene regulatory element evolution across mammals, including four species of cetaceans. This study identified highly conserved gene regulatory elements based on their histone modifications (H3K27ac and H3K4me3), showed that recently evolved enhancers were associated with genes under positive selection in marine mammals, and identified unique *Delphinus*-specific enhancers. Finally, reduced-representation epigenomic approaches have also been developed (Gu et al. 2011), and although they have not yet been used in marine mammals to our knowledge, these techniques could facilitate future studies of how changes in DNA methylation patterns affect other biological processes, such as stress levels or pregnancy.

Data analysis

Following the generation of genomic data, researchers must select the most appropriate genomic analysis (i.e., bioinformatics) pipelines, which often differ significantly from those used in traditional genetic studies of NMOs. The choice of analysis pipeline will depend on multiple factors including the availability of a reference genome, the level of diversity within the dataset (e.g., single- or multi-species), the type of data generated (e.g., single- or paired-end), and the computing resources available. The computational needs, both in terms of hardware and competency in computer science, for analysis of genomic data typically far exceed those necessary for traditional genetic markers. On the smaller end of the spectrum, one lane of 50 bp single-end sequencing on an Illumina HiSeq 2500 can produce tens of gigabytes of data, while data files associated with a single high-quality vertebrate genome may reach hundreds of gigabytes in size (Ekblom and Wolf 2014). Computing resources necessary for the analysis of these genomic datasets can range from ~10 gigabytes for a pilot study using a reduced-representation sequencing approach to over a terabyte for whole genome sequence assembly (Ekblom and Wolf 2014). Fortunately, university computing clusters, cloud-based (Stein 2010) and high-performance computing clusters (e.g., XSEDE; Towns et al. 2014), and open web-based platforms for genomic research (e.g., Galaxy; Goecks et al. 2010) are becoming increasingly accessible. Furthermore, new pipelines are continuously being developed and improved, and there are a growing number of resources aimed at training molecular ecologists and evolutionary biologists in computational large-scale data analysis (Andrews and Luikart 2014; Belcaid and Toonen 2015; Benestan et al. 2016). We provide an indicative list of the current, most commonly used analysis pipelines that are specific to each data generation method in Table 1. Here we briefly summarize current genomic data analysis pipelines and discuss considerations that are likely to be similar across multiple data generation methods.

Genomic data analysis often involves multiple steps, and the choice of analysis tool for each step can greatly affect the outcome, with different tools producing different (though usually overlapping) sets of results (e.g., Schurch et al. 2015). All analyses begin by evaluating data quality, trimming sequences if necessary to remove erroneous nucleotides (MacManes 2014), and implementing appropriate data quality filters (e.g., phred scores, read length, and/or read depth). Raw reads also need to be demultiplexed based on unique barcodes if pools of individuals were sequenced in a single lane. Analyses then usually proceed in a *de novo* or genome-enabled manner, depending on available resources. Briefly, sequences can be compared (e.g., to identify variants) by mapping all reads to a reference genome or *de novo* assembling stacks of sequences putatively derived from the same locus based on sequence similarity. *De novo* methods are sensitive to sequencing error, as well as true genetic variation, and therefore can erroneously assemble polymorphic sequences as separate loci or transcripts, requiring further filtering to remove redundancy. The opposite problem can also occur in both *de novo* and reference mapping approaches, where two distinct loci (e.g., paralogous loci) may assemble as a single locus or map to the same reference location. Researchers should therefore recognize the inherent trade-offs when carefully selecting their thresholds for acceptable levels of variation within and among loci.

Considerations relevant to the selection of subsequent downstream analyses are specific to the type of data generated and the research objective. For example, RADseq analysis pipelines differ in the algorithms used to genotype variants (Table 1). Similarly, there are several gene expression analysis pipelines for RNAseq data that compare transcript abundance between samples (Table 1). Analysis of TSC data usually uses standard *de novo* assemblers (e.g., Trinity, Velvet); these assemblers can be run using packages such as PHYLUCE (Faircloth 2015), which is designed specifically for use with ultraconserved elements. Unfortunately, for most analyses, there are no unifying recommendations currently available and researchers must evaluate several approaches, each with their own advantages and disadvantages, in order to select the most appropriate tool for their particular experiment and system. Furthermore, we can expect that the recommendations for analysis tools will continue to evolve as new programs become available in the future.

Guidelines for data quality control and sharing

With rapid growth in sequencing platforms and bioinformatics analysis pipelines comes the need to extend existing principles (e.g., Bonin et al. 2004) on quality control, analysis, and transparency. General recommendations for sample and data handling, library preparation, and sequencing have been discussed elsewhere (Paszkiewicz et al. 2014). We therefore focus on the need to produce guidelines on data quality evaluation and reporting for genomic data (e.g., Morin et al. 2010b). A primary challenge in this area is that quality metrics vary widely across sequencing technologies. Yet, regardless of sequencing platform, the quality of sequencing reads must be evaluated (e.g., using FastQC; Andrews 2010) and reported.

Best practices guidelines for reference genome sequencing and RNAseq data generation, analysis, and reporting are available from the human-centric ENCODE consortium (www.encodeproject.org). These include minimum depth of sequencing and number and reproducibility of biological replicates. For RNAseq experiments, evaluation of *de novo* assembly quality remains a challenge. Suggested quality metrics include percentage of raw reads mapping back to the assembly and number of assembled transcripts with homology to known proteins (MacManes 2016). Emerging tools such as Transrate (Smith-Unna et al. 2015) attempt to integrate these and other metrics into a comprehensive assembly quality score.

In contrast, there is not yet any standard way to estimate or report error rates with RADseq or genome re-sequencing methods (but see Mastretta-Yanes et al. 2015; Fountain et al. 2016). Recommendations to improve confidence in genotyping include using methods that account for population-level allele frequencies when calling individual genotypes, mapping reads to reference genomes rather than *de novo* assembly (Nadeau et al. 2014; Fountain et al. 2016), filtering out PCR duplicates (Andrews et al. 2014), identifying and removing markers in possible repeat regions, and filtering data to include only those with high read depth (>10-20x per locus per individual) (Nielsen et al. 2011). Other analysis methods, such as robust Bayesian methods and likelihood-based approaches that account for read quality in calculations of posterior probabilities of genotypes and per-site allele frequencies utilizing the sample mean site frequency spectrum as a prior (Fumagalli et al. 2013), can account for uncertainty and/or error in the data, and are therefore suitable for use with low to moderate read depths (2-20x per locus; e.g., Han et al. 2015; O'Rawe et al. 2015).

Due to the large number of analysis tools that are available, data quality and reproducibility ultimately depend on methods and data transparency. All raw sequencing reads should be publicly archived, for example deposited in the NCBI Sequence Read Archive. Many journals, including the *Journal of Heredity* (Baker 2013), now also require that primary data supporting the published results and conclusions (e.g., SNP genotypes, assemblies) be publicly archived in online data repositories (e.g., Dryad). We further recommend making public the analysis pipelines, scripts (e.g., using GitHub), and additional outputs, as appropriate, in order for analyses to be fully reproducible and transparent, which is the cornerstone of the scientific method (Nosek et al. 2015).

Future directions

As demonstrated here for one group of mammalian taxa, the rapid growth of the field of non-model genomics has been both impressive and empowering. As we approach a point of relative saturation in reference genomes, we anticipate an increase in population-scale genomic studies that produce lower depth or coverage datasets per individual but across larger sample sizes. In addition (or alternatively), we hope to see increasing efforts to sequence reference transcriptomes and improve NMO genome annotation in ways beyond the inherently limited approach of comparison to gene lists from a few model organisms. Population-scale genomic studies will facilitate greater ecological understanding of natural populations, while efforts to improve annotation will address persistent limitations in our understanding of gene function for NMOs. Ultimately, improving our understanding of local adaptation, adaptive potential, and demographic history through the use of genomic toolkits such as those described here is likely to have important implications for the future conservation of these populations.

Advances in sequencing technologies and analytical tools will no doubt continue, in some cases drawing on established techniques in model organisms, posing both new opportunities and new challenges for researchers in NMO genomics. Likely the most persistent challenge will remain selecting the data generation and experimental design that is most appropriate for the respective research objective. Our review identified few cases that exhibit relative dominance of a single methodology and analytical pipeline (e.g., RADseq and STACKS, RNAseq and Trinity); rather, more often we found a diversity of approaches even within each category of data generation. In fact, such diversity of approaches has its benefits, with each approach promoting its own advantages (and limitations). Overall, our reflections on lessons learned from the past decade of NMO genomics in one well-studied group of mammalian taxa clearly demonstrate the value, increased ease, and future promise of applying genomic techniques across a wide range of non-model species to gain previously unavailable insights into evolution, population biology, and physiology on a genome-wide scale.

Acknowledgements

This review paper is the outcome of two international workshops held in 2013 and 2015 on marine mammal genomics. The workshops were organized by KMC, AF, and C. Scott Baker and hosted by the Society for Marine Mammalogy, with support from a Special Event Award from the American Genetic Association. We sincerely thank all the workshop participants for their contributions to inspiring discussions on marine mammal genomics. We would also like to thank two anonymous reviewers and C. Scott Baker for their helpful feedback on an earlier version of this manuscript. Illustrations are by C. Buell with permission for use granted by J. Gatesy.

Funding

The authors involved in this work were supported by a National Science Foundation Postdoctoral Research Fellowship in Biology (Grant No. 1523568) to KMC; an Office of Naval Research Award (No. N00014-15-1-2773) to JIK; a Marie Slodowska Curie Fellowship to ELC (Behaviour-Connect) funded by the EU Horizon2020 program; Royal Society Newton International Fellowships to ELC and MRM; a Deutsche Forschungsgemeinschaft studentship to EH; a Fyssen Foundation postdoctoral fellowship to ML; postdoctoral funding from the University of Idaho College of Natural Resources to KRA; a short visit grant from the European Science Foundation-Research Networking Programme ConGenOmics to ADF; and a Swiss National Science Foundation grant (31003A-143393) to L. Excoffier that further supported ADF. The first marine mammal genomics workshop we held to begin discussions towards this review was supported by a Special Event Award from the American Genetic Association.

References

Albrechtsen A, Nielsen FC, Nielsen R. 2010. Ascertainment biases in SNP chips affect measures of population divergence. *Mol Biol Evol*. 27:2534-2547.

Alexander A, Steel D, Hoekzema K, Mesnick S, Engelhaupt D, Kerr I, Payne R, Baker CS. 2016. What influences the worldwide genetic structure of sperm whales (*Physeter macrocephalus*)? *Mol Ecol*.

Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlstrom T, Vinner L*, et al.* 2015. Population genomics of Bronze Age Eurasia. *Nature*. 522:167-172.

Alvarez M, Schrey AW, Richards CL. 2015. Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? *Mol Ecol*. 24:710-725.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol*. 11:R106.

Andrews K, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet*. 17:81-92.

Andrews KR, Hohenlohe PA, Miller MR, Hand BK, Seeb JE, Luikart G. 2014. Trade-offs and utility of alternative RADseq methods: Reply to Puritz *et al*. 2014. *Mol Ecol*. 23:5943-5946.

Andrews KR, Luikart G. 2014. Recent novel approaches for population genomics data analysis. *Mol Ecol*. 23:1661-1667.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Ankeny RA, Leonelli S. 2011. What's so special about model organisms? *Studies in History and Philosophy of Science*. 42:313-323.

Armengaud J, Trapp J, Pible O, Geffard O, Chaumot A, Hartmann EM. 2014. Non-model organisms, a species endangered by proteogenomics. *J Proteomics*. 105:5-18.

Arnason U, Adegoke JA, Bodin K, Born EW, Esa YB, Gullberg A, Nilsson M, Short RV, Xu X, Janke A. 2002. Mammalian mitogenomic relationships and the root of the eutherian tree. *Proc Natl Acad Sci USA*. 99:8151-8156.

Arnason U, Gullberg A, Widegren B. 1991. The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. *J Mol Evol*. 33:556-568.

Ávila-Arcos M, Cappellini E, Romero-Navarro JA, Wales N, Moreno-Mayar JV, Rasmussen M, Fordyce SL, Montiel R, Vielle-Calzada J-P, Willerslev E*, et al.* 2011. Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Sci Rep*. 1:74.

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:e3376.

Baker CS. 2013. *Journal of Heredity* adopts Joint Data Archiving Policy. *J Hered*. 104:1.

Barrett RDH, Rogers SM, Schluter D. 2008. Natural selection on a major armor gene in threespine stickleback. *Science*. 322:255-257.

Bashiardes S, Veile R, Helms C, Mardis ER, Bowcock AM, Lovett M. 2005. Direct genomic selection. *Nat Methods*. 2:63-69.

Belcaid M, Toonen RJ. 2015. Demystifying computer science for molecular ecologists. *Mol Ecol*. 24:2619-2640.

Benestan LM, Ferchaud A-L, Hohenlohe PA, Garner BA, Naylor GJP, Baums IB, Schwartz MK, Kelley JL, Luikart G. 2016. Conservation genomics of natural and managed populations: building a conceptual and practical framework. *Mol Ecol*.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina Sequence Data. *Bioinformatics*. 30:2114-2120.

Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P. 2004. How to track and assess genotyping errors in population genetics studies. *Mol Ecol*. 13:3261-3273.

Brown CT, Howe A, Zhang Q, Pyrkosz AB, Brom TH. 2012. A reference-free algorithm for computational normalization of shotgun sequencing data. *arXive*. 1203:4802.

Cammen KM, Schultz TF, Rosel PE, Wells RS, Read AJ. 2015. Genomewide investigation of adaptation to harmful algal blooms in common bottlenose dolphins (*Tursiops truncatus*). *Mol Ecol*. 24:4697-4710.

Campbell NR, Harmon SA, Narum SR. 2015. Genotyping-in-Thousands by sequencing (GT-seq): a cost effective SNP genotyping method based on custom amplicon sequencing. *Mol Ecol Resour*. 15:855-867.

Carroll EL, Baker CS, Watson M, Alderman R, Bannister J, Gaggiotti OE, Gröcke DR, Patenaude N, Harcourt R. 2015. Cultural traditions across a migratory network shape the genetic structure of southern right whales around Australia and New Zealand. *Sci Rep*. 5:16182.

Catchen JM, Amores A, Hohenlohe PA, Cresko WA, Postlethwait JH. 2011. *Stacks*: building and genotyping loci *de novo* from short-read sequences. *G3*. 1:171-182.

Catchen JM, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool set for population genomics. *Mol Ecol*. 22:3124-2140.

Chancerel E, Lepoittevin C, Le Provost G, Lin Y-C, Jaramillo-Correa JP, Eckert AJ, Wegrzyn JL, Zelenika D, Boland A, Frigerio J-M*, et al.* 2011. Development and implementation of a highly-multiplexed SNP array for genetic mapping in maritime pine and comparative mapping with loblolly pine. *BMC Genomics*. 12:368.

Chen H, Patterson N, Reich D. 2010. Population differentiation as a test for selective sweeps. *Genome Res*. 20:393-402.

Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*. 307:1928-1933.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 21:3674-3676.

Corander J, Majander KK, Cheng L, Merilä J. 2013. High degree of cryptic population differentiation in the Baltic Sea herring *Clupea harengus*. *Mol Ecol*. 22:2931-2940.

Cummings N, King R, Rickers A, Kaspi A, Lunke S, Haviv I, Jowett JBM. 2010. Combining target enrichment with barcode multiplexing for high throughput SNP discovery. *BMC Genomics*. 11:641.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet*. 12:499-510.

De Mita S, Thuillet A-C, 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.

De Wit P, Pespeni MH, Palumbi SR. 2015. SNP genotyping and population genomics from expressed sequences - current advances and future possibilities. *Mol Ecol*. 24:2310-2323.

Deagle BE, Kirkwood R, Jarman SN. 2009. Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Mol Ecol*. 18:2022-2038.

Deméré TA, McGowen MR, Berta A, Gatesy J. 2008. Morphological and molecular evidence for a stepwise evolutionary transition from teeth to baleen in mysticete whales. *Syst Biol*. 57:15-37.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M*, et al.* 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 43:491-498.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 29:15-21.

Eaton DAR. 2014. PyRAD: assembly of *de novo* RADseq loci for phylogenetic analysis. *Bioinformatics*. 30:1844-1849.

Ekblom R, Galindo J. 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*. 107:1-15.

Ekblom R, Wolf JBW. 2014. A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*. 7:1026-1042.

Ellegren H. 2014. Genome sequencing and population genomics in non-model organisms. *Trends Ecol Evol*. 29:51-63.

Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, Künstner A, Mäkinen H, Nadachowska-Brzyska K, Qvarnström A*, et al.* 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*. 491:756-760.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*. 6:e19379.

Enk J, Devault A, Kuch M, Murgha Y, Rouillard J-M, Poinar H. 2014. Ancient whole genome enrichment using baits built from modern DNA. *Mol Biol Evol*. 31:1292-1294.

Evans TG. 2015. Considerations for the use of transcriptomics in identifying the 'genes that matter' for environmental adaptation. *J Exp Biol*. 218:1925-1935.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genetics*. 9:e1003905.

Faircloth BC. 2015. PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*. 32:786-788.

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.

Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. 2014. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol Biol Evol*. 31:1275-1291.

Flicek P, Birney E. 2009. Sense from sequence reads: methods for alignment and assembly. *Nat Methods*. 6:S6-S12.

Foote AD, Liu Y, Thomas GWC, Vinař Ts, Alföldi J, Deng J, Dugan S, van Elk CE, Hunter ME, Joshi V*, et al.* 2015. Convergent evolution of the genomes of marine mammals. *Nat Genet*. 47:272-275.

Foote AD, Newton J, Ávila-Arcos MC, Kampmann M-L, Samaniego JA, Post K, Rosing-Asvid A, Sinding M-HS, Gilbert MTP. 2013. Tracking niche variation over millennial timescales in sympatric killer whale lineages. *Proc R Soc Lond B Biol Sci*. 280:20131481.

Foote AD, Thomsen PF, Sveegaard S, Wahlberg M, Kielgast J, Kyhn LA, Salling AB, Galatius A, Orlando L, Gilbert MTP. 2012. Investigating the potential use of environmental DNA (eDNA) for genetic monitoring of marine mammals. *PLoS One*. 7:e41781.

Foote AD, Vijay N, Ávila-Arcos M, Baird RW, Durban JW, Fumagalli M, Gibbs RA, Hanson MB, Korneliussen TS, Martin MD*, et al.* 2016. Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nat Commun*. 7:11693.

Fountain ED, Pauli JN, Reid BN, Palsbøll PJ, Peery MZ. 2016. Finding the right coverage: the impact of coverage and sequence quality on single nucleotide polymorphism genotyping error rates. *Mol Ecol Resour*.

Fumagalli M, Vieira FG, Korneliussen TS, Linderoth T, Huerta-Sánchez E, Albrechtsen A, Nielsen R. 2013. Quantifying population genetic differentiation from next-generation sequencing data. *Genetics*. 195:979-992.

Fumagalli M, Vieira FG, Linderoth T, Nielsen R. 2014. *ngsTools*: methods for population genetics analyses from Next-Generation Sequencing data. *Bioinformatics*. 30:1486-1487.

Gao X, Han J, Lu Z, Li Y, He C. 2013. *De novo* assembly and characterization of spotted seal *Phoca largha* transcriptome using Illumina paired-end sequencing. *Comp Biochem Physiol D Genom Proteom*. 8:103-110.

Garner BA, Hand BK, Amish SJ, Bernatchez L, Foster JT, Miller KM, Morin PA, Narum SR, O'Brien SJ, Roffler G*, et al.* 2016. Genomics in conservation: case studies and bridging the gap between data and application. *Trends Ecol Evol*. 31:81-83.

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES. 2014. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS One*. 9:e90346.

Gnerre S, MacCallum I, Przbylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S*, et al.* 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci USA*. 108:1513-1518.

Goecks J, Nekrutenko A, Taylor J, The Galaxy Team. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol*. 11:R86.

Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A. 2011. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc*. 6:468-481.

Gui D, Jia K, Xia J, Yang L, Chen J, Wu Y, Yi M. 2013. *De novo* assembly of the Indo-Pacific humpback dolphin leucocyte transcriptome to identify putative genes involved in the aquatic adaptation and immune response. *PLoS One*. 8:e72417.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*. 5:e1000695.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M*, et al.* 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 8:1494-1512.

Han E, Sinsheimer JS, Novembre J. 2015. Fast and accurate site frequency spectrum estimation from low coverage sequence data. *Bioinformatics*. 31:720-727.

Hancock-Hanser BL, Frey A, Leslie MS, Dutton PH, Archer FI, Morin PA. 2013. Targeted multiplex next-generation sequencing: advances in techniques of mitochondrial and nuclear DNA sequencing for population genomics. *Mol Ecol Resour*. 13:254-268.

Harris K, Nielsen R. 2013. Inferring demographic history from a spectrum of shared haplotype lengths. *PLoS Genetics*. 9:e1003521.

Hayden S, Bekaert M, Crider TA, Mariani S, Murphy WJ, Teeling EC. 2010. Ecological adaptation determines functional mammalian olfactory subgenomes. *Genome Res*. 20:1-9.

Hedrick PW. 2000 *Genetics of Populations*. Jones and Bartlett Publishers, Sudbury, MA.

Helyar SJ, Hemmer-Hansen J, Bekkevold D, Taylor MI, Ogden R, Limborg MT, Cariani A, Maes GE, Diopere E, Carvalho GR*, et al.* 2011. Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Mol Ecol Resour*. 11:123-136.

Higdon JW, Bininda-Emonds ORP, Beck RMD, Ferguson SH. 2007. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. *BMC Evol Biol*. 7:216.

Hodges E, Rooks M, Xuan Z, Bhattacharjee A, Gordon DB, Brizuela L, McCombie WR, Hannon GJ. 2009. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nat Protoc*. 4:960-974.

Hoffman JI. 2011. Gene discovery in the Antarctic fur seal (*Arctocephalus gazella*) skin transcriptome. *Mol Ecol Resour*. 11:703-710.

Hoffman JI, Nicholas HJ. 2011. A novel approach for mining polymorphic microsatellite markers *in silico*. *PLoS One*. 6:e23283.

Hoffman JI, Simpson F, David P, Rijks JM, Kuiken T, Thorne MAS, Lacy RC, Dasmahapatra KK. 2014. High-throughput sequencing reveals inbreeding depression in a natural population. *Proc Natl Acad Sci USA*. 111:3775-3780.

Hoffman JI, Thorne MAS, Trathan PN, Forcada J. 2013. Transcriptome of the dead: characterisation of immune genes and marker development from necropsy samples in a free-ranging marine mammal. *BMC Genomics*. 14:52.

Hoffman JI, Tucker R, Bridgett SJ, Clark MS, Forcada J, Slate J. 2012. Rates of assay success and genotyping error when single nucleotide polymorphism genotyping in non-model organisms: a case study in the Antarctic fur seal. *Mol Ecol Resour*. 12:861-872.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet*. 6:e1000862.

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

Humble E, Martinez-Barrio A, Forcada J, Trathan PN, Thorne MAS, Hoffmann M, Wolf JBW, Hoffman JI. 2016. A draft fur seal genome provides insights into factors affecting SNP validation and how to mitigate them. *Mol Ecol Resour*.

