Resource

A chromosome-level assembly of the Atlantic herring genome—detection of a supergene and other signals of selection

Mats E. Pettersson,[1] Christina M. Rochus,[1,13] Fan Han,[1,13] Junfeng Chen,[1] Jason Hill,[1] Ola Wallerman,[1] Guangyi Fan,[2,3] Xiaoning Hong,[2,4] Qiwu Xu,[2] He Zhang,[2] Shanshan Liu,[2] Xin Liu,[2,5,6] Leanne Haggerty,[7] Toby Hunt,[7] Fergal J. Martin,[7] Paul Flicek,[7] Ignas Bunikis,[8] Arild Folkvord,[9,10] and Leif Andersson[1,11,12]

[1]Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, SE-75123 Uppsala, Sweden; [2]BGI-Qingdao, BGI-Shenzhen, Qingdao 266555, China; [3]State Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese Medical Sciences, University of Macau, Macao 999078, China; [4]BGI Education Center, University of Chinese Academy of Sciences, Shenzhen 518083, China; [5]BGI-Shenzhen, Shenzhen 518083, China; [6]China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China; [7]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom; [8]Science for Life Laboratory Uppsala, Department of Immunology, Genetics and Pathology, Uppsala University, SE-75123 Uppsala, Sweden; [9]Department of Biological Sciences, University of Bergen, 5020 Bergen, Norway; [10]Institute of Marine Research, 5018 Bergen, Norway; [11]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, SE-75007 Uppsala, Sweden; [12]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, Texas 77843, USA

The Atlantic herring is a model species for exploring the genetic basis for ecological adaptation, due to its huge population size and extremely low genetic differentiation at selectively neutral loci. However, such studies have so far been hampered because of a highly fragmented genome assembly. Here, we deliver a chromosome-level genome assembly based on a hybrid approach combining a de novo Pacific Biosciences (PacBio) assembly with Hi-C-supported scaffolding. The assembly comprises 26 autosomes with sizes ranging from 12.4 to 33.1 Mb and a total size, in chromosomes, of 726 Mb, which has been corroborated by a high-resolution linkage map. A comparison between the herring genome assembly with other high-quality assemblies from bony fishes revealed few inter-chromosomal but frequent intra-chromosomal rearrangements. The improved assembly facilitates analysis of previously intractable large-scale structural variation, allowing, for example, the detection of a 7.8-Mb inversion on Chromosome 12 underlying ecological adaptation. This supergene shows strong genetic differentiation between populations. The chromosome-based assembly also markedly improves the interpretation of previously detected signals of selection, allowing us to reveal hundreds of independent loci associated with ecological adaptation.

[Supplemental material is available for this article.]

The Atlantic herring (*Clupea harengus*) is a model system for ecological adaptation and the consequences of natural selection (Martinez Barrio et al. 2016; Lamichhaney et al. 2017; Hill et al. 2019), largely due to the enormous population size minimizing genetic drift. The Atlantic herring is, in fact, one of the most abundant vertebrates on Earth, with schools comprising more than a billion individuals and an estimated global population in excess of 10^{11} fish (Feng et al. 2017). It is also one of few marine species to successfully colonize the Baltic Sea, a brackish body of water formed after the last Ice Age, giving rise to the phenotypically distinct subspecies Baltic herring.

Earlier work provided the first draft version of the herring genome (Martinez Barrio et al. 2016) and revealed regions with strong signals of selection related to both adaptation to the brackish Baltic Sea and differences in spawning time between herring populations (Martinez Barrio et al. 2016; Lamichhaney et al. 2017). In contrast, there is essentially no genetic differentiation at selectively neutral loci even between geographically distant populations, a fact documented already by isozyme and microsatellite analyses (Andersson et al. 1981; Ryman et al. 1984; Larsson et al. 2010; Limborg et al. 2012) and verified by whole-genome sequencing (Lamichhaney et al. 2012, 2017). However, while the signals of selection are strong, the fragmented draft genome made it challenging to determine the number of independent loci under selection and made it difficult to study the impact of large-scale inversions and other structural variations.

Here, by combining a de novo long-read assembly of an Atlantic herring with long-range chromatin interaction (Hi-C)

---

[13]These authors contributed equally to this work.
Corresponding author: leif.andersson@imbim.uu.se

information (Lieberman-Aiden et al. 2009), we remedy this fragmentation and deliver a chromosome-level assembly of the herring genome comprising 26 autosomes with sizes ranging from 12.4 to 33.1 Mb and a total size of 726 Mb. We also show how this new assembly has a major impact on our ability to interpret signals of selection.

Results

The assembly is based on a new, de novo assembly of an Atlantic herring, as opposed to a Baltic herring used for the previously published version 1.2 (Martinez Barrio et al. 2016). The assembly workflow is outlined in Figure 1, and the final version, with the parts that could not be linked to the 26 chromosomes included as unplaced scaffolds, is publicly available via the European Nucleotide Archive (https://www.ebi.ac.uk/ena/data/view/GCA_900700415). Summary statistics of the involved assemblies are in Table 1, while the sizes of the assembled chromosomes are in Table 2. We assume that the 26 superscaffolds correspond to chromosomes and have named these Chr 1–Chr 26 based on size.

A comprehensive linkage map

We used pedigree data from two full-sib families, one Baltic herring and one Atlantic herring family, bred in captivity. The parents and 45 Baltic and 50 Atlantic full-sibs were genotyped for approximately 45,000 markers across the genome using a previously described SNP chip (Martinez Barrio et al. 2016). The markers formed 26 linkage groups in perfect agreement with the genome assembly, and the linkage map confirmed the linear order of genomic segments along the chromosomes. The total length of the sex-average linkage map was 1660 cM, and the average recombination rate was 2.54 cM/Mb (Table 2; sex-specific maps in Supplemental Table S1); the male map was 13% longer than the female map. Since the herring genome is composed of 26 chromosomes, the total map distance of 1660 cM implies that there is slightly more than one recombination event per chromosome pair in each meiosis (26×50 cM = 1300 cM). There was a consistent pattern of nonuniform recombination rate across chromosomes, with the typical case being an L-shaped map where one section, in many cases, >10 Mb, displays little to no recombination in the pedigree (Fig. 2A). In essence, it seems most chromosomes have one hot side and one cold side in terms of recombination rate. However, on the population level, the linkage disequilibrium (LD)-based recombination frequencies across the cold regions are above zero, indicating that there is not a complete repression of recombination

Table 1. Summary statistics for different assemblies of the herring genome

	v1.2[a]	PacBio	v2.02
Scaffolds	6915	–	1723
Scaffold N50 (Mb)	1.84	–	29.85
Scaffold L50	113	–	13
Contigs	73687	1356	–
Contig N50 (kb)	22.3	1615.0	–
Contig L50	8566	119	–
Size (Mb)	808	793	726/786[b]
GC (%)	44.1	44.2	44.1
N (%)	9.98	0.00	0.11

[a]v1.2 is the version published in Martinez-Barrio et al. (2016), Pacific Biosciences (PacBio) is the FALCON-unzip assembly, and v2.0.2 is the final hybrid assembly presented here.
[b]The 26 chromosomes total 726 Mb; unplaced scaffolds constitute 61 Mb.

but rather a moderation of its frequency. The linkage map for Chromosome 8 is shown in Figure 2B, and all maps are in Supplemental Figure S1.

