Title: High quality whole genome sequence of an abundant Holarctic odontocete, the harbour porpoise (*Phocoena phocoena*)

Authors: Marijke Autenrieth[1], Stefanie Hartmann[2], Ljerka Lah[1,3], Anna Roos[4], Alice B. Dennis[1], Ralph Tiedemann[1]*

[1] University of Potsdam, Institute of Biochemistry and Biology, Evolutionary Biology & Systematic Zoology, 14476 Potsdam, Germany

[2] University of Potsdam, Institute of Biochemistry and Biology, Evolution and Adaptive Genomics, 14476 Potsdam, Germany

[3] current address: Novartis BTDM Mengeš, Kolodvorska 27, SI-1234 Mengeš, Slovenia (All work in regards to this manuscript was performed at University of Potsdam[1])

[4] Swedish Museum of Natural History, SE-104 05 Stockholm, Sweden

***Corresponding author**: Prof. Dr. Ralph Tiedemann tiedeman@uni-potsdam.de

Abstract

The harbour porpoise (*Phocoena phocoena*) is a highly mobile cetacean found in waters across the Northern hemisphere. It occurs in coastal water and inhabits water basins that vary broadly in salinity, temperature, and food availability. These diverse habitats could drive differentiation among populations; population structure within the north Atlantic (north of 51° latitude) is not fully resolved, particularly in relation to Baltic Sea populations. Here we report the first harbour porpoise genome, assembled *de novo* from a Swedish Kattegat individual. The genome is one of the most complete cetacean genomes currently available, with a total size of 2.7 Gb, and 50% of the total length found in just 34 scaffolds. Using the largest scaffolds, we were able to examine chromosome-level rearrangements relative to the genome of the closest related species available, domestic cattle (*Bos taurus*). The draft annotation comprises 22,154 predicted gene models, which we further annotated through matches to NCBI nucleotide database, GO categorization, and motif prediction. To infer the adaptive abilities of this species, as well as their population history, we performed Bayesian skyline analysis of the genome, which is concordant with the demographic history of this species, including expansion and fragmentation events. Overall, this genome assembly, together with the draft annotation, represents a crucial addition to the limited genetic markers currently available for the study of porpoise and cetacean conservation, phylogeny, and evolution.

Introduction

As an apex predator, the harbour porpoise (*Phocoena phocoena*) is a key indicator for conservation and biodiversity measurements in the Nordic Seas (Hooker & Gerber, 2004; Lawrence et al., 2016; Sergio et al., 2008). Marine mammals in particular face many threats from their environment (Fietz et al., 2013; Godard-Codding et al., 2011) including noise pollution (Dyndo et al., 2015; Nabe-Nielsen et al., 2014), marine debris and by-catch (Scheidat et al., 2008; Unger et al., 2017), predation by grey seals (Leopold et al., 2014), and infectious diseases (Siebert et al., 2001; van Beurden et al., 2017). These threats impact structure, boundaries, and stability of populations. This is especially true in the Kattegat/Baltic Sea area, where broad ecological shifts have occurred on a relatively short time scale. Since forming 15,000 years ago, the Baltic has undergone periods of brackish, marine, and completely fresh water, and encountered increasing and continuous humans impacts including eutrophication, pollution and overharvesting (Korpinen et al., 2012; Paasche et al., 2015; Ukkonen et al., 2014; Varjopuro et al., 2014). This geological history has created a series of challenges to marine species, and has likely fostered local adaption and population differentiation.

Harbour porpoises are the most abundant costal cetaceans across their wide distribution from sub-polar to temperate waters in the Northern hemisphere (Fontaine et al., 2017; Gaskin, 1984). As one of the smallest marine mammals, they belong to the *Delphinoida* and are the sister group to the *Monodontidae* (Gatesy et al., 2013; Geisler et al., 2011; Hassanin et al., 2012). Three subspecies of harbour porpoise, *P. p. vomerina* (North Pacific), *P. p. relicta* (Black Sea) and *P. p. phocoena* (North Atlantic), can be differentiated genetically (Rosel et al., 1999), but also by morphological traits including body size and diet (Fontaine et al., 2017; Galatius et al., 2012).

The population size of harbour porpoises in European Atlantic Shelf waters is estimated to be 375,000 with shifts across the last decade in the exact regions they occupy (e.g. in the North Sea; (Hammond et al., 2013). Estimates of population size in the western Baltic Sea are smaller, approximately 40,000 animals (Benke et al., 2014; Scheidat et al., 2008; Viquerat et al., 2014). The Baltic Sea proper population, which is not included in the former surveys, has very low estimates (below 500 individuals; Amundin, 2016) and is considered critically endangered (Benke et al., 2014; Hammond et al., 2008; Scheidat et al., 2008).

As with other marine mammals in the Northern Atlantic, e.g. grey and harbour seals, subpopulations of the harbor porpoise arose during the end of the last glacial period as North Sea populations recolonized the Baltic Sea (Fietz et al., 2016). Now these different populations show shifts in habitat use based largely on food availability (Hammond et al., 2013) and activity patterns (Nuuttila et al., 2017), and display fine scale morphological and genetic differences (Fontaine et al., 2012, 2014; Wiemann et al., 2010) and significant isolation by distance (Lah et al., 2016). Recent studies based on morphometric and genetic data suggest that different ecotypes of harbour porpoise in the North Atlantic and Baltic Sea exist and may need further conservation measures (Fontaine et al., 2014, 2017; Galatius et al., 2012).

These fine scale differences in morphology and behavior may constitute local adaptation, yet the genes underlying such a potentially adaptive differentiation are still unknown and would be best investigated on a whole-genome scale. To examine this, there is a need for high quality genomic resources for this species. A genome will also allow for a broader investigation of population structure, demographic history, functional, and evolutionary questions, as has been shown for other cetacean species in recent studies (Foote et al., 2016; Keane et al., 2015; Nery et al., 2013; Sun et al., 2013; Yim et al., 2013; Zhou et al., 2013). To this end, a full genome will enable mapping of so far anonymous nuclear microsatellite (Wiemann et al. 2010) and SNP (Lah et al. 2016) loci, thus facilitating population genomic inference.

We present here the first *de novo* assembly of the full genome of the harbor porpoise, scaffolded with *in vitro* proximity ligation data (hereafter "Chicago" library), and draft-annotated to predict its coding proteins and their functions (Deposited at NCBI as BioProject: PRJNA417595 with BioSample-ID: SAMN08000480). We demonstrate chromosome-level homology with other Cetartiodactyla (Gatesy et al., 2013), and insight into past population dynamics using a Bayesian skyline plot (Li & Durbin, 2011).

