Invited Review

Genomic Methods Take the Plunge: Recent Advances in High-Throughput Sequencing of Marine Mammals


From the School of Marine Sciences, University of Maine, Orono, ME 04469 (Cammen); Department of Fish and Wildlife Sciences, University of Idaho, 875 Perimeter Drive MS 1136, Moscow, ID 83844-1136 (Andrews); Scottish Oceans Institute, University of St Andrews, East Sands, St Andrews, Fife KY16 8LB, UK (Carroll and Louis); Computational and Molecular Population Genetics Lab, Institute of Ecology and Evolution, University of Bern, Bern CH-3012, Switzerland (Foote); Department of Animal Behaviour, University of Bielefeld, Postfach 100131, 33501 Bielefeld, Germany (Humble); British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 OET, UK (Humble); Department of Biology, Sonoma State University, Rohnert Park, CA 94928 (Khudyakov); School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK (Mcgowen); Evolutionary Genomics Section, Natural History Museum of Denmark, University of Copenhagen, DK-1353 Copenhagen K, Denmark (Olsen); and Scripps Institution of Oceanography, University of California San Diego, 8622 Kennel Way, La Jolla, CA 92037 (Van Cise).

Address correspondence to Kristina M. Cammen at the address above, or e-mail: kristina.cammen@maine.edu

Received May 9, 2016; First decision May 30, 2016; Accepted July 12, 2016.

Corresponding editor: C Scott Baker

Abstract

The dramatic increase in the application of genomic techniques to non-model organisms (NMOs) 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 NMOs 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 NMOs. 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 NMOs.

Subject areas: Genomics and gene mapping

Key words: non-model organisms, RADseq, RNAseq, SNP array, target sequence capture, whole genome sequencing
Recent advances in sequencing technologies, coincident with dramatic declines in cost, have increasingly enabled the application of genomic sequencing in non-model systems (Ekbloom 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 (NMOs) (Armengaud et al. 2014). 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 (Figure 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 many 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 both in publication rate (Figure 2) and the rise in number and size of genomic sequences deposited in public resources (Figure 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 (Mircetta 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 the evolution and conservation of marine mammals. In the past decade, many marine mammal species have been characterized by high-quality reference genomes, and the use of high-throughput sequencing has further enabled genomic and population genetic analyses of many species. As the number and size of genomic sequences deposited in public resources continues to grow rapidly, marine mammal genomics has the potential to provide a wealth of information for understanding marine mammal evolution, population genetics, and conservation biology.
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 11 December 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 Supplementary Table S1.

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 (Figures 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 single nucleotide polymorphism (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 genomics studies thus far have primarily contributed to the fields of phylogenomics and comparative genomics (Figure 2, Supplementary 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 (Supplementary 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...
Figure 4. Timelines depicting the independent progression of genomic studies for 4 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.

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 500–1000 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

Restriction Site-Associated DNA Sequencing

Reduced-representation sequencing methods evaluate only a small portion of the genome, allowing researchers to sequence samples 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.

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.
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 lower 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 7431 RADseq SNPs (Cammen et al. 2015), and genomic regions associated with habitat use and resource specialization in killer whales (Orcinus orca) were identified using 3281 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 (Foose 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).

**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 transcriptions (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 (Louis M et al., 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 Fu 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 2 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.

**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 (Bashirades 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 (Avila-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) specimens 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 (Supplementary 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.

Reference Genome Sequencing

At the time of publication, there are 12 publicly available whole, or near-whole, marine mammal genomes of varying quality representing 10 families, including 7 cetaceans (Figure 1A), 3 pinnipeds (Figure 1B), the West Indian manatee (Trichechus manatus), and the polar bear. These genomes are available on NCBI’s online genome database or Dryad, but they have not all been published. Note that 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. The first sequenced marine mammal genome was that of the common bottlenose dolphin, which was originally sequenced to ~2.5× 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 (Supplementary 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, enamelegenesis, 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) 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 provides 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.