Chromosome-wise recombination profiles

The chromosomal-level assembly provides an opportunity to investigate the variation of recombination rates across the genome using population genetic data. Here, we constructed a recombination rate profile based on patterns of linkage disequilibrium using LDhat (Auton and McVean 2007). We estimated crossover events among ~1.7 million SNPs from 14 Baltic herring that have been individually sequenced (Martinez Barrio et al. 2016; Lamichhaney et al. 2017). A fine-scale recombination map was generated with mean recombination rate of $\rho = 31.3$/kb, corresponding to 2.1 cM/Mb, given nucleotide diversity (π) = 0.3% (Martinez Barrio et al. 2016) and mutation rate (μ) = 2×10^{-9} (Feng et al. 2017), similar to the rate in zebrafish (1.6 cM/Mb) (Bradley et al. 2011). Overall, there was excellent agreement between pedigree-based (linkage) and population-based (LD) estimates of recombination rates along herring chromosomes (Fig. 2B; Supplemental Fig. S1).

Correspondence to karyotype and positioning of the centromeres

The genome assembly is consistent with the observed karyotype for the sister species Pacific herring (*Clupea pallasi*), both in terms of the number of chromosomes and their size distribution. Ida et al. (1991) showed that the diploid genome consisted of 26 pairs ($2n = 52$), where most were of similar size, with two smaller pairs that were speculated to be the result of a recent chromosome fission event. Out of the 26 pairs, three were determined to be metacentric or submetacentric, while the remaining 23 were reported to be acrocentric. The recombination profile observed across the Atlantic herring chromosomes suggests, based on the situation in many other species, that the recombination rate is relatively low toward centromeres and relatively high toward telomeres. Therefore, we propose that Chromosomes 3, 20, and 22 are metacentric, as their profile is shaped like a "U" rather than an "L" (Fig. 2). For the "L"-shaped chromosomes, we expect the centromere to be located at the flat end, not necessarily at the beginning, as the assembly was agnostic to the linkage map.
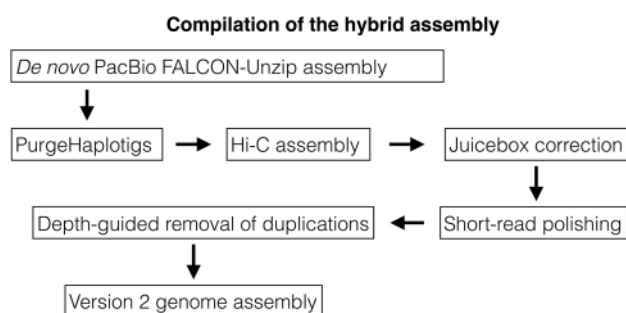
Compilation of the hybrid assembly

Figure 1. Flow chart of the assembly compilation process (see Methods for further details).

Table 2. Physical and genetic sizes of the chromosomes

Name	Size (Mb)	Size (cM)
Chr 1	33.1	46.8
Chr 2	33.0	57.9
Chr 3	32.5	75.7
Chr 4	32.3	62.4
Chr 5	31.6	46.8
Chr 6	31.5	85.5
Chr 7	31.0	50.8
Chr 8	30.7	57.9
Chr 9	30.5	107.1
Chr 10	30.2	60.5
Chr 11	30.1	54.8
Chr 12	30.0	79.9
Chr 13	29.8	94.0
Chr 14	29.3	44.4
Chr 15	28.7	60.4
Chr 16	27.8	95.1
Chr 17	27.6	56.6
Chr 18	27.2	51.3
Chr 19	27.1	52.2
Chr 20	26.7	62.4
Chr 21	26.5	26.8
Chr 22	25.7	60.6
Chr 23	25.3	100.8
Chr 24	20.1	55.3
Chr 25	14.9	56.4
Chr 26	12.4	57.5
Total	725.7	1659.9

(2) spanning at least 100 bases, and (3) separated from the next peak by at least 100 kb, we find 125 independent loci associated with adaptation to the Baltic Sea and 22 loci associated with different spawning times (Fig. 4A,B). Using a lower cut-off of 10^{-15} yields 195 and 47 independent loci, respectively. Thus, the qualitative result is not sensitive to the choice of cut-off value. While both these adaptations have a complex genetic background, there is a four- to fivefold difference in the number of loci reaching statistical significance, which makes sense given that adaptation to the Baltic Sea is likely a more complex process than adaptation to different spawning times; the Baltic Sea differs from the Atlantic Ocean as regards salinity, optic characteristics, depth, higher seasonal variation in temperature, plankton production, predators present, and pollution.

Resolving spawning time-associated variation at the *TSHR* locus

Our previous work (Martinez Barrio et al. 2016; Lamichhaney et al. 2017) showed that the SNPs most strongly associated with variation in spawning time are located immediately adjacent to the thyroid stimulating hormone receptor (TSHR) gene (Fig. 4C). However, the previous gene model was truncated compared to homologs from other species, missing the first three exons. This led to difficulties in interpreting the signal of selection, as it was unclear how the differentiated SNPs were positioned in relation to the *TSHR* start of transcription (TSS). In order to improve the gene model, we performed 5′ and 3′ RACE experiments, which, together with subsequent PCR validation of the entire coding region,
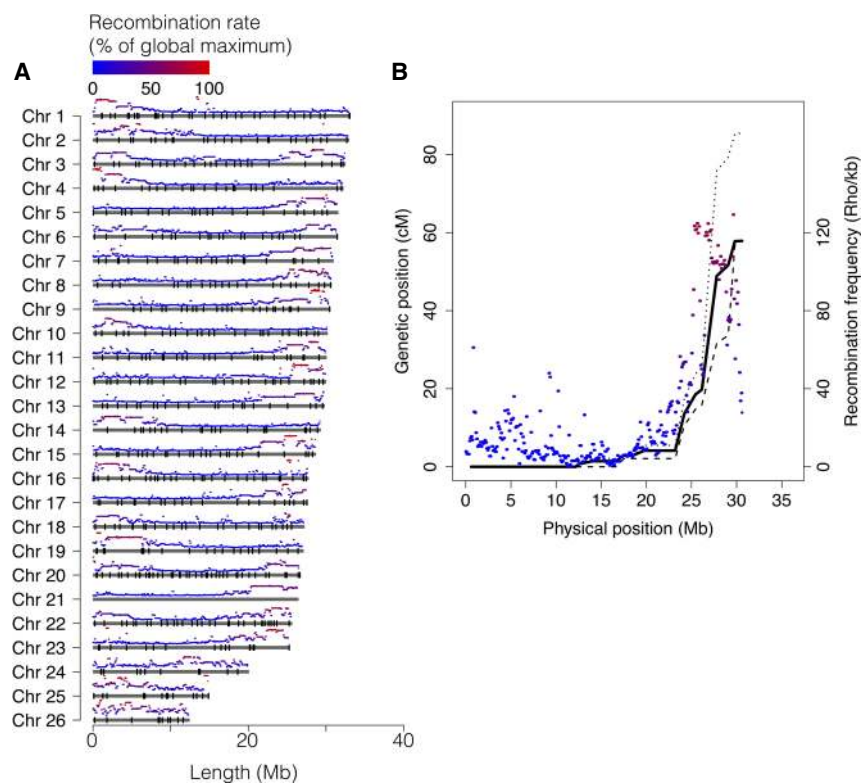
Inter-chromosomal rearrangements are rare, but intra-chromosomal rearrangements are frequent among teleosts

The chromosome-level assembly allows comparisons of the herring genomic organization with that in other species. We performed pair-wise whole-genome alignments between the new herring assembly and five other teleosts with chromosome-level assemblies publicly available. These comparisons revealed a very high degree of conserved synteny among teleosts, as illustrated by the comparison of the Atlantic herring and stickleback genomes (Fig. 3A). However, the linear orders along chromosomes are highly rearranged (Fig. 3B).