Materials and Methods

DNA sampling

Tissue for whole genome sequencing came from a single individual from the Kattegat (Glommen - Falkenberg), Sweden (ID: C2009/02665). Muscle tissue was sampled in July 2009 from a by-caught female of probably young age (22.4kg, 110.5m), frozen, and transported to Potsdam, Germany for DNA extraction. Sample preparation and Genomic DNA isolation were performed following the Quiagen DNeasy Blood & Tissue Kit (Cat 69506, Hilden). Successful high molecular weight DNA-isolation was confirmed by Sanger sequencing of the mitochondrial control region, and visualization of fragment sizes of the entire extraction using the Tape Station (Agilent 2200, Santa Clara, CA 95051). By mtDNA sequencing, we verified that the analyzed specimen carried haplotype PHO7 (Tiedemann et al., 1996), indicative of the separated Beltsea population of the Kattegat/Western Baltic Sea region (Lah et al., 2016; Wiemann et al., 2010).

Genome sequencing and assembly

The draft *de novo* assembly was constructed from two libraries (insert sizes ca. 300 and ca. 500bp); sequenced in 125bp PE on the Illumina HiSeq 2500 at EUROFINS Genomics. Reads were trimmed using CUTADAPT v1.10 (Martin, 2011) and an initial assembly was made using SOAPDENOVO2 (Luo et al., 2015). DNA from the same sample was used by Dovetail Genomics for construction of a Chicago library (Putnam et al., 2016), and sequenced in 150bp PE reads on an Illumina NextSeq500 at the University of Potsdam. The draft assembly was then scaffolded with the Chicago library results for the final HiRise assembly, performed by Dovetail Genomics.

Presence of core, single copy, and orthologous genes was measured using CEGMA and BUSCO, run in the genome mode for the Laurasiatheria database (Simão et al., 2015). BLOBTOOLS was run to examine potential contaminants, based on divergence in GC-content and read coverage variation across the assembly (Laetsch & Blaxter, 2017).

Genome annotation

Genome annotation was performed by MAKER2 (Holt & Yandell, 2011) in two steps. MAKER2 makes use of different programs and draws from several lines of evidence. Prior to annotation, repetitive elements were soft-masked with REPEATMASKER (Smit et al., 2013-2015) using the te_protein repeat database (Smith et al., 2007). In the first MAKER2 run, three gene predictors were used: SNAP (Bromberg et al., 2008) was *ab initio* trained with the CEGMA results (Parra et al., 2007), GENEMARK-ES (Ter-Hovhannisyan et al., 2008) was run using an HMM produced by *ab initio* training on the whole *P. phocoena* genome, and AUGUSTUS was run using the presets for human, as is recommended for vertebrates (Stanke et al., 2004). Protein sequences, supplied as evidence were obtained from the complete

SwissProt database (553,941 Proteins) plus NCBI entries of 184,527 proteins from eight different cetacean groups (On 20 March 2017, all hits to following keywords: "*Balaenopteridae*", "*Lipotes vexillifer*", "*Neophocaena*", "*Orcinus orca*", "*Phocoena*", "*Physeter catodon*", "*Pontoporia blainvillei*", "*Tursiops truncatus*").

For the second MAKER2 run, we created a new SNAP-HMM based on the first MAKER2 output, and ran it with the same parameters as the first run, exchanging only the SNAP HMM and excluding the protein evidence. The resulting CDS predictions were extracted from the final gff file, which was created by *fathom* implemented in SNAP (Bromberg et al., 2008). These gene predictions were further verified by a BLASTN search against the entire GENBANK non-redundant nucleotide sequence database (date downloaded 21.07.2017). Summary statistics were generated using GENOME ANNOTATION GENERATOR (Hall et al., 2014). We then used all CDS and their BLAST results in BLAST2GO (Goetz et al., 2008) to identify conserved protein domains with INTERPROSCAN (including a Pfam comparison). We functional annotated the CDS with GO terms, which are a controlled vocabulary to describe gene function constantly actualized by the Gene Ontology Consortium (Ashburner et al., 2000; Carbon et al., 2017).

Comparative genomics

The closest relative with a chromosome-level assembly currently available is the domestic cattle, *Bos taurus*. To validate our assembly, we compared our scaffolds to the *B. taurus* chromosomes (assembly UMD 3.1.1 downloaded from NCBI, ACCESSION DAAA00000000). Specifically, the 122 *P. phocoena* scaffolds of at least 1Mbp were aligned to the *B. taurus* chromosomes using the nucmer software of the MUMMER package v. 3.23 (Kurtz et al., 2004). From the coordinates of these alignments, runs of ten or more consecutive matches of each at least 250bp between a given *P. phocoena* scaffold and a *B. taurus* chromosome were extracted using custom perl scripts. Their start and end positions were used to generate a CIRCOS (http://circos.ca/) plot that shows regions of collinearity as well as rearrangements. For the CIRCOS plot, separate ribbons are displayed between a *B. taurus* chromosome and a *P. phocoena* scaffold for consecutive hits that were each no more than 20,000 bp apart. If a hit is more than 20,000bp from the next run of consecutive hits, a new ribbon was started; in total 24,394 separate ribbons were constructed (Figure 1).

Population genomics

In using genome-wide diploid sequence data it is possible to reconstruct the population history in estimating population sizes through the past (Li et al., 2011). To estimate the demographic history of the individual sequenced, we used the SNP Frequency spectra based on our genome assembly, which is a haploid sequence, and the PE reads used to construct the *de novo* assembly, prior to Chicago scaffolding (described above, we used both insert sizes). These reads were first mapped back to the final assembly using BWA (Li & Durbin, 2009). SNP data was extracted from the resulting bam files, and variants were extracted using SAMTOOLS vs.1.6. (Li, Handsaker, et al., 2009), and BCFTOOLS (Li, Handsaker, et al., 2009), implemented with the script vcfutils.pl (Li, Handsaker, et al., 2009). This generated a final *.fq.gz file, which was then used to generate the final Bayesian skyline plot in the PSMC package, using perl scripts psmc2history.pl and psmc_plot.pl (Li et al., 2011). The parameters of the PSMC analysis were set following the recommendation from the authors (Li & Durbin, 2011, https://github.com/lh3/psmc) and we applied a generation time of 10 years (Birkun Jr. & Frantzis, 2008) and a mutation rate of 2.2×10^{-9} year/site (Taylor et al., 2007).

Results

De novo *assembly of the* P. phocoena *genome*

Shotgun sequencing produced a total number of 1,268M reads (Table 1), these were used to generate a draft assembly with 2.4M scaffolds and an N50 of 33.1kb. This assembly was combined with the Chicago library data (556M read) for final scaffolding by Dovetail Genomics (Putnam et al., 2016). The final HiRise assembly from Dovetail contains ca. 2M scaffolds (Table 2) and has a total length of 2.7Gb (N50 of 23.8Mb). The greatest improvements from the addition of the Chicago libraries is in building up the 34 longest scaffolds, which make up approximately half of the entire assembly (Table 2). The CIRCOS plot illustrates the near-completeness of these long scaffolds. We observe almost complete coverage of the cow chromosomes by scaffolds bigger than 1Mb in our assembly (Figure 1). The BUSCO and CEGMA analyses also suggests that we have largely reconstructed the entire genome, and identified 96.9% (91.3% complete) of the 2,586 Eukaryotic and 94.2 (88.7% complete) of the 6,253 Laurasiatheria BUSCO core genes and 90% of the 248 ultra-conserved CEGs (54% complete).