Jackson JA, Baker CS, Vant M, Steel DJ, Medrano-González L, Palumbi SR. 2009. Big and slow: phylogenetic estimates of molecular evolution in baleen whales (suborder Mysticeti). *Mol Biol Evol*. 26:2427-2440.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S*, et al.* 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*. 484:55-61.

Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, Maruyama H*, et al.* 2014. Efficient *de novo* assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res*. 24:1384-1395.

Keane M, Semeiks J, Webb AE, Li YI, Quesada V, Craig T, Madsen LB, van Dam S, Brawand D, Marques PI*, et al.* 2015. Insights into the evolution of longevity from the bowhead whale genome. *Cell Reports*. 10:112-122.

Khudyakov JI, Champagne CD, Preeyanon L, Ortiz RM, Crocker DE. 2015a. Muscle transcriptome response to ACTH administration in a free-ranging marine mammal. *Physiol Genomics*. 47:318-330.

Khudyakov JI, Preeyanon L, Champagne CD, Ortiz RM, Crocker DE. 2015b. Transcriptome analysis of northern elephant seal (*Mirounga angustirostris*) muscle tissue provides a novel molecular resource and physiological insights. *BMC Genomics*. 16:64.

Kishida T, Thewissen JGM, Hayakawa T, Imai H, Agata K. 2015. Aquatic adaptation and the evolution of smell and taste in whales. *Zoolog Lett*. 1:9.

Koepfli K-P, Paten B, Genome 10K Community of Scientists, O'Brien SJ. 2015. The Genome 10K Project: a way forward. *Annu Rev Anim Biosci*. 3:57-111.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*. 15:356.

Künstner A, Wolf JBW, Backström N, Whitney O, Balakrishnan CN, Day L, Edwards SV, Janes DE, Schlinger BA, Wilson RK*, et al.* 2010. Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 bird species. *Mol Ecol*. 19:266-276.

Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerová M, Rubin C-J, Wang C, Zamani N*, et al.* 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*. 518:371-375.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 10:R25.

Lemmon AR, Emme SA, Lemmon EM. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst Biol*. 61:727-744.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 12:323.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25:1754-1760.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature*. 475:493-496.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map form and SAMtools. *Bioinformatics*. 25:2078-2079.

Li S, Jakobsson M. 2012. Estimating demographic paramaters from large-scale population genomic data using Approximate Bayesian Computation. *BMC Genet*. 13:22.

Li Y, Hu Y, Bolund L, Wang J. 2010. State of the art *de novo* assembly of human genomes from massively parallel sequencing data. *Human Genomics* 4:271-277.

Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E*, et al.* 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*. 478:476-482.

Lindqvist C, Schuster SC, San Y, Talbot SL, Qi J, Ratan A, Tomsho LP, Kasson L, Zeyl E, Aars J*, et al.* 2010. Complete mitochondrial genome of a Pleistocene jawbown unveils the origin of polar bear. *Proc Natl Acad Sci USA*. 107:5053-5057.

Liu S, Lorenzen ED, Fumagalli M, Li B, Harris K, Xiong Z, Zhou L, Korneliussen TS, Somel M, Babbitt C*, et al.* 2014a. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell*. 157:785-794.

Liu X, Fun Y-X. 2015. Exploring population size changes using SNP frequency spectra. *Nat Genet*. 47:555-559.

Liu Y, Zhou J, White KP. 2014b. RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics*. 30:301-304.

Lotterhos KE, Whitlock MC. 2014. Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Mol Ecol*. 23:2178-2192.

Louis M, Viricel A, Lucas T, Peltier H, Alfonsi E, Berrow S, Brownlow A, Covelo P, Dabin W, Deaville R*, et al.* 2014. Habitat-driven population structure of bottlenose dolphins, *Tursiops truncatus*, in the North-east Atlantic. *Mol Ecol*. 23:857-874.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 15:550.

MacManes MD. 2014. On the optimal trimming of high-throughput mRNA sequence data. *Front Genet*. 5:13.

MacManes MD. 2016. Establishing evidence-based best practice for the *de novo* assembly and evaluation of transcriptomes from non-model organisms. *bioRxiv*. doi: http://dx.doi.org/10.1101/035642.

Magera AM, Mills Flemming JE, Kaschner K, Christensen LB, Lotze HK. 2013. Recovery trends in marine mammal populations. *PLoS One*. 8:e77908.

Malenfant RM, Coltman DW, Davis CS. 2015. Design of a 9K Illumina BeadChip for polar bears (*Ursus maritimus*) from RAD and transcriptome sequencing. *Mol Ecol Resour*. 15:587-600.

Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ. 2010. Target-enrichment strategies for next-generation sequencing. *Nat Methods*. 7:111-118.

Mancia A, Abelli L, Kucklick JR, Rowles TK, Wells RS, Balmer BC, Hohn AA, Baatz JE, Ryan JC. 2015. Microarray applications to understand the impact of exposure to environmental contaminants in wild dolphins (*Tursiops truncatus*). *Mar Genomics*. 19:47-57.

Mancia A, Lundqvist ML, Romano TA, Peden-Adams MM, Fair PA, Kindy MS, Ellis BC, Gattoni-Celli S, McKillen DJ, Trent HF*, et al.* 2007. A dolphin peripheral blood leukocyte cDNA microarray for studies of immune function and stress reactions. *Dev Comp Immunol*. 31:520-529.

Mancia A, Ryan JC, Chapman RW, Wu Q, Warr GW, Gulland FMD, Van Dolah FM. 2012. Health status, infection and disease in California sea lions (*Zalophus californianus*) studied using a canine microarray platform and machine-learning approaches. *Dev Comp Immunol*. 36:629-637.

Mancia A, Warr GW, Chapman RW. 2008. A transcriptomic analysis of the stress induced by capture-release health assessment studies in wild dolphins (*Tursiops truncatus*). *Mol Ecol*. 17:2581-2589.

Mastretta-Yanes A, Arrigo N, Alvarez N, Jorgensen TH, Piñero D, Emerson BC. 2015. Restriction site-associated DNA sequencing, genotyping error estimation and *de novo* assembly optimization for population genetic inference. *Mol Ecol Resour*. 15:28-41.

McCormack JE, Faircloth BC, Crawford NG, Gowaty PA, Brumfield RT, Glenn TC. 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Res*. 22:746-754.

McGowen MR. 2011. Toward the resolution of an explosive radiation - a multilocus phylogeny of oceanic dolphins (Delphinidae). *Mol Phylogenet Evol*. 60:345-357.

McGowen MR, Clark C, Gatesy J. 2008. The vestigial olfactory receptor subgenome of odontocete whales: phylogenetic congruence between gene-tree reconciliation and supermatrix methods. *Syst Biol*. 57:574-590.

McGowen MR, Gatesy J, Wildman DE. 2014. Molecular evolution tracks macroevolutionary transitions in Cetacea. *Trends Ecol Evol*. 29:336-346.

McGowen MR, Grossman LI, Wildman DE. 2012. Dolphin genome provides evidence for adaptive evolution of nervous system genes and a molecular rate slowdown. *Proc R Soc Lond B Biol Sci*. 279:3643-3651.

McGowen MR, Spaulding M, Gatesy J. 2009. Divergence date estimation and a comprehensive molecular tree of extant cetaceans. *Mol Phylogenet Evol*. 53:891-906.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M*, et al.* 2010. The Genome Analysis Toolkit: A

MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 20:1297-1303.

McTavish EJ, Hillis DM. 2015. How do SNP ascertainment schemes and population demographics affect inferences about population history? *BMC Genomics*. 16:266.

Meredith RW, Gatesy J, Emerling CA, York VM, Springer MS. 2013. Rod monochromacy and the coevolution of cetacean retinal opsins. *PLoS Genetics*. 9:e1003432.

Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, de Filippo C*, et al.* 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science*. 338:222-226.

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.

Miller W, Schuster SC, Welch AJ, Ratan A, Bedoya-Reina OC, Zhao F, Kim HL, Burhans RC, Drautz DI, Wittekindt NE*, et al.* 2012. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proc Natl Acad Sci USA*. 109:E2382-E2390.

Mirceta S, Signore AV, Burns JM, Cossins AR, Campbell KL, Berenbrink M. 2013. Evolution of mammalian diving capacity traced by myoglobin net surface charge. *Science*. 340:1234192.

Morin PA, Archer FI, Foote AD, Vilstrup J, Allen EE, Wade P, Durban JW, Parsons K, Pitman R, Li L*, et al.* 2010a. Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Res*. 20:908-916.

Morin PA, Luikart G, Wayne RK, SNP workshop group. 2004. SNPs in ecology, evolution and conservation. *Trends Ecol Evol*. 19:208-216.

Morin PA, Martien KK, Archer FI, Cipriano F, Steel D, Jackson J, Taylor BL. 2010b. Applied conservation genetics and the need for quality control and reporting of genetic data used in fisheries and wildlife management. *J Hered*. 101:1-10.

Morin PA, Parsons KM, Archer FI, Ávila-Arcos M, Barrett-Lennard LG, Dalla Rosa L, Duchêne S, Durban JW, Ellis GM, Ferguson SH*, et al.* 2015. Geographic and temporal dynamics of a global radiation and diversification in the killer whale. *Mol Ecol*. 24:3964-3979.

Moura AE, Kenny JG, Chaudhuri R, Hughes MA, Welch AJ, Reisinger RR, de Bruyn PJN, Dahlheim ME, Hall N, Hoelzel AR. 2014a. Population genomics of the killer whale indicates ecotype evolution in sympatry involving both selection and drift. *Mol Ecol*. 23:5179-5192.

Moura AE, Nielsen SCA, Vilstrup JT, Moreno-Mayar JV, Gilbert MTP, Gray HWI, Natoli A, Möller L, Hoelzel AR. 2013. Recent diversification of a marine genus (*Tursiops* spp.) tracks habitat preference and environmental change. *Syst Biol*. 62:865-877.

Moura AE, van Rensburg CJ, Pilot M, Tehrani A, Best PB, Thornton M, Plön S, de Bruyn PJN, Worley KC, Gibbs RA*, et al.* 2014b. Killer whale nuclear genome and mtDNA reveal widespread population bottleneck during the last glacial maximum. *Mol Biol Evol*. 31:1121-1131.

Nadeau NJ, Ruiz M, Salazar P, Counterman B, Alejandro Medina J, Ortiz-Zuazaga H, Morrison A, McMillan WO, Jiggins CD, Papa R. 2014. Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. *Genome Res*. 24:1316-1333.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA. 2013. Genotyping-by-sequencing in ecological and conservation genomics. *Mol Ecol*. 22:2841-2847.

Narum SR, Hess JE. 2011. Comparison of F_{ST} outlier tests for SNP loci under selection. *Mol Ecol Resour*. 11:184-194.

Nelson TM, Apprill A, Mann J, Rogers TL, Brown MV. 2015. The marine mammal microbiome: current knowledge and future directions. *Microbiology Australia*. 36:8-13.

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE*, et al.* 2009. Targeted capture and massively parallel sequencing of twelve human exomes. *Nature*. 461:272-276.

Nielsen R, Paul JS, Anders A, Song YS. 2011. Genotype and SNP calling from next-generation sequencing data. *Nat Rev Genet*. 12:433-451.

Noonan JP, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, Chen F, Platt D, Pääbo S, Pritchard JK*, et al.* 2006. Sequencing and analysis of Neanderthal genomic DNA. *Science*. 314:1113-1118.

Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, Buck S, Chambers CD, Chin G, Christensen G*, et al.* 2015. Promoting an open research culture: Author guidelines for journals could help to promote transparency, openness, and reproducibility. *Science*. 348:1422-1425.

O'Rawe JA, Ferson S, Lyon GJ. 2015. Accounting for uncertainty in DNA sequencing data. *Trends Genet*. 31:61-66.

Olsen MT, Volny VH, Bérubé M, Dietz R, Lydersen C, Kovacs KM, Dodd RS, Palsbøll PJ. 2011. A simple route to single-nucleotide polymorphisms in a nonmodel species: identification and characterization of SNPs in the Arctic ringed seal (*Pusa hispida hispida*). *Mol Ecol Resour*. 11:9-19.

Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I*, et al.* 2013. Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*. 499:74-78.

Pabuwal V, Boswell M, Pasquali A, Wise SS, Kumar S, Shen Y, Garcia T, Lacerte C, Wise JP, Jr., Wise JP, Sr.*, et al.* 2013. Transcriptomic analysis of cultured whale skin cells exposed to hexavalent chromium [Cr(VI)]. *Aquat Toxicol*. 134-135:74-81.

Parker J, Tsagkogeorga G, Cotton JA, Liu Y, Provero P, Stupka E, Rossiter SJ. 2013. Genome-wide signatures of convergent evolution in echolocating mammals. *Nature*. 502:228-231.

Paszkiewicz KH, Farbox A, O'Neill P, Moore K. 2014. Quality control on the frontier. *Front Genet*. 5:157.

Patro R, Duggal G, Kingsford C. 2015. Accurate, fast, and model-aware transcript expression quantification with Salmon. *bioRxiv*. doi: http://dx.doi.org/10.1101/021592.

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:e37135.

Poh Y-P, Domingues VS, Hoekstra HE, Jensen JD. 2014. On the prospect of identifying adaptive loci in recently bottlenecked populations. *PLoS One*. 9:e110579.

Poland JA, Brown PJ, Sorrells ME, Jannink J-L. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One*. 7:e32253.

Polanowski AM, Robbins J, Chandler D, Jarman SN. 2014. Epigenetic estimation of age in humpback whales. *Mol Ecol Resour*. 14:976-987.

Puritz JB, Hollenbeck CM, Gold JR. 2014. *dDocent*: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*. 2:e431.

Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R*, et al.* 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*. 463:757-762.

Riesch R, Barrett-Lennard LG, Ellis GM, Ford JKB, Deecke VB. 2012. Cultural traditions and the evolution of reprodutive isolation: ecological speciation in killer whales? *Biol J Linn Soc Lond*. 2012:1-17.

Robinson JD, Coffman AJ, Hickerson MJ, Gutenkunst RN. 2014. Sampling strategies for frequency spectrum-based population genomic inference. *BMC Evol Biol*. 14:254.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 26:139-140.

Ruan R, Guo A-H, Hao Y-J, Zheng J-S, Wang D. 2015. *De novo* assembly and characterization of narrow-ridged finless porpoise renal transcriptome and identification of candidate genes involved in osmoregulation. *Int J Mol Sci*. 16:2220-2238.

Ruegg K, Rosenbaum HC, Anderson EC, Engel M, Rothschild A, Baker CS, Palumbi SR. 2013. Long-term population size of the North Atlantic humpback whale within the context of worldwide population structure. *Cons Gen*. 14:103-114.

Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nat Genet*. 46:919-925.

Schubert M, Lindgreen S, Orlando L. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes*. 9:88.

Schurch NJ, Schofield P, Gierlinski M, Cole C, Sherstnev A, Singh V, Wrobel N, Gharbi K, Simpson GG, Owen-Hughes T*, et al.* 2015. Evaluation of tools for differential gene expression analysis by RNA-seq on a 48 biological replicate experiment. *arXive*. 1505:02017.

Seim I, Ma S, Zhou X, Gerashchenko MV, Lee SG, Suydam R, George JC, Bickham JW, Gladyshev VN. 2014. The transcriptome of the bowhead whale *Balaena mysticetus* reveals adaptations of the longest-lived mammal. *Aging*. 6:879-899.

Shafer ABA, Cullingham CI, Côté SD, Coltman DW. 2010. Of glaciers and refugia: a decade of study sheds new light on the phylogeographic patterns of northwestern North America. *Mol Ecol*. 19:4589-4621.

Shafer ABA, Davis CS, Coltman DW, Stewart REA. 2014. Microsatellite assessment of walrus (*Odobenus rosmarus rosmarus*) stocks in Canada. *NAMMCO Scientific Publications*. 9.

Shafer ABA, Gattepaille LM, Stewart REA, Wolf JBW. 2015. Demographic inferences using short-read genomic data in an approximate Bayesian computation framework: *in silico* evaluation of power, biases and proof of concept in Atlantic walrus. *Mol Ecol*. 24:328-345.

Shen Y-Y, Zhou W-P, Zhou T-C, Zeng Y-N, Li G-M, Irwin DM, Zhang Y-P. 2012. Genome-wide scan for bats and dolphin to detect their genetic basis for new locomotive styles. *PLoS One*. 7:e46455.

Smith-Unna RD, Boursnell C, Patro R, Hibberd JM, Kelly S. 2015. TransRate: reference free quality assessment of *de-novo* transcriptome assemblies. *bioRxiv*.

Spies D, Ciaudo C. 2015. Dynamics in transcriptomics: advancements in RNA-seq time course and downstream analysis. *Comput Struct Biotechnol J*. 13:469-477.

Springer MS, Signore AV, Paijmans JLA, Vélez-Juarbe J, Domning DP, Bauer CE, He K, Crerar L, Campos PF, Murphy WJ*, et al.* 2015. Interordinal gene capture, the phylogenetic position of Steller's sea cow based on molecular and morphological data, and the macroevolutionary history of Sirenia. *Mol Phylogenet Evol*. 91:178-193.

Springer MS, Starrett J, Morin PA, Lanzetti A, Hayashi C, Gatesy J. 2016. Inactivation of *C4orf26* in toothless placental mammals. *Mol Phylogenet Evol*. 95:34-45.

Sremba AL, Martin AR, Baker CS. 2015. Species identification and likely catch time preiod of whale bones from South Georgia. *Mar Mamm Sci*. 31:122-132.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Res*. 34:W435-W439.

Stein LD. 2010. The case for cloud computing in genome informatics. *Genome Biol*. 11:207.

Stinchcombe JR, Hoekstra HE. 2008. Combining population genomics and quantitative genetics: finding genes underlying ecologically important traits. *Heredity*. 100:158-170.

Tabuchi M, Veldhoen N, Dangerfield N, Jeffries S, Helbing CC, Ross PS. 2006. PCB-related alteration of thyroid hormones and thyroid hormone receptor gene expression in free-ranging harbor seals (*Phoca vitulina*). *Environ Health Perspect*. 114:1024-1031.

Taylor BL, Gemmell NJ. 2016. Emerging technologies to conserve biodiversity: further opportunities via genomics. Response to Pimm *et al. Trends Ecol Evol*. 31:171-172.

The *Heliconius* Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*. 487:94-98.

Thomsen PF, Kielgast J, Iversen LL, Møller PR, Rasmussen M, Willerslev E. 2012. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS One*. 7:e41732.

Towns J, Cockerill T, Dahan M, Foster I, Gaither K, Grimshaw A, Hazlewood V, Lathrop S, Lifka D, Peterson GD*, et al.* 2014. XSEDE: accelerating scientific discovery. *Computing in Science and Engineering*. 16:62-74.

Tsagkogeorga G, McGowen MR, Davies KT, Jarman S, Polanowski A, Bertelsen MF, Rossiter SJ. 2015. A phylogenomic analysis of the role and timing of molecular adaptation in the aquatic transition of cetartiodactyl mammals. *R Soc Open Sci*. 2:150156.

van Dijk EL, Auger H, Jaszczyzyn Y, Thermes C. 2014. Ten years of next-generation sequencing technology. *Trends Genet*. 30:418-426.

VanRaden PM, Sun C, O'Connell JR. 2015. Fast imputation using medium or low-coverage sequence data. *BMC Genet*. 16:82.

Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ*, et al.* 2015. Enhancer evolution across 20 mammalian species. *Cell*. 160:554-566.

Viricel A, Pante E, Dabin W, Simon-Bouhet B. 2014. Applicability of RAD-tag genotyping for interfamilial comparisons: empirical data from two cetaceans. *Mol Ecol Resour*. 14:597-605.

Viricel A, Rosel PE. 2014. Hierarchical population structure and habitat differences in a highly mobile marine species: the Atlantic spotted dolphin. *Mol Ecol*. 23:5018-5035.

Wolf JB. 2013. Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Mol Ecol Resour*. 13:559-572.

Xiong Y, Brandley MC, Xu S, Zhou K, Yang G. 2009. Seven new dolphin mitochondrial genomes and a time-calibrated phylogeny of whales. *BMC Evol Biol*. 9:20.

Yandell M, Ence D. 2012. A beginner's guide to eukaryotic genome annotation. *Nat Rev Genet*. 13:329-342.

Yeh R-F, Lim LP, Burge CB. 2001. Computational inference of homologous gene structures in the human genome. *Genome Res*. 11:803-816.

Yim H-S, Cho YS, Guang X, Kang SG, Jeong J-Y, Cha S-S, Oh H-M, Lee J-H, Yang EC, Kwon KK, *et al.* 2014. Minke whale genome and aquatic adaptation in cetaceans. *Nat Genet*. 46:88-92.

Zhao Q-Y, Wang Y, Kong Y-M, Luo D, Li X, Hao P. 2011. Optimizing *de novo* transcriptome assembly from short-read RNA-Seq data: a comparative study. *BMC Bioinformatics*. 12:S2.

Zhou X, Sun F, Xu S, Fan G, Zhu K, Liu X, Chen Y, Shi C, Yang Y, Huang Z*, et al.* 2013. Baiji genomes reveal low genetic variability and new insights into secondary aquatic adaptations. *Nat Commun*. 4:2708.

Zou Z, Zhang J. 2015. No genome-wide protein sequence convergence for echolocation. *Mol Biol Evol*. 32:1237-1241.

Table 1. Current and commonly used tools for analysis of genomic data generated in non-model organisms. Please note that this list is not exhaustive and new computational tools are continuously being developed.