Number of independent signals of selection associated with ecological adaptation

Our previous estimates of the number of independent loci associated with ecological adaptation to the brackish Baltic Sea, and to spring versus autumn (Martinez Barrio et al. 2016), were uncertain due to the fragmented nature of the previous assembly. Now, by transferring the data set to the new assembly and defining an independent peak as (1) at least two SNPs with χ^2-test *P*-values $< 10^{-20}$, as before (Martinez Barrio et al. 2016),

Figure 2. Chromosome size distribution and recombination rate. (*A*) Physical extent of the assembly for each chromosome, with average recombination rate, in 100-kb windows, shown on *top* of each chromosome and markers used in the linkage map indicated as black bars. (*B*) Linkage map data (black lines) and recombination-rate profile (colored segments) for Chromosome 8. Solid line: sex-average; dashed line: male; dotted line: female.
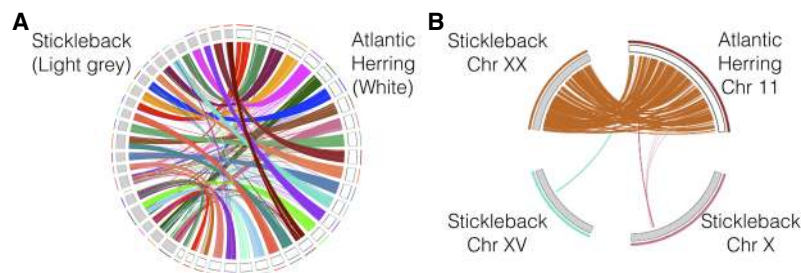
Figure 3. Conserved synteny between Atlantic herring and stickleback. (*A*) Whole-genome alignment between the Atlantic herring and threespine stickleback assemblies. (*B*) Detailed view of sequence homologies between herring Chromosome 11 and stickleback Chromosome XX, indicating many intra-chromosomal rearrangements.

these to genotype unrelated individuals by similarity to the two reference haplotypes; these were denoted Northern (N) and Southern (S) based on their geographic distribution (see below). Finally, by analyzing LD patterns in the groups of individuals determined to be homozygous for the two haplotypes separately, we noted a lack of strong LD across the entire region within both groups (Fig. 5B, middle and bottom). This indicates that recombination occurs between chromosomes of the same class but is
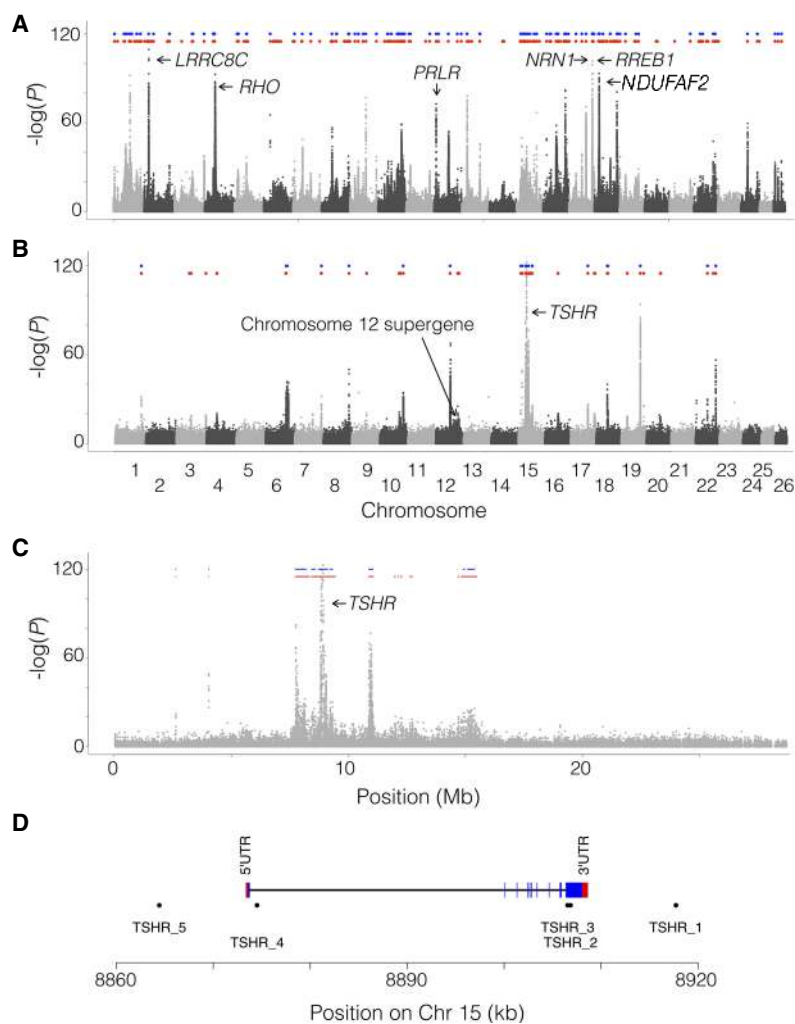
allowed us to generate a complete gene model that shows that the two most differentiated SNPs are located in a 10-kb region around the TSS of *TSHR* (Fig. 4D).

Identification of a supergene on Chromosome 12

The LD pattern across the herring genome was analyzed in more than 1170 individuals, genotyped for approximately 45,000 SNPs, collected from around the Swedish coast by high school students (project Forskarhjälpen). This analysis revealed an extensive LD block, from 17.8 Mb to 25.6 Mb on Chromosome 12, that was divided into four different scaffolds in the previous assembly. This region also stood out in screens for ecological adaptation, since the genetic differentiation was more or less equally strong across the block (Fig. 5A), lacking the typical pyramidal peak shape. Inside the block there was no correlation between the strength of LD and physical distance (Fig. 5B, top). This indicated the presence of a possible supergene, either in the form of an inversion or a block of otherwise repressed recombination. However, the pattern contained inconsistencies; specifically, there were moderately linked markers interspersed with virtually perfectly linked ones across the full extent of the block.