Genome completeness and annotation

The MAKER2 annotation resulted in the prediction of 22,154 coding genes (Table 3). In total 21,750 CDS had a BLAST hit against the nucleotide database, which accounts for 98% of the total CDSs. Of these BLAST hits, 99% account for vertebrate, and these were dominated (90%) by hits to Cetacea (thereof 59% *Tursiops truncatus*, 27% *Orcinus orca*). Further annotation with INTERPROSCAN revealed 250,126 features of these predicted proteins. These comprise hits in several protein domain databases, e.g. 23,319 PFAM protein domains, 37,046

PANTHER gene families, 24,538 SUPERFAMILY annotations and 31,114 GENE3D domains. Assignment of the BLAST results to Gene Ontology (GO) categories resulted in 55,143 hits across the GO categories (

Figure ***2***).

Inference of Kattegat/Baltic population history

We inferred the population history of the harbour porpoise *P. phocoena* based on one single individual (Li et al., 2011) using the PSMC algorithm, which combines all generated PE read data generated. Between eight and four million years ago the inferred population size (N_e) was low, around 10,000 individuals (Figure 3). It began to increase slightly at 3Myr, and rose more rapidly around 2Myr, reaching an N_e of 45,000 during the following 1.5 Myr. The estimated population size peaked approximates 400kyra before it dropped to a quarter of the original size around 100kyrs ago, leading to a very low N_e, similar to that seen in present day populations (Hammond et al., 2013).

Discussion

We present here a high quality *de novo* genome assembly for the harbour porpoise *Phocoena phocoena*. With a GC-content of 41.4% and a total length of 2.7 GB, this assembly is comparable to other high quality genomes (Groenen et al., 2013; Zimin et al., 2009). BUSCO and CEGMA gene scans support a near completeness of core genes in the assembly, and support that we have largely reconstructed the entire genome. For almost completely covering the chromosomes of the *B. taurus* genome (Figure 1), only 122 scaffolds are needed, including the 34 largest scaffolds representing 50% of the whole genome. Of these largest scaffolds some completely match single *B. taurus* chromosomes, e.g., chromosome 25. Other *B. taurus* chromosomes are in only 2-3 pieces in our scaffolds, e.g. chromosomes 12, 24. Based on this comparison, we infer that our assembly represents a nearly complete genome of *P. phocoena*, and that our largest scaffolds are nearly-complete chromosomes. The CIRCOS plot also illustrates chromosomal rearrangements between domestic cattle and the harbour porpoise, two species diverged approximately 60Myrs ago within the Cetartiodactyla (Gatesy et al., 2013). These chromosomal rearrangements are seen several times among distinct lineage of Cetartiodactyla (Avila et al., 2015; Kulemzina et al., 2009, 2011; Pauciullo et al., 2014), e.g., comparison between camel, pig and domestic cattle (Balmus et al., 2007).

The number of annotated genes (22,154) is comparable to other published cetacean genomes: 21,459 bottlenose dolphin (Lindblad-Toh et al., 2011), 20,605 minke whale (Yim et al., 2013), 22,711 grey whale (DeWoody et al., 2017). They appear to broadly span key functional gene categories, e.g. biological processes, cellular components and molecular function, both across the annotated GO terms and the INTERPROSCAN analysis. With this information we can directly search for known, respectively key genes, for further investigations, e.g. selection or adaptive traits.

The harbor porpoise is estimated to have split from is closest relative ca. 5Myr ago (Gatesy et al., 2013). Interestingly our Bayesian skyline plot (Figure 3) coincides with this date by starting a population expansion around that time point. Around 4.5 Myr ago an expansion occurred, during which time the North Atlantic is known to have cooled, leading to an extinction of 65% of the marine organisms (Stanley, 1995). The harbour porpoise is well known in subarctic regions and some populations (e.g. Greenland) occur in areas which freeze to a large extent during winter (Tolley & Rosel, 2006). Therefore, an extinction of other marine species during a cold water period does not preclude that the harbour porpoise could increase its population size and expand through the Atlantic. During the last interglacial period, Eemian, the inferred N_e remained relatively high at around 50,000 individuals before, dropping dramatically with the beginning of the last glacial period 100kya. When comparing this pattern to the demographic history of other cetaceans, it is most similar to the bottlenose dolphin (*Tursiops truncatus*), a related species with a similar North Atlantic distribution (Brüniche-Olsen et al., n.d.; Foote et al., 2016; Yim et al., 2013; Zhou et al., 2013). The newly forming sea ice areas, around 400kya ago, could have led to fragmentation of different populations, and therefore lead to a drop in regional total effective population size in regards to our sample. A potential low population size we see postulated for today would fit to the history of the Baltic Sea and the population status of *P. phocoena* (Johannesson et al., 2011; Johannesson & André, 2006; Ukkonen et al., 2014). Specifically, there is strong evidence for a Western Baltic/Kattegat (i.e., Beltsea) population separated from the North Sea/North Atlantic (Hammond et al., 2013; Lah et al., 2016), which currently counts approximately 40,000 animals (Benke et al., 2014; Scheidat et al., 2008; Viquerat et al., 2014). Our sequenced specimen was assigned with high likelihood to this Beltsea population by mtDNA analysis (exhibiting haplotype PHO 7; cf. Tiedemann et al., 1996; Wiemann et al., 2010).

In this study we present the first whole genome assembly and annotation of the harbour porpoise, at this point the most complete assembly for the Family *Phocoenidae*. This genome adds to the Cetacean genome collection by supplying important resources for further investigation within the *Odontoceti* as well as outside the *Cetacea*. This will provide an invaluable resource for further genetic studies within the harbour porpoise itself, both as a resource for whole-genome investigations into population structure and to identify key genes associated with local adaptation. This genome represents a crucial genetic resource for further investigation in the population genetics and phylogeny on other species of the *Phocoenidae* including the currently most rare marine mammal, the almost extinct Vaquita (*Phocoena sinus*) (Taylor et al., 2017), and is hence especially important for conservation efforts.

Acknowledgments

Financial support came from the Bundesamt für Naturschutz (FKZ # 3514824600), as part of a larger study of population genomics. We thank Prof. Dr. Michael Hofreiter for providing access to the Illumina NextSeq platform. Additional support came from the University of Potsdam. Large-scale computational effort was made possible by computing resources provided by the department of Genetics at University of Potsdam and the High Performance Computing Cluster Orson2, managed by ZIM (Zentrum für Informationstechnologie und Medienmanagement) at the University of Potsdam.

