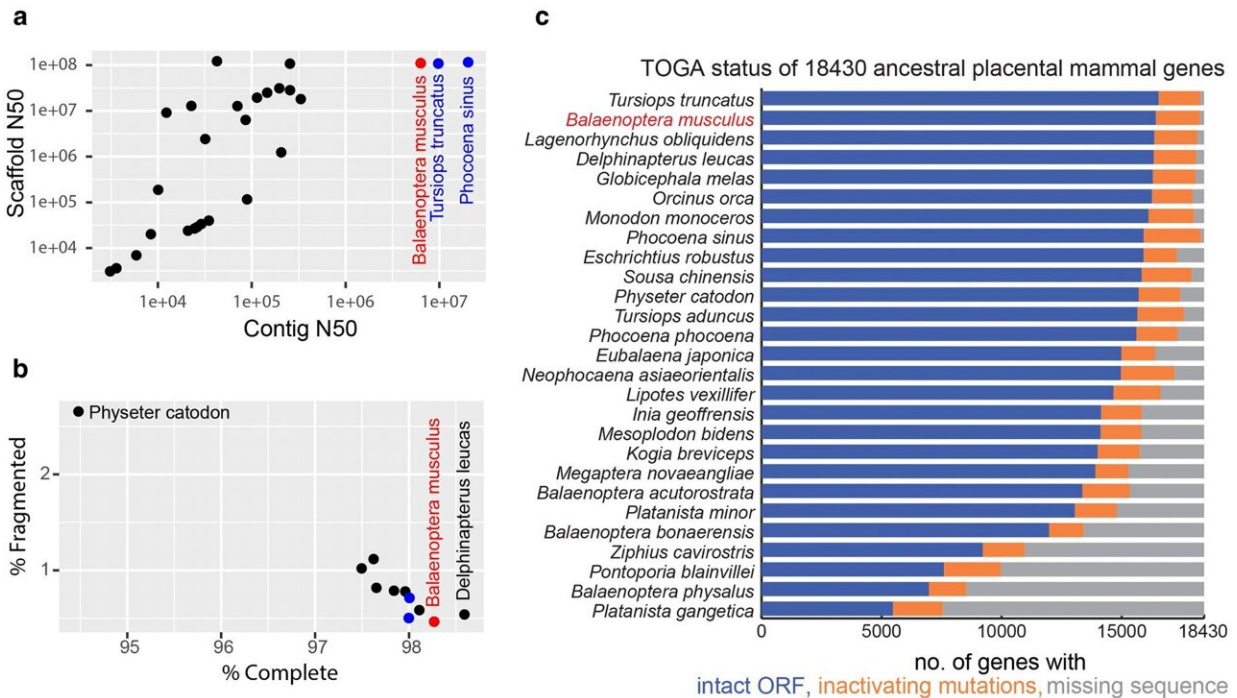


Sequencing the blue whale and Etruscan shrew genomes

March 18 2024, by Mariel Mohns



Assembly quality metrics. Blue whale (*Balaenoptera musculus*) data are shown in red; the 2 other VGP assemblies, vaquita (*Phocoena sinus*) and bottlenose dolphin (*Tursiops truncatus*), are in blue. a) Assembly contig and scaffold N50 metrics. Contigs are segments of contiguous, i.e. gapless sequence. Scaffolds are sets of contigs that have been ordered and oriented using long-range mapping data such as optical maps and Hi-C with gaps between contigs. N50 is a measure of average length, e.g. 50% of all bases are contained in contigs of length N50 or

longer. b) % of complete and fragmented universal single copy BUSCO orthologs found in an annotated genome. Universal single copy orthologs are genes that are present in a single copy in all or most genomes within a phylogenetic group. A high % complete score is an indication that a genome assembly is not missing a large amount of gene-coding sequence (Simão et al. 2015; Manni et al. 2021). C) TOGA status of 18,430 ancestral placental mammal genes. Note: For 2 species, different assemblies were used in panel C compared to panel A: GCA_004363415.1 instead of GCA_002189225.1 for *Eschrichtius robustus* and GCA_008795845.1 instead of GCA_023338255.1 for *Balaenoptera physalus*. Credit: *Molecular Biology and Evolution* (2024). DOI: 10.1093/molbev/msae036

The [blue whale genome](#) was published in the journal *Molecular Biology and Evolution*, and the [Etruscan shrew genome](#) was published in the journal *Scientific Data*.

Research models using animal cell cultures can help navigate big biological questions, but these tools are only useful when following the right map.

"The [genome](#) is a blueprint of an organism," says Yury Bukhman, first author of the published research and a computational biologist in the Ron Stewart Computational Group at the Morgridge Institute, an independent research organization that works in affiliation with the University of Wisconsin–Madison in emerging fields such as regenerative biology, metabolism, virology, and biomedical imaging.

"In order to manipulate [cell cultures](#) or measure things like [gene expression](#), you need to know the genome of the species—it makes more research possible."

The Morgridge team's interest in the blue whale and the Etruscan shrew began with research on the biological mechanisms behind the "developmental clock" from James Thomson, emeritus director of regenerative biology at Morgridge and longtime professor of cell and regenerative biology in the UW School of Medicine and Public Health.

It's generally understood that larger organisms take longer to develop from a fertilized egg to a full-grown adult than smaller creatures, but the reason why remains unknown.

"It's important just for fundamental biological knowledge from that perspective. How do you build such a large animal? How can it function?" says Bukhman.