Computational Tool	Purpose	Strengths/Weaknesses	Reference
*RADseq**			
STACKS	quality filtering, *de novo* assembly or reference-aligned read mapping, variant genotyping	scalable (new data can be compared against existing locus catalog); flexible filtering and export options; recently implemented a gapped alignment algorithm to process insertion-deletion (indel) mutations; secondary algorithm adjusts SNP calls using population-level allele frequencies; compatible with input data from multiple RADseq methods	Catchen et al. (2011; 2013), http://catchenlab.life.illinois.edu/stacks/
PyRAD	quality filtering, *de novo* assembly, read mapping, variant genotyping	efficiently processes indel mutations, thus optimal for analysis of highly divergent species; high speed and quality of paired-end library assemblies; compatible with input data from multiple RADseq methods	Eaton (2014)
TASSEL-GBS	quality filtering, reference-aligned read mapping, variant genotyping	optimized for single-end data from large sample sizes (tens of thousands of individuals) with a reference genome; performs genome-wide association studies	Glaubitz et al. (2014)
dDocent	quality trimming, *de novo* assembly, read mapping, variant genotyping	beneficial in analysis of paired-end data; identifies both SNP and indel variants; most appropriate for ezRAD and ddRAD data	Puritz et al. (2014)
AftrRAD	quality filtering, *de novo* assembly, read mapping, variant genotyping	identifies both SNP and indel variants; computationally faster than STACKS and PyRAD	Sovic et al. (2015)
Array-based high-throughput sequencing			
Affymetrix Axiom™ Analysis Suite, Illumina® GenomeStudio	genotype scoring	visualization of genotype clusters; quality scores assigned to genotype calls allow user-specific filtering; manual editing possible	
Whole genome sequencing			
AdapterRemoval v2, Trimmomatic	trim raw sequences	remove adapter sequences and low-quality bases prior to assembly	Bolger et al. (2014), Schubert et al. (2016)
ALLPATHS-LG, PLATANUS, SOAPdenovo	*de novo* genome assembly	designed for short-read sequences of large heterozygous genomes	Li et al. (2010), Gnerre et al. (2011), Kajitani et al. (2014)
AUGUSTUS, GenomeScan, MAKER2	gene annotation	highly accurate evidence-driven or BLASTX-guided gene prediction (Yandell and Ence 2012)	Yeh et al. (2001), Stanke et al. (2006), Holt and Yandell (2011)

Bowtie, bwa	read mapping	rapid short-read alignment with compressed reference genome index, but limited number of acceptable mismatches per alignment (Flicek and Birney 2009)	Langmead et al. (2009), Li and Durbin (2009)
SAMtools	data processing, variant calling	multi-purpose tool that conducts file conversion, alignment sorting, PCR duplicate removal, and variant (SNP and indel) calling for SAM/BAM/CRAM files	Li et al. (2009)
GATK	data processing and quality control, variant calling	suitable for data with low to high mean read depth across the genome; initially optimized for large human datasets, then modified for use with non-model organisms	McKenna et al. (2010), DePristo et al. (2011)
ANGSD/NGStools	data processing, variant calling, estimation of diversity metrics, population genomic analyses	suitable for data with low mean read depth, including palaeogenomic data; allow downstream analyses such as D-statistics and SFS estimation	Fumagalli et al. (2014), Korneliussen et al. (2014)
RNAseq			
Fastx Toolkit, Trimmomatic	trim raw sequences	remove erroneous nucleotides from reads prior to assembly	MacManes (2014)
khmer diginorm, Trinity normalization	*in silico* read normalization	reduce memory requirements for assembly, but can result in fragmented assemblies and collapse heterozygosity	Brown et al. (2012); Haas et al. (2013)
Trinity	*de novo* and genome-guided transcriptome assembly	accurate assembly across conditions, but requires long runtime if normalization is not used (Zhao et al. 2011)	Haas et al. (2013)
bowtie, bowtie2, STAR	read alignment to genome or transcriptome assembly	required for many downstream analyses, but bowtie is computationally intensive and all produce very large output BAM files	Langmead et al. (2009), Dobin et al. (2013)
eXpress, kallisto, RSEM, Sailfish, Salmon	estimation of transcript abundance	RSEM requires computationally intensive read mapping back to the assembly; the others are faster streaming alignment, quasi-alignment, or alignment-free algorithms	Li and Dewey (2011), Patro et al. (2015)
DESeq, DESeq2, edgeR	differential expression analysis	exhibit highest true positive and lowest false positive rates in experiments with smaller sample sizes (Schurch et al. 2015)	Anders and Huber (2010), Robinson et al. (2010), Love et al. (2014)
blast2GO, Trinotate	functional annotation of assembled transcripts	complete annotation pipelines including gene ontology and pathway enrichment analyses	Conesa et al. (2005), Haas et al. (2013)

* This is a non-exhaustive list of software that focuses on *de novo* loci assembly and genotype calling for RADseq data, as many practitioners working on NMOs will not have access to a reference genome. Other programs (e.g., GATK and ANGSD) that undertake genotype calling using reference-aligned loci are described in the whole genome sequencing section.

Figure 1

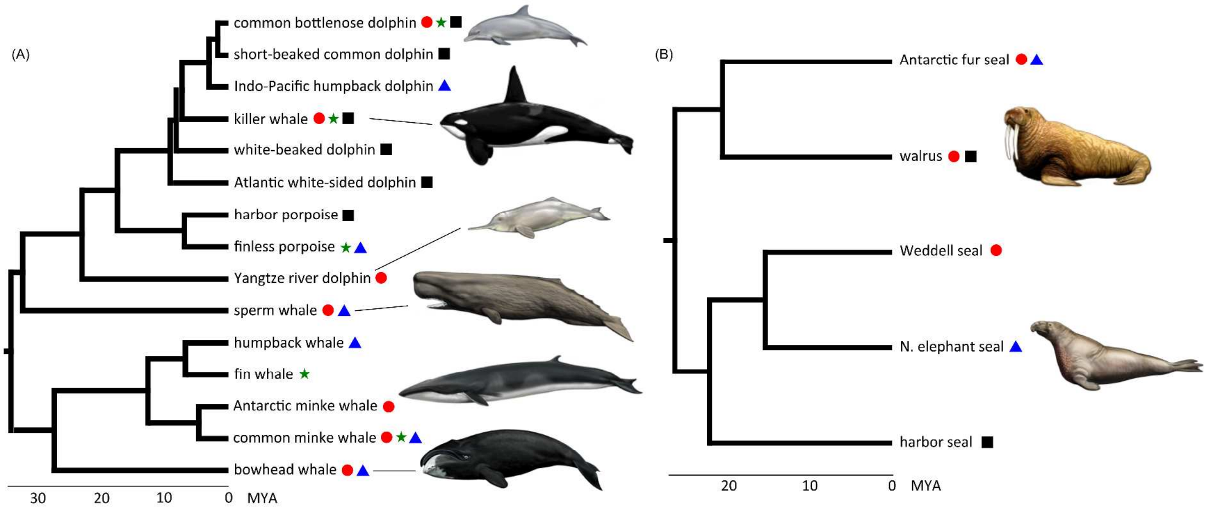


Figure 1. Phylogenetic tree showing current genomic resources available for (A) cetaceans and (B) pinnipeds; relationships and branch lengths are based on molecular dating estimates from McGowen et al. (2009), McGowen (2011), and Higdon et al. (2007). Scale is in millions of years ago (MYA). Red circles indicate species with high-quality reference genomes; green stars indicate whole genome re-sequencing data; blue triangles indicate transcriptomes (generated by microarray or RNAseq); and black squares indicate RADseq data.

Figure 2

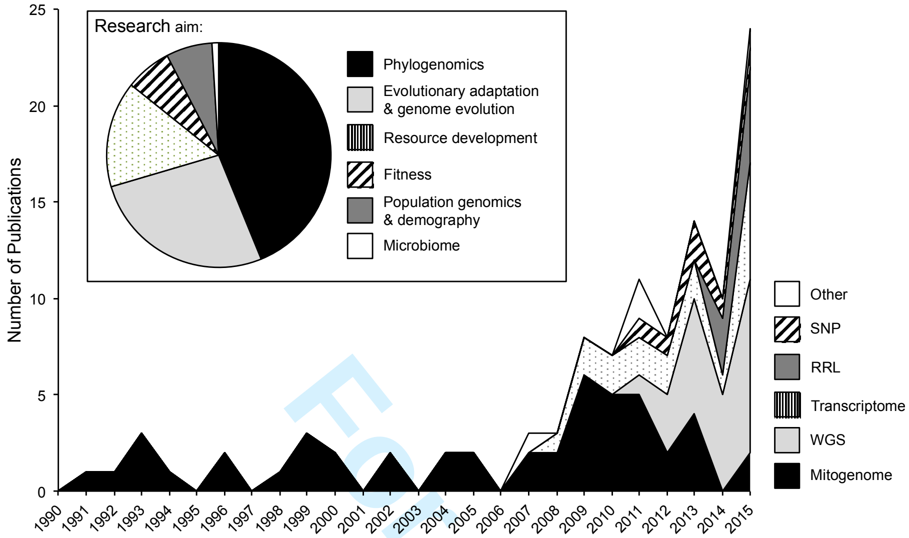


Figure 2. Number of marine mammal genomics publications from 1990 to 2015, categorized by primary methodology and research aim. Genomic methodologies include high-throughput single nucleotide polymorphism (SNP) genotyping and sequencing of mitogenomes, whole genomes (WGS), transcriptomes (generated by microarray or RNAseq), and reduced-representation genomic libraries (RRL). The "Other" category includes studies of microbiomes, BAC libraries, and large (~100) gene sets.

Figure 3

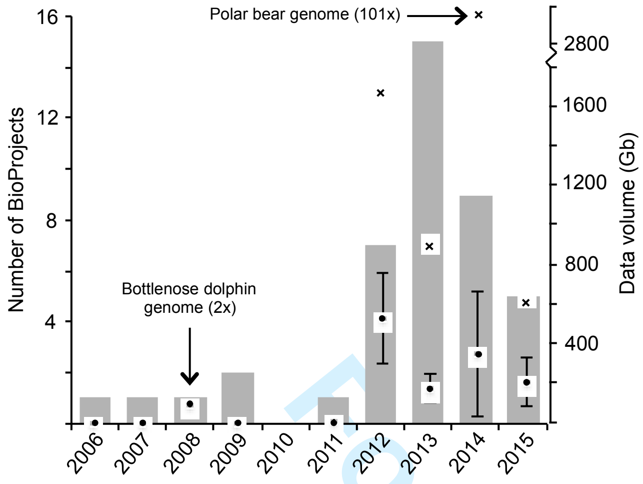
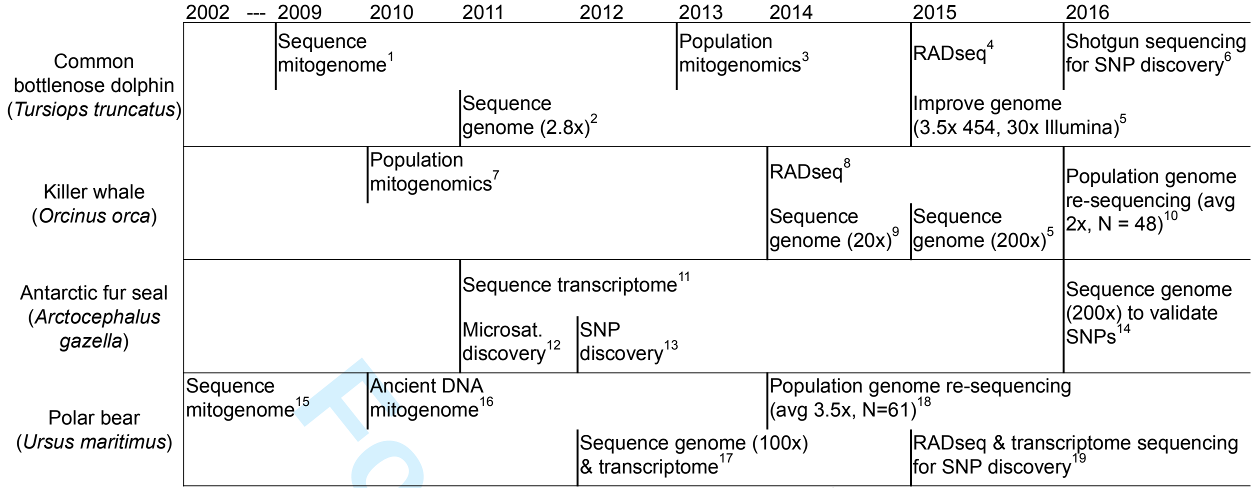


Figure 3. Number of BioProjects (gray bars) related to marine mammal genomics submitted from 2006 to 2015 to an online public database maintained by NCBI. Early BioProjects were largely microarray datasets. The number of projects created each year, as well as the yearly average (black dots ± SE) and maximum (×) size of data submitted in each BioProject, increased dramatically after 2011, reflecting advances in high-throughput sequencing technologies that facilitated their use in non-model systems.

Figure 4



2002 --- 2009 2010 2011 2012 2013 2014 2015 2016

Common bottlenose dolphin (*Tursiops truncatus*)

Sequence mitogenome[1]

Sequence genome (2.8x)[2]

Population mitogenomics[3]

RADseq[4]

Improve genome (3.5x 454, 30x Illumina)[5]

Shotgun sequencing for SNP discovery[6]

Killer whale (*Orcinus orca*)

Population mitogenomics[7]

RADseq[8]

Sequence genome (20x)[9]

Sequence genome (200x)[5]

Population genome re-sequencing (avg 2x, N = 48)[10]

Antarctic fur seal (*Arctocephalus gazella*)

Sequence transcriptome[11]

Microsat. discovery[12]

SNP discovery[13]

Sequence genome (200x) to validate SNPs[14]

Polar bear (*Ursus maritimus*)

Sequence mitogenome[15]

Ancient DNA mitogenome[16]

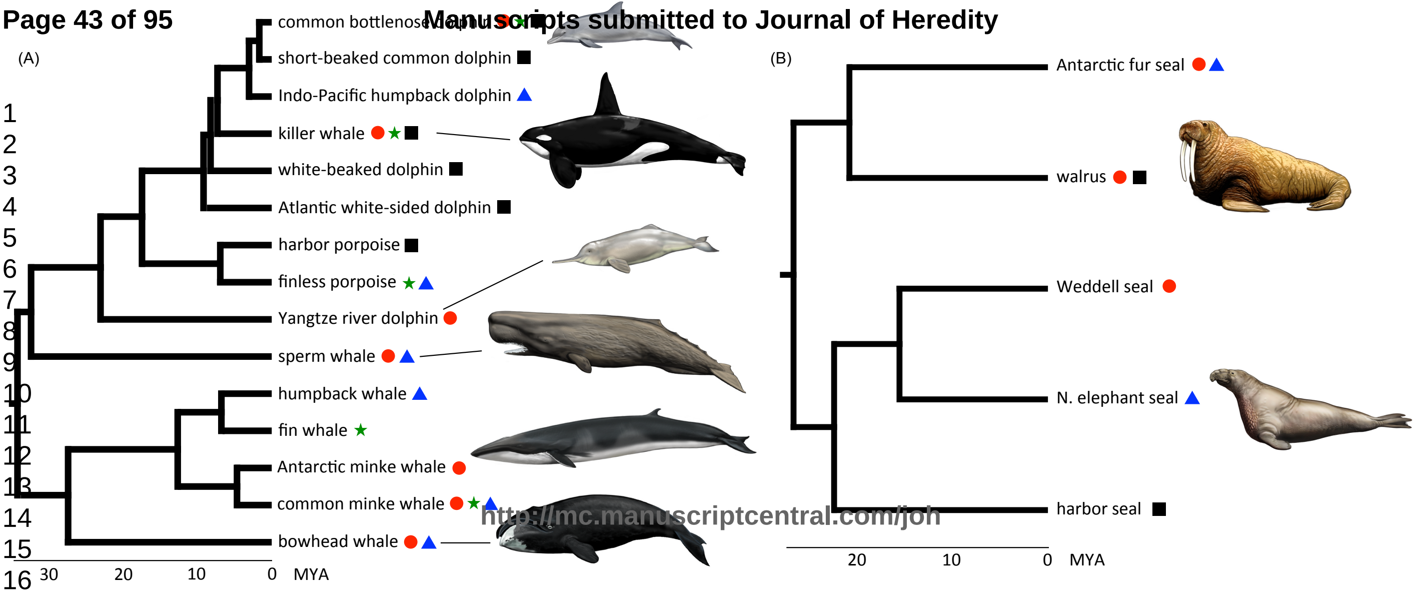
Sequence genome (100x) & transcriptome[17]

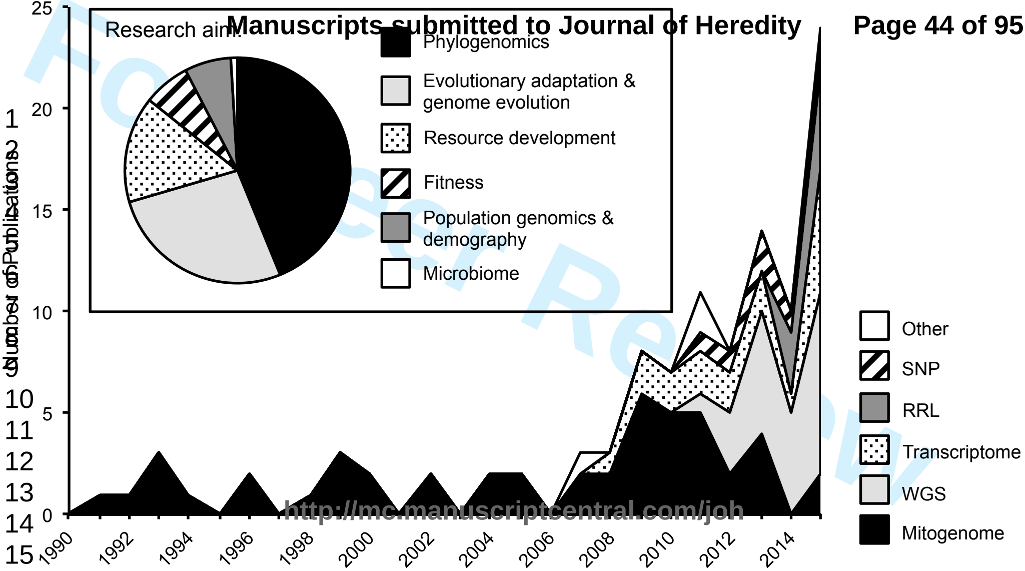
Population genome re-sequencing (avg 3.5x, N=61)[18]

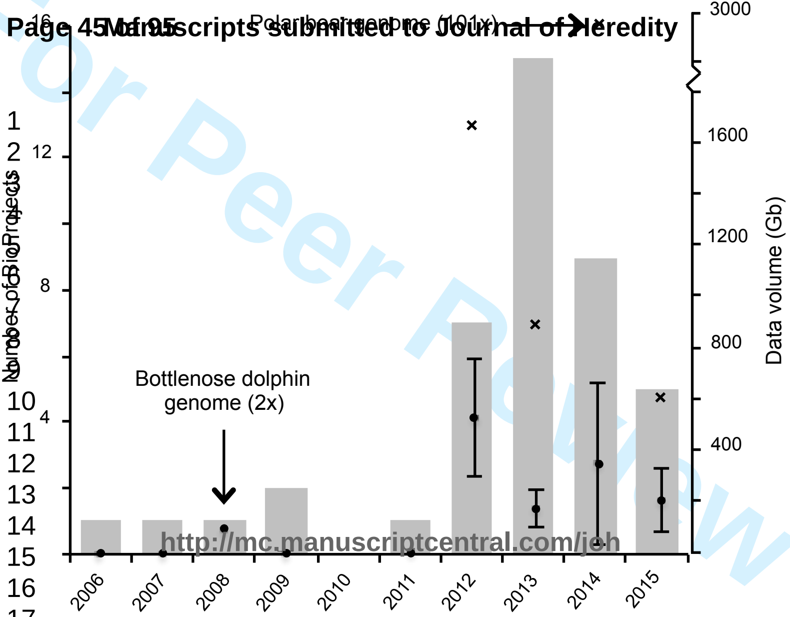
RADseq & transcriptome sequencing for SNP discovery[19]

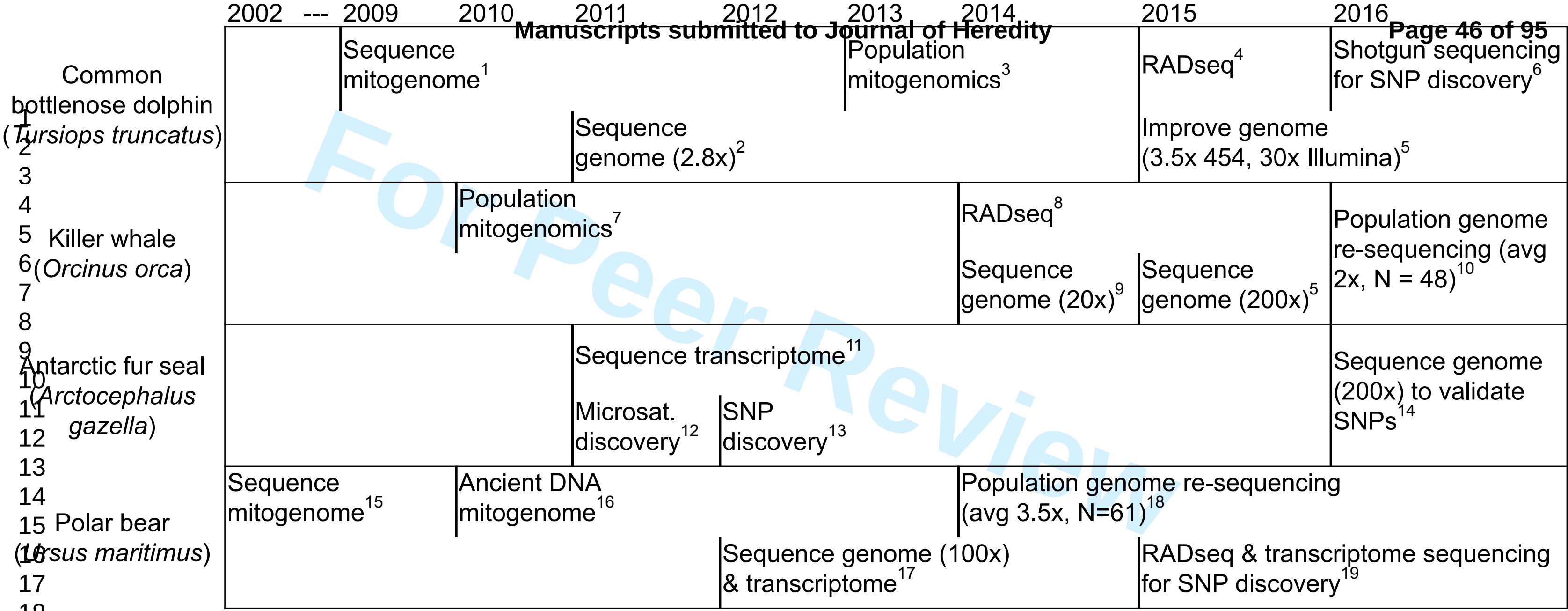
1) Xiong et al. 2009; 2) Lindblad-Toh et al. 2011; 3) Moura et al. 2013; 4) Cammen et al. 2015; 5) Foote et al. 2015; 6) Louis et al. unpubl. data; 7) Morin et al. 2010; 8) Moura et al. 2014a; 9) Moura et al. 2014b; 10) Foote et al. 2016; 11) Hoffman 2011; 12) Hoffman and Nicholas 2011; 13) Hoffman et al. 2012; 14) Humble et al. 2016; 15) Arnason et al. 2002; 16) Lindqvist et al. 2010; 17) Miller et al. 2012; 18) Liu et al. 2014; 19) Malenfant et al. 2015

Figure 4. Timelines depicting the independent progression of genomic studies for four representative marine mammal species. Trajectories show the common progression for non-model species from mitogenome sequencing to whole genome sequencing, as well as from sequencing reference specimens to population-scale genomic sequencing. In addition, the timelines reveal the utility of genomic and transcriptomic sequencing for subsequent genetic marker development.