References

Amundin, M. (2016). SAMBAH - Static Acoustic Monitoring of the Baltic Sea Harbour porpoise. LIFE Project Number LIFE 08 NAT/S/000261 European Commission, 77pp. available at http://www.sambah.org/SAMBAH-Final-Report-FINAL-for-website-April-2017.pdf.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2000). Gene ontologie: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. doi:10.1038/75556.Gene

Avila, F., Baily, M. P., Merriwether, D. A., Trifonov, V. A., Rubes, J., Kutzler, M. A., … Raudsepp, T. (2015). A cytogenetic and comparative map of camelid chromosome 36 and the minute in alpacas. *Chromosome Research*, *23*(2), 237–251. doi:10.1007/s10577-014-9463-3

Balmus, G., Trifonov, V. A., Biltueva, L. S., O'Brien, P. C. M., Alkalaeva, E. S., Fu, B., … Ferguson-Smith, M. A. (2007). Cross-species chromosome painting among camel, cattle, pig and human: Further insights into the putative Cetartiodactyla ancestral karyotype. *Chromosome Research*, *15*(4), 499–515. doi:10.1007/s10577-007-1154-x

Benke, H., Bräger, S., Dähne, M., Gallus, A., Hansen, S., Honnef, C. G., … Verfuß, U. K. (2014). Baltic Sea harbour porpoise populations: Status and conservation needs derived from recent survey results. *Marine Ecology Progress Series*, *495*, 275–290. doi:10.3354/meps10538

Birkun Jr., A. A., & Frantzis, A. (2008). *Phocoena phocoena ssp. relicta*. The IUCN Red List of Threatened Species 2008: e.T17030A6737111. Downloaded on 12 October 2017.

Bromberg, Y., Yachdav, G., & Rost, B. (2008). SNAP predicts effect of mutations on protein function. *Bioinformatics*, *24*(20), 2397–2398. doi:10.1093/bioinformatics/btn435

Brüniche-Olsen, A., Westerman, R., Kazmierczyk, Z., Vertyankin, V. V., Godard-Codding, C., Bickham, J. W., & DeWoody, J. A. (n.d.). The inference of gray whale (Eschrichtius robustus) population attributes from whole-genome sequences.

Carbon, S., Dietze, H., Lewis, S. E., Mungall, C. J., Munoz-Torres, M. C., Basu, S., … Westerfield, M. (2017). Expansion of the gene ontology knowledgebase and resources: The gene ontology consortium. *Nucleic Acids Research*, *45*(D1), D331–D338. doi:10.1093/nar/gkw1108

DeWoody, J. A., Fernandez, N. B., Brüniche-Olsen, A., Antonides, J. D., Doyle, J. M., San Miguel, P., … Bickham, J. (2017). Characterization of the gray whale (Eschrichtius robustus) genome and a genotyping array based on single nucleotide polymorphisms in candidate genes. *Biological Bulletin*, *232*(June), 186–197.

Dyndo, M., Wiśniewska, D. M., Rojano-Doñate, L., & Madsen, P. T. (2015). Harbour porpoises react to low levels of high frequency vessel noise. *Scientific Reports*, *5*, 11083. doi:10.1038/srep11083

Fietz, K., Galatius, A., Teilmann, J., Dietz, R., Frie, A. K., Klimova, A., … Olsen, M. T. (2016). Shift of grey seal subspecies boundaries in response to climate, culling and conservation. *Molecular Ecology*, *25*(17), 4097–4112. doi:10.1111/mec.13748

Fietz, K., Graves, J. A., & Olsen, M. T. (2013). Control Control Control: A Reassessment and Comparison of GenBank and Chromatogram mtDNA Sequence Variation in Baltic Grey Seals (*Halichoerus grypus*). *PLoS ONE*, *8*(8), 1–7. doi:10.1371/journal.pone.0072853

Fontaine, M. C., Roland, K., Calves, I., Austerlitz, F., Palstra, F. P., Tolley, K. A., … Aguilar, A. (2014). Postglacial climate changes and rise of three ecotypes of harbour porpoises, *Phocoena phocoena*, in western Palearctic waters. *Molecular Ecology*, *23*(13), 3306–3321. doi:10.1111/mec.12817

Fontaine, M. C., Snirc, A., Frantzis, A., Koutrakis, E., Oztürk, B., Oztürk, A. a, & Austerlitz, F. (2012). History of expansion and anthropogenic collapse in a top marine predator of the Black Sea estimated from genetic data. *Proceedings of the National Academy of Sciences of the USA*, *109*(38), E2569-76. doi:10.1073/pnas.1201258109

Fontaine, M. C., Thatcher, O., Ray, N., Piry, S., Brownlow, A., Davison, N. J., … Goodman, S. J. (2017). Mixing of porpoise ecotypes in southwestern UK waters revealed by genetic profiling. *Royal Society Open Science*, *4*(3), 160992. doi:10.1098/rsos.160992

Foote, A. D., Vijay, N., Ávila-Arcos, M. C., Baird, R. W., Durban, J. W., Fumagalli, M., … Wolf, J. B. W. (2016). Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nature Communications*, *7*(May), 11693. doi:10.1038/ncomms11693

Galatius, A., Kinze, C. C., & Teilmann, J. (2012). Population structure of harbour porpoises in the Baltic region: evidence of separation based on geometric morphometric comparisons. *Journal of the Marine Biological Association of the United Kingdom*, *92*(8), 1669–1676. doi:10.1017/S0025315412000513

Gaskin, D. (1984). The harbour porpoise *Phocoena phocoena* (L.): regional populations, status, and infromation on direct and indirect catches. *Reports of the International Whaling Commission*, *34*, 569–586.

Gatesy, J., Geisler, J. H., Chang, J., Buell, C., Berta, A., Meredith, R. W., … McGowen, M. R. (2013). A phylogenetic blueprint for a modern whale. *Molecular Phylogenetics and Evolution*, *66*(2), 479–506. doi:10.1016/j.ympev.2012.10.012

Geisler, J. H., McGowen, M. R., Yang, G., & Gatesy, J. (2011). A supermatrix analysis of genomic, morphological, and paleontological data from crown Cetacea. *BMC Evolutionary Biology*, *11*(1), 112. doi:10.1186/1471-2148-11-112

Godard-Codding, C. A. J., Clark, R., Fossi, M. C., Marsili, L., Maltese, S., West, A. G., … Stegeman, J. J. (2011). Pacific ocean-wide profile of CYP1A1 expression, stable carbon and nitrogen isotope ratios, and organic contaminant burden in sperm whale skin biopsies. *Environmental Health Perspectives*, *119*(3), 337–343. doi:10.1289/ehp.0901809

Goetz, S., Garccia-Gomez, M. J., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., … Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, *36*(10), 3420–3435. doi:10.1093/nar/gkn176

Groenen, M. A. M., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., … Hunt, T. (2013). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*, *491*(7424), 393–398. doi:10.1038/nature11622

Hall, B., DeRego, T., & Geib, S. (2014). GAG: the Genome Annotation Generator (Version 1.0) [Software]. doi:Available from http://genomeannotation.github.io/GAG

Hammond, P. S., Bearzi, G., Bjørge, A., Forney, K. A., Karczmarski, L., Kasuya, T., … Wilson, B. (2008). *Phocoena phocoena* (Baltic Sea subpopulation). (errata version published in 2016) The IUCN Red List of Threatened Species 2008: e.T17031A98831650. Downloaded on 10 October 2017. Retrieved from http://www.iucnredlist.org/details/17031/0