Population-Level Genome Re-sequencing

In contrast to reference genome sequencing that today often exceeds 100× 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 ≥2× 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 5 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 threespine 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 2 marine mammal population genomics studies using whole genome re-sequencing have been published. These studies involved re-sequencing the genomes of 79 individuals from 3 populations of polar bears (Liu et al. 2014a) and 48 individuals from 5 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 ancient 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) genome re-sequencing data in pairwise comparisons between the same 2 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 2 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 2 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]) (Supplementary 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 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 (ChIP-seq)—to examine gene regulatory element evolution across mammals, including 4 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 analysis 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 2 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,
<table>
<thead>
<tr>
<th>Computational tool</th>
<th>Purpose</th>
<th>Strengths/weaknesses</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>RADseq’ STACKS</td>
<td>Quality filtering, <em>de novo</em> assembly or reference-aligned read mapping, variant genotyping</td>
<td>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</td>
<td>Catchen et al. (2011; 2013), <a href="http://catchenlab.life.illinois.edu/stacks/">http://catchenlab.life.illinois.edu/stacks/</a></td>
</tr>
<tr>
<td>PyRAD</td>
<td>Quality filtering, <em>de novo</em> assembly, read mapping, variant genotyping</td>
<td>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</td>
<td>Eaton (2014)</td>
</tr>
<tr>
<td>TASSEL-GBS</td>
<td>Quality filtering, reference-aligned read mapping, variant genotyping</td>
<td>Optimized for single-end data from large sample sizes (tens of thousands of individuals) with a reference genome; performs genome-wide association studies</td>
<td>Glaubitz et al. (2014)</td>
</tr>
<tr>
<td>dDocent</td>
<td>Quality trimming, <em>de novo</em> assembly, read mapping, variant genotyping</td>
<td>Beneficial in analysis of paired-end data; identifies both SNP and indel variants; most appropriate for eRAD and ddRAD data</td>
<td>Puritz et al. (2014)</td>
</tr>
<tr>
<td>AfrRAD</td>
<td>Quality filtering, <em>de novo</em> assembly, read mapping, variant genotyping</td>
<td>Identifies both SNP and indel variants; computationally faster than STACKS and PyRAD</td>
<td>Sovic et al. (2015)</td>
</tr>
<tr>
<td>Array-based high-throughput sequencing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affymetrix Axiom™ Analysis Suite, Illumina® GenomeStudio</td>
<td>Genotype scoring</td>
<td>Visualization of genotype clusters; quality scores assigned to genotype calls allow user-specific filtering; manual editing possible</td>
<td></td>
</tr>
<tr>
<td>Whole genome sequencing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AdapterRemoval v2, Trimmomatic ALLPATHS-LG, PLATANUS, SOAPdenovo AUGUSTUS, GenomeScan, MAKER2 Bowtie, bwa</td>
<td>Trim raw sequences</td>
<td>Remove adapter sequences and low-quality bases prior to assembly</td>
<td>Bolger et al. (2014), Schubert et al. (2016)</td>
</tr>
<tr>
<td>De novo genome assembly</td>
<td>Designed for short-read sequences of large heterozygous genomes</td>
<td>Li et al. (2010), Gnerre et al. (2011), Kajitani et al. (2014)</td>
<td></td>
</tr>
<tr>
<td>Gene annotation</td>
<td>Highly accurate evidence-driven or BLASTX-guided gene prediction (Yandell and Ence 2012)</td>
<td>Yeh et al. (2001), Stanke et al. (2006), Holt and Yandell (2011)</td>
<td></td>
</tr>
<tr>
<td>Read mapping</td>
<td>Rapid short-read alignment with compressed reference genome index, but limited number of acceptable mismatches per alignment (Flicek and Birney 2009)</td>
<td>Langmead et al. (2009), Li and Durbin (2009)</td>
<td></td>
</tr>
<tr>
<td>SAMtools</td>
<td>Data processing, variant calling</td>
<td>Multi-purpose tool that conducts file conversion, alignment sorting, PCR duplicate removal, and variant (SNP and indel) calling for SAM/BAM/CRAM files</td>
<td>Li et al. (2009)</td>
</tr>
<tr>
<td>GATK</td>
<td>Data processing and quality control, variant calling</td>
<td>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</td>
<td>McKenna et al. (2010), DePristo et al. (2011)</td>
</tr>
<tr>
<td>ANGSD/ngsTools</td>
<td>Data processing, variant calling, estimation of diversity metrics, population genomic analyses</td>
<td>Suitable for data with low mean read depth, including palaeogenomic data; allows downstream analyses such as D-statistics and SFS estimation</td>
<td>Fumagalli et al. (2014), Korneliussen et al. (2014)</td>
</tr>
<tr>
<td>RNAseq Fastx Toolkit, Trimmomatic khmer diginorm, Trinity normalization</td>
<td>Trim raw sequences</td>
<td>Remove erroneous nucleotides from reads prior to assembly</td>
<td>MacManes (2014)</td>
</tr>
<tr>
<td>De novo and genome-guided transcriptome assembly</td>
<td>In silico read normalization</td>
<td>Reduce memory requirements for assembly, but can result in fragmented assemblies and collapse heterozygosity</td>
<td>Brown et al. (2012); Haas et al. (2013)</td>
</tr>
<tr>
<td>De novo and genome-guided transcriptome assembly</td>
<td>Accurate assembly across conditions, but requires long runtime if normalization is not used (Zhao et al. 2011)</td>
<td>Haas et al. (2013)</td>
<td></td>
</tr>
</tbody>
</table>
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 PHYLUXE (Faircloth 2016), which is designed specifically for use with UCEs. 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 (>10–20× 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 (2–20× 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 a 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 genomics 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.

Supplementary Material
Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

Funding
The authors involved in this work were supported by a National Science Foundation Postdoctoral Research Fellowship in Biology (1523568) to K.M.C.; an Office of Naval Research Award (N00014-15-1-2773) to J.L.K.; a Marie Sladowska Curie Fellowship to E.L.C. (Behaviour-Connect) funded by the EU Horizon2020 program; Royal Society Newton International Fellowships to E.L.C. and M.R.M.; a Swiss National Science Foundation Postdoctoral Research Fellowship in Biology (Behaviour-Connect) funded by the EU Horizon2020 program; Royal Society Newton International Fellowships to E.L.C. and M.R.M.; a Swiss National Science Foundation Award (31003A-143393) to L. Excoffier that further supported A.D.F. The first marine mammal genomics workshop was held to begin discussions towards this review was supported by a short visit grant from the European Science Foundation-Research Networking Programme ConGenOmics to A.D.F.; and a Swiss National Foundation Postdoctoral Fellowship to M.L.; postdoctoral funding from the University of Idaho College of Natural Resources to K.R.A.; an Office of Naval Research Award (N00014-15-1-2773) to J.L.K.; a Marie Sladowska Curie Fellowship to E.L.C. (Behaviour-Connect) funded by the EU Horizon2020 program; Royal Society Newton International Fellowships to E.L.C. and M.R.M.; a Swiss National Science Foundation Award (31003A-143393) to L. Excoffier that further supported A.D.F. 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.

Acknowledgments
This review paper is the outcome of 2 international workshops held in 2013 and 2015 on marine mammal genomics. The workshops were organized by K.M.C., A.F., 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 2 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.

References