(A)
common bottlenose dolphin ● ★ ■
short-beaked common dolphin ■
Indo-Pacific humpback dolphin ▲
killer whale ● ★ ■
white-beaked dolphin ■
Atlantic white-sided dolphin ■
harbor porpoise ■
finless porpoise ★ ▲
Yangtze river dolphin ●
sperm whale ● ▲
humpback whale ▲
fin whale ★
Antarctic minke whale ●
common minke whale ● ★ ▲
bowhead whale ● ▲
30
20
10
0
MYA
(B)
Antarctic fur seal ● ▲
walrus ● ■
Weddell seal ●
N. elephant seal ▲
harbor seal ■
20
10
0
MYA


Research aim:
Phylogenomics
Evolutionary adaptation & genome evolution
Resource development
Fitness
Population genomics & demography
Microbiome
Number of Publications
0
5
10
15
20
25
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
Other
SNP
RRL
Transcriptome
WGS
Mitogenome


Polar bear genome (101x)
Bottlenose dolphin genome (2x)
Number of BioProjects
16
12
8
4
Data volume (Gb)
3000
1600
1200
800
400
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015


2002

2009
2010
2011
2012
2013
2014
2015
2016
Common bottlenose dolphin (*Tursiops truncatus*)
Sequence mitogenome1
Population mitogenomics3
RADseq4
Shotgun sequencing for SNP discovery6
Sequence genome (2.8x)2
Improve genome (3.5x 454, 30x Illumina)5
Killer whale (*Orcinus orca*)
Population mitogenomics7
RADseq8
Sequence genome (20x)9
Sequence genome (200x)5
Population genome re-sequencing (avg 2x, N = 48)10
Antarctic fur seal (*Arctocephalus gazella*)
Sequence transcriptome11
Microsat. discovery12
SNP discovery13
Sequence genome (200x) to validate SNPs14
Polar bear (*Ursus maritimus*)
Sequence mitogenome15
Ancient DNA mitogenome16
Population genome re-sequencing (avg 3.5x, N=61)18
Sequence genome (100x) & transcriptome17
RADseq & transcriptome sequencing for SNP discovery19

1) Xiong et al. 2009; 2) Lindblad-Toh et al. 2011; 3) Moura et al. 2013; 4) Cammen et al. 2015; 5) Foote et al. 2015; 6) Louis et al. unpubl. data; 7) Morin et al. 2010; 8) Moura et al. 2014a; 9) Moura et al. 2014b; 10) Foote et al. 2016; 11) Hoffman 2011; 12) Hoffman and Nicholas 2011; 13) Hoffman et al. 2012; 14) Humble et al. 2016; 15) Arnason et al. 2002; 16) Lindqvist et al. 2010; 17) Miller et al. 2012; 18) Liu et al. 2014; 19) Malenfant et al. 2015

Table 1. Current and commonly used tools for analysis of genomic data generated in non-model organisms. Please note that this list is not exhaustive and new computational tools are continuously being developed.

Computational Tool	Purpose	Strengths/Weaknesses	Reference
*RADseq**			
STACKS	quality filtering, *de novo* assembly or reference-aligned read mapping, variant genotyping	scalable (new data can be compared against existing locus catalog); flexible filtering and export options; recently implemented a gapped alignment algorithm to process insertion-deletion (indel) mutations; secondary algorithm adjusts SNP calls using population-level allele frequencies; compatible with input data from multiple RADseq methods	Catchen et al. (2011; 2013), http://catchenlab.life.illinois.edu/stacks/
PyRAD	quality filtering, *de novo* assembly, read mapping, variant genotyping	efficiently processes indel mutations, thus optimal for analysis of highly divergent species; high speed and quality of paired-end library assemblies; compatible with input data from multiple RADseq methods	Eaton (2014)
TASSEL-GBS	quality filtering, reference-aligned read mapping, variant genotyping	optimized for single-end data from large sample sizes (tens of thousands of individuals) with a reference genome; performs genome-wide association studies	Glaubitz et al. (2014)
dDocent	quality trimming, *de novo* assembly, read mapping, variant genotyping	beneficial in analysis of paired-end data; identifies both SNP and indel variants; most appropriate for ezRAD and ddRAD data	Puritz et al. (2014)
AftrRAD	quality filtering, *de novo* assembly, read mapping, variant genotyping	identifies both SNP and indel variants; computationally faster than STACKS and PyRAD	Sovic et al. (2015)
Array-based high-throughput sequencing			
Affymetrix Axiom™ Analysis Suite, Illumina® GenomeStudio	genotype scoring	visualization of genotype clusters; quality scores assigned to genotype calls allow user-specific filtering; manual editing possible	
Whole genome sequencing			
AdapterRemoval v2, Trimmomatic	trim raw sequences	remove adapter sequences and low-quality bases prior to assembly	Bolger et al. (2014), Schubert et al. (2016)
ALLPATHS-LG, PLATANUS, SOAPdenovo	*de novo* genome assembly	designed for short-read sequences of large heterozygous genomes	Li et al. (2010), Gnerre et al. (2011), Kajitani et al. (2014)
AUGUSTUS, GenomeScan, MAKER2	gene annotation	highly accurate evidence-driven or BLASTX-guided gene prediction (Yandell and Ence 2012)	Yeh et al. (2001), Stanke et al. (2006), Holt and Yandell (2011)

Bowtie, bwa	read mapping	rapid short-read alignment with compressed reference genome index, but limited number of acceptable mismatches per alignment (Flicek and Birney 2009)	Langmead et al. (2009), Li and Durbin (2009)
SAMtools	data processing, variant calling	multi-purpose tool that conducts file conversion, alignment sorting, PCR duplicate removal, and variant (SNP and indel) calling for SAM/BAM/CRAM files	Li et al. (2009)
GATK	data processing and quality control, variant calling	suitable for data with low to high mean read depth across the genome; initially optimized for large human datasets, then modified for use with non-model organisms	McKenna et al. (2010), DePristo et al. (2011)
ANGSD/NGStools	data processing, variant calling, estimation of diversity metrics, population genomic analyses	suitable for data with low mean read depth, including palaeogenomic data; allow downstream analyses such as D-statistics and SFS estimation	Fumagalli et al. (2014), Korneliussen et al. (2014)
RNAseq			
Fastx Toolkit, Trimmomatic	trim raw sequences	remove erroneous nucleotides from reads prior to assembly	MacManes (2014)
khmer diginorm, Trinity normalization	*in silico* read normalization	reduce memory requirements for assembly, but can result in fragmented assemblies and collapse heterozygosity	Brown et al. (2012); Haas et al. (2013)
Trinity	*de novo* and genome-guided transcriptome assembly	accurate assembly across conditions, but requires long runtime if normalization is not used (Zhao et al. 2011)	Haas et al. (2013)
bowtie, bowtie2, STAR	read alignment to genome or transcriptome assembly	required for many downstream analyses, but bowtie is computationally intensive and all produce very large output BAM files	Langmead et al. (2009), Dobin et al. (2013)
eXpress, kallisto, RSEM, Sailfish, Salmon	estimation of transcript abundance	RSEM requires computationally intensive read mapping back to the assembly; the others are faster streaming alignment, quasi-alignment, or alignment-free algorithms	Li and Dewey (2011), Patro et al. (2015)
DESeq, DESeq2, edgeR	differential expression analysis	exhibit highest true positive and lowest false positive rates in experiments with smaller sample sizes (Schurch et al. 2015)	Anders and Huber (2010), Robinson et al. (2010), Love et al. (2014)
blast2GO, Trinotate	functional annotation of assembled transcripts	complete annotation pipelines including gene ontology and pathway enrichment analyses	Conesa et al. (2005), Haas et al. (2013)

* This is a non-exhaustive list of software that focuses on *de novo* loci assembly and genotype calling for RADseq data, as many practitioners working on NMOs will not have access to a reference genome. Other programs (e.g., GATK and ANGSD) that undertake genotype calling using reference-aligned loci are described in the whole genome sequencing section.

Cammen_SupMat_TableS1 - Marine mammal genomics - *JHered*

Table S1. Broad applications of genomic tools in studies of non-model organisms are provided with concrete examples of research areas drawn from the field of marine mammal genomics. The number of loci used in each study provides an estimate of the scope of the respective genomic tools and study, but represents the outcome of several filtering steps from raw sequence data that vary across studies. Further details of each method can be found in the listed references. Please note that this is not an exhaustive list. GBS: Genotyping by Sequencing; RADseq: restriction site-associated DNA sequencing; SNP: single nucleotide polymorphism; TSC: target sequence capture; WGS: whole genome sequencing.

Method	# loci	Research area	Reference
Evolutionary genomics: describe evolutionary history and adaptation			
Mitogenome sequencing	Mitogenome	Cetacean phylogenomics	McGowen et al. (2009)
TSC	Mitogenome	Comparison of sub-fossil and modern killer whales	Foote et al. (2013)
TSC	>30kb coding sequence	Evolution of Sirenia	Springer et al. (2015)
WGS	Whole genome	Yangtze river dolphin genome analysis	Zhou et al. (2013)
WGS	Whole genome	Minke whale genome analysis	Yim et al. (2014)
WGS	Whole genome	Bowhead whale genome analysis	Keane et al. (2015)
WGS	Whole genome	Analysis of convergent evolution in marine mammal lineages	Foote et al. (2015)
WGS	10,025 coding sequences	Positive selection in common bottlenose dolphin genome	McGowen et al. (2012)
WGS	Sensory genes	Analysis of gene loss in olfaction and taste in Antarctic minke whale	Kishida et al. (2015)
Genome re-seq	Whole genome	Speciation and adaptation in brown and polar bears	Liu et al. (2014)
Transcriptomics	9,395 genes	Evolution of longevity in bowhead whales	Seim et al. (2014)
Transcriptomics	103,077 unigenes	Osmoregulatory divergence in narrow-ridged finless porpoise	Ruan et al. (2015)
Population genomics: characterize population structure and investigate demography			
RADseq	3,281 SNPs	Killer whale ecotype divergence	Moura et al. (2014)
RADseq (GBS)	24,996 loci; 4,854 SNPs	Historical demography in Atlantic walrus	Shafer et al. (2015)
TSC	Mitogenome and 43-118 nuclear loci	Phylogeography and population genomics of cetaceans	Hancock-Hanser et al. (2013); Morin et al. (2015)
Genome re-seq	Whole genome	Demographic history, population differentiation, and ecotype divergence in killer whales	Foote et al. (2016)

Cammen_SupMat_TableS1 - Marine mammal genomics - *JHered*

Adaptation genomics: describe relationships between genomic variation and fitness			
RADseq	83,148 loci; 14,585 SNPs	Effect of inbreeding depression on parasite infection in harbor seals	Hoffman et al. (2014)
RADseq	129,494 loci; 7,431 SNPs	Common bottlenose dolphin adaptation to harmful algal blooms	Cammen et al. (2015)
Transcriptomics	11,286 contigs	Sperm whale skin cell response to hexavalent chromium	Pabuwal et al. (2013)
Transcriptomics	164,966 contigs	Physiological stress response in northern elephant seals	Khudyakov et al. (2015a; 2015b)
Develop molecular resources			
RADseq	3,595 loci	Comparison of short-beaked common dolphin and harbor porpoise	Viricel et al. (2014)
Shotgun sequencing	440,718 SNPs	SNP discovery in Northeast Atlantic common bottlenose dolphins	M. Louis (unpubl. data)
WGS	144 SNPs	SNP validation in Antarctic fur seal	Humble et al. (2016)
Transcriptomics	23,096 contigs; 144 SNPs	Gene and SNP discovery in Antarctic fur seal	Hoffman et al. (2011; 2012; 2013)
Transcriptomics & RADseq	9,000 SNPs	Development of SNP array for polar bear and demonstration of utility in population genomics	Malenfant et al. (2015)

References

Cammen KM, Schultz TF, Rosel PE, Wells RS, Read AJ. 2015. Genomewide investigation of adaptation to harmful algal blooms in common bottlenose dolphins (*Tursiops truncatus*). *Mol Ecol*. 24:4697-4710.

Foote AD, Liu Y, Thomas GWC, Vinař Ts, Alföldi J, Deng J, Dugan S, van Elk CE, Hunter ME, Joshi V*, et al.* 2015. Convergent evolution of the genomes of marine mammals. *Nat Genet*. 47:272-275.

Foote AD, Newton J, Ávila-Arcos MC, Kampmann M-L, Samaniego JA, Post K, Rosing-Asvid A, Sinding M-HS, Gilbert MTP. 2013. Tracking niche variation over millennial timescales in sympatric killer whale lineages. *Proc R Soc Lond B Biol Sci*. 280:20131481.

Foote AD, Vijay N, Ávila-Arcos M, Baird RW, Durban JW, Fumagalli M, Gibbs RA, Hanson MB, Korneliussen TS, Martin MD*, et al.* 2016. Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nat Commun*. 7:11693.

Hancock-Hanser BL, Frey A, Leslie MS, Dutton PH, Archer FI, Morin PA. 2013. Targeted multiplex next-generation sequencing: advances in techniques of mitochondrial and nuclear DNA sequencing for population genomics. *Mol Ecol Resour*. 13:254-268.

Hoffman JI. 2011. Gene discovery in the Antarctic fur seal (*Arctocephalus gazella*) skin transcriptome. *Mol Ecol Resour*. 11:703-710.

Hoffman JI, Simpson F, David P, Rijks JM, Kuiken T, Thorne MAS, Lacy RC, Dasmahapatra KK. 2014. High-throughput sequencing reveals inbreeding depression in a natural population. *Proc Natl Acad Sci USA*. 111:3775-3780.

Hoffman JI, Thorne MAS, Trathan PN, Forcada J. 2013. Transcriptome of the dead: characterisation of immune genes and marker development from necropsy samples in a free-ranging marine mammal. *BMC Genomics*. 14:52.

Hoffman JI, Tucker R, Bridgett SJ, Clark MS, Forcada J, Slate J. 2012. Rates of assay success and genotyping error when single nucleotide polymorphism genotyping in non-model organisms: a case study in the Antarctic fur seal. *Mol Ecol Resour*. 12:861-872.

Humble E, Martinez-Barrio A, Forcada J, Trathan PN, Thorne MAS, Hoffmann M, Wolf JBW, Hoffman JI. 2016. A draft fur seal genome provides insights into factors affecting SNP validation and how to mitigate them. *Mol Ecol Resour*.

Keane M, Semeiks J, Webb AE, Li YI, Quesada V, Craig T, Madsen LB, van Dam S, Brawand D, Marques PI*, et al.* 2015. Insights into the evolution of longevity from the bowhead whale genome. *Cell Reports*. 10:112-122.

Khudyakov JI, Champagne CD, Preeyanon L, Ortiz RM, Crocker DE. 2015a. Muscle transcriptome response to ACTH administration in a free-ranging marine mammal. *Physiol Genomics*. 47:318-330.

Khudyakov JI, Preeyanon L, Champagne CD, Ortiz RM, Crocker DE. 2015b. Transcriptome analysis of northern elephant seal (*Mirounga angustirostris*) muscle tissue provides a novel molecular resource and physiological insights. *BMC Genomics*. 16:64.

Kishida T, Thewissen JGM, Hayakawa T, Imai H, Agata K. 2015. Aquatic adaptation and the evolution of smell and taste in whales. *Zoolog Lett*. 1:9.

Liu S, Lorenzen ED, Fumagalli M, Li B, Harris K, Xiong Z, Zhou L, Korneliussen TS, Somel M, Babbitt C*, et al.* 2014. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell*. 157:785-794.

Malenfant RM, Coltman DW, Davis CS. 2015. Design of a 9K Illumina BeadChip for polar bears (*Ursus maritimus*) from RAD and transcriptome sequencing. *Mol Ecol Resour*. 15:587-600.

McGowen MR, Grossman LI, Wildman DE. 2012. Dolphin genome provides evidence for adaptive evolution of nervous system genes and a molecular rate slowdown. *Proc R Soc Lond B Biol Sci*. 279:3643-3651.

McGowen MR, Spaulding M, Gatesy J. 2009. Divergence date estimation and a comprehensive molecular tree of extant cetaceans. *Mol Phylogenet Evol*. 53:891-906.

Morin PA, Parsons KM, Archer FI, Ávila-Arcos M, Barrett-Lennard LG, Dalla Rosa L, Duchêne S, Durban JW, Ellis GM, Ferguson SH*, et al.* 2015. Geographic and temporal dynamics of a global radiation and diversification in the killer whale. *Mol Ecol*. 24:3964-3979.

Moura AE, Kenny JG, Chaudhuri R, Hughes MA, Welch AJ, Reisinger RR, de Bruyn PJN, Dahlheim ME, Hall N, Hoelzel AR. 2014. Population genomics of the killer whale indicates ecotype evolution in sympatry involving both selection and drift. *Mol Ecol*. 23:5179-5192.

Pabuwal V, Boswell M, Pasquali A, Wise SS, Kumar S, Shen Y, Garcia T, Lacerte C, Wise JP, Jr., Wise JP, Sr.*, et al.* 2013. Transcriptomic analysis of cultured whale skin cells exposed to hexavalent chromium [Cr(VI)]. *Aquat Toxicol*. 134-135:74-81.

Ruan R, Guo A-H, Hao Y-J, Zheng J-S, Wang D. 2015. *De novo* assembly and characterization of narrow-ridged finless porpoise renal transcriptome and identification of candidate genes involved in osmoregulation. *Int J Mol Sci*. 16:2220-2238.

Seim I, Ma S, Zhou X, Gerashchenko MV, Lee S-G, Suydam R, George JC, Bickham JW, Gladyshev VN. 2014. The transcriptome of the bowhead whale *Balaena mysticetus* reveals adaptations of the longest-lived mammal. *Aging*. 6:879-899.

Shafer ABA, Gattepaille LM, Stewart REA, Wolf JBW. 2015. Demographic inferences using short-read genomic data in an approximate Bayesian computation framework: *in silico* evaluation of power, biases and proof of concept in Atlantic walrus. *Mol Ecol*. 24:328-345.

Springer MS, Signore AV, Paijmans JLA, Vélez-Juarbe J, Domning DP, Bauer CE, He K, Crerar L, Campos PF, Murphy WJ*, et al.* 2015. Interordinal gene capture, the phylogenetic position of Steller's sea cow based on molecular and morphological data, and the macroevolutionary history of Sirenia. *Mol Phylogenet Evol*. 91:178-193.

Viricel A, Pante E, Dabin W, Simon-Bouhet B. 2014. Applicability of RAD-tag genotyping for interfamilial comparisons: empirical data from two cetaceans. *Mol Ecol Resour*. 14:597-605.

Yim H-S, Cho YS, Guang X, Kang SG, Jeong J-Y, Cha S-S, Oh H-M, Lee J-H, Yang EC, Kwon KK*, et al.* 2014. Minke whale genome and aquatic adaptation in cetaceans. *Nat Genet*. 46:88-92.

Zhou X, Sun F, Xu S, Fan G, Zhu K, Liu X, Chen Y, Shi C, Yang Y, Huang Z*, et al.* 2013. Baiji genomes reveal low genetic variability and new insights into secondary aquatic adaptations. *Nat Commun*. 4:2708.

Genomic methods take the plunge: recent advances in high-throughput sequencing of marine mammals

KRISTINA M. CAMMEN[1]*, KIMBERLY R. ANDREWS[2], EMMA L. CARROLL[3], ANDREW D. FOOTE[4], EMILY HUMBLE[5,6], JANE I. KHUDYAKOV[7], MARIE LOUIS[3], MICHAEL R. MCGOWEN[8], MORTEN TANGE OLSEN[9], AND AMY M. VAN CISE[10]

[1]School of Marine Sciences, University of Maine, Orono, Maine 04469, USA
[2]Department of Fish and Wildlife Sciences, University of Idaho, 875 Perimeter Drive MS 1136, Moscow, Idaho 83844-1136, USA
[3]Scottish Oceans Institute, University of St Andrews, East Sands, St Andrews, Fife KY16 8LB, UK
[4]Computational and Molecular Population Genetics ~~CMPG~~ L~~l~~ab, Institute of Ecology and Evolution, University of Bern, Bern CH-3012, Switzerland
[5]Department of Animal Behaviour, University of Bielefeld, Postfach 100131, 33501 Bielefeld, Germany
[6]British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 OET, UK
[7]Department of Biology, Sonoma State University, Rohnert Park, California 94928, USA
[8]School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK
[9]Evolutionary Genomics Section, Natural History Museum of Denmark, University of Copenhagen, DK-1353 Copenhagen K, Denmark
[10]Scripps Institution of Oceanography, 8622 Kennel Way, La Jolla, California 92037, USA

*Corresponding author: kristina.cammen@maine.edu

Running title: Marine mammal genomics

Abstract

The dramatic increase in the application of genomic techniques to non-model organisms over the past decade has yielded numerous valuable contributions to evolutionary biology and ecology, many of which would not have been possible with traditional genetic markers. We review this recent progression with a particular focus on genomic studies of marine mammals, a group of taxa that represent key macroevolutionary transitions from terrestrial to marine environments and for which available genomic resources have recently undergone notable rapid growth. Genomic studies of non-model organisms utilize an expanding range of approaches, including ~~low- and high-coverage~~ whole genome sequencing, restriction site-associated DNA sequencing, array-based ~~high-throughput~~ sequencing of single nucleotide polymorphisms and target sequence probes (e.g., exomes), and transcriptome sequencing. These approaches generate different types and quantities of data, and many can be applied with limited or no prior genomic resources, thus overcoming one traditional limitation~~s~~ of research on non-model organisms. Within marine mammals, such studies have thus far yielded significant contributions to the fields of phylogenomics and comparative genomics, as well as enabled investigations of fitness, demography, and population structure~~ in natural populations~~. Here~~,~~ we review the primary options for generating genomic data, introduce several emerging techniques, and discuss the suitability of each approach for different applications in the study of non-model organisms.

Keywords: RADseq, SNP array, target sequence capture, whole genome sequencing, RNAseq, non-model organisms

Introduction

Recent advances in sequencing technologies, coincident with dramatic declines in cost, have increasingly enabled the application of genomic sequencing in non-model systems (Ekblom and Galindo 2011; Ellegren 2014). These advances in molecular technologies have in many ways begun to blur the distinction between model and non-model organisms (Armengaud et al. 2014). Non-model organisms (NMOs) have traditionally been defined as those for which whole-organism experimental manipulation is rarely, if ever, possible due to logistical and/or ethical constraints (Ankeny and Leonelli 2011). Further, NMOs have typically been characterized by limited genomic resources, but this is becoming increasingly less so as the number of NMO reference genomes grows rapidly, for example through efforts like the Genome 10K Project (Koepfli et al. 2015). In fact, in some taxonomic orders, we are approaching the point at which all species have at least one representative reference genome available for a closely related species (Fig 1).

Despite the limitations of working with NMOs, including potentially small sample sizes, low DNA quantity, and limited information on gene function, genetic and genomic investigations of NMOs have yielded numerous valuable contributions to understanding their evolutionary biology and ecology. For the past several decades, traditional genetic markers such as microsatellites and short fragments of mitochondrial DNA (e.g., the control region) have been extensively used in molecular ecology. These markers, which typically evolve under neutral expectations, have proven useful for identifying population structure and reconstructing population demographic history (Hedrick 2000). However, the power of such studies is limited by the number of markers that can feasibly be evaluated using traditional approaches. The advent of low-cost high-throughput sequencing has led to dramatic increases in the number of neutral markers that can be evaluated, in many cases improving our power to resolve fine-scale or cryptic population structure in species with high dispersal capability (e.g., Corander et al. 2013) and improving the accuracy of estimating some (though not all) demographic parameters (Li and Jakobsson 2012; Shafer et al. 2015). Importantly, high-throughput sequencing has also further enabled genomic studies of non-neutral processes in NMOs, for example, characterizing both deleterious and adaptive variation within and across species (Stinchcombe and Hoekstra 2008;

Künstner et al. 2010). It is increasingly evident that genomic analyses of NMOs can and have provided important insights that could not be identified with traditional genetic markers.