Hammond, P. S., Macleod, K., Berggren, P., Borchers, D. L., Burt, L., Canadas, A., … Vazquez, J. A. (2013). Cetacean abundance and distribution in European Atlantic shelf waters to inform conservation and management. *Biological Conservation*, *164*, 107–122. doi:10.1016/j.biocon.2013.04.010

Hassanin, A., Delsuc, F., Ropiquet, A., Hammer, C., Jansen Van Vuuren, B., Matthee, C., … Couloux, A. (2012). Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *Comptes Rendus - Biologies*, *335*(1), 32–50. doi:10.1016/j.crvi.2011.11.002

Holt, C., & Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, *12*(1), 491. doi:10.1186/1471-2105-12-491

Hooker, S. K., & Gerber, L. (2004). Marine Reserves as a Tool for Ecosystem-Based Management : The Potential Importance of Megafauna. *BioScience*, *54*(1), 27–39. doi:10.1641/0006-3568(2004)054[0027:MRAATF]2.0.CO;2

Johannesson, K., & André, C. (2006). Life on the margin: Genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Molecular Ecology*, *15*(8), 2013–2029. doi:10.1111/j.1365-294X.2006.02919.x

Johannesson, K., Smolarz, K., Grahn, M., & André, C. (2011). The future of baltic sea populations: Local extinction or evolutionary rescue? *Ambio*, *40*(2), 179–190. doi:10.1007/s13280-010-0129-x

Keane, M., Semeiks, J., Webb, A. E., Li, Y. I., Quesada, V., Craig, T., … deMagalhães, J. P. (2015). Insights into the evolution of longevity from the bowhead whale genome. *Cell Reports*, *10*(1), 112–122. doi:10.1016/j.celrep.2014.12.008

Korpinen, S., Meski, L., Andersen, J. H., & Laamanen, M. (2012). Human pressures and their potential impact on the Baltic Sea ecosystem. *Ecological Indicators*, *15*(1), 105–114. doi:10.1016/j.ecolind.2011.09.023

Kulemzina, A. I., Trifonov, V. A., Perelman, P. L., Rubtsova, N. V., Volobuev, V., Ferguson-Smith, M. A., … Graphodatsky, A. S. (2009). Cross-species chromosome painting in Cetartiodactyla: Reconstructing the karyotype evolution in key phylogenetic lineages. *Chromosome Research*, *17*(3), 419–436. doi:10.1007/s10577-009-9032-3

Kulemzina, A. I., Yang, F., Trifonov, V. A., Ryder, O. A., Ferguson-Smith, M. A., & Graphodatsky, A. S. (2011). Chromosome painting in Tragulidae facilitates the reconstruction of Ruminantia ancestral karyotype. *Chromosome Research*, *19*(4), 531–539. doi:10.1007/s10577-011-9201-z

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, *5*(2), R12. doi:10.1186/gb-2004-5-2-r12

Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research*, *6*(1287), 1–15. doi:10.12688/f1000research.12232.1 and doi:10.5281/zenodo.845347

Lah, L., Trense, D., Benke, H., Berggren, P., Gunnlaugsson, Þ., Lockyer, C., … Tiedemann, R. (2016). Spatially Explicit Analysis of Genome-Wide SNPs Detects Subtle Population Structure in a Mobile Marine Mammal, the Harbor Porpoise. *PLoS ONE*, *11*(10), 1–23. doi:10.1371/journal.pone.0162792

Lawrence, J. M., Armstrong, E., Gordon, J., Lusseau, S. M., & Fernandes, P. G. (2016). Passive and active, predator and prey: using acoustics to study interactions between cetaceans and forage fis. *ICES Journal of Marine Science*, *73*(8), 2075–2084. doi:10.1093/icesjms/fsw013

Leopold, M. F., Begeman, L., van Bleijswijk, J. D. L., IJsseldijk, L. L., Witte, H. J., & Gröne, A. (2014). Exposing the grey seal as a major predator of harbour porpoises. *Proc. R. Soc. B*, *282*, 20142429. doi:10.1098/rspb.2014.2429

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. doi:10.1093/bioinformatics/btp324

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493–496. doi:10.1038/nature10231

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. doi:10.1093/bioinformatics/btp352

Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M. F., Parker, B. J., Washietl, S., … Kellis, M. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, *478*(7370), 476–482. doi:10.1038/nature10530

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., … Wang, J. (2015). Erratum: SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*, *4*(1), 30. doi:10.1186/s13742-015-0069-2

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*(1), 10. doi:10.14806/ej.17.1.200

Nabe-Nielsen, J., Sibly, R. M., Tougaard, J., Teilmann, J., & Sveegaard, S. (2014). Effects of noise and by-catch on a Danish harbour porpoise population. *Ecological Modelling*, *272*, 242–251. doi:10.1016/j.ecolmodel.2013.09.025

Nery, M. F., Gonzalez, D. J., & Opazo, J. C. (2013). How to Make a Dolphin: Molecular Signature of Positive Selection in Cetacean Genome. *PLoS ONE*, *8*(6), 2–8. doi:10.1371/journal.pone.0065491

Nuuttila, H. K., Courtene-Jones, W., Baulch, S., Simon, M., & Evans, P. G. H. (2017). Don't forget the porpoise: acoustic monitoring reveals fine scale temporal variation between bottlenose dolphin and harbour porpoise in Cardigan Bay SAC. *Marine Biology*, *164*(3), 1–16. doi:10.1007/s00227-017-3081-5

Paasche, Ø., Österblom, H., Neuenfeldt, S., Bonsdorff, E., Brander, K., Conley, D. J., … Stenseth, N. C. (2015). Connecting the Seas of Norden. *Nature Climate Change*, *5*(2), 89–92. doi:10.1038/nclimate2471

Parra, G., Bradnam, K., & Korf, I. (2007). CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, *23*(9), 1061–1067. doi:10.1093/bioinformatics/btm071

Pauciullo, A., Perucatti, A., Cosenza, G., Iannuzzi, A., Incarnato, D., Genualdo, V., … Iannuzzi, L. (2014). Sequential cross-species chromosome painting among river buffalo, cattle, sheep and goat: A useful tool for chromosome abnormalities diagnosis within the family bovidae. *PLoS ONE*, *9*(10). doi:10.1371/journal.pone.0110297

Putnam, N. H., Connell, B. O., Stites, J. C., Rice, B. J., Hartley, P. D., Sugnet, C. W., … Rokhsar, D. S. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Research*, *26*, 342–350. doi:10.1101/gr.193474.115

Rosel, P. E., Tiedemann, R., & Walton, M. (1999). Genetic evidence for limited trans-Atlantic movements of the harbor porpoise, *Phocoena phocoena*. *Marine Biology (Berlin)*, *133*(4), 583–591.