Led by the LD patterns revealed in the population data, detailed examination of the pedigree data used for linkage mapping revealed an elevated frequency of heterozygous SNPs in one out of four parents across the putative inversion, a pattern that was repeated in 28 out of 50 offspring in that family, consistent with proper Mendelian inheritance. Assuming that the high-heterozygosity parent and offspring were heterozygous for the proposed inversion allowed us to deduce supergene haplotypes and use

Figure 4. Signals of selections associated with ecological adaptation. The panels show the association, measured as –log(*P*) from χ^2 tests on read counts of previously published data (Martinez Barrio et al. 2016) replotted along the new assembly. (*A*) Genetic differentiation between seven populations of Atlantic herring from the Atlantic Ocean, North Sea, Skagerrak, and Kattegat (salinity in the range 20–35 psu) and 10 populations from the Baltic Sea (salinity in the range 3–12 psu). (*B*) Genetic differentiation between 10 populations of spring-spawning herring versus three populations of autumn-spawning herring. In both panels, the blue and red dots indicate identified, independent regions of selection at a *P*-value cut-off of either 10^{-20} (blue) or 10^{-15} (red). (*C*) Zoom in on Chromosome 15 for the contrast based on differences in spawning time, which contains the most significant peak, located around *TSHR*. (*D*) Improved gene model of *TSHR*. TSHR_1 to 5 are selected SNPs showing highly significant association ($P < 10^{-95}$) to spawning time and/or being nonsynonymous coding, while covering the extent of the *THSR* gene model.
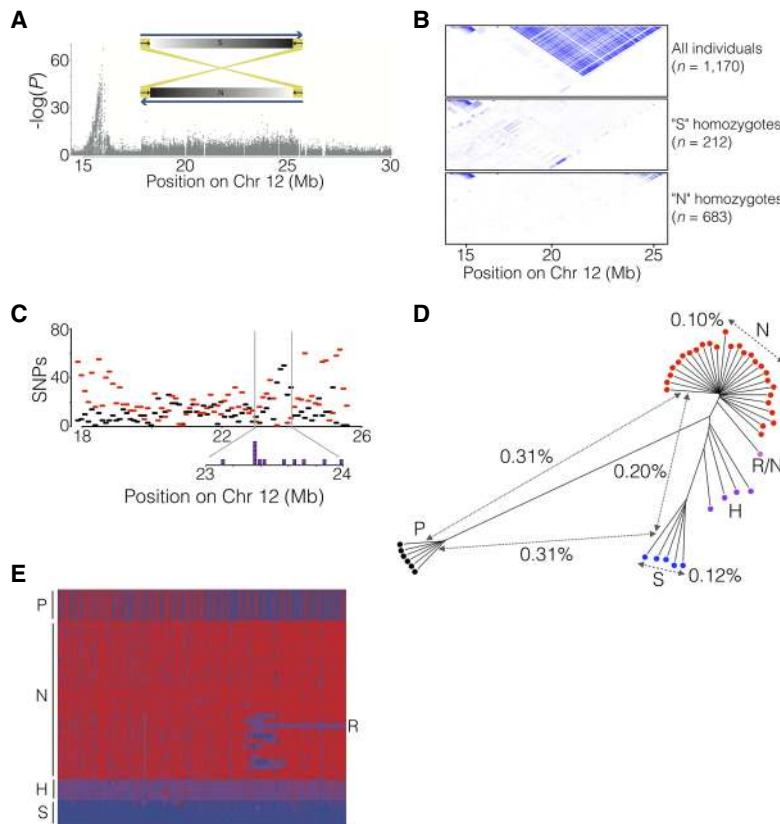
Figure 5. Identification of a 7.8-Mb inversion on herring Chromosome 12. (*A*) The spawning time contrast for Chromosome 12 highlights the block-like association pattern for the region from 17.9 to 25.6 Mb. The sketch shows the location of inverted repeats flanking the supergene. (*B*) LD patterns across the region in different groups of individuals sorted according to genotype for the putative inversion. (*C*) The distribution, as number of SNPs per 100 kb, of shared (black) and diagnostic (red) SNPs across the inversion region. Purple boxes (*inset*) are estimated locations of breakpoints in individuals that appear to carry a recombinant chromosome (see text). (*D*) Neighbor-joining tree based on genotypes for all SNPs in the inversion region called from individual whole-genome sequencing. The distances indicated across the tree are average nucleotide differences between haplotypes, either within groups (dashed) or between groups (solid). The letters designate the supergene type of the individual as follows: N = Northern homozygote; S = Southern homozygote; H = N/S heterozygote; R/N = individual carrying an N haplotype and a recombinant haplotype (see text); P = Pacific herring. (*E*) Heat map of the genotypes for diagnostic SNPs based on individual whole-genome sequencing. Supergene type of the samples is indicated as in *D*.

mismatched reads at the other breakpoint (around 25.6 Mb) due to the high repeat content. These observations support the hypothesis that the extremely strong LD in this region is caused by an inversion and that the inverted repeats have facilitated its occurrence.

Examining individual SNP allele frequencies in each group of locus-wide homozygotes, we were able to classify a fraction of the SNPs within the interval as shared, defined as having allele frequencies in the range 10%–90% in both haplotype-groups. This is not expected of a canonical inversion with a single origination event and complete suppression of recombination. Thus, some amount of genetic exchange must be ongoing. In an attempt to quantify the amount of genetic exchange across the supergene, we tallied both diagnostic, defined as having allele frequency >90% in one group and <10% in the other, and shared SNPs in 100-kb windows (Fig 5C). Diagnostic (red) SNPs are enriched at each end of the interval, with the left-hand side having stronger enrichment. The shared SNPs (black) have a peak, matched by a corresponding lack of diagnostic SNPs, at around position 23.5 Mb. A similar pattern is seen for the absolute delta allele frequencies for all SNPs (Supplemental Fig. S3A).

It seems likely that this pattern is linked to a rare class of individuals (12 out of 1170) that carry a third haplotype where the segment leading up 23.5 Mb follows the "Southern" haplotype, while the block beyond that is of the "Northern" type. The estimated switching-points of these 12 individuals are shown as purple blocks in the inset of Figure 5C. We can also detect a lack of shared

SNPs close to both edges of the supergene, in particular, the right-hand one (Supplemental Fig. S3B), a pattern that is consistent with genetic exchange between inversion haplotypes due to twisting of the chromosomes.

We constructed a genetic distance tree for the supergene region based on individual whole-genome sequence data (Fig. 5D); the color of each leaf in the tree corresponds to the supergene type. This shows the expected clustering of homozygotes for the two supergene types (N or S), while the heterozygotes (H) are positioned in between. Notably, there is one individual that carries one copy of the partial inversion haplotype (R) discussed above, with an estimated breakpoint in the same region found in individuals genotyped using the SNP chip. The tree also reveals that the inversion must have occurred after the divergence from the Pacific herring, because the two alleles are equidistant from alleles found in Pacific herring (Fig. 5D). The nucleotide diversity inside each haplotype group is lower than the genomic average of 0.3%, which is consistent with both lower effective population size, due to restricted recombination, for this region compared to the rest of

strongly repressed in heterozygotes, the expected pattern for an inversion but not if recombination is suppressed in the region for other reasons.

In an attempt to identify inversion break points, we examined reads from whole-genome sequence data from individuals with different genotypes and found inverted repeat patterns at both ends of the putative inversion block (Fig. 5A; Supplemental Fig. S2). For the putative breakpoint at 17.9 Mb, we found short-read mismatch patterns at the edge of this repeat (position 17,826,318 bp) that correlated perfectly with the SNP-based supergene genotype (Supplemental Fig. S2). No short-read mismatches (e.g., soft-clipped reads) were identified in individuals classified as SS homozygotes, consistent with the fact that the reference assembly contains the S haplotype. In contrast, 50% and 100% of the short reads from NS heterozygotes and NN homozygotes, respectively, were mismatched at this position. Thus, we consider this as a putative inversion breakpoint, but at present we cannot exclude the possibility that this is a structural variant in complete LD with the true breakpoint. We were not able to identify similar

the genome, as well as a bottleneck when the inversion was formed. Supplemental Figure S3, C and D, shows the allele frequency distributions among 11,965 typed SNPs from the inversion region in "S" and "N" homozygotes. The higher number of SNPs close to fixation (MAF < 5%) in the "N" group indicates that it is the derived version, which could be correlated to northward expansion of the Atlantic herring in response to receding glaciation.