Bukhman suggests that a practical application of this knowledge is in the emerging area of stem cell-based therapies. To heal an injury, stem cells must differentiate into specialized cell types of the relevant organ or tissue. The speed of this process is controlled by some of the same molecular mechanisms that underlie the developmental clock.

What genomes from animals of different sizes can tell us about our own health

Understanding the genomes of the largest and smallest of mammals may also help unravel the biomedical mystery known as Peto's paradox. This is a curious phenomenon in which large mammals such as whales and elephants live longer and are less likely to develop cancer—often caused by DNA replication errors that occasionally happen during cell division—despite having a greater number of cells (and therefore more cell divisions) than smaller mammals like humans or mice.

Meanwhile, knowledge of the Etruscan shrew genome will enable new

insights in the field of metabolism. The shrew has an extremely high surface-to-volume ratio and fast metabolic rate. These high energy demands are a product of its tiny size—no bigger than a human thumb and weighing less than a penny—making it an interesting model to understand the regulation of metabolism better.

The blue whale and Etruscan shrew genome projects are part of a large collaborative effort involving dozens of contributors from institutions across North America and several European countries in conjunction with the Vertebrate Genomes Project.

The mission of the VGP is to assemble high-quality reference genomes for all living vertebrate species on Earth. This international consortium of researchers includes top experts in genome assembly and curation.

"The VGP has established a set of methods and criteria for producing a reference genome," Bukhman says. "Accuracy, contiguity, and completeness are three measures of quality."

Previous methods to sequence genomes used short read technologies, which produce short lengths of the DNA sequence 150 to 300 base pairs long, called reads. Overlapping reads are then assembled into longer contiguous sequences called contigs.

Contigs assembled from short reads tend to be relatively small compared to mammalian chromosomes. As a result, draft genomes reconstructed from such contigs tend to be very fragmented and have a lot of gaps.

Instead, the team used long-read sequencing, with reads around 10,000 base pairs in length, with the principal advantage being longer contigs and fewer gaps.

"Then you can use other methods such as optical mapping and Hi-C to

assemble contigs into bigger structures called scaffolds, and those can be as big as an entire chromosome," Bukhman explains.

The researchers also analyzed segmental duplications, large regions of duplicated sequence that often contain genes and can provide insight into evolutionary processes when compared to other species, either closely or distantly related.

They found that the blue whale had a large burst of segmental duplications in the recent past, with larger numbers of copies than the bottlenose dolphin and the vaquita (the world's smallest cetacean, the order of mammals including whales, dolphins, and porpoises). While most of the copies of genes created this way are likely non-functional, or their function is still unknown, the team did identify several known genes.

One encodes the protein metallothionein, which is known to bind heavy metals and sequester their toxicity—a useful mechanism for large animals that accumulate heavy metals while living in the ocean.

How reference genomes can help with wildlife conservation

A reference genome is also useful for species conservation. The blue whale was hunted almost to extinction in the first half of the 20th century. It is now protected by an international treaty and the populations are recovering.

"In the world's oceans, the blue whale is basically everywhere except for the high Arctic. So, if you have a reference genome, then you can make comparisons and can better understand the population structure of the different blue whale groups in different parts of the globe," Bukhman

says. "The blue whale genome is highly heterozygous, there's still a lot of genetic diversity, which has important implications for conservation."

This begs the question: how do you go about acquiring samples from a large, endangered creature that exists in the vastness of the oceans?

"The logistics posed several challenges, including the fact that blue whale sightings in our area are very rare and almost unpredictable," says Susanne Meyer, a research specialist at the University of California Santa Barbara, who spent over a year coordinating the permits, personnel, and resources needed to procure the samples.

Once their local whale-watching team determined the timing and coordinates of the whale sightings, they brought in licensed whale researcher Jeff K. Jacobsen to perform the whale biopsies using an approved standard [cetacean skin biopsy technique](#), which involves a custom stainless steel biopsy tube fitted to a crossbow arrow.

The team acquired samples from four blue whales, which Meyer used to develop and expand fibroblasts in cell culture for genome sequencing and further research use.

Size doesn't matter when it comes to an animal's genome

While the Etruscan shrew genome wasn't studied as extensively as the [blue whale](#) genome, the team reported an interesting finding.

"We found that there are relatively few segmental duplications in the shrew genome," Bukhman says, while emphasizing that this result does not necessarily correlate to the diminutive size of the shrew itself.

"While shrews belong to a different mammalian order, some similarly

small rodents have lots of segmental duplications, and the house mouse is kind of a champion in that sense that it has the most. So, it's not a matter of size."

As the Vertebrate Genomes Project makes strides in producing more high-quality reference genomes for all vertebrates, Bukhman is hopeful that contributions to those efforts will continue to advance biological research in the future.

More information: Yury V Bukhman et al, A High-Quality Blue Whale Genome, Segmental Duplications, and Historical Demography, *Molecular Biology and Evolution* (2024). [DOI: 10.1093/molbev/msae036](https://doi.org/10.1093/molbev/msae036)

Yury V. Bukhman et al, Chromosome level genome assembly of the Etruscan shrew *Suncus etruscus*, *Scientific Data* (2024). [DOI: 10.1038/s41597-024-03011-x](https://doi.org/10.1038/s41597-024-03011-x)

Provided by University of Wisconsin-Madison

Citation: Sequencing the blue whale and Etruscan shrew genomes (2024, March 18) retrieved 27 April 2024 from <https://phys.org/news/2024-03-sequencing-blue-whale-etruscan-shrew.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.