Scheidat, M., Gilles, A., Kock, K. H., & Siebert, U. (2008). Harbour porpoise *Phocoena phocoena* abundance in the southwestern Baltic Sea. *Endangered Species Research*, *5*(2–3), 215–223. doi:10.3354/esr00161

Sergio, F., Caro, T., Brown, D., Clucas, B., Hunter, J., Ketchum, J., … Hiraldo, F. (2008). Top Predators as Conservation Tools: Ecological Rationale, Assumptions, and Efficacy. *Annual Review of Ecology, Evolution, and Systematics*, *39*(1), 1–19. doi:10.1146/annurev.ecolsys.39.110707.173545

Siebert, U., Wünschmann, A., Weiss, R., Frank, H., Benke, H., & Frese, K. (2001). Post-mortem findings in harbour porpoises (*phocoena phocoena*) from the German North and Baltic Seas. *Journal of Comparative Pathology*, *124*(2–3), 102–114. doi:10.1053/jcpa.2000.0436

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. doi:10.1093/bioinformatics/btv351

Smit, A. F. A., Hubley, R., & Green, P. (n.d.). RepeatMasker Open-4.0. <http://www.repeatmasker.org>.

Smith, C. D., Edgar, R. C., Yandell, M. D., Smith, D. R., Celniker, S. E., Myers, E. W., & Karpen, G. H. (2007). Improved repeat identification and masking in Dipterans. *Gene*, *389*(1), 1–9. doi:10.1016/j.gene.2006.09.011

Stanke, M., Steinkamp, R., Waack, S., & Morgenstern, B. (2004). AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Research*, *32*(Web Server issue), W309–W312. doi:10.1093/nar/gkh379

Stanley, S. M. (1995). 7 Neogene Ice Age in the North Atlantic Region: Climatic Changes, Biotic Effects, and Forcing Factors. In *Effects of Past Global Change on Life*. Washington (DC): National Academies: National Research Council (US) Panel on Effects of Past Global Change on Life.

Sun, Y. B., Zhou, W. P., Liu, H. Q., Irwin, D. M., Shen, Y. Y., & Zhang, Y. P. (2013). Genome-wide scans for candidate genes involved in the aquatic adaptation of dolphins. *Genome Biology and Evolution*, *5*(1), 130–139. doi:10.1093/gbe/evs123

Taylor, B. L., Chivers, S. J., Larese, J., & Perrin, W. F. (2007). Generation length and percent mature estimates for IUCN assessments of cetaceans. *Administrative Report LJ-07-01 National Marine Fisheries*, 24. doi:10.1.1.530.4789

Taylor, B. L., Rojas-Bracho, L., Moore, J., Jaramillo-Legorreta, A., Ver Hoef, J. M., Cardenas-Hinojosa, G., … Hammond, P. S. (2017). Extinction is Imminent for Mexico's Endemic Porpoise Unless Fishery Bycatch is Eliminated. *Conservation Letters*, *10*(5), 588–595. doi:10.1111/conl.12331

Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O., & Borodovsky, M. (2008). Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training, 1979–1990. doi:10.1101/gr.081612.108

Tiedemann, R., Harder, J., Gmeiner, C., & Haase, E. (1996). Mitochondrial DNA sequence patterns of Harbour porpoises (*Phocoena phocoena*) from the North and the Baltic Sea. *Zeitschrift Für Säugetierkunde*, *61*, 104–111.

Tolley, K., & Rosel, P. (2006). Population structure and historical demography of eastern North Atlantic harbour porpoises inferred through mtDNA sequences. *Marine Ecology Progress Series*, *327*, 297–308. doi:10.3354/meps327297

Ukkonen, P., Aaris-Sorensen, K., Arppe, L., Daugnora, L., Halkka, A., Lougas, L., … Stora, J. (2014). An Arctic seal in temperate waters: History of the ringed seal (*Pusa hispida*) in the Baltic Sea and its adaptation to the changing environment. *The Holocene*, *24*(12), 1694–1706. doi:10.1177/0959683614551226

Unger, B., Herr, H., Benke, H., Böhmert, M., Burkhardt-Holm, P., Dähne, M., … Siebert, U. (2017). Marine debris in harbour porpoises and seals from German waters. *Marine Environmental Research*, 1–8. doi:10.1016/j.marenvres.2017.07.009

van Beurden, S. J., Ijsseldijk, L. L., van de Bildt, M. W. G., Begeman, L., Wellehan, J. F. X., Waltzek, T. B., … Penzes, J. J. (2017). A novel cetacean adenovirus in stranded harbour porpoises from the North Sea: detection and molecular characterization. *Archives of Virology*, *162*(7), 2035–2040. doi:10.1007/s00705-017-3310-8

Varjopuro, R., Andrulewicz, E., Blenckner, T., Dolch, T., Heiskanen, A. S., Pihlajamäki, M., … Psuty, I. (2014). Coping with persistent environmental problems: Systemic delays in reducing eutrophication of the Baltic Sea. *Ecology and Society*, *19*(4). doi:10.5751/ES-06938-190448

Viquerat, S., Herr, H., Gilles, A., Peschko, V., Siebert, U., Sveegaard, S., & Teilmann, J. (2014). Abundance of harbour porpoises (*Phocoena phocoena*) in the western Baltic, Belt Seas and Kattegat. *Marine Biology*, *161*(4), 745–754. doi:10.1007/s00227-013-2374-6

Wiemann, A., Andersen, L. W., Berggren, P., Siebert, U., Benke, H., Teilmann, J., … Tiedemann, R. (2010). Mitochondrial Control Region and microsatellite analyses on harbour porpoise (*Phocoena phocoena*) unravel population differentiation in the Baltic Sea and adjacent waters. *Conservation Genetics*, *11*(1), 195–211. doi:10.1007/s10592-009-0023-x

Yim, H.-S., Cho, Y. S., Guang, X., Kang, S. G., Jeong, J.-Y., Cha, S.-S., … Lee, J.-H. (2013). Minke whale genome and aquatic adaptation in cetaceans. *Nature Genetics*, *46*(1), 88–92. doi:10.1038/ng.2835

Zhou, X., Sun, F., Xu, S., Fan, G., Zhu, K., Liu, X., … Yang, G. (2013). Baiji genomes reveal low genetic variability and new insights into secondary aquatic adaptations. *Nature Communications*, *4*(2708), 1–6. doi:10.1038/ncomms3708

Zimin, A. V, Delcher, A. L., Florea, L., Kelley, D. R., Schatz, M. C., Puiu, D., … Salzberg, S. L. (2009). A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biology*, *10*(4), R42. doi:10.1186/gb-2009-10-4-r42

Data Accessibility

The genome assembly, finale genome sequence and the draft annotation are deposit on NCBI under BioProject-ID: PRJNA417595 and BioSample-ID: SAMN08000480).

Authors Contributions

R.T. and L.L. designed the study; A.R. provided the sample and associated biological information. L.L. performed molecular lab work, S.H. performed initial *de novo* assembly, M.A. executed all genome annotations and analyses, M.A., S.H., and, A.B.D. analyzed and interpreted the results, M.A. wrote the manuscript. All authors edited and approved the final manuscript.