Many molecular ecologists now face the challenge of deciding which of the broad range of genomic approaches to apply to their study systems. Here we review the primary options for generating genomic data and their relative suitability for different applications in the study of NMOs. We focus on marine mammals, which represent several mammalian clades with notably rapid growth in available genomic resources in recent years. This growth is clearly evident in both publication rate (Fig 2) and the rise in number and size of genomic sequences deposited in public resources (Fig 3). We comprehensively review the literature on marine mammal genomics, highlighting recent trends in methodology and applications, and then describe in detail the molecular approaches that are most commonly applied to studies of ~~non-model~~NMO genomics. Our hope is that this review will highlight the promise of genomics for NMOs and offer guidance to researchers considering the application of genomic techniques in their non-model study system of choice.

Why study marine mammal genomics?

Marine mammals represent key macroevolutionary transitions from terrestrial to marine environments (McGowen et al. 2014) and accordingly are an exemplary system for investigating the evolution of several morphological and physiological adaptations (Foote et al. 2015) associated with locomotion (Shen et al. 2012), sight (Meredith et al. 2013), echolocation (Parker et al. 2013; Zou and Zhang 2015), deep diving (Mirceta et al. 2013), osmoregulation (Ruan et al. 2015), and cognition (McGowen et al. 2012). Furthermore, studies of marine mammal evolution to date have characterized several unique aspects of their genome evolution that merit further investigation, including low genomic diversity and a relatively slow molecular clock, especially in cetaceans (Jackson et al. 2009; McGowen et al. 2012; Zhou et al. 2013). As many cetacean species are highly mobile with no obvious physical geographic barriers to dispersal, they provide a unique opportunity to study the role of behavior and culture in shaping population structure and genetic diversity (Riesch et al. 2012; Carroll et al. 2015; Alexander et al. 2016).~~ Finally,~~ T~~t~~hough highly mobile, many marine mammals exhibit evidence of local adaptation; for example, several species show parallel divergent morphological and behavioral adaptations to coastal and pelagic environments (Moura et al. 2013; Louis et al. 2014; Viricel and Rosel 2014). These species may be studied across ocean basins as emerging examples of ecological adaptation and speciation (Morin et al. 2010a).

Beyond their value as systems of evolutionary study, many marine mammals are also of broader interest relating to their historical and present conservation status. Many marine mammal populations share histories of dramatic decline due to hunting and other human impacts. Genomics provides a promising tool with which to expand our insights into these historical population changes, which so far primarily have been derived from archival review and traditional genetic approaches (Ruegg et al. 2013; Sremba et al. 2015). More recently, since the implementation of national and international protections, many marine mammal populations have partially or fully recovered (Magera et al. 2013), yet the conservation status of certain marine mammal populations remains of concern. Such vulnerable populations could benefit greatly from an improved understanding of their genetic diversity and evolution, especially in ways that can inform predictions of adaptive capacity to anthropogenic pressures and expand the toolkit for conservation policy (Garner et al. 2016; Taylor and Gemmell 2016).

Recent trends in marine mammal genomics

We conducted a meta-analysis of the peer-reviewed marine mammal genomics literature to evaluate trends in publication rates across research methodologies and aims. A search of the Web of Science database using the term “genom*” and one of the following terms indicating study species - “marine mammal”, “pinniped”, “seal”, “sea lion”, “sea otter”, “whale”, “dolphin”, “polar bear”, “manatee” - identified 825 records on December 11, 2015. We excluded 77% of the search results that were not directly related to genomic studies in marine mammal systems. The remaining 101 articles that were relevant to marine mammal genomics were further categorized by primary research methodology and general research aim. A subset of these articles is described briefly in Supplemental Table 1.

From the early 1990s through 2015, published literature in the field ~~has~~ shifted from an early focus on mitogenome sequencing to more sequence-intensive approaches, such as transcriptome and whole genome sequencing (Figs 2 and 4). This trajectory closely follows trends in sequencing technologies, from Sanger sequencing of short- and long-range PCR products for mitogenome sequencing (Arnason et al. 1991) and SNP discovery (Olsen et al. 2011), to high-throughput sequencing of reduced-representation genomic libraries (RRLs) that consist of selected subsets of the genome (e.g., Viricel et al. 2014), to high-throughput sequencing of whole genomes with varying levels of depth, ~~of~~ coverage, and contiguity. Today, high-throughput sequencing can be used both to generate high-~~-~~quality reference genome assemblies (Yim et al. 2014; Foote et al. 2015; Humble et al. 2016) and to re-sequence whole genomes at a population scale (Liu et al. 2014a; Foote et al. 2016). Similarly, the scale of gene expression studies has increased from quantitative real-time PCR of candidate genes (Tabuchi et al. 2006) to microarrays containing hundreds to thousands of genes (Mancia et al. 2007) and high-throughput RNAseq that evaluates hundreds of thousands of contigs across the genome (Khudyakov et al. 2015b). As the cost of high-throughput sequencing continues to decline, we anticipate an increase in studies that sequence RRLs, whole genomes, and transcriptomes in NMOs at a population scale.

Marine mammal genomic studies thus far have primarily contributed to the fields of phylogenomics and comparative genomics (Fig 2, Table S1). Several of these comparative genomics studies have aimed to improve our understanding of the mammalian transition to an aquatic lifestyle and describe the evolutionary relationships within and among marine mammals and their terrestrial relatives (McGowen et al. 2014; Foote et al. 2015). Whereas such studies require only a single representative genome per species, an emerging class of studies applying genomic techniques at a population scale enables further investigations of fitness, demography, and population structure within ~~a~~ species (Table S1). However, expanding the scale of genomic studies requires careful selection of an appropriate method for data generation and analysis~~,~~ from a growing number of approaches that are becoming available to non-model systems.

Data generation

Our review of marine mammal genomics highlights an increasing number of options for the generation and analysis of genomic data. Choosing which of these sequencing strategies to apply is a key step in any genomics study. Here~~,~~ we describe approaches that have been used successfully in order to help guide future studies of ecological, physiological, and evolutionary genomics in NMOs. Across data generation methods, we highlight approaches that can be used with limited or no prior genomic resources, overcoming one traditional challenge of genomic studies of NMOs (the need for a reference genome to which sequencing reads can be mapped). These methods produce a range in quantity and type of data output, from hundreds of SNPs to whole genome sequences, and from single individuals to population samples, reflecting the trade-off between number of samples and amount of data generated per sample.

Sample collection, storage and extraction

Prior to starting a genomic study, researchers must recognize that many recent methods for high-throughput sequencing require genetic material of much higher quality and quantity than techniques used to characterize traditional genetic markers. These more stringent sample requirements necessitate new standards for tissue sampling, storage, and DNA/RNA extraction. Ideally, samples should be collected from live or newly deceased individuals and stored at -80°C, or when this is not possible at -20°C in RNAlater, Trizol, ethanol, salt-saturated DMSO, or dry, depending on the intended application. Given the sensitivity of new sequencing methods, great care should be taken to minimize cross-contamination during sampling, as even minute amounts of genetic material from another individual can bias downstream analyses, for example variant genotyping and gene expression profiles. Choice of extraction method varies with sample type and study aim, but typically genomic methods require cleanup and treatment with RNase to yield pure extracts, whereas RNAseq methods require rigorous DNase treatment to remove genomic contamination that can bias expression results. Depending on the genomic methodology, target quantities for a final sample may range from as low as 50 ng of DNA for some RRL sequencing methods (Andrews et al. 2016) up to ~1 mg for sequencing the full set of libraries (of different insert sizes) necessary for high-quality genome assemblies (Ekblom and Wolf 2014). Most commercial RNAseq library preparation services require at least 500-1,000 ng of pure total RNA that shows minimal degradation as measured by capillary gel electrophoresis (RNA Integrity Number (RIN) ≥ 8). Samples should ideally consist of high molecular weight genetic material (with little shearing), though continuing molecular advances enable genomic sequencing even of low quantity or poor quality starting material. Extreme examples of the latter include successfully sequenced whole genomes from ancient material (e.g., Rasmussen et al. 2010;

Meyer et al. 2012; Allentoft et al. 2015), including a more than 500,000-year-old horse (Orlando et al. 2013).

Reduced-representation genome sequencing

i. RADseq

Reduced-representation sequencing methods evaluate only a small portion of the genome, allowing researchers to sequence samples from a larger number of individuals within a given budget in comparison to sequencing whole genomes. Restriction site-associated DNA sequencing (RADseq) is currently the most widely ~~-~~used RRL sequencing method for NMOs (Davey et al. 2011; Narum et al. 2013; Andrews et al. 2016). RADseq generates sequence data from short regions adjacent to restriction cut sites and therefore targets markers that are distributed relatively randomly across the genome and occur primarily in non-coding regions. This method allows simultaneous discovery and genotyping of thousands of genetic markers for virtually any species, regardless of availability of prior genomic resources. Of greatest interest are variable markers, characterized either as single SNPs or phased alleles that can be resolved from the identification of several ~~SNPs~~ variants within a single locus.

The large number of markers generated by RADseq dramatically increases genomic resolution and statistical power for addressing many ecological and evolutionary questions when compared to studies using traditional markers (Table S1). For example, heterozygosity ~~-~~fitness ~~associations~~ correlations in harbor seals (*Phoca vitulina*) were nearly fivefold higher when using 14,585 RADseq SNPs than when using 27 microsatellite loci (Hoffman et al. 2014). A recent study on the Atlantic walrus (*Odobenus rosmarus rosmarus*) using 4,854 RADseq SNPs to model demographic changes in connectivity and effective population size associated with the Last Glacial Maximum (Shafer et al. 2015) both supported and extended inferences from previous studies using traditional markers (Shafer et al. 2010; Shafer et al. 2014).

Furthermore, RADseq can provide sufficient numbers of markers across the genome to identify genomic regions influenced by natural selection~~ in some cases~~. These analyses require large numbers (thousands to tens of thousands) of markers to ensure that some markers will be in linkage disequilibrium with genomic regions under selection and to minimize false positives, particularly under non-equilibrium demographic scenarios (Narum and Hess 2011; De Mita et al. 2013; Lotterhos and Whitlock 2014). Extreme demographic shifts, as experienced by many marine mammal populations (e.g., killer whales, Foote et al. 2016), can drive shifts in allele frequencies that confound the distinction of drift and selection and make it difficult to detect genomic signatures of selection (Poh et al. 2014). Proof of concept of the application of RADseq for identifying genomic signatures of selection in wild populations was demonstrated in three-spined sticklebacks (*Gasterosteus aculeatus*), for which analyses of over 45,000 SNPs (Hohenlohe et al. 2010) identified genomic regions of known evolutionary importance associated with differences between marine and freshwater forms (Colosimo et al. 2005; Barrett et al. 2008). RADseq studies with similar aims in marine mammals have resulted in comparatively sparser sampling of SNPs (<10,000), likely due to both methodological differences and generally low genetic diversity particularly among cetaceans. Nonetheless, genomic regions associated with resistance to harmful algal blooms in common bottlenose dolphins (*Tursiops truncatus*) were identified across multiple pairwise comparisons using 7,431 RADseq SNPs (Cammen et al. 2015), and genomic regions associated with habitat use and resource specialization in killer whales (*Orcinus orca*) were identified using 3,281 RADseq SNPs (Moura et al. 2014a). Some of these RADseq SNPs associated with diet in killer whales were later also confirmed as occurring in genomic regions of high differentiation and reduced diversity consistent with a signature of selection identified in a study utilizing ~~low-coverage~~ whole genome re-sequencing (Foote et al. 2016). It will remain important for further studies of genomic signatures of selection in NMOs to carefully consider which approach~~es~~ will generate a sufficiently large number of SNPs to accurately identify the range of putatively neutral F_{ST} values (and thus outliers) given the demographic history of the population (Lotterhos and Whitlock 2014).

Numerous laboratory methods have been developed for generating RADseq data (reviewed in Andrews et al. 2016), with the most popular library preparation methods currently being the original RAD (Miller et al. 2007; Baird et al. 2008), Genotyping by Sequencing (GBS, Elshire et al. 2011; Poland et al. 2012), and double digest RAD (ddRAD, Peterson et al. 2012). All RADseq methods share the common goal of sequencing regions adjacent to restriction cut sites across the genome, but differ in technical details, such as the number and type of restriction enzymes used, the mechanisms for reducing genomic DNA fragment sizes, and the strategies for attaching sequencing adapters to the target DNA fragments. For example, both the original RAD method and GBS use a single enzyme digest, but the original RAD ~~protocol~~ method uses a rare-cutting enzyme and mechanical shearing to reduce DNA fragment size (Baird et al. 2008), whereas GBS uses a more frequent-cutting enzyme and relies on preferential PCR amplification of shorter fragments for indirect size selection (Elshire et al. 2011). These ~~types of variation~~modifications lead to differences across methods in the time and cost of library preparation, the number and lengths of loci produced, and the types of error and bias present in the resulting data. Different RADseq methods will be better suited to different research questions, study species, and research budgets, and therefore researchers embarking on a RADseq study should carefully consider the suitability of each method for their individual projects. Further details on the advantages and disadvantages of each method are described in Andrews et al. (2016).

ii. SNP arrays

An alternative high-throughput reduced-representation genotyping approach involves the use of custom arrays designed to capture and sequence targeted regions of the genome. Such array-based approaches may provide certain advantages over RADseq, including the ability to easily estimate genotyping error rates, scalability to thousands of samples, lower requirements for DNA quantity/quality and technical effort, greater comparability of markers across studies, and the ability to genotype SNPs within candidate genomic regions. However, unlike RADseq, array-based techniques require prior knowledge of the study system's genome or the genome of a closely related species, which remains unavailable for some NMOs. Furthermore, SNP arrays must take into account the potential for ascertainment bias (e.g., Malenfant et al. 2015), whereas RADseq avoids ascertainment bias by simultaneously discovering and genotyping markers.

To identify SNPs for NMO array development, researchers must rely on existing genomic resources or generate new reference sequences, in the form of whole or reduced-representation genomes or transcriptomes (Hoffman et al. 2012; Malenfant et al. 2015). When a whole genome reference assembly is available for the target species or a related species, multiplex shotgun sequencing can facilitate the rapid discovery of hundreds of thousands of SNPs for array development. This SNP discovery approach involves high-throughput sequencing of sheared genomic DNA~~, which~~ that can be sequenced at a low depth of coverage (i.e., low mean read depth across the genome) if suitable genotype likelihood-based methods (O'Rawe et al. 2015) are used to identify polymorphic sites. Thus, this approach is less restrictive in terms of DNA quality. For example, shotgun sequencing of 33 Northeast Atlantic common bottlenose dolphins, which included degraded DNA collected from stranded specimens, on one Illumina HiSeq2000 lane of 100-bp single-end sequencing identified 440,718 high-quality SNPs (M. Louis unpublished data). Such dense sampling of SNPs is essential for studies of population genomics that require a large number of markers, such as for inferences of demographic history (Gutenkunst et al. 2009; Excoffier et al. 2013; Liu and Fun 2015) and selective sweeps (Chen et al. 2010). Once a set of putative markers has been identified, hybridization probes can be designed from their flanking sequences and printed onto a SNP array. The two principal SNP genotyping platforms supporting thousands to millions of SNPs are the Illumina Infinium iSelect® and Affymetrix Axiom® arrays.

The use of SNP arrays in NMOs has thus far been somewhat limited, potentially due to low SNP validation rates (Chancerel et al. 2011; Helyar et al. 2011), issues of ascertainment bias (Albrechtsen et al. 2010; McTavish and Hillis 2015), and cost of SNP discovery. However, using both SNP data and whole genome sequence from the Antarctic fur seal (*Arctocephalus gazella*), Humble et al. (2016) recently demonstrated that careful filtering based on SNP genomic context prior to array development has the potential to substantially increase assay success rates. Further, ascertainment bias can be reduced by selecting samples for SNP discovery that span the geographic range of populations that will be target-sequenced (Morin et al. 2004). By accounting for ascertainment bias, Malenfant et al. (2015) were able to demonstrate population structure in Canadian polar bears (*Ursus maritimus*) more clearly using a 9K SNP array than 24 microsatellite markers.

iii. Target sequence capture

Target sequence capture (TSC, also called target enrichment, direct selection, or Hyb-seq) has many of the same advantages and disadvantages as the array-based SNP approaches described above, but differs in library preparation, sequencing platform, and resulting sequence data. While SNP arrays genotype single variable positions, TSC can be used to sequence selected short fragments. With TSC, researchers can amplify and sequence up to a million target probes on solid-state arrays, and even more if in-solution arrays are used. This gives the user the ability to choose to sequence many samples in parallel (Cummings et al. 2010), as many as 100-150 per Illumina HiSeq lane, or to sequence many regions per individual. Recent advances in target enrichment, such as genotyping in thousands (Campbell et al. 2015), anchored hybrid enrichment (Lemmon et al. 2012), and target capture of ultra~~-~~conserved elements (UCEs, Faircloth et al. 2012; McCormack et al. 2012), have further increased the number of regions and individuals that can be sampled in a single lane. In addition, UCEs overcome the need for a reference genome, enabling their wide application across many NMOs (though designing custom probe sets from closely related species will remain preferable in many cases (Hancock-Hanser et al. 2013)). Although a number of methodological variants have been developed and optimized (Bashiardes et al. 2005; Noonan et al. 2006; Hodges et al. 2009; Cummings et al. 2010; Mamanova et al. 2010; Hancock-Hanser et al. 2013), TSC generally relies on hybridization and amplification of specially prepared libraries consisting of fragmented genomic DNA. Many companies offer kits for TSC, such as Agilent (SureSelect) and MYcroarray (MYbaits), with MYcroarray specifically marketing their kits for use with NMOs.

The most common use of TSC has been the capture of whole exomes in model organisms, including humans (Ng et al. 2009). However, as costs have plummeted, TSC is increasingly being used in investigations of NMOs. TSC is particularly useful in sequencing ancient DNA, where it can enrich the sample for endogenous DNA content relative to exogenous DNA (i.e., contamination) and thereby increase the relative DNA yield (Ávila-Arcos et al. 2011; Enk et al. 2014). For example, TSC has been used to generate mitogenome sequences from subfossil killer whale specimens originating from the mid-Holocene~~,~~ for comparison with modern lineages (Foote et al. 2013). TSC was also recently utilized to compare >30 kb of exonic sequence from museum specimens of the extinct Steller’s sea cow (*Hydrodamalis gigas*) and a modern dugong (*Dugong dugon*) specimen to investigate evolution within Sirenia (Springer et al. 2015). Springer et al. (2016) further used TSC to examine gene evolution related to dentition across edentulous mammals, including mysticetes. Finally, TSC of both exonic and intronic regions has been used to assess genetic divergence across cetacean species (Hancock-Hanser et al. 2013; Morin et al. 2015). These studies show the potential use of TSC across evolutionary time~~-~~scales for population genomics, phylogenomics, and studies of selection and gene loss across divergent lineages (Table S1).

Whole genome sequencing

Beyond advances enabled by the reduced-representation methods presented above, our power and resolution to elucidate evolutionary processes, including selection and demographic shifts, can be further increased by sequencing whole genomes.

i. ~~High-coverage R~~reference genome sequencing

At the time of publication, there ~~exist~~ are 12 publicly available[1] whole (or near-whole) marine mammal genomes of varying quality representing 10 families, including 7 cetaceans (Fig 1A), 3 pinnipeds (Fig 1B), the West Indian manatee (*Trichechus manatus*), and the polar bear. The first sequenced marine mammal genome was that of the common bottlenose dolphin, which was originally sequenced to ~2.5x depth of coverage using Sanger sequencing (Lindblad-Toh et al. 2011). This genome was later improved upon by adding both 454 and Illumina HiSeq data (Foote et al. 2015). Other subsequent marine mammal genomes were produced solely using Illumina sequencing and mate-paired or paired-end libraries with varied insert sizes (Miller et al. 2012; Zhou et al. 2013; Yim et al. 2014; Foote et al. 2015; Keane et al. 2015; Kishida et al. 2015; Humble et al. 2016).

Whole genome sequencing has been used to address many issues in marine mammal genome evolution, usually by comparison with other existing mammalian genomes. Biological insights discussed in the genome papers listed above include the evolution of transposons and repeat elements, gene evolution and positive selection, predicted population structure through time, SNP validation, molecular clock rates, and convergent molecular evolution (Table S1). For example, analyses of the Yangtze river dolphin (*Lipotes vexillifer*) genome confirmed that a bottleneck occurred in this species during the last period of deglaciation (Zhou et al. 2013). In addition, following upon earlier smaller-scale studies (e.g., Deméré et al. 2008; McGowen et al.

[1] These genomes are available on NCBI's online genome database or Dryad, but they have not all been published. As agreed upon in the Fort Lauderdale Convention, the community standard regarding such unpublished genomic resources is to respect the data generators' right to publish with these data first.

2008; Hayden et al. 2010), genomic analyses have confirmed the widespread decay of gene families involved in olfaction, gustation, enamelogenesis, and hair growth in some cetaceans (Yim et al. 2014; Kishida et al. 2015). Perhaps the most widespread use of whole genome studies has been the use of models of selection to detect protein-coding genes that show evidence of natural selection in specific lineages. A recent study by Foote et al. (2015) ~~has~~ extended this approach to investigate convergent positive selection among cetaceans, pinnipeds, and sirenians. This study exemplifies a trend in recent genomic studies~~, which~~ that sequence multiple genomes to address a predetermined evolutionary question, in this case, the molecular signature of aquatic adaptation.

In addition to these evolutionary insights that typically stem from a comparative genomics approach, the development of high-~~-~~quality reference genome assemblies provide an important resource that facilitates mapping of reduced-representation genomic data (see previous section) as well as ~~relatively low-coverage,~~ short-read sequencing data with relatively low depth of coverage (see following section). These data types can be generated at relatively low cost on larger sample sizes enabling population-~~-~~scale genomic studies. In many cases, genome assemblies from closely related species are sufficient for use as a reference. Particularly among marine mammals, given their generally slow rate of nucleotide divergence, it is therefore likely unnecessary to sequence a high-~~-~~quality reference genome assembly for every species. Instead, resources could be allocated toward population-~~-~~scale studies, including ~~low-coverage~~ genome re-sequencing efforts.

ii. Population-level ~~low-coverage~~ genome re-sequencing

In contrast to ~~high-coverage~~reference genome sequencing that today often exceeds 100x mean read ~~coverage~~ depth and typically combines long- and short-insert libraries to generate high-quality assemblies for one to a few individuals, ~~low-coverage~~ genome re-sequencing studies ~~capitalize on existing reference assemblies and~~ aim to achieve only ≥2x ~~coverage~~ mean read depth on tens to hundreds of individuals from short-insert libraries ~~which~~ that ~~are then~~whose reads are anchored to ~~the~~ existing reference assemblies~~y~~. ~~Given~~ Despite the inherent trade-offs between cost, read depth, coverage, and sample size, ~~low-coverage~~ genome re-sequencing of large numbers of individuals for population-level inference can be conducted at a relatively low cost. In the past five years, several influential studies have used genome re-sequencing to advance our understanding of the genomic underpinnings of different biological questions in model systems. For example, population genomics of *Heliconius* butterflies highlighted the exchange of genes between species that exhibit convergent wing patterns (The *Heliconius* Genome Consortium 2012); whole genome re-sequencing of three-spined sticklebacks highlighted the re-use of alleles in replicated divergences associated with ecological speciation and local adaptation (Jones et al. 2012); and combined population genomics and phylogenomics have identified regions of the genome associated with variation in beak shape and size in Darwin's finches (Lamichhaney et al. 2015).