A heat map based on diagnostic SNPs, deduced from individual whole-genome sequence data, supports the notion that the two haplotype groups evolved subsequent to the split between Atlantic and Pacific herring and illustrates the extreme LD across the region (Fig. 5E). This analysis provides further evidence for ongoing recombination between haplotypes in the interval from 23.1 to 24.0 Mb. Additionally, the heat map makes it clear that the individual labeled "R/N" in Figure 5D carries one copy of the recombinant haplotype described above.

The supergene on Chromosome 12 underlies ecological adaptation

Across the supergene region, allele frequencies differ substantially for virtually all SNPs (Fig. 6A), allowing estimation of supergene allele frequency even in pooled sequencing data. Using the SNPs found to be essentially fixed for different alleles in the Northern and Southern haplotype groups, i.e., those SNPs found in the lower-right corner of Figure 6A, we estimated haplotype frequencies in pooled samples based on the average allele frequency at diagnostic SNPs.

The estimated haplotype frequencies in the pooled data, which covers a wide range of herring populations, revealed a highly significant genetic differentiation among populations (Fig. 6B, C). There was a consistent trend in West Atlantic, East Atlantic, and in the Baltic Sea in that the populations spawning most northerly had a high frequency of the Northern haplotype while the Southern haplotype dominated in populations spawning more southerly, with the exception being a few populations in the Southern Baltic Sea. The most extreme population, almost fixed for the Southern haplotype, was the one representing autumn-spawning North Sea herring (NS in Fig. 6B). This strong genetic differentiation is never observed at selectively neutral loci among the populations included in this analysis (Martinez Barrio et al. 2016; Lamichhaney et al. 2017). Thus, this supergene polymorphism must be under selection, possibly related to temperature at spawning, which is known to be a major stressor for the southernmost and high temperature-exposed herring populations (Peck et al. 2012; Ojaveer et al. 2015).

Discussion

Integrity of the assembly

The overall organization of the assembly presented here is mainly supported by two separate observations: the one-to-one correlation between putative chromosomes and independently determined linkage groups, and the discrete blocks detected in the Hi-C contact map. Based on these two data sets, we are confident that the 26 superscaffolds present in the assembly match the 26 physical chromosomes identified by karyotyping of the sister taxon Pacific herring (Ida et al. 1991). Furthermore, the high quality of the assembly is strongly supported by the very high degree of conserved synteny between our Atlantic herring genome and those of other teleosts.
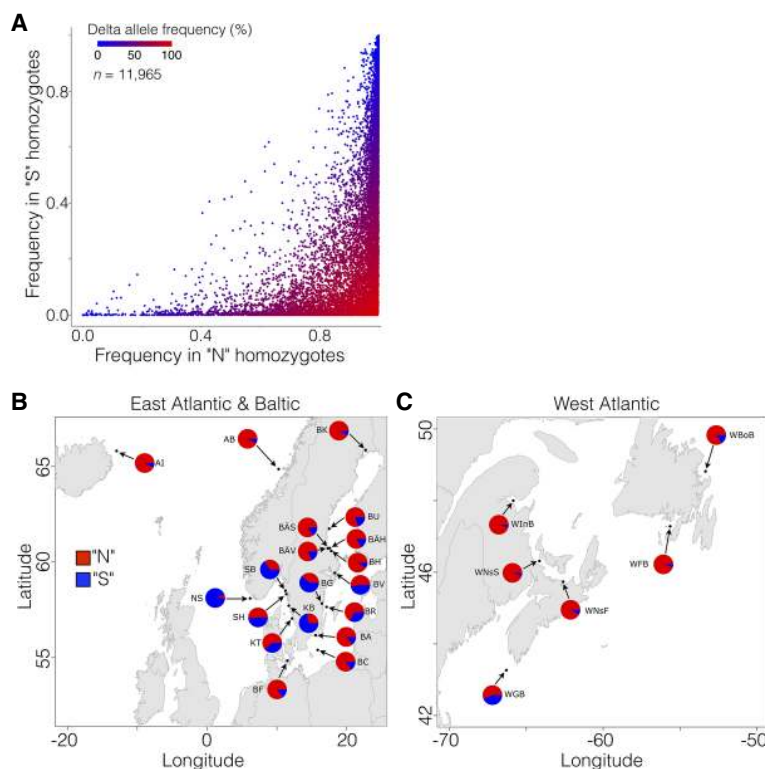
Inter-chromosomal rearrangements are rare, while intra-chromosomal rearrangements are abundant in teleosts

This study revealed a contrast between conserved synteny, with very few inter-chromosomal rearrangements, between Atlantic herring and even distantly related teleosts that have been separated from herring for hundreds of millions of years, and the frequent occurrence of intra-chromosomal rearrangements, a finding consistent with previous studies (Amores et al. 2014; Rondeau et al. 2014). There is a difference among vertebrate groups wherein fishes and birds usually show few inter-chromosomal rearrangements but frequent intra-chromosomal rearrangements. In contrast, there is an opposite trend among

Figure 6. Genetic differentiation in the region encompassing the putative inversion on Chromosome 12. (A) Allele frequencies for all SNPs (n = 11,965) inside the inversion among individuals homozygous for the Southern (S) (y-axis) or Northern (N) haplotypes (x-axis). All frequencies are expressed in terms of the allele that is more common in the "N-"context than in the "S-"context. (B,C) Estimated frequencies for the Northern and Southern haplotypes in pooled population samples in the Baltic Sea and East Atlantic (B) and in the West Atlantic (C). The location and date of capture of the pooled samples are listed in Lamichhaney et al. (2017).

mammals, where often fairly closely related species like mouse and rat show many inter-chromosomal rearrangements (Coghlan et al. 2005).

Large-scale inversions and their evolutionary significance

In our previous studies (Martinez Barrio et al. 2016; Lamichhaney et al. 2017), using the scaffold-level assembly available at the time, it was indicated that small inversions were not a major contributor to differences between herring populations while the issue of larger, megabase-scale inversions was intractable given the scaffold length distribution. Using the improved assembly presented here, we now have clear evidence of selection acting on a supergene that is, in fact, a 7.8-Mb inversion. We were even able to identify a putative inversion breakpoint at position 17,826,318 bp on Chromosome 12. However, our data also illustrate how difficult it is to exactly define inversion breakpoints because they are often embedded in repeat regions.

Our observations fit a supergene model, where large essentially nonrecombining haplotypes allow accumulation of multiple causal variants synergistically affecting fitness. The divergence time between the two haplotype groups (Fig. 5D) cannot be accurately determined due to the ongoing genetic exchange between haplotype groups. However, the finding that the Pacific herring carries a separate haplotype limits the maximum age of the inversion to after the split of the two species, and the observed nucleotide divergence (0.2%) is lower than the genomic average (about 0.3%), which indicates that the putative inversion may be fairly recent. There is apparently ongoing genetic exchange between the inversion alleles as illustrated by the presence of many shared polymorphisms as well as recombinant haplotypes (Fig. 5E). It is likely that both double recombination and gene conversion contributes to this genetic exchange. It is, in fact, a characteristic feature of inversions that some recombination occurs between alleles, although it is severely suppressed. This feature is well documented for the inversion underlying the Rosecomb phenotype in domestic chicken (Imsland et al. 2012) and the inversion associated with variant mating strategies in the ruff (Küpper et al. 2016; Lamichhaney et al. 2016). In the herring, it is possible that the flanking inverted repeats have promoted recurrent inversion events, which would facilitate genetic exchange between haplotype groups.