Tables and Figures

Table 1 Sequencing statistics of libraries used for the two assemblies.

	Insert (bp)	n reads	Coverage
Illumina Library 300	300	794 M	74.1 X
Illumina Library 500	500	473 M	44.2 X
Chicago library	1-50kb	528 M	87.2 X

Table 2 Assembly statistics of the harbour porpoise genome.

	SOAPdenovo assembly	Dovetail HiRise Assembly
Total length	2,669.6 Mb	2,681.2 Mb
Scaffolds		
N50 (number/length)	23,685 / 0.032Mb	34 / 23.8Mb
N90 (number)	159,889	43,146
Longest	304,733	67,078,619
Number	2,139,681	2,025,248

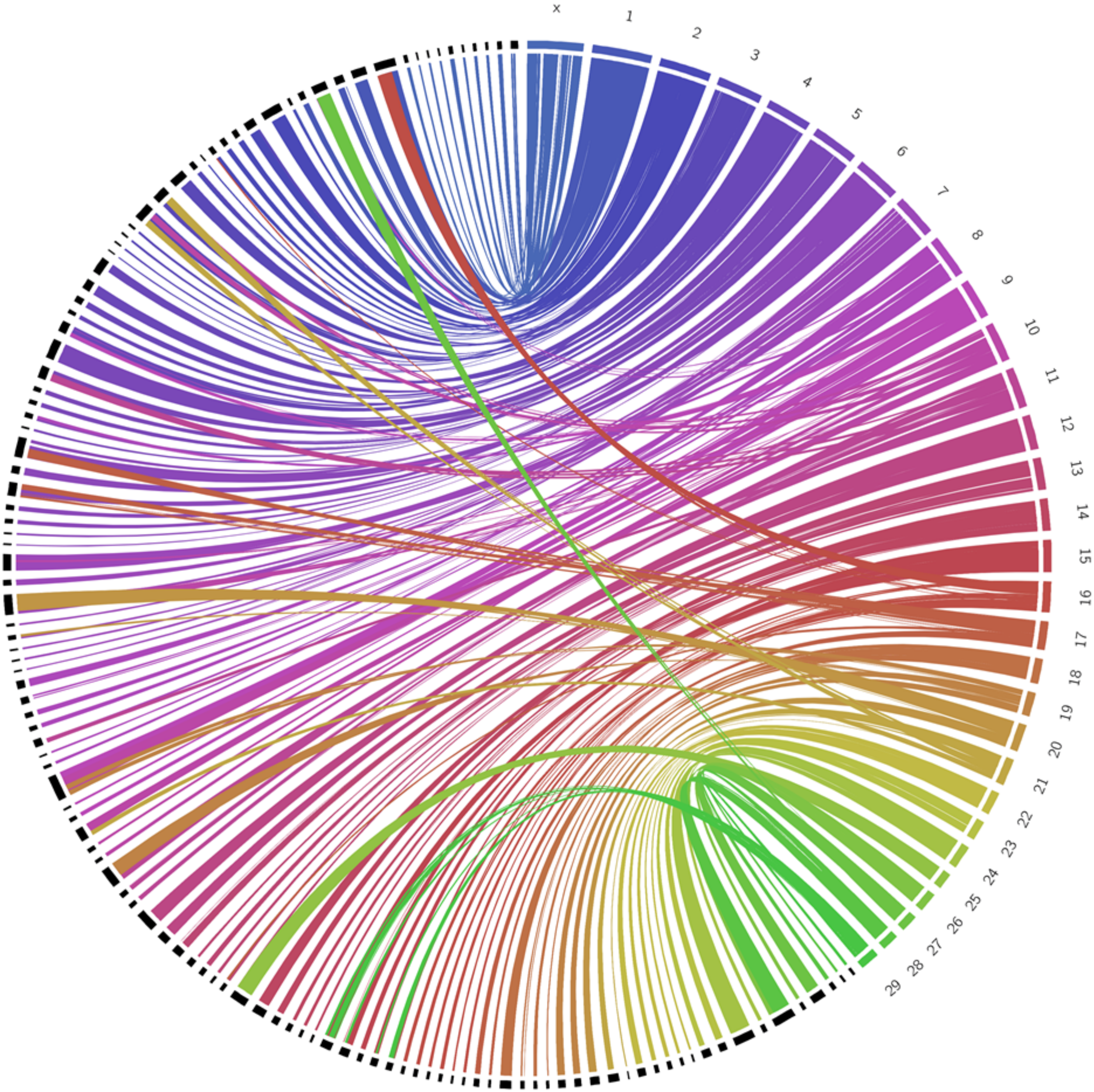
Table 3 Genome annotation statistics

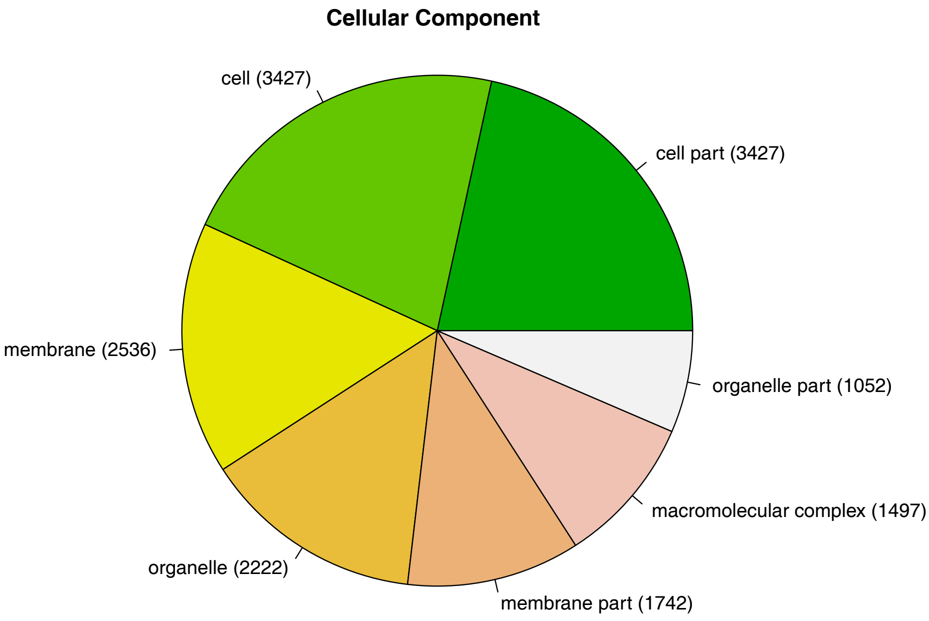
	Exons	Introns	Genes	CDS
Number	171,735	149,581	22,154	22,154
Longest	10,332	1,532,135	2,299,565	20,613
Mean length	166	3,966	28,051	1,282
% genome covered by	-	-	23,2%	1.1%
GC% in CDS	-	-	-	54.48%