To date only two marine mammal population genomics studies using whole genome re-sequencing have been published. These studies involved re-sequencing the genomes of 79 individuals from three populations of polar bears (Liu et al. 2014a) and 48 individuals from five evolutionarily divergent ecotypes of killer whale (Foote et al. 2016). The findings of Foote et al. (2016) confirmed results of population differentiation that had previously been established using traditional genetic markers (Morin et al. 2010a). However, the study also provided new insights into the demographic history, patterns of selection associated with ecological niche, and evidence of episodic ancestral admixture that could not have been obtained using traditional markers.

Several new resources have made such population genomic studies economically possible for a greater number of NMOs, including the availability of~~ a~~ reference genome assemblies~~y~~ (see section above), relatively low-cost high-throughput sequencing (further increases in throughput expected with the new Illumina HiSeq X Ten (van Dijk et al. 2014)), and crucially, the development of likelihood-based methods that allow estimation of population genetic metrics from ~~low-coverage ~~re-sequencing data (Fumagalli et al. 2013; O'Rawe et al. 2015). One last consideration is the ease of laboratory methods necessary to generate whole genome re-sequencing data when compared to other methods such as RADseq or TSC. DNA simply needs to be extracted from the samples and, using proprietary kits, built into individually index-amplified libraries~~ using proprietary kits, which~~ that are ~~then ~~equimolarly pooled and submitted for sequencing.

Many population genomic analyses are based on the coalescent model that gains most information from the number of independent genetic markers, not the number of individuals sampled. Sample sizes of ~10 individuals are usually considered sufficient (Robinson et al. 2014) and have been standard in many genome-wide studies in the eco-evolutionary sciences (Ellegren et al. 2012; Jones et al. 2012). Thus, sampling fewer individuals ~~at lower coverage but for orders of magnitude more data~~by whole genome re-sequencing is a salient approach~~, which~~ that allows us to consider many more gene trees, whilst continuing to provide robust estimates of per-site genetic metrics (e.g., F_{ST}). The robustness of inference from ~~low-coverage~~ data with low mean read depth across the genome was recently confirmed using a comparison of per-site F_{ST} estimates for the same sites from high~~- coverage~~depth (≥20x) RADseq data and low-~~coverage~~depth (≈2x) whole genome re-sequencing data in pairwise comparisons between the same two killer whale ecotypes (Foote et al. 2016).

Beyond the increased power afforded by sequencing more polymorphic sites, whole genome re-sequencing also allows inference of demographic history from the genome of even just a single individual by identifying Identical By Descent (IBD) segments and runs of homozygosity (Li and Durbin 2011; Harris and Nielsen 2013). For example, Liu et al. (2014a) found evidence for ongoing gene flow from polar bears into brown bears after the two species initially diverged. Genome re-sequencing of sufficient numbers of individuals also facilitates haplotype phasing, which has many applications, including the detection of ongoing selective sweeps (Ferrer-Admetlla et al. 2014) and the inference of demographic history of multiple populations based on coalescence of pairs of haplotypes in different individuals (Schiffels and Durbin 2014). However, haplotype phasing ~~has~~ typically require~~s~~d genomic ~~higher coverage~~ data with higher mean read depth (~20x) from tens of individuals (though recent advances in genotype imputation suggest success with ~~lower coverage~~ data of lower mean read depth (VanRaden et al. 2015)). Thus far, phasing has been restricted to relatively few NMO studies, and no marine mammal studies to the best of our knowledge.

Transcriptome sequencing

In comparison with the DNA-based genomic approaches described above, RNA-based genomic approaches are a relatively new and emerging application in NMOs such as marine mammals.

Transcriptomics by RNA sequencing (RNAseq) can rapidly generate vast amounts of information regarding genes and gene expression without any prior genomic resources. This approach can resolve differences in global gene expression patterns between populations, individuals, tissues, cells, and physiological or environmental conditions, and can yield insights into the molecular basis of environmental adaptation and speciation in wild animals (Wolf 2013; Alvarez et al. 2015). Furthermore, RNAseq is a valuable tool for resource development, for example as a precursor to designing SNP and TSC arrays (e.g., Hoffman et al. 2012). However, applying RNAseq to NMOs requires several unique considerations in comparison to the DNA-based methods described above. Most importantly, the labile nature of gene transcription and high detection sensitivity of RNAseq have the potential to amplify transcriptional "noise" and are thus extremely sensitive to experimental design.

If the experimental goal is to capture a comprehensive transcriptome profile for a study organism, multiple tissues from individuals of varied life history stages should be sampled. However, if the aim is to characterize transcriptional responses to physiological or environmental stimuli, efforts should focus on minimizing variability in individuals and sampling conditions (Wolf 2013). For differential expression analyses, pairwise comparisons should be made within the same individual if at all possible (e.g., before and after treatment, between two developmental stages). As RNAseq only captures a 'snapshot' of gene expression in time, repeated sampling or time-course studies are necessary to obtain a more complete picture of cellular responses to the condition(s) in question (Spies and Ciaudo 2015). Sampling and sequencing depth requirements will depend on the study design. Simulation studies have shown that a minimum of 5-6 biological replicates sequenced at a depth of 10-20 million reads per sample is necessary for differential expression analysis (Liu et al. 2014b; Schurch et al. 2015). RNAseq can also be used for biomarker development to expand molecular toolkits for NMOs without sequenced genomes (Hoffman et al. 2013). In this case, higher sequencing depths of 30-60 million reads per sample are recommended for SNP discovery and genotyping (De Wit et al. 2015).

Following sequence generation, transcript annotation remains a challenge for NMOs without reference transcriptomes or genomes. *De novo* transcriptomes can be annotated through detection of assembled orthologs of highly conserved proteins, but these analyses remain limited by the quality of reference databases. As a result, NMO transcriptomes are biased in favor of highly conserved terrestrial mammal genes and therefore provide an incomplete understanding of animal adaptations to natural environments (Evans 2015). For example, while 70.0% of northern elephant seal (*Mirounga angustirostris*) skeletal muscle transcripts had BLASTx hits to mouse genes, only 54.1% of blubber transcripts could be annotated due to poor representation of this tissue in terrestrial mammal reference proteomes (Khudyakov et al. 2015b).

To date, RNAseq has been used for gene discovery and phylogenomics analyses in Antarctic fur seal (Hoffman 2011; Hoffman et al. 2013), polar bear (Miller et al. 2012), Indo-Pacific humpback dolphin (*Sousa chinensis* (Gui et al. 2013)), spotted seal (*Phoca largha* (Gao et al. 2013)), bowhead whale (*Balaena mysticetus* (Seim et al. 2014)), narrow-ridged finless porpoise (*Neophocaena asiaeorientalis* (Ruan et al. 2015)), and humpback whale (*Megaptera novaeangliae* (Tsagkogeorga et al. 2015)) (Table S1). Due to the challenges of repeated sampling of wild marine mammals, few studies have examined cetacean or pinniped transcriptome responses to environmental or experimental stimuli. The majority of such functional gene expression studies have used microarrays (Mancia et al. 2008; Mancia et al. 2012; Mancia et al. 2015); however, RNAseq has been employed to profile sperm whale (*Physeter macrocephalus*) skin cell response to hexavalent chromium (Pabuwal et al. 2013) and free-ranging northern elephant seal skeletal muscle response to an acute stress challenge (Khudyakov et al. 2015a; Khudyakov et al. 2015b). With decreasing sequencing costs and improvements in bioinformatics tools, RNAseq has the potential to accelerate molecular discoveries in marine mammal study systems and supplement existing functional genomics approaches.

Emerging techniques

In addition to the relatively proven NMO genomic data generation techniques described above, a suite of emerging techniques is entering the field, with exciting promise for exploration of existing and new research areas. For example, high-throughput shotgun sequencing is increasingly being used to identify genetic material from multiple species in a single sample (metagenomics and metatranscriptomics), rather than focus on characterizing variation in a single target individual. These multi-species approaches can be used, for example, to characterize diet from fecal samples (Deagle et al. 2009) and to investigate microbiomes (Nelson et al. 2015), objectives with implications for improving our understanding of both basic ecology and health in natural populations of NMOs. Furthermore, high-throughput sequencing of environmental DNA dramatically increases the throughput of NMO detection in environmental (e.g., seawater) samples (Thomsen et al. 2012), using degenerate primers for multi-species detection rather than requiring the design and implementation of numerous single-species protocols (Foote et al. 2012).

A second broad area of emerging interest moves beyond the study of variation at the DNA and RNA levels to examine epigenetic effects of histone modification on gene regulation and evolution. Epigenomic studies often examine changes in DNA methylation in association with processes such as cancer and ageing. Such approaches, from targeted gene to genome-wide, have only very recently and not yet frequently been applied in NMOs. Polanowski et al. (2014) used a targeted gene approach to examine changes in DNA methylation in age-associated genes, previously identified in humans and mice, in humpback whales of known age. The most informative markers were able to estimate humpback whale ages with standard deviations of approximately 3-5 years, demonstrating the potential transferability of these approaches from model to non-model organism. Villar et al. (2015) utilized a genome-wide approach – chromatin immunoprecipitation followed by high-throughput sequencing (ChIPseq) – to examine gene-regulatory element evolution across mammals, including four species of cetaceans. This study identified highly conserved gene-regulatory elements based on their histone modifications (H3K27ac and H3K4me3), showed that recently evolved enhancers were associated with genes under positive selection in marine mammals, and identified unique *Delphinus*-specific enhancers. Finally, reduced-representation epigenomic approaches have also been developed (Gu et al. 2011), and although they have not yet been used in marine mammals to our knowledge, these techniques could facilitate future studies of how changes in DNA methylation patterns affect other biological processes, such as stress levels or pregnancy.

Data analysis

Following the generation of genomic data, researchers must select the most appropriate genomic analysis (i.e., bioinformatics) pipelines, which often differ significantly from those used in traditional genetic studies of NMOs. The choice of analysis pipeline will depend on multiple factors including the availability of a reference genome, the level of diversity within the dataset (e.g., single- or multi~~-ple ~~species), the type of data generated (e.g., single~~-end vs.~~- or paired-end), and the computing resources available. The computational needs, both in terms of hardware and competency in computer science, for analysis of genomic data typically far exceed those necessary for traditional genetic markers. On the smaller end of the spectrum, one lane of 50 bp single-end sequencing on an Illumina HiSeq 2500 can produce tens of gigabytes of data, while data files associated with a single high-~~coverage ~~quality vertebrate genome may reach hundreds of gigabytes in size (Ekblom and Wolf 2014). Computing resources necessary for the analysis of these genomic datasets can range from ~10 gigabytes for a pilot study using a reduced-representation sequencing approach to over a terabyte for whole ~~-~~genome sequence assembly (Ekblom and Wolf 2014). Fortunately, university computing clusters, cloud-based (Stein 2010) and high-performance computing clusters (e.g., XSEDE; Towns et al. 2014), and open web-based platforms for genomic research (e.g., Galaxy; Goecks et al. 2010) are becoming increasingly accessible. Furthermore, new pipelines are continuously being developed and improved, and there are a growing number of resources aimed at training molecular ecologists and evolutionary biologists in computational large-scale data analysis (Andrews and Luikart 2014; Belcaid and Toonen 2015; Benestan et al. 2016). We provide ~~a limited~~an indicative list of the current, most commonly used analysis pipelines that are specific to each data generation method in ~~Supplemental ~~Table 1~~2~~. Here~~,~~ we briefly summarize current genomic data analysis pipelines and discuss considerations that are likely to be similar across multiple data generation methods.

Genomic data analysis often involves multiple steps, and the choice of analysis tool for each step can greatly affect the outcome, with different tools producing different (though usually overlapping) sets of results (e.g., Schurch et al. 2015). All analyses begin by evaluating data quality, trimming sequences if necessary to remove erroneous nucleotides (MacManes 2014), and implementing appropriate data quality filters (e.g., phred scores, read length, and/or read depth). Raw reads also need to be demultiplexed based on unique barcodes if pools of individuals were sequenced in a single lane. Analyses then usually proceed in a *de novo* or genome-enabled manner, depending on available resources. Briefly, sequences can be compared (e.g., to identify variants) by mapping all reads to a reference genome or *de novo* assembling stacks of sequences putatively derived from the same locus~~,~~ based on sequence similarity. *De novo* methods are sensitive to sequencing error, as well as true genetic variation, and therefore can erroneously assemble polymorphic sequences as separate loci or transcripts, requiring further filtering to remove redundancy. The opposite problem can also occur in both *de novo* and reference mapping approaches, where two distinct loci (e.g., paralogous loci) may assemble as a single locus or map to the same reference location. Researchers should therefore recognize the inherent trade-offs when carefully selecting their thresholds for acceptable levels of variation within and among loci.

Considerations relevant to the selection of subsequent downstream analyses are specific to the type of data generated and the research objective. For example, RADseq analysis pipelines differ in the algorithms used to genotype variants (Table 1~~S2~~). Similarly, there are several gene expression analysis pipelines for RNAseq data that compare transcript abundance between samples (Table 1~~S2~~). Analysis of TSC data usually uses standard *de novo* assemblers (e.g., Trinity, Velvet); these assemblers can be run using packages such as PHYLUCE (Faircloth 2015), which is designed specifically for use with ultraconserved elements. Unfortunately, for most analyses, there are no unifying recommendations currently available and researchers must evaluate several approaches, each with their own advantages and disadvantages, in order to select the most appropriate tool for their particular experiment and system. Furthermore, we can expect that the recommendations for analysis tools will continue to evolve as new programs become available in the future.

Guidelines for data quality control and sharing

With rapid growth in sequencing platforms and bioinformatics analysis pipelines comes the need to extend existing principles (e.g., Bonin et al. 2004) on quality control, analysis, and transparency. General recommendations for sample and data handling, library preparation, and sequencing have been discussed elsewhere (Paszkiewicz et al. 2014). We therefore focus on the need to produce guidelines on data quality evaluation and reporting for genomic data (e.g., Morin et al. 2010b). A primary challenge in this area is that quality metrics vary widely across sequencing technologies. Yet, regardless of sequencing platform, the quality of sequencing reads must be evaluated (e.g., using FastQC; Andrews 2010) and reported.

Best practices guidelines for ~~high-coverage whole~~reference genome sequencing and RNAseq data generation, analysis, and reporting are available from the human-centric ENCODE consortium (www.encodeproject.org). These include minimum depth of sequencing and number and reproducibility of biological replicates. For RNAseq experiments, evaluation of *de novo* assembly quality remains a challenge. Suggested quality metrics include percentage of raw reads mapping back to the assembly and number of assembled transcripts with homology to known proteins (MacManes 2016). Emerging tools such as Transrate (Smith-Unna et al. 2015) attempt to integrate these and other metrics into a comprehensive assembly quality score.

In contrast, there is not yet any standard way to estimate or report error rates with RADseq or ~~low-coverage~~ genome re-sequencing methods (but see Mastretta-Yanes et al. 2015; Fountain et al. 2016). Recommendations to improve confidence in genotyping include using methods that account for population--level allele frequencies when calling individual genotypes, mapping reads to reference genomes rather than *de novo* assembly (Nadeau et al. 2014; Fountain et al. 2016), filtering out PCR duplicates (Andrews et al. 2014), identifying and removing markers in possible repeat regions, and filtering data to include only those with high read depth (>10-20x per locus per individual) (Nielsen et al. 2011). Other analysis methods, such as robust Bayesian methods and likelihood--based approaches that account for read quality in calculations of posterior probabilities of genotypes and per-site allele frequencies utilizing the sample mean site frequency spectrum as a prior (Fumagalli et al. 2013), can account for uncertainty and/or error in the data, and are therefore suitable for use with low to moderate read depths (2-20x per locus; e.g., Han et al. 2015; O'Rawe et al. 2015).

Due to the large number of analysis tools that are available, data quality and reproducibility ultimately depend on methods and data transparency. All raw sequencing reads should be publicly archived, for example deposited in the NCBI Sequence Read Archive. Many journals, including the *Journal of Heredity* (Baker 2013), now also require that primary data supporting the published results and conclusions (e.g., SNP genotypes, assemblies) be publicly archived in online data repositories (e.g., Dryad). We further recommend making public the analysis pipelines, scripts (e.g., using GitHub), and additional outputs, as appropriate, in order for analyses to be fully reproducible and transparent, which is the cornerstone of the scientific method (Nosek et al. 2015).

Future directions

As demonstrated here for one group of mammalian taxa, the rapid growth of the field of non-model genomics has been both impressive and empowering. As we approach a point of relative saturation in reference genomes, we anticipate an increase in population-scale genomic studies that produce lower depth or coverage datasets per individual but across larger sample sizes ~~relative to high-coverage sequencing of a few individuals of each species~~. In addition (or alternatively), we hope to see increasing efforts to sequence reference transcriptomes and improve NMO genome annotation in ways beyond the inherently limited approach of comparison to gene lists from a few model organisms. Population-scale genomic studies will facilitate greater ecological understanding of natural populations, while efforts to improve annotation will address persistent limitations in our understanding of gene function for NMOs. Ultimately, improving our understanding of local adaptation, adaptive potential, and demographic history through the use of genomic toolkits such as those described here is likely to have important implications for the future conservation of these populations.

Advances in sequencing technologies and analytical tools will no doubt continue, in some cases drawing on established techniques in model organisms, posing both new opportunities and new challenges for researchers in NMO genomics. Likely the most persistent challenge will remain selecting the data generation and experimental design that is most appropriate for the respective research objective. Our review identified few cases that exhibit relative dominance of a single methodology and analytical pipeline (e.g., RADseq and STACKS, RNAseq and Trinity); rather, more often we found a diversity of approaches even within each category of data generation. In fact, such diversity of approaches has its benefits, with each approach promoting its own advantages (and limitations). Overall, our reflections on lessons learned from the past decade of NMO genomics in one well-studied group of mammalian taxa clearly demonstrate the value, increased ease, and future promise of applying genomic techniques across a wide range of non-model species to gain previously unavailable insights into evolution, population biology, and physiology on a genome-wide scale.

Acknowledgements

This review paper is the outcome of two international workshops held in 2013 and 2015 on marine mammal genomics. The workshops were organized by KMC, AF, and C. Scott Baker and hosted by the Society for Marine Mammalogy, with support from a Special Event Award from the American Genetic Association. We sincerely thank all the workshop participants for their contributions to inspiring discussions on marine mammal genomics. We would also like to thank two anonymous reviewers and C. Scott Baker for their helpful feedback on an earlier version of this manuscript. Illustrations are by C. Buell with permission for use granted by J. Gatesy.

Funding

The authors involved in this work were supported by a National Science Foundation Postdoctoral Research Fellowship in Biology (Grant No. 1523568) to KMC; an Office of Naval Research Award (No. N00014-15-1-2773) to JIK; a Marie Slodowska Curie Fellowship to ELC (Behaviour-Connect) funded by the EU Horizon2020 program; Royal Society Newton International Fellowships to ELC and MRM; a Deutsche Forschungsgemeinschaft studentship to EH; a Fyssen Foundation postdoctoral fellowship to ML; postdoctoral funding from the University of Idaho College of Natural Resources to KRA; a short visit grant from the European Science Foundation-Research Networking Programme ConGenOmics to ADF; and a Swiss National Science Foundation grant (31003A-143393) to L. Excoffier that further supported ADF. The first marine mammal genomics workshop we held to begin discussions towards this review was supported by a Special Event Award from the American Genetic Association.

References

Albrechtsen A, Nielsen FC, Nielsen R. 2010. Ascertainment biases in SNP chips affect measures of population divergence. *Mol Biol Evol*. 27:2534-2547.

Alexander A, Steel D, Hoekzema K, Mesnick S, Engelhaupt D, Kerr I, Payne R, Baker CS. 2016. What influences the worldwide genetic structure of sperm whales (*Physeter macrocephalus*)? *Mol Ecol*.

Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlstrom T, Vinner L*, et al.* 2015. Population genomics of Bronze Age Eurasia. *Nature*. 522:167-172.

Alvarez M, Schrey AW, Richards CL. 2015. Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? *Mol Ecol*. 24:710-725.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol*. 11:R106.

Andrews K, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet*. 17:81-92.

Andrews KR, Hohenlohe PA, Miller MR, Hand BK, Seeb JE, Luikart G. 2014. Trade-offs and utility of alternative RADseq methods: Reply to Puritz *et al*. 2014. *Mol Ecol*. 23:5943-5946.

Andrews KR, Luikart G. 2014. Recent novel approaches for population genomics data analysis. *Mol Ecol*. 23:1661-1667.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Ankeny RA, Leonelli S. 2011. What's so special about model organisms? *Studies in History and Philosophy of Science*. 42:313-323.

Armengaud J, Trapp J, Pible O, Geffard O, Chaumot A, Hartmann EM. 2014. Non-model organisms, a species endangered by proteogenomics. *J Proteomics*. 105:5-18.

Arnason U, Adegoke JA, Bodin K, Born EW, Esa YB, Gullberg A, Nilsson M, Short RV, Xu X, Janke A. 2002. Mammalian mitogenomic relationships and the root of the eutherian tree. *Proc Natl Acad Sci USA*. 99:8151-8156.

Arnason U, Gullberg A, Widegren B. 1991. The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. *J Mol Evol*. 33:556-568.

Ávila-Arcos M, Cappellini E, Romero-Navarro JA, Wales N, Moreno-Mayar JV, Rasmussen M, Fordyce SL, Montiel R, Vielle-Calzada J-P, Willerslev E*, et al.* 2011. Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Sci Rep*. 1:74.

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:e3376.

Baker CS. 2013. *Journal of Heredity* adopts Joint Data Archiving Policy. *J Hered*. 104:1.

Barrett RDH, Rogers SM, Schluter D. 2008. Natural selection on a major armor gene in threespine stickleback. *Science*. 322:255-257.

Bashiardes S, Veile R, Helms C, Mardis ER, Bowcock AM, Lovett M. 2005. Direct genomic selection. *Nat Methods*. 2:63-69.

Belcaid M, Toonen RJ. 2015. Demystifying computer science for molecular ecologists. *Mol Ecol*. 24:2619-2640.

Benestan LM, Ferchaud A-L, Hohenlohe PA, Garner BA, Naylor GJP, Baums IB, Schwartz MK, Kelley JL, Luikart G. 2016. Conservation genomics of natural and managed populations: building a conceptual and practical framework. *Mol Ecol*.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina Sequence Data. *Bioinformatics*. 30:2114-2120.

Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P. 2004. How to track and assess genotyping errors in population genetics studies. *Mol Ecol*. 13:3261-3273.

Brown CT, Howe A, Zhang Q, Pyrkosz AB, Brom TH. 2012. A reference-free algorithm for computational normalization of shotgun sequencing data. *arXive*. 1203:4802.