The supergene contains 225 genes, with an additional approximately 10 genes located in flanking positions on the outside of the estimated breakpoint positions. Thus, it is difficult to determine which genes and/or variants contribute to the fitness effects of the inversion. However, based on the disruption of the local context, genes near the breakpoints are more likely to show altered gene regulation and are therefore listed in Supplemental Table S2. While it is currently not known what causes the fitness differences between the inversion variants in herring, it appears highly likely that it is related to ecological adaptation in relation to the water temperature during gonadal maturation before spawning or the water temperature at spawning/early larval development. There is a clear clinal variation of an increasing frequency of the Southern variant from north to south both in the East and West Atlantic (Fig. 6B,C).

The supergene on herring Chromosome 12 adds to the growing list of supergenes associated with morphological variation and ecological adaptation (Schwander et al. 2014). The first very early examples of supergenes under balancing selection were those detected by cytogenetic studies of polytene chromosomes in

Drosophila (Dobzhansky and Sturtevant 1938). More recent examples include supergenes controlling mimicry in butterflies (Zhang et al. 2017), social behavior in fire ants (Wang et al. 2013), plumage variation and mating preferences in white-throated sparrows (Tuttle et al. 2016), and alternative male mating strategies in ruff (Küpper et al. 2016; Lamichhaney et al. 2016). Furthermore, five putative inversions ranging in size from 3.5 to 18.5 Mb have recently been associated with migratory behavior and geographical distribution in the Atlantic cod (*Gadus morhua*) (Berg et al. 2017). The present study shows how chromosome-based assemblies will facilitate the identification of many other similar examples of supergenes.

Methods

FALCON de novo assembly

Genomic DNA was fragmented to 20 kb using a DNA shearing device (Hydroshear, Digilab), and the sheared fragments were size-selected for the 7- to 50-kb size range using Blue Pippin (Sage Science). The sequencing library was prepared following the standard SMRTbell construction protocol (PacBio). The library was sequenced on 100 PacBio RSII SMRT cells using the P6-C4 chemistry. Raw data were imported into SMRT Analysis software 2.3.0 (PacBio). Subreads shorter than 500 bp or with a quality (QV) <80 were filtered out. The final data set contained 63.1 Gb of filtered subreads with N50 of 15.6 kb and was used for de novo assembly with FALCON (pb-falcon 0.2.4) (Chin et al. 2016). To further improve the assembly, we ran FALCON Unzip (pb-falcon 0.2.4) (Chin et al. 2016) followed by consensus calling using the Arrow algorithm. In order to remove highly heterozygous haplotypes assembled as separate primary contigs, we ran the Purge Haplotigs pipeline (v1.0.4) (Roach et al. 2018), which identifies and reassigns allelic contigs. The configuration file used for assembly constitutes Supplemental Text S1.

Hi-C library construction and sequencing

In situ Hi-C was conducted following the protocol provided by Rao et al. (2014) with minor modifications. The restriction endonuclease MboI was used to digest DNA, followed by biotinylated residue labeling. The Hi-C library was then sequenced on a BGISEQ-500 platform with pair-end sequencing using a read length of 50 bp. The raw number of reads was 656,695,125, out of which 98,838,909 provided useful Hi-C contact information.

Compiling the hybrid assembly

The FALCON-unzip assembly was processed through the Purge Haplotigs pipeline (Roach et al. 2018) in order to remove redundant sequences from the primary assembly. This procedure resulted in a de novo assembly with a total size of 792.6 Mb, a contig N50 of 1.61 Mb, comparable to the scaffold N50 (1.84 Mb) of the published v1.2 genome. Thus, the PacBio assembly achieves a similar level of organization while eliminating a substantial degree of uncertainty, as v1.2 contains close to 10% undetermined bases (Ns) as compared to zero Ns in the FALCON-unzip assembly.

Mapping with Juicer v1.5.6 (Durand et al. 2016b) yielded 99 million informative Hi-C read-pairs, which were used to scaffold the PacBio de novo assembly into chromosome-level organization using the 3D-DNA workflow pipeline (Dudchenko et al. 2017), followed by manual correction using Juicebox v1.9.8 (Durand et al. 2016a). The output assembly was polished using Pilon v1.22 (Walker et al. 2014), based on 50× Illumina paired-end coverage from the same individual. Finally, a custom R script

(Supplemental Code S1; R Core Team 2015) was applied to eliminate a set of small, nearly identical repeats that were deemed likely to be redundant haplotypes based on analysis of the mapped read depth of a set of Illumina short reads from a previously sequenced herring population (Martinez Barrio et al. 2016). This procedure eliminated in total 6.9 Mb (removed fragments are available as Supplemental Data S1).

Annotation

The herring gene set was generated via the Ensembl Gene Annotation (Aken et al. 2016) and has been made available as part of Ensembl release 98 (expected October 2019). A detailed description of the annotation is available as Supplemental Text S2.

SNP chip genotyping

We previously designed a 60k Affymetrix SNP chip (Martinez Barrio et al. 2016). For this study, this SNP chip has been used to genotype two data sets: (1) a pedigree comprising two families with two parents and approximately 50 offspring each (Feng et al. 2017); and (2) 1170 individuals collected in the school project "Forskarhjälpen" in which students from 20 junior high schools from across Sweden contributed to research by collecting a sample of approximately 50 herring from one locality per school (Supplemental Table S3).

Construction of a high-resolution linkage map

Fifty full-sib progeny from the Atlantic family and 45 from the Baltic family were genotyped for about 45,000 SNPs using our previously described SNP-chip (Martinez Barrio et al. 2016). Linkage groups spanning the herring genome were constructed using these data and the Lep-MAP 2 (Rastas et al. 2013) software. The raw linkage groups were in overall concordance with the Hi-C assemblies, allowing each chromosome to be conclusively associated with a distinct linkage group. Based on this association, we were able to prune the marker set based on physical position, with the intent of controlling artificial map expansion while marinating coverage across the entire chromosome. The final, ordered linkage maps of each chromosome were calculated using CRI-MAP v2.5 (Green et al. 1990), and the locations of the markers used are shown in Figure 2, with detailed versions found in Supplemental Figure S1 and sex-specific maps found in Supplemental Table S1.

Linkage disequilibrium analysis and chromosome-wise recombination profiles

LD between markers was measured as correlation between genotypes, calculated using the "r2fast" method from the R-package GenABEL (Aulchenko et al. 2007). To estimate the recombination rates from population data, we applied the LDhat v2.2 package (Auton and McVean 2007) on genetic markers from 14 Baltic individuals phased using Beagle 4.0 (Browning and Browning 2007). The expected crossover events (ρ) between each pair of neighboring markers were calculated with the interval program for 1,000,000 iterations of the rjMCMC procedure with sampling every 2000 iterations. The first 50 iterations were discarded as burn-in. The block penalty 5 was determined after comparing output from simulations with block penalties of 5, 20, 50, and 100. The population recombination map was summarized by summing up the ρ from every 100-kb window, and only windows containing 50–2000 variable sites were included in the final map.