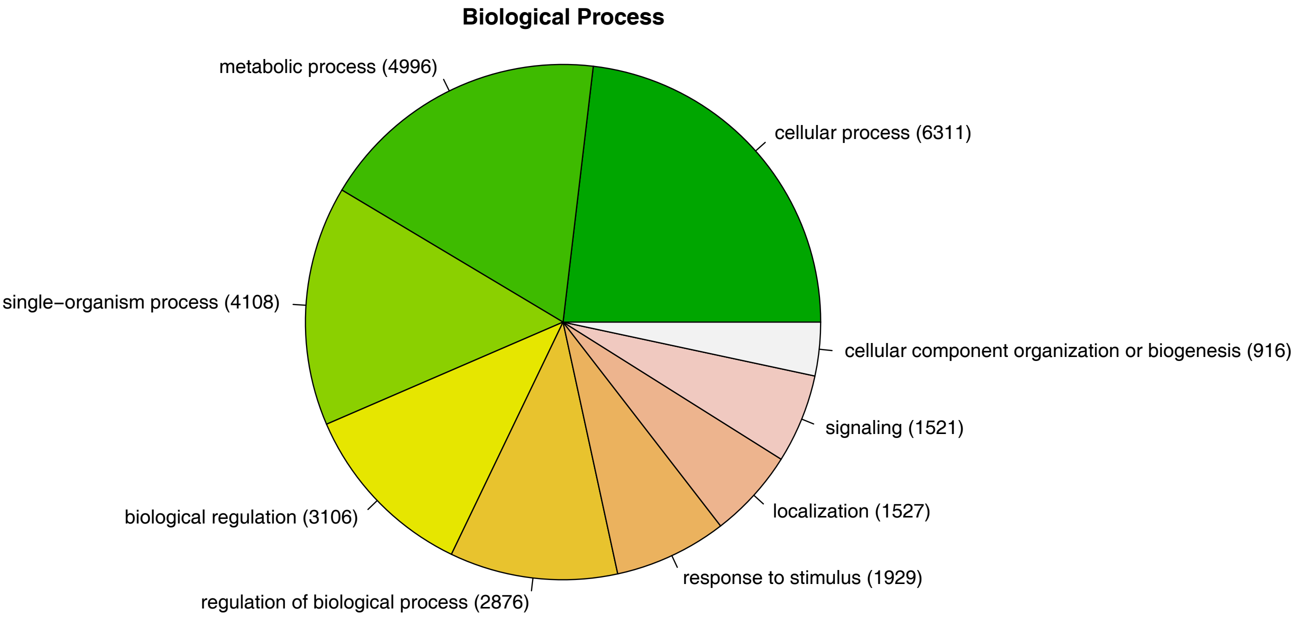
Figure 1 CIRCOS plot of harbour porpoise scaffolds (at least 1Mbp, black on the left outer rim) mapped against the cow (*Bos taurus*) chromosomes (colored bars, labeled X and 1-29). The *B. taurus* autosomal chromosomes (1-29) as well as the X chromosome (x) are shown in different colors. The 122 largest *P. phocoena* scaffolds with consecutive MUMMER hits between a *B. taurus* chromosome of at least 250bp that are no more than 20,000bp apart are shown in black. Matches between *B. taurus* chromosomes and *P. phocoena* scaffolds are shown in the color of the *B. taurus* chromosomes.

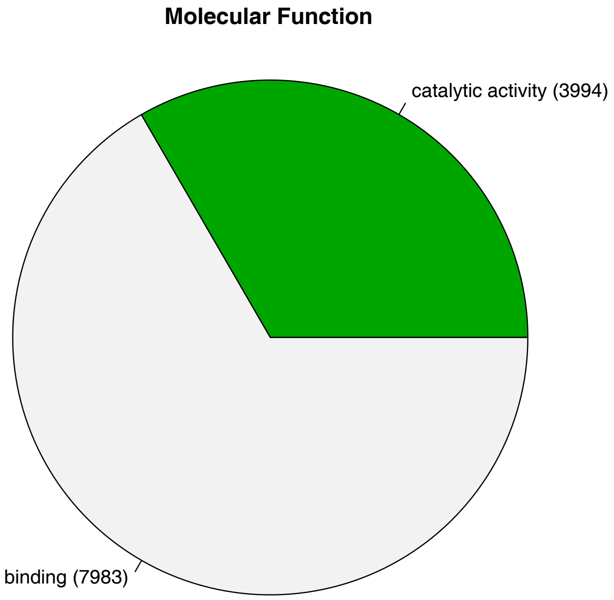
Figure 2 GO-Terms; GO annotation level 2, separated by gene ontology terms: “Cellular Component”, “Biological Process”, and “Molecular Function”. Separate categories are listed, with the number of hits in parentheses.

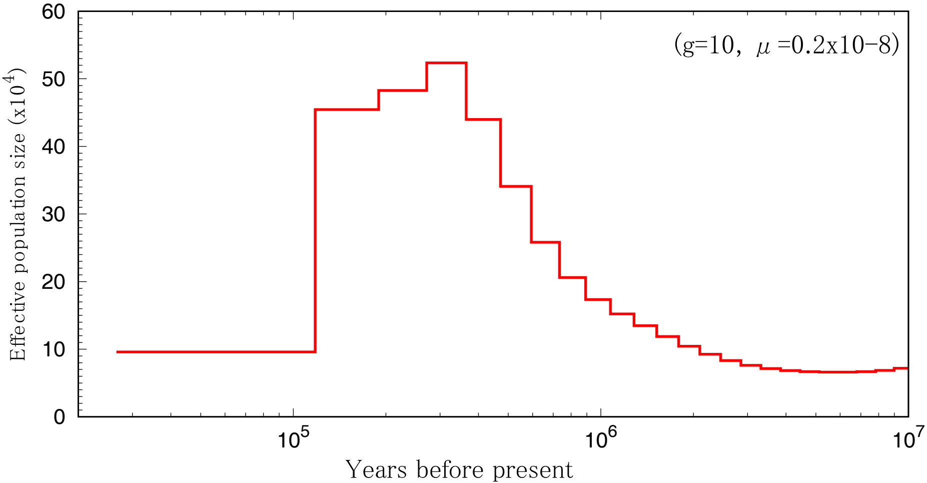
Figure 3 PSMC estimated harbour porpoise population size changes over time for the Baltic Sea. g = generation time; μ = mutation rate (per site, per year). Porpoise data generated on the basis of mapping PE reads to whole genome scaffolds during SNP calling.


x
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29


Cellular Component
cell (3427)
cell part (3427)
membrane (2536)
organelle part (1052)
macromolecular complex (1497)
organelle (2222)
membrane part (1742)


Biological Process
metabolic process (4996)
cellular process (6311)
single−organism process (4108)
cellular component organization or biogenesis (916)
signaling (1521)
localization (1527)
biological regulation (3106)
response to stimulus (1929)
regulation of biological process (2876)


Molecular Function
catalytic activity (3994)
binding (7983)


(g=10, μ =0.2x10−8)
Effective population size (x10^4)
60
50
40
30
20
10
0
10^5
10^6
10^7
Years before present