Cammen KM, Schultz TF, Rosel PE, Wells RS, Read AJ. 2015. Genomewide investigation of adaptation to harmful algal blooms in common bottlenose dolphins (*Tursiops truncatus*). *Mol Ecol*. 24:4697-4710.

Campbell NR, Harmon SA, Narum SR. 2015. Genotyping-in-Thousands by sequencing (GT-seq): a cost effective SNP genotyping method based on custom amplicon sequencing. *Mol Ecol Resour*. 15:855-867.

Carroll EL, Baker CS, Watson M, Alderman R, Bannister J, Gaggiotti OE, Gröcke DR, Patenaude N, Harcourt R. 2015. Cultural traditions across a migratory network shape the genetic structure of southern right whales around Australia and New Zealand. *Sci Rep*. 5:16182.

Catchen JM, Amores A, Hohenlohe PA, Cresko WA, Postlethwait JH. 2011. *Stacks*: building and genotyping loci *de novo* from short-read sequences. *G3*. 1:171-182.

Catchen JM, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool set for population genomics. *Mol Ecol*. 22:3124-2140.

Chancerel E, Lepoittevin C, Le Provost G, Lin Y-C, Jaramillo-Correa JP, Eckert AJ, Wegrzyn JL, Zelenika D, Boland A, Frigerio J-M*, et al.* 2011. Development and implementation of a highly-multiplexed SNP array for genetic mapping in maritime pine and comparative mapping with loblolly pine. *BMC Genomics*. 12:368.

Chen H, Patterson N, Reich D. 2010. Population differentiation as a test for selective sweeps. *Genome Res*. 20:393-402.

Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*. 307:1928-1933.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 21:3674-3676.

Corander J, Majander KK, Cheng L, Merilä J. 2013. High degree of cryptic population differentiation in the Baltic Sea herring *Clupea harengus*. *Mol Ecol*. 22:2931-2940.

Cummings N, King R, Rickers A, Kaspi A, Lunke S, Haviv I, Jowett JBM. 2010. Combining target enrichment with barcode multiplexing for high throughput SNP discovery. *BMC Genomics*. 11:641.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet*. 12:499-510.

De Mita S, Thuillet A-C, 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.

De Wit P, Pespeni MH, Palumbi SR. 2015. SNP genotyping and population genomics from expressed sequences - current advances and future possibilities. *Mol Ecol*. 24:2310-2323.

Deagle BE, Kirkwood R, Jarman SN. 2009. Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Mol Ecol*. 18:2022-2038.

Deméré TA, McGowen MR, Berta A, Gatesy J. 2008. Morphological and molecular evidence for a stepwise evolutionary transition from teeth to baleen in mysticete whales. *Syst Biol*. 57:15-37.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M*, et al.* 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 43:491-498.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 29:15-21.

Eaton DAR. 2014. PyRAD: assembly of *de novo* RADseq loci for phylogenetic analysis. *Bioinformatics*. 30:1844-1849.

Ekblom R, Galindo J. 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*. 107:1-15.

Ekblom R, Wolf JBW. 2014. A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*. 7:1026-1042.

Ellegren H. 2014. Genome sequencing and population genomics in non-model organisms. *Trends Ecol Evol*. 29:51-63.

Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, Künstner A, Mäkinen H, Nadachowska-Brzyska K, Qvarnström A*, et al.* 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*. 491:756-760.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*. 6:e19379.

Enk J, Devault A, Kuch M, Murgha Y, Rouillard J-M, Poinar H. 2014. Ancient whole genome enrichment using baits built from modern DNA. *Mol Biol Evol*. 31:1292-1294.

Evans TG. 2015. Considerations for the use of transcriptomics in identifying the 'genes that matter' for environmental adaptation. *J Exp Biol*. 218:1925-1935.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genetics*. 9:e1003905.

Faircloth BC. 2015. PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*. 32:786-788.

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.

Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. 2014. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol Biol Evol*. 31:1275-1291.

Flicek P, Birney E. 2009. Sense from sequence reads: methods for alignment and assembly. *Nat Methods*. 6:S6-S12.

Foote AD, Liu Y, Thomas GWC, Vinař Ts, Alföldi J, Deng J, Dugan S, van Elk CE, Hunter ME, Joshi V*, et al.* 2015. Convergent evolution of the genomes of marine mammals. *Nat Genet*. 47:272-275.

Foote AD, Newton J, Ávila-Arcos MC, Kampmann M-L, Samaniego JA, Post K, Rosing-Asvid A, Sinding M-HS, Gilbert MTP. 2013. Tracking niche variation over millennial timescales in sympatric killer whale lineages. *Proc R Soc Lond B Biol Sci*. 280:20131481.

Foote AD, Thomsen PF, Sveegaard S, Wahlberg M, Kielgast J, Kyhn LA, Salling AB, Galatius A, Orlando L, Gilbert MTP. 2012. Investigating the potential use of environmental DNA (eDNA) for genetic monitoring of marine mammals. *PLoS One*. 7:e41781.

Foote AD, Vijay N, Ávila-Arcos M, Baird RW, Durban JW, Fumagalli M, Gibbs RA, Hanson MB, Korneliussen TS, Martin MD*, et al.* 2016. Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nat Commun*. 7:11693.

Fountain ED, Pauli JN, Reid BN, Palsbøll PJ, Peery MZ. 2016. Finding the right coverage: the impact of coverage and sequence quality on single nucleotide polymorphism genotyping error rates. *Mol Ecol Resour*.

Fumagalli M, Vieira FG, Korneliussen TS, Linderoth T, Huerta-Sánchez E, Albrechtsen A, Nielsen R. 2013. Quantifying population genetic differentiation from next-generation sequencing data. *Genetics*. 195:979-992.

Fumagalli M, Vieira FG, Linderoth T, Nielsen R. 2014. *ngsTools*: methods for population genetics analyses from Next-Generation Sequencing data. *Bioinformatics*. 30:1486-1487.

Gao X, Han J, Lu Z, Li Y, He C. 2013. *De novo* assembly and characterization of spotted seal *Phoca largha* transcriptome using Illumina paired-end sequencing. *Comp Biochem Physiol D Genom Proteom*. 8:103-110.

Garner BA, Hand BK, Amish SJ, Bernatchez L, Foster JT, Miller KM, Morin PA, Narum SR, O'Brien SJ, Roffler G*, et al.* 2016. Genomics in conservation: case studies and bridging the gap between data and application. *Trends Ecol Evol*. 31:81-83.

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES. 2014. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS One*. 9:e90346.

Gnerre S, MacCallum I, Przbylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S*, et al.* 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci USA*. 108:1513-1518.

Goecks J, Nekrutenko A, Taylor J, The Galaxy Team. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol*. 11:R86.

Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A. 2011. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc*. 6:468-481.

Gui D, Jia K, Xia J, Yang L, Chen J, Wu Y, Yi M. 2013. *De novo* assembly of the Indo-Pacific humpback dolphin leucocyte transcriptome to identify putative genes involved in the aquatic adaptation and immune response. *PLoS One*. 8:e72417.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*. 5:e1000695.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M*, et al.* 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 8:1494-1512.

Han E, Sinsheimer JS, Novembre J. 2015. Fast and accurate site frequency spectrum estimation from low coverage sequence data. *Bioinformatics*. 31:720-727.

Hancock-Hanser BL, Frey A, Leslie MS, Dutton PH, Archer FI, Morin PA. 2013. Targeted multiplex next-generation sequencing: advances in techniques of mitochondrial and nuclear DNA sequencing for population genomics. *Mol Ecol Resour*. 13:254-268.

Harris K, Nielsen R. 2013. Inferring demographic history from a spectrum of shared haplotype lengths. *PLoS Genetics*. 9:e1003521.

Hayden S, Bekaert M, Crider TA, Mariani S, Murphy WJ, Teeling EC. 2010. Ecological adaptation determines functional mammalian olfactory subgenomes. *Genome Res*. 20:1-9.

Hedrick PW. 2000 *Genetics of Populations*. Jones and Bartlett Publishers, Sudbury, MA.

Helyar SJ, Hemmer-Hansen J, Bekkevold D, Taylor MI, Ogden R, Limborg MT, Cariani A, Maes GE, Diopere E, Carvalho GR*, et al.* 2011. Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Mol Ecol Resour*. 11:123-136.

Higdon JW, Bininda-Emonds ORP, Beck RMD, Ferguson SH. 2007. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. *BMC Evol Biol*. 7:216.

Hodges E, Rooks M, Xuan Z, Bhattacharjee A, Gordon DB, Brizuela L, McCombie WR, Hannon GJ. 2009. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nat Protoc*. 4:960-974.

Hoffman JI. 2011. Gene discovery in the Antarctic fur seal (*Arctocephalus gazella*) skin transcriptome. *Mol Ecol Resour*. 11:703-710.

Hoffman JI, Nicholas HJ. 2011. A novel approach for mining polymorphic microsatellite markers *in silico*. *PLoS One*. 6:e23283.

Hoffman JI, Simpson F, David P, Rijks JM, Kuiken T, Thorne MAS, Lacy RC, Dasmahapatra KK. 2014. High-throughput sequencing reveals inbreeding depression in a natural population. *Proc Natl Acad Sci USA*. 111:3775-3780.

Hoffman JI, Thorne MAS, Trathan PN, Forcada J. 2013. Transcriptome of the dead: characterisation of immune genes and marker development from necropsy samples in a free-ranging marine mammal. *BMC Genomics*. 14:52.

Hoffman JI, Tucker R, Bridgett SJ, Clark MS, Forcada J, Slate J. 2012. Rates of assay success and genotyping error when single nucleotide polymorphism genotyping in non-model organisms: a case study in the Antarctic fur seal. *Mol Ecol Resour*. 12:861-872.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet*. 6:e1000862.

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

Humble E, Martinez-Barrio A, Forcada J, Trathan PN, Thorne MAS, Hoffmann M, Wolf JBW, Hoffman JI. 2016. A draft fur seal genome provides insights into factors affecting SNP validation and how to mitigate them. *Mol Ecol Resour*.

Jackson JA, Baker CS, Vant M, Steel DJ, Medrano-González L, Palumbi SR. 2009. Big and slow: phylogenetic estimates of molecular evolution in baleen whales (suborder Mysticeti). *Mol Biol Evol*. 26:2427-2440.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S*, et al.* 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*. 484:55-61.

Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, Maruyama H*, et al.* 2014. Efficient *de novo* assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res*. 24:1384-1395.

Keane M, Semeiks J, Webb AE, Li YI, Quesada V, Craig T, Madsen LB, van Dam S, Brawand D, Marques PI*, et al.* 2015. Insights into the evolution of longevity from the bowhead whale genome. *Cell Reports*. 10:112-122.

Khudyakov JI, Champagne CD, Preeyanon L, Ortiz RM, Crocker DE. 2015a. Muscle transcriptome response to ACTH administration in a free-ranging marine mammal. *Physiol Genomics*. 47:318-330.

Khudyakov JI, Preeyanon L, Champagne CD, Ortiz RM, Crocker DE. 2015b. Transcriptome analysis of northern elephant seal (*Mirounga angustirostris*) muscle tissue provides a novel molecular resource and physiological insights. *BMC Genomics*. 16:64.

Kishida T, Thewissen JGM, Hayakawa T, Imai H, Agata K. 2015. Aquatic adaptation and the evolution of smell and taste in whales. *Zoolog Lett*. 1:9.

Koepfli K-P, Paten B, Genome 10K Community of Scientists, O'Brien SJ. 2015. The Genome 10K Project: a way forward. *Annu Rev Anim Biosci*. 3:57-111.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*. 15:356.

Künstner A, Wolf JBW, Backström N, Whitney O, Balakrishnan CN, Day L, Edwards SV, Janes DE, Schlinger BA, Wilson RK*, et al.* 2010. Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 bird species. *Mol Ecol*. 19:266-276.

Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerová M, Rubin C-J, Wang C, Zamani N*, et al.* 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*. 518:371-375.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 10:R25.

Lemmon AR, Emme SA, Lemmon EM. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst Biol*. 61:727-744.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 12:323.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25:1754-1760.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature*. 475:493-496.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map form and SAMtools. *Bioinformatics*. 25:2078-2079.

Li S, Jakobsson M. 2012. Estimating demographic paramaters from large-scale population genomic data using Approximate Bayesian Computation. *BMC Genet*. 13:22.

Li Y, Hu Y, Bolund L, Wang J. 2010. State of the art *de novo* assembly of human genomes from massively parallel sequencing data. *Human Genomics* 4:271-277.

Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E*, et al.* 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*. 478:476-482.

Lindqvist C, Schuster SC, San Y, Talbot SL, Qi J, Ratan A, Tomsho LP, Kasson L, Zeyl E, Aars J*, et al.* 2010. Complete mitochondrial genome of a Pleistocene jawbown unveils the origin of polar bear. *Proc Natl Acad Sci USA*. 107:5053-5057.

Liu S, Lorenzen ED, Fumagalli M, Li B, Harris K, Xiong Z, Zhou L, Korneliussen TS, Somel M, Babbitt C*, et al.* 2014a. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell*. 157:785-794.

Liu X, Fun Y-X. 2015. Exploring population size changes using SNP frequency spectra. *Nat Genet*. 47:555-559.

Liu Y, Zhou J, White KP. 2014b. RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics*. 30:301-304.

Lotterhos KE, Whitlock MC. 2014. Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Mol Ecol*. 23:2178-2192.

Louis M, Viricel A, Lucas T, Peltier H, Alfonsi E, Berrow S, Brownlow A, Covelo P, Dabin W, Deaville R*, et al.* 2014. Habitat-driven population structure of bottlenose dolphins, *Tursiops truncatus*, in the North-east Atlantic. *Mol Ecol*. 23:857-874.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 15:550.

MacManes MD. 2014. On the optimal trimming of high-throughput mRNA sequence data. *Front Genet*. 5:13.

MacManes MD. 2016. Establishing evidence-based best practice for the *de novo* assembly and evaluation of transcriptomes from non-model organisms. *bioRxiv*. doi: http://dx.doi.org/10.1101/035642.

Magera AM, Mills Flemming JE, Kaschner K, Christensen LB, Lotze HK. 2013. Recovery trends in marine mammal populations. *PLoS One*. 8:e77908.

Malenfant RM, Coltman DW, Davis CS. 2015. Design of a 9K Illumina BeadChip for polar bears (*Ursus maritimus*) from RAD and transcriptome sequencing. *Mol Ecol Resour*. 15:587-600.

Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ. 2010. Target-enrichment strategies for next-generation sequencing. *Nat Methods*. 7:111-118.

Mancia A, Abelli L, Kucklick JR, Rowles TK, Wells RS, Balmer BC, Hohn AA, Baatz JE, Ryan JC. 2015. Microarray applications to understand the impact of exposure to environmental contaminants in wild dolphins (*Tursiops truncatus*). *Mar Genomics*. 19:47-57.

Mancia A, Lundqvist ML, Romano TA, Peden-Adams MM, Fair PA, Kindy MS, Ellis BC, Gattoni-Celli S, McKillen DJ, Trent HF*, et al.* 2007. A dolphin peripheral blood leukocyte cDNA microarray for studies of immune function and stress reactions. *Dev Comp Immunol*. 31:520-529.

Mancia A, Ryan JC, Chapman RW, Wu Q, Warr GW, Gulland FMD, Van Dolah FM. 2012. Health status, infection and disease in California sea lions (*Zalophus californianus*) studied using a canine microarray platform and machine-learning approaches. *Dev Comp Immunol*. 36:629-637.

Mancia A, Warr GW, Chapman RW. 2008. A transcriptomic analysis of the stress induced by capture-release health assessment studies in wild dolphins (*Tursiops truncatus*). *Mol Ecol*. 17:2581-2589.

Mastretta-Yanes A, Arrigo N, Alvarez N, Jorgensen TH, Piñero D, Emerson BC. 2015. Restriction site-associated DNA sequencing, genotyping error estimation and *de novo* assembly optimization for population genetic inference. *Mol Ecol Resour*. 15:28-41.

McCormack JE, Faircloth BC, Crawford NG, Gowaty PA, Brumfield RT, Glenn TC. 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Res*. 22:746-754.

McGowen MR. 2011. Toward the resolution of an explosive radiation - a multilocus phylogeny of oceanic dolphins (Delphinidae). *Mol Phylogenet Evol*. 60:345-357.

McGowen MR, Clark C, Gatesy J. 2008. The vestigial olfactory receptor subgenome of odontocete whales: phylogenetic congruence between gene-tree reconciliation and supermatrix methods. *Syst Biol*. 57:574-590.

McGowen MR, Gatesy J, Wildman DE. 2014. Molecular evolution tracks macroevolutionary transitions in Cetacea. *Trends Ecol Evol*. 29:336-346.

McGowen MR, Grossman LI, Wildman DE. 2012. Dolphin genome provides evidence for adaptive evolution of nervous system genes and a molecular rate slowdown. *Proc R Soc Lond B Biol Sci*. 279:3643-3651.

McGowen MR, Spaulding M, Gatesy J. 2009. Divergence date estimation and a comprehensive molecular tree of extant cetaceans. *Mol Phylogenet Evol*. 53:891-906.

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

McTavish EJ, Hillis DM. 2015. How do SNP ascertainment schemes and population demographics affect inferences about population history? *BMC Genomics*. 16:266.

Meredith RW, Gatesy J, Emerling CA, York VM, Springer MS. 2013. Rod monochromacy and the coevolution of cetacean retinal opsins. *PLoS Genetics*. 9:e1003432.

Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, de Filippo C*, et al.* 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science*. 338:222-226.

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.

Miller W, Schuster SC, Welch AJ, Ratan A, Bedoya-Reina OC, Zhao F, Kim HL, Burhans RC, Drautz DI, Wittekindt NE*, et al.* 2012. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proc Natl Acad Sci USA*. 109:E2382-E2390.

Mirceta S, Signore AV, Burns JM, Cossins AR, Campbell KL, Berenbrink M. 2013. Evolution of mammalian diving capacity traced by myoglobin net surface charge. *Science*. 340:1234192.

Morin PA, Archer FI, Foote AD, Vilstrup J, Allen EE, Wade P, Durban JW, Parsons K, Pitman R, Li L*, et al.* 2010a. Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Res*. 20:908-916.

Morin PA, Luikart G, Wayne RK, SNP workshop group. 2004. SNPs in ecology, evolution and conservation. *Trends Ecol Evol*. 19:208-216.

Morin PA, Martien KK, Archer FI, Cipriano F, Steel D, Jackson J, Taylor BL. 2010b. Applied conservation genetics and the need for quality control and reporting of genetic data used in fisheries and wildlife management. *J Hered*. 101:1-10.

Morin PA, Parsons KM, Archer FI, Ávila-Arcos M, Barrett-Lennard LG, Dalla Rosa L, Duchêne S, Durban JW, Ellis GM, Ferguson SH*, et al.* 2015. Geographic and temporal dynamics of a global radiation and diversification in the killer whale. *Mol Ecol*. 24:3964-3979.

Moura AE, Kenny JG, Chaudhuri R, Hughes MA, Welch AJ, Reisinger RR, de Bruyn PJN, Dahlheim ME, Hall N, Hoelzel AR. 2014a. Population genomics of the killer whale indicates ecotype evolution in sympatry involving both selection and drift. *Mol Ecol*. 23:5179-5192.

Moura AE, Nielsen SCA, Vilstrup JT, Moreno-Mayar JV, Gilbert MTP, Gray HWI, Natoli A, Möller L, Hoelzel AR. 2013. Recent diversification of a marine genus (*Tursiops* spp.) tracks habitat preference and environmental change. *Syst Biol*. 62:865-877.

Moura AE, van Rensburg CJ, Pilot M, Tehrani A, Best PB, Thornton M, Plön S, de Bruyn PJN, Worley KC, Gibbs RA*, et al.* 2014b. Killer whale nuclear genome and mtDNA reveal widespread population bottleneck during the last glacial maximum. *Mol Biol Evol*. 31:1121-1131.

Nadeau NJ, Ruiz M, Salazar P, Counterman B, Alejandro Medina J, Ortiz-Zuazaga H, Morrison A, McMillan WO, Jiggins CD, Papa R. 2014. Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. *Genome Res*. 24:1316-1333.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA. 2013. Genotyping-by-sequencing in ecological and conservation genomics. *Mol Ecol*. 22:2841-2847.

Narum SR, Hess JE. 2011. Comparison of F_{ST} outlier tests for SNP loci under selection. *Mol Ecol Resour*. 11:184-194.

Nelson TM, Apprill A, Mann J, Rogers TL, Brown MV. 2015. The marine mammal microbiome: current knowledge and future directions. *Microbiology Australia*. 36:8-13.

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE*, et al.* 2009. Targeted capture and massively parallel sequencing of twelve human exomes. *Nature*. 461:272-276.

Nielsen R, Paul JS, Anders A, Song YS. 2011. Genotype and SNP calling from next-generation sequencing data. *Nat Rev Genet*. 12:433-451.

Noonan JP, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, Chen F, Platt D, Pääbo S, Pritchard JK*, et al.* 2006. Sequencing and analysis of Neanderthal genomic DNA. *Science*. 314:1113-1118.

Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, Buck S, Chambers CD, Chin G, Christensen G*, et al.* 2015. Promoting an open research culture: Author guidelines for journals could help to promote transparency, openness, and reproducibility. *Science*. 348:1422-1425.

O'Rawe JA, Ferson S, Lyon GJ. 2015. Accounting for uncertainty in DNA sequencing data. *Trends Genet*. 31:61-66.

Olsen MT, Volny VH, Bérubé M, Dietz R, Lydersen C, Kovacs KM, Dodd RS, Palsbøll PJ. 2011. A simple route to single-nucleotide polymorphisms in a nonmodel species: identification and characterization of SNPs in the Arctic ringed seal (*Pusa hispida hispida*). *Mol Ecol Resour*. 11:9-19.

Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I*, et al.* 2013. Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*. 499:74-78.

Pabuwal V, Boswell M, Pasquali A, Wise SS, Kumar S, Shen Y, Garcia T, Lacerte C, Wise JP, Jr., Wise JP, Sr., *et al.* 2013. Transcriptomic analysis of cultured whale skin cells exposed to hexavalent chromium [Cr(VI)]. *Aquat Toxicol*. 134-135:74-81.

Parker J, Tsagkogeorga G, Cotton JA, Liu Y, Provero P, Stupka E, Rossiter SJ. 2013. Genome-wide signatures of convergent evolution in echolocating mammals. *Nature*. 502:228-231.

Paszkiewicz KH, Farbox A, O'Neill P, Moore K. 2014. Quality control on the frontier. *Front Genet*. 5:157.

Patro R, Duggal G, Kingsford C. 2015. Accurate, fast, and model-aware transcript expression quantification with Salmon. *bioRxiv*. doi: http://dx.doi.org/10.1101/021592.

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:e37135.

Poh Y-P, Domingues VS, Hoekstra HE, Jensen JD. 2014. On the prospect of identifying adaptive loci in recently bottlenecked populations. *PLoS One*. 9:e110579.

Poland JA, Brown PJ, Sorrells ME, Jannink J-L. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One*. 7:e32253.

Polanowski AM, Robbins J, Chandler D, Jarman SN. 2014. Epigenetic estimation of age in humpback whales. *Mol Ecol Resour*. 14:976-987.

Puritz JB, Hollenbeck CM, Gold JR. 2014. *dDocent*: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*. 2:e431.

Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R*, et al.* 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*. 463:757-762.

Riesch R, Barrett-Lennard LG, Ellis GM, Ford JKB, Deecke VB. 2012. Cultural traditions and the evolution of reprodutive isolation: ecological speciation in killer whales? *Biol J Linn Soc Lond*. 2012:1-17.

Robinson JD, Coffman AJ, Hickerson MJ, Gutenkunst RN. 2014. Sampling strategies for frequency spectrum-based population genomic inference. *BMC Evol Biol*. 14:254.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 26:139-140.

Ruan R, Guo A-H, Hao Y-J, Zheng J-S, Wang D. 2015. *De novo* assembly and characterization of narrow-ridged finless porpoise renal transcriptome and identification of candidate genes involved in osmoregulation. *Int J Mol Sci*. 16:2220-2238.

Ruegg K, Rosenbaum HC, Anderson EC, Engel M, Rothschild A, Baker CS, Palumbi SR. 2013. Long-term population size of the North Atlantic humpback whale within the context of worldwide population structure. *Cons Gen*. 14:103-114.

Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nat Genet*. 46:919-925.

Schubert M, Lindgreen S, Orlando L. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes*. 9:88.

Schurch NJ, Schofield P, Gierlinski M, Cole C, Sherstnev A, Singh V, Wrobel N, Gharbi K, Simpson GG, Owen-Hughes T*, et al.* 2015. Evaluation of tools for differential gene expression analysis by RNA-seq on a 48 biological replicate experiment. *arXive*. 1505:02017.

Seim I, Ma S, Zhou X, Gerashchenko MV, Lee SG, Suydam R, George JC, Bickham JW, Gladyshev VN. 2014. The transcriptome of the bowhead whale *Balaena mysticetus* reveals adaptations of the longest-lived mammal. *Aging*. 6:879-899.

Shafer ABA, Cullingham CI, Côté SD, Coltman DW. 2010. Of glaciers and refugia: a decade of study sheds new light on the phylogeographic patterns of northwestern North America. *Mol Ecol*. 19:4589-4621.

Shafer ABA, Davis CS, Coltman DW, Stewart REA. 2014. Microsatellite assessment of walrus (*Odobenus rosmarus rosmarus*) stocks in Canada. *NAMMCO Scientific Publications*. 9.

Shafer ABA, Gattepaille LM, Stewart REA, Wolf JBW. 2015. Demographic inferences using short-read genomic data in an approximate Bayesian computation framework: *in silico* evaluation of power, biases and proof of concept in Atlantic walrus. *Mol Ecol*. 24:328-345.

Shen Y-Y, Zhou W-P, Zhou T-C, Zeng Y-N, Li G-M, Irwin DM, Zhang Y-P. 2012. Genome-wide scan for bats and dolphin to detect their genetic basis for new locomotive styles. *PLoS One*. 7:e46455.

Smith-Unna RD, Boursnell C, Patro R, Hibberd JM, Kelly S. 2015. TransRate: reference free quality assessment of *de-novo* transcriptome assemblies. *bioRxiv*.

Spies D, Ciaudo C. 2015. Dynamics in transcriptomics: advancements in RNA-seq time course and downstream analysis. *Comput Struct Biotechnol J*. 13:469-477.

Springer MS, Signore AV, Paijmans JLA, Vélez-Juarbe J, Domning DP, Bauer CE, He K, Crerar L, Campos PF, Murphy WJ*, et al.* 2015. Interordinal gene capture, the phylogenetic position of Steller's sea cow based on molecular and morphological data, and the macroevolutionary history of Sirenia. *Mol Phylogenet Evol*. 91:178-193.

Springer MS, Starrett J, Morin PA, Lanzetti A, Hayashi C, Gatesy J. 2016. Inactivation of *C4orf26* in toothless placental mammals. *Mol Phylogenet Evol*. 95:34-45.

Sremba AL, Martin AR, Baker CS. 2015. Species identification and likely catch time preiod of whale bones from South Georgia. *Mar Mamm Sci*. 31:122-132.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Res*. 34:W435-W439.

Stein LD. 2010. The case for cloud computing in genome informatics. *Genome Biol*. 11:207.

Stinchcombe JR, Hoekstra HE. 2008. Combining population genomics and quantitative genetics: finding genes underlying ecologically important traits. *Heredity*. 100:158-170.

Tabuchi M, Veldhoen N, Dangerfield N, Jeffries S, Helbing CC, Ross PS. 2006. PCB-related alteration of thyroid hormones and thyroid hormone receptor gene expression in free-ranging harbor seals (*Phoca vitulina*). *Environ Health Perspect*. 114:1024-1031.

Taylor BL, Gemmell NJ. 2016. Emerging technologies to conserve biodiversity: further opportunities via genomics. Response to Pimm *et al. Trends Ecol Evol*. 31:171-172.

The *Heliconius* Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*. 487:94-98.

Thomsen PF, Kielgast J, Iversen LL, Møller PR, Rasmussen M, Willerslev E. 2012. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS One*. 7:e41732.

Towns J, Cockerill T, Dahan M, Foster I, Gaither K, Grimshaw A, Hazlewood V, Lathrop S, Lifka D, Peterson GD*, et al.* 2014. XSEDE: accelerating scientific discovery. *Computing in Science and Engineering*. 16:62-74.

Tsagkogeorga G, McGowen MR, Davies KT, Jarman S, Polanowski A, Bertelsen MF, Rossiter SJ. 2015. A phylogenomic analysis of the role and timing of molecular adaptation in the aquatic transition of cetartiodactyl mammals. *R Soc Open Sci*. 2:150156.

van Dijk EL, Auger H, Jaszczyzyn Y, Thermes C. 2014. Ten years of next-generation sequencing technology. *Trends Genet*. 30:418-426.

VanRaden PM, Sun C, O'Connell JR. 2015. Fast imputation using medium or low-coverage sequence data. *BMC Genet*. 16:82.

Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ*, et al.* 2015. Enhancer evolution across 20 mammalian species. *Cell*. 160:554-566.

Viricel A, Pante E, Dabin W, Simon-Bouhet B. 2014. Applicability of RAD-tag genotyping for interfamilial comparisons: empirical data from two cetaceans. *Mol Ecol Resour*. 14:597-605.

Viricel A, Rosel PE. 2014. Hierarchical population structure and habitat differences in a highly mobile marine species: the Atlantic spotted dolphin. *Mol Ecol*. 23:5018-5035.

Wolf JB. 2013. Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Mol Ecol Resour*. 13:559-572.

Xiong Y, Brandley MC, Xu S, Zhou K, Yang G. 2009. Seven new dolphin mitochondrial genomes and a time-calibrated phylogeny of whales. *BMC Evol Biol*. 9:20.

Yandell M, Ence D. 2012. A beginner's guide to eukaryotic genome annotation. *Nat Rev Genet*. 13:329-342.

Yeh R-F, Lim LP, Burge CB. 2001. Computational inference of homologous gene structures in the human genome. *Genome Res*. 11:803-816.

Yim H-S, Cho YS, Guang X, Kang SG, Jeong J-Y, Cha S-S, Oh H-M, Lee J-H, Yang EC, Kwon KK*, et al.* 2014. Minke whale genome and aquatic adaptation in cetaceans. *Nat Genet*. 46:88-92.

Zhao Q-Y, Wang Y, Kong Y-M, Luo D, Li X, Hao P. 2011. Optimizing *de novo* transcriptome assembly from short-read RNA-Seq data: a comparative study. *BMC Bioinformatics*. 12:S2.

Zhou X, Sun F, Xu S, Fan G, Zhu K, Liu X, Chen Y, Shi C, Yang Y, Huang Z*, et al.* 2013. Baiji genomes reveal low genetic variability and new insights into secondary aquatic adaptations. *Nat Commun*. 4:2708.

Zou Z, Zhang J. 2015. No genome-wide protein sequence convergence for echolocation. *Mol Biol Evol*. 32:1237-1241.

Table 1. Current and commonly used tools for analysis of genomic data generated in non-model organisms. Please note that this list is not exhaustive and new computational tools are continuously being developed.

Computational Tool	Purpose	Strengths/Weaknesses	Reference
*RADseq**			
STACKS	quality filtering, *de novo* assembly or reference-aligned read mapping, variant genotyping	scalable (new data can be compared against existing locus catalog); flexible filtering and export options; recently implemented a gapped alignment algorithm to process insertion-deletion (indel) mutations; secondary algorithm adjusts SNP calls using population-level allele frequencies; compatible with input data from multiple RADseq methods	Catchen et al. (2011; 2013), http://catchenlab.life.illinois.edu/stacks/
PyRAD	quality filtering, *de novo* assembly, read mapping, variant genotyping	efficiently processes indel mutations, thus optimal for analysis of highly divergent species; high speed and quality of paired-end library assemblies; compatible with input data from multiple RADseq methods	Eaton (2014)
TASSEL-GBS	quality filtering, reference-aligned read mapping, variant genotyping	optimized for single-end data from large sample sizes (tens of thousands of individuals) with a reference genome; performs genome-wide association studies	Glaubitz et al. (2014)
dDocent	quality trimming, *de novo* assembly, read mapping, variant genotyping	beneficial in analysis of paired-end data; identifies both SNP and indel variants; most appropriate for ezRAD and ddRAD data	Puritz et al. (2014)
AftrRAD	quality filtering, *de novo* assembly, read mapping, variant genotyping	identifies both SNP and indel variants; computationally faster than STACKS and PyRAD	Sovic et al. (2015)
Array-based high-throughput sequencing			
Affymetrix Axiom™ Analysis Suite, Illumina® GenomeStudio	genotype scoring	visualization of genotype clusters; quality scores assigned to genotype calls allow user-specific filtering; manual editing possible	
Whole genome sequencing			
AdapterRemoval v2, Trimmomatic	trim raw sequences	remove adapter sequences and low-quality bases prior to assembly	Bolger et al. (2014), Schubert et al. (2016)
ALLPATHS-LG, PLATANUS, SOAPdenovo	*de novo* genome assembly	designed for short-read sequences of large heterozygous genomes	Li et al. (2010), Gnerre et al. (2011), Kajitani et al. (2014)
AUGUSTUS, GenomeScan, MAKER2	gene annotation	highly accurate evidence-driven or BLASTX-guided gene prediction (Yandell and Ence 2012)	Yeh et al. (2001), Stanke et al. (2006), Holt and Yandell (2011)

Bowtie, bwa	read mapping	rapid short-read alignment with compressed reference genome index, but limited number of acceptable mismatches per alignment (Flicek and Birney 2009)	Langmead et al. (2009), Li and Durbin (2009)
SAMtools	data processing, variant calling ~~(SNP and indel discovery)~~	multi-purpose tool that conducts file conversion, alignment sorting, PCR duplicate removal, and variant (SNP and indel) calling for SAM/BAM/CRAM files	Li et al. (2009)
GATK	data processing and quality control, variant calling	suitable for ~~processing and analyses of~~ data with low to high mean read depth across the genome~~coverage data~~; initially optimized for large human datasets, then modified for use with non-model organisms	McKenna et al. (2010), DePristo et al. (2011)
ANGSD/NGStools	data processing, variant calling, estimation of diversity metrics, population genomic analyses	suitable for ~~processing and analyses of~~ data with low mean read depth, including ~~coverage and~~ palaeogenomic data; allow downstream analyses such as D-statistics and SFS estimation	Fumagalli et al. (2014), Korneliussen et al. (2014)
RNAseq			
Fastx Toolkit, Trimmomatic	trim raw sequences	remove erroneous nucleotides from reads prior to assembly	MacManes (2014)
khmer diginorm, Trinity normalization	*in silico* read normalization	reduce~~s~~ memory requirements for assembly, but can result in fragmented assemblies and collapse heterozygosity	Brown et al. (2012); Haas et al. (2013)
Trinity	*de novo* and genome-guided transcriptome assembly	accurate assembly across conditions, but requires long runtime if normalization is not used (Zhao et al. 2011)	Haas et al. (2013)
bowtie, bowtie2, STAR	read alignment to genome or transcriptome assembly	required for many downstream analyses, but bowtie is computationally intensive and all produce very large output BAM files	Langmead et al. (2009), Dobin et al. (2013)
eXpress, kallisto, RSEM, Sailfish, Salmon	estimation of transcript abundance	RSEM requires computationally intensive read mapping back to the assembly; the others are faster streaming alignment, quasi-alignment, or alignment-free algorithms	Li and Dewey (2011), Patro et al. (2015)
DESeq, DESeq2, edgeR	differential expression analysis	exhibit highest true positive and lowest false positive rates in experiments with smaller sample sizes (Schurch et al. 2015)	Anders and Huber (2010), Robinson et al. (2010), Love et al. (2014)
blast2GO, Trinotate	functional annotation of assembled transcripts	complete annotation pipelines including gene ontology and pathway enrichment analyses	Conesa et al. (2005), Haas et al. (2013)

* This is a non-exhaustive list of software that ~~include~~ focuses on *de novo* loci assembly and genotype calling for RADseq data, as many practitioners working on NMOs will not have access to a reference genome. Other programs (e.g., GATK and ANGSD) that undertake genotype calling using reference-aligned loci ~~only~~ are described in the whole genome sequencing section.

Figure 1

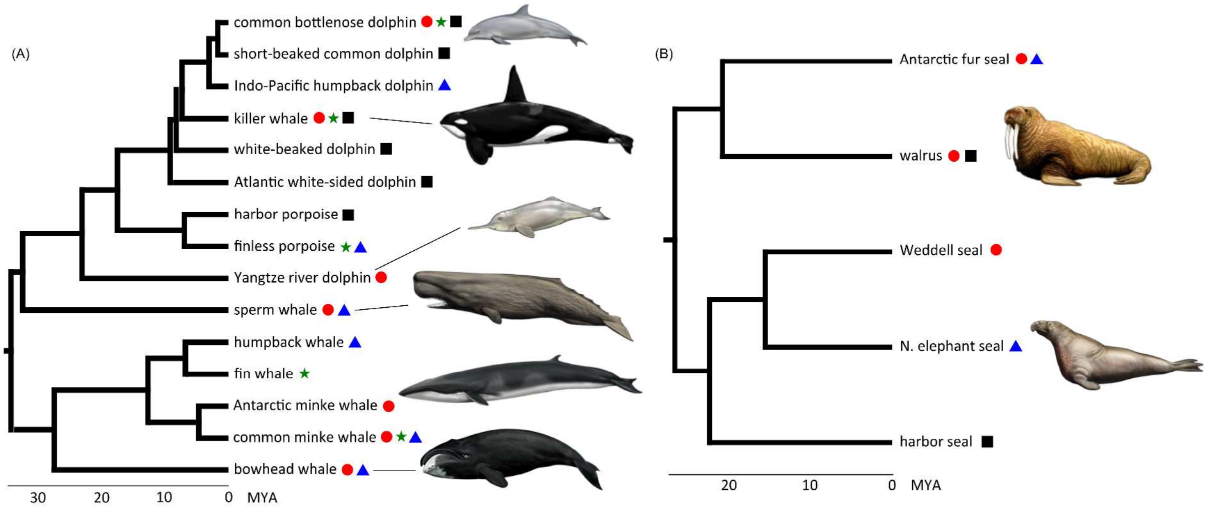


Figure 1. Phylogenetic tree showing current genomic resources available for (A) cetaceans and (B) pinnipeds; relationships and branch lengths are based on molecular dating estimates from McGowen et al. (2009), McGowen (2011), and Higdon et al. (2007). Scale is in millions of years ago (MYA). Red circles indicate species with high-~~coverage~~ quality ~~whole~~ reference genomes; green stars indicate ~~low-coverage~~ whole genome re-sequencing data; blue triangles indicate transcriptomes (generated by microarray or RNAseq); and black squares indicate RADseq data.

Figure 2

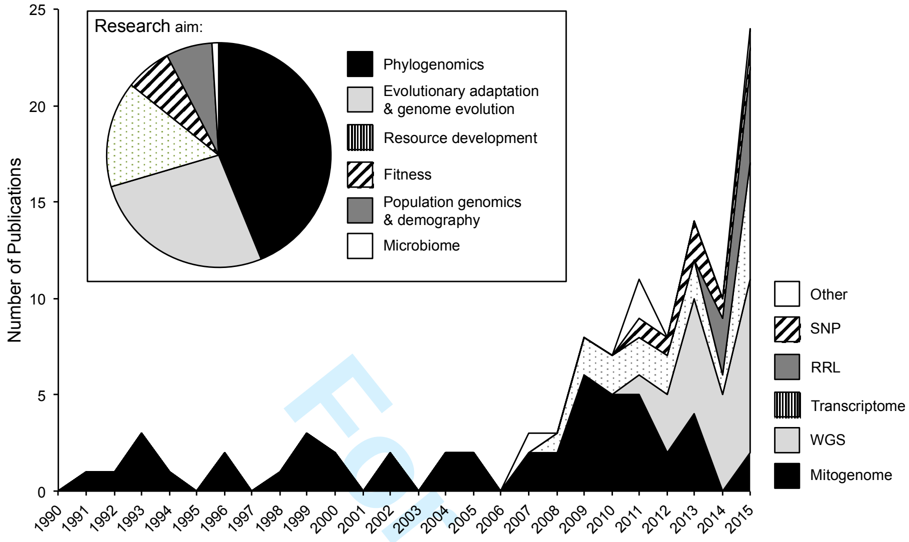


Figure 2. Number of marine mammal genomics publications from 1990 to 2015, categorized by primary methodology and research aim. Genomic methodologies include high-throughput single nucleotide polymorphism (SNP) genotyping and sequencing of mitogenomes, whole genomes (WGS), transcriptomes (generated by microarray or RNAseq), and reduced-representation genomic libraries (RRL). The "Other" category includes studies of microbiomes, BAC libraries, and large (~100) gene sets.

Figure 3

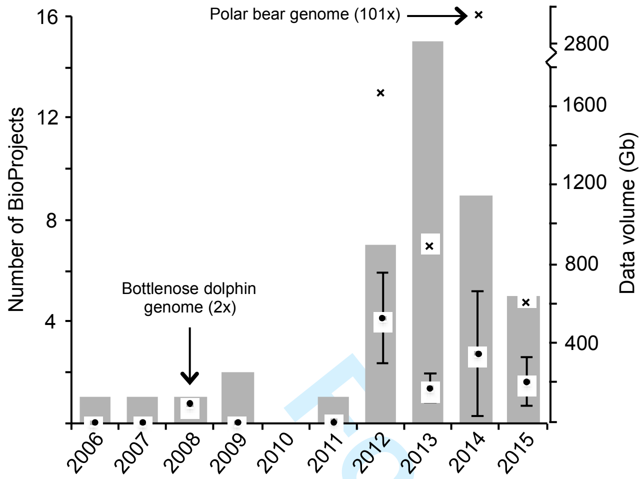
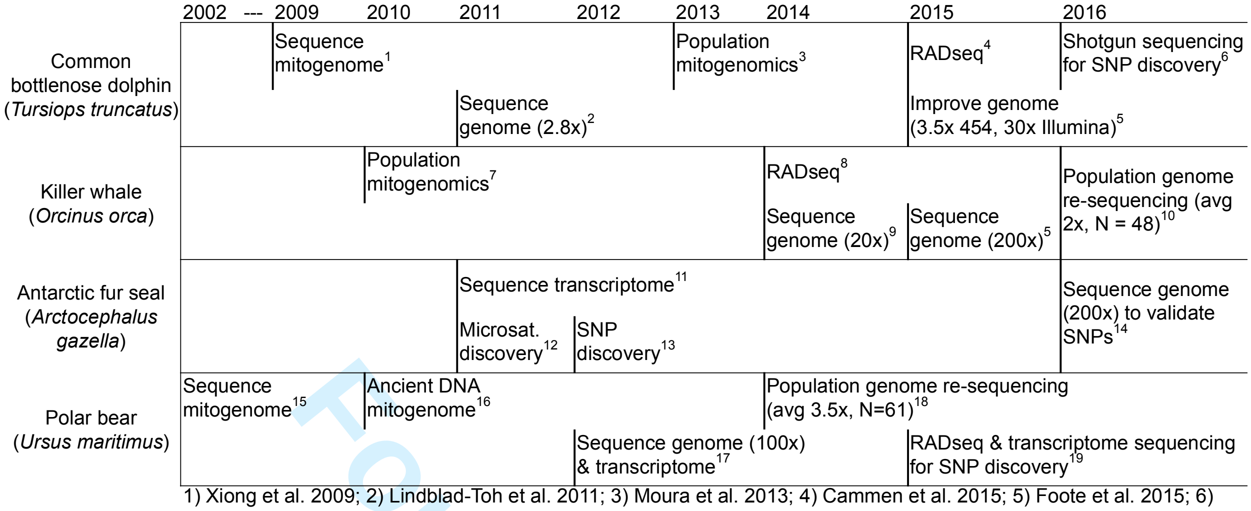


Figure 3. Number of BioProjects (~~shaded~~ gray bars) related to marine mammal genomics submitted from 2006 to 2015 to an online public database maintained by NCBI. Early BioProjects were largely microarray datasets. The number of projects created each year, as well as the yearly average (black dots ± SE) and maximum (×) size of data submitted in each BioProject, increased dramatically after 2011, reflecting advances in high-throughput sequencing technologies that facilitated their use in non-model systems.

Figure 4


2002

2009
2010
2011
2012
2013
2014
2015
2016
Common bottlenose dolphin (*Tursiops truncatus*)
Sequence mitogenome1
Population mitogenomics3
RADseq4
Shotgun sequencing for SNP discovery6
Sequence genome (2.8x)2
Improve genome (3.5x 454, 30x Illumina)5
Killer whale (*Orcinus orca*)
Population mitogenomics7
RADseq8
Sequence genome (20x)9
Sequence genome (200x)5
Population genome re-sequencing (avg 2x, N = 48)10
Antarctic fur seal (*Arctocephalus gazella*)
Sequence transcriptome11
Microsat. discovery12
SNP discovery13
Sequence genome (200x) to validate SNPs14
Polar bear (*Ursus maritimus*)
Sequence mitogenome15
Ancient DNA mitogenome16
Population genome re-sequencing (avg 3.5x, N=61)18
Sequence genome (100x) & transcriptome17
RADseq & transcriptome sequencing for SNP discovery19

1) Xiong et al. 2009; 2) Lindblad-Toh et al. 2011; 3) Moura et al. 2013; 4) Cammen et al. 2015; 5) Foote et al. 2015; 6) Louis et al. unpubl. data; 7) Morin et al. 2010; 8) Moura et al. 2014a; 9) Moura et al. 2014b; 10) Foote et al. 2016; 11) Hoffman 2011; 12) Hoffman and Nicholas 2011; 13) Hoffman et al. 2012; 14) Humble et al. 2016; 15) Arnason et al. 2002; 16) Lindqvist et al. 2010; 17) Miller et al. 2012; 18) Liu et al. 2014; 19) Malenfant et al. 2015

Figure 4. Timelines depicting the independent progression of genomic studies for four representative marine mammal species. Trajectories show the common progression for non-model species from mitogenome sequencing to whole genome sequencing, as well as from sequencing reference specimens to population-scale genomic sequencing. In addition, the timelines reveal the utility of genomic and transcriptomic sequencing for subsequent genetic marker development.