Identification of conserved synteny across teleosts

Whole-genome alignments were performed using Satsuma Chromosemble (Grabherr et al. 2010). The following genome versions were used: northern pike (*Esox lucius*): GCA_000721915.3 (Rondeau et al. 2014); threespine stickleback (*Gasterosteus aculeatus*): Gac-HiC_revised_genome_assembly (Peichel et al. 2017); guppy (*Poecilia reticulata*): GCA_000633615.2 (Künstner et al. 2016); zebrafish (*Danio rerio*): GCA_000002035.4; medaka (*Oryzias latipes*): GCA_002234695.1 (Ichikawa et al. 2017). The zebrafish genome is from the Genome Reference Consortium (Church et al. 2011).

Improvement of the *TSHR* gene model

Total RNA was extracted from the eye of a spring-spawning Atlantic herring using the RNeasy Mini kit (QIAGEN). Six micrograms of the isolated RNA was used for 5′ and 3′ RACE reactions with a FirstChoice RLM-RACE kit (Thermo Fisher Scientific). Nested RACE PCRs were performed in a 25-μL reaction containing 5× KAPA2G Buffer B, 0.24 mM dNTPs, 0.5 μM each of the forward and reverse primer, 1 U KAPA2G Robust DNA Polymerase (Kapa Biosystems), and 1 μL of the cDNA or Outer RACE PCR product as PCR template. Amplification was carried out with the following program: 95°C for 3 min, 35 cycles of 95°C for 15 sec, 58°C for 30 sec and 72°C for 30 sec, and a final extension of 5 min at 72°C. In order to confirm the obtained 5′ and 3′ cDNA ends from RACE reactions, cDNA was prepared using Oligo $(dT)_{18}$ primer with a High-Capacity cDNA Reverse Transcription kit (Thermo Fisher Scientific). Then, nested PCR primers were designed in the 5′ and 3′ UTRs to amplify the whole coding region of *TSHR*. Targeted PCR products were purified from 1% agarose gel using a QIAquick Gel Extraction kit (QIAGEN) and Sanger-sequenced (Eurofins Genomics) with five primers to span the entire PCR product. All primers used for RACE reactions and full-length transcript validation are listed in Supplemental Table S4. In line with previous work on the herring genome, we follow the human gene nomenclature (https://www.genenames.org) in this paper such as *TSHR*.

Characterization of inversion breakpoints

In order to identify the breakpoints of the inversion, we used the BreakDancer software (Chen et al. 2009) on 46 individual sequenced samples independently. Reads with mapping quality above 30 were retained in the analysis. BreakDancer helped to narrow down the range of the potential breakpoints into 5 kb around the ends of the inversion. In an attempt to find the breakpoints at single-base level, we extracted soft-clipped reads and compared the normalized depths between samples carrying the Southern and Northern haplotypes around such reads. The visualization of short reads and clipped reads was performed in Integrative Genomics Viewer (IGV) (Robinson et al. 2017).

Data access

The assembly and RNA reads for annotations generated in this study have been submitted to the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena) under accession number PRJEB31270. SNP-chip genotypes and associated auxiliary information constitute Supplemental Data S2, while Supplemental Code S1 contains the custom R scripts used.

Acknowledgments

References

Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S, Fernandez Banet J, Billis K, García Girón C, Hourlier T, et al. 2016. The Ensembl gene annotation system. *Database (Oxford)* **2016:** baw093. doi:10.1093/database/baw093

Amores A, Catchen J, Nanda I, Warren W, Walter R, Schartl M, Postlethwait JH. 2014. A RAD-tag genetic map for the platyfish (*Xiphophorus maculatus*) reveals mechanisms of karyotype evolution among teleost fish. *Genetics* **197:** 625–641. doi:10.1534/genetics.114.164293

Andersson L, Ryman N, Rosenberg R, Ståhl G. 1981. Genetic variability in Atlantic herring (*Clupea harengus harengus*): description of protein loci and population data. *Hereditas* **95:** 69–78. doi:10.1111/j.1601-5223.1981.tb01330.x

Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. 2007. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23:** 1294–1296. doi:10.1093/bioinformatics/btm108

Auton A, McVean G. 2007. Recombination rate estimation in the presence of hotspots. *Genome Res* **17:** 1219–1227. doi:10.1101/gr.6386707

Berg PR, Star B, Pampoulie C, Bradbury IR, Bentzen P, Hutchings JA, Jentoft S, Jakobsen KS. 2017. Trans-oceanic genomic divergence of Atlantic cod ecotypes is associated with large inversions. *Heredity (Edinb)* **119:** 418–428. doi:10.1038/hdy.2017.54

Bradley KM, Breyer JP, Melville DB, Broman KW, Knapik EW, Smith JR. 2011. An SNP-based linkage map for zebrafish reveals sex determination loci. *G3 (Bethesda)* **1:** 3–9. doi:10.1534/g3.111.000190

Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* **81:** 1084–1097. doi:10.1086/521987

Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, McGrath SD, Wendl MC, Zhang Q, Locke DP, et al. 2009. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods* **6:** 677–681. doi:10.1038/nmeth.1363

Chin CS, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, et al. 2016. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods* **13:** 1050–1054. doi:10.1038/nmeth.4035

Church DM, Schneider VA, Graves T, Auger K, Cunningham F, Bouk N, Chen HC, Agarwala R, McLaren WM, Ritchie GR, et al. 2011. Modernizing reference genome assemblies. *PLoS Biol* **9:** e1001091. doi:10.1371/journal.pbio.1001091

Coghlan A, Eichler EE, Oliver SG, Paterson AH, Stein L. 2005. Chromosome evolution in eukaryotes: a multi-kingdom perspective. *Trends Genet* **21:** 673–682. doi:10.1016/j.tig.2005.09.009

Dobzhansky T, Sturtevant AH. 1938. Inversions in the chromosomes of *Drosophila pseudoobscura*. *Genetics* **23:** 28–64.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356:** 92–95. doi:10.1126/science.aal3327

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. 2016a. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst* **3:** 99–101. doi:10.1016/j.cels.2015.07.012

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016b. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst* **3:** 95–98. doi:10.1016/j.cels.2016.07.002

Feng CG, Pettersson M, Lamichhaney S, Rubin CJ, Rafati N, Casini M, Folkvord A, Andersson L. 2017. Moderate nucleotide diversity in the Atlantic herring is associated with a low mutation rate. *eLife* **6:** e23907. doi:10.7554/eLife.23907

Grabherr MG, Russell P, Meyer M, Mauceli E, Alfoldi J, Di Palma F, Lindblad-Toh K. 2010. Genome-wide synteny through highly sensitive sequence alignment: *Satsuma*. *Bioinformatics* **26:** 1145–1151. doi:10.1093/bioinformatics/btq102

Green P, Falls K, Crooks S. 1990. *Documentation for CRI-MAP, version 2.4*. Washington University School of Medicine, St. Louis.

Hill J, Enbody ED, Pettersson ME, Sprehn CG, Bekkevold D, Folkvord A, Laikre L, Kleinau G, Scheerer P, Andersson L. 2019. Recurrent convergent evolution at amino acid residue 261 in fish rhodopsin. *Proc Natl Acad Sci* **116:** 18473–18478. doi:10.1073/pnas.1908332116

Ichikawa K, Tomioka S, Suzuki Y, Nakamura R, Doi K, Yoshimura J, Kumagai M, Inoue Y, Uchida Y, Irie N, et al. 2017. Centromere evolution and CpG methylation during vertebrate speciation. *Nat Commun* **8:** 1833. doi:10.1038/s41467-017-01982-7

Ida H, Oka N, Hayashigaki K. 1991. Karyotypes and cellular DNA contents of three species of the subfamily *Clupeinae*. *Jpn J Ichthyol* **38:** 289–294. doi:10.1007/BF02905574

Imsland F, Feng C, Boije H, Bed'hom B, Fillon V, Dorshorst B, Rubin CJ, Liu R, Gao Y, Gu X, et al. 2012. The *Rose-comb* mutation in chickens constitutes a structural rearrangement causing both altered comb morphology and defective sperm motility. *PLoS Genet* **8:** e1002775. doi:10.1371/journal.pgen.1002775

Künstner A, Hoffmann M, Fraser BA, Kottler VA, Sharma E, Weigel D, Dreyer C. 2016. The genome of the Trinidadian guppy, *Poecilia reticulata*, and variation in the Guanapo population. *PLoS One* **11:** e0169087. doi:10.1371/journal.pone.0169087

Küpper C, Stocks M, Risse JE, Dos Remedios N, Farrell LL, McRae SB, Morgan TC, Karlionova N, Pinchuk P, Verkuil YI, et al. 2016. A supergene determines highly divergent male reproductive morphs in the ruff. *Nat Genet* **48:** 79–83. doi:10.1038/ng.3443

Lamichhaney S, Barrio AM, Rafati N, Sundstrom G, Rubin CJ, Gilbert ER, Berglund J, Wetterbom A, Laikre L, Webster MT, et al. 2012. Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proc Natl Acad Sci* **109:** 19345–19350. doi:10.1073/pnas.1216128109

Lamichhaney S, Fan G, Widemo F, Gunnarsson U, Thalmann DS, Hoeppner MP, Kerje S, Gustafson U, Shi C, Zhang H, et al. 2016. Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nat Genet* **48:** 84–88. doi:10.1038/ng.3430

Lamichhaney S, Fuentes-Pardo AP, Rafati N, Ryman N, McCracken GR, Bourne C, Singh R, Ruzzante DE, Andersson L. 2017. Parallel adaptive evolution of geographically distant herring populations on both sides of the North Atlantic Ocean. *Proc Natl Acad Sci* **114:** E3452–E3461. doi:10.1073/pnas.1617728114

Larsson LC, Laikre L, André C, Dahlgren TG, Ryman N. 2010. Temporally stable genetic structure of heavily exploited Atlantic herring (*Clupea harengus*) in Swedish waters. *Heredity (Edinb)* **104:** 40–51. doi:10.1038/hdy.2009.98

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326:** 289–293. doi:10.1126/science.1181369

Limborg MT, Helyar SJ, De Bruyn M, Taylor MI, Nielsen EE, Ogden R, Carvalho GR, Consortium FPT, Bekkevold D. 2012. Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Mol Ecol* **21:** 3686–3703. doi:10.1111/j.1365-294X.2012.05639.x

Martinez Barrio A, Lamichhaney S, Fan G, Rafati N, Pettersson M, Zhang H, Dainat J, Ekman D, Höppner M, Jern P, et al. 2016. The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. *eLife* **5:** e12081. doi:10.7554/eLife.12081

Ojaveer H, Tomkiewicz J, Arula T, Klais R. 2015. Female ovarian abnormalities and reproductive failure of autumn-spawning herring (*Clupea harengus membras*) in the Baltic Sea. *ICES J Mar Sci* **72:** 2332–2340. doi:10.1093/icesjms/fsv103

Peck MA, Kanstinger P, Holste L, Martin M. 2012. Thermal windows supporting survival of the earliest life stages of Baltic herring (*Clupea harengus*). *ICES J Mar Sci* **69:** 529–536. doi:10.1093/icesjms/fss038

Peichel CL, Sullivan ST, Liachko I, White MA. 2017. Improvement of the threespine stickleback genome using a Hi-C-based proximity-guided assembly. *J Hered* **108:** 693–700. doi:10.1093/jhered/esx058

R Core Team. 2015. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159:** 1665–1680. doi:10.1016/j.cell.2014.11.021

Rastas P, Paulin L, Hanski I, Lehtonen R, Auvinen P. 2013. Lep-MAP: fast and accurate linkage map construction for large SNP datasets. *Bioinformatics* **29:** 3128–3134. doi:10.1093/bioinformatics/btt563

Roach MJ, Schmidt SA, Borneman AR. 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19:** 460. doi:10.1186/s12859-018-2485-7

Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. 2017. Variant review with the integrative genomics viewer. *Cancer Res* **77:** e31–e34. doi:10.1158/0008-5472.CAN-17-0337

Rondeau EB, Minkley DR, Leong JS, Messmer AM, Jantzen JR, von Schalburg KR, Lemon C, Bird NH, Koop BF. 2014. The genome and linkage map of the northern pike (*Esox lucius*): conserved synteny revealed between the salmonid sister group and the neoteleostei. *PLoS One* **9:** e102089. doi:10.1371/journal.pone.0102089

Ryman N, Lagercrantz U, Andersson L, Chakraborty R, Rosenberg R. 1984. Lack of correspondence between genetic and morphologic variability patterns in Atlantic herring (*Clupea harengus*). *Heredity (Edinb)* **53:** 687–704. doi:10.1038/hdy.1984.127

Schwander T, Libbrecht R, Keller L. 2014. Supergenes and complex phenotypes. *Curr Biol* **24:** R288–R294. doi:10.1016/j.cub.2014.01.056

Tuttle EM, Bergland AO, Korody ML, Brewer MS, Newhouse DJ, Minx P, Stager M, Betuel A, Cheviron ZA, Warren WC, et al. 2016. Divergence and functional degradation of a sex chromosome-like supergene. *Curr Biol* **26:** 344–350. doi:10.1016/j.cub.2015.11.069

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9:** e112963. doi:10.1371/journal.pone.0112963

Wang J, Wurm Y, Nipitwattanaphon M, Riba-Grognuz O, Huang YC, Shoemaker D, Keller L. 2013. A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493:** 664–668. doi:10.1038/nature11832

Zhang W, Westerman E, Nitzany E, Palmer S, Kronforst MR. 2017. Tracing the origin and evolution of supergene mimicry in butterflies. *Nat Commun* **8:** 1269. doi:10.1038/s41467-017-01370-1

A chromosome-level assembly of the Atlantic herring genome—detection of a supergene and other signals of selection

Mats E. Pettersson, Christina M. Rochus, Fan Han, et al.

Supplemental Material	http://genome.cshlp.org/content/suppl/2019/10/24/gr.253435.119.DC1
References	This article cites 45 articles, 9 of which can be accessed free at: http://genome.cshlp.org/content/29/11/1919.full.html#ref-list-1
Open Access	Freely available online through the *Genome Research* Open Access option.
Creative Commons License	This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**