GWAS identifies a single selective sweep for age of maturation in wild and cultivated Atlantic salmon males.

Fernando Ayllon[1], Erik Kjærner-Semb[1,3], Tomasz Furmanek[1], Vidar Wennevik[1], Monica Solberg[1], Harald Sægrov[2], Kurt Urdal[2], Geir Dahle[1], Geir Lasse Taranger[1], Kevin A. Glover[1,3], Markus S Almén[4], Carl J Rubin[4], Rolf B.Edvardsen[1], Anna Wargelius[1*]
[1] Institute of Marine Research, Bergen, Norway
[2] Rådgivende Biologer AS, Bergen, Norway
[3] Department of Biology, University of Bergen, Norway
[4] Uppsala Universitet, Uppsala, Sweden

*Corresponding author

Keywords. SNPs, puberty, teleost, salmonides, *vgll3*, *akap11*, *chmp2B*, salmon genome, pool sequencing, aquaculture

Abstract

Background

Sea age at sexual maturation displays large plasticity for wild Atlantic salmon males and varies between 1-5 years. This flexibility can also be observed in domesticated salmon. Previous studies have uncovered a genetic predisposition for age at maturity with moderate heritability, thus suggesting a polygenic nature of this trait. The aim with this study was to identify genomic regions and associated SNPs and genes conferring age at maturity in salmon.

Results

We performed a GWAS using a pool sequencing approach (n=20 per river and trait) of salmon returning as sexually mature either after one sea winter (2009) or after three sea winters (2011) in six rivers in Norway. The study revealed one major selective sweep, which covered 76 significant SNP in a 230 kb region of Chr 25. A SNP assay of other year classes of wild salmon and from cultivated fish supported this finding. The assay in cultivated fish reduced the haplotype conferring the trait to a region which covered 4 SNPs of a 2386 bp region containing the *vgll3* gene. 2 of these SNPs caused miss-sense mutations in *vgll3*.

Conclusions

This study presents a single selective region in the genome for age at maturation in male Atlantic salmon. The SNPs identified may be used as QTLs to prevent early maturity in aquaculture and in monitoring programs of wild salmon. Interestingly, the identified *vgll3* gene has previously been linked to time of puberty in humans, suggesting a conserved mechanism for time of puberty in vertebrates.

Background

Both wild and cultured stocks of Atlantic salmon (*Salmo salar*) show large genetic and phenotypic plasticity for sea age at sexual maturity [1]. Wild salmon males can stay in the sea from 1-5 years before they initiate sexual maturation and return to their native river to spawn, while females usually return to the river after 2-3 years in the sea. This plasticity in age of sexual maturity, especially for males, can also be observed in salmon aquaculture, where precocious puberty of males has been a major problem due to negative effects on somatic growth, flesh quality, animal welfare and susceptibility to disease [1]. Early maturation in cultured salmon can also increase the risk of further genetic introgression of escaped salmon in wild populations [2, 3], as maturing fish will have a high likelihood of migrating to a nearby river to spawn, whereas immature fish most likely will migrate to sea where mortality is large before reaching maturity [4]. The impact of precocious maturity in farms has been greatly reduced using continuous light treatment which can override the genetic predisposition for early puberty [1]. Further improvement of selective breeding for late puberty is, on the other hand, hampered as the current production conditions, using continuous light, can mask genetic traits related to early puberty, making it difficult to gain further improvements of this trait by classical breeding protocols. In recent years sea temperatures have been higher than usual in Norwegian waters [5]. Increased sea temperature can override the effect of continuous light treatment in delaying maturation, causing unwanted male precocious maturation at the postsmolt stage [6], and possibly also increase incidence of unwanted sexual maturation after one sea winter. Increased water temperatures associated with climate change therefore demand for better genetic selection programs, in which production fish ultimately show a more robust genetic trait for late maturation

It is well known that salmonids display moderately high heritability for sexual maturity [7-11] and QTLs relating to this trait have been identified [12]. Also three recent papers used SNP arrays to identify regions under selection for age of sea water return in both an aquaculture strain using a low density SNP array [15, 16] and wild populations using a high density SNP array [14]. These three studies revealed multiple regions in the genome which adds to the age at maturity trait. However, no clear answer regarding possible mechanism/genes/regions behind time of maturity was revealed. These studies have searched for regions in the genome under selection using an already defined set of SNPs. This strategy may fail to identify the causative SNPs, since these SNPs may not be present on the array [17]. The recent sequencing of the salmon genome enables fine tuned selection with both SNP arrays and especially genome re-sequencing [13, 14], which gives the opportunity to not only identify novel SNPs, but also indels and insertions in the genome which may explain the underlying genetic mechanisms behind traits [18]. The use of this method in salmon may enable linking of novel SNPs to regulatory regions and/or genes in the genome to, sea age of maturity trait in salmon.

Many molecules and pathways are already known to be involved in the male maturation process in vertebrates. The brain pituitary gonad axis (BPG) is strongly involved in this process in vertebrates and also in fish [1, 19]. However it is unknown how this axis is

modulated to trigger puberty at different ages. As the Atlantic salmon displays large plasticity for this trait, it represents an ideal species in which to identify factors that are generally used by vertebrates to control maturation. Thus, the study of the genetic mechanisms controlling age of maturation in Atlantic salmon may also provide information that is relevant to general vertebrate biology.

This study aims to elucidate which genes and genomic regions regulate the sea age of maturation in male Atlantic salmon. To elucidate the underlying genomic basis of sea age of maturation in salmon, we used scale samples from wild fish returning to rivers either after 1 year at sea (1SW, 2009) or after 3 years at sea (3SW, 2011) from six rivers in Western Norway (Figure 1). In each river we used pooled DNA from 20 1SW fish and 20 3SW fish, which in total gave a material of 240 fish. Each pool was sequenced to a depth of ~12.3 X coverage, using deep sequencing. Sequenced pools were mapped to the salmon genome and SNP calling was performed. Significantly differing SNPs among age-groups were identified using the Mantel test. Using sliding windows approach we identified a region in chromosome 25 (Chr 25) displaying a dense selection of significant SNPs covering 230 kb. This region was found to cover three genes (*chmp2b*, *vgll3* and *akap11*). Interestingly, the identified haplotype Chr 25 were either linked to late or early maturity also in other year-classes (1999, 2004) and in an aquaculture strain, confirming the importance of this single genomic region in determining age at maturity in salmon. The aquaculture strain data revealed a shorter haplotype covering only 4 SNPs found in a 2386 bp region comprising the *vgll3* gene.

Results and Discussion

To find significant SNPs for the age at maturation in salmon males, we sequenced 20 salmon per river and sea age (1SW and 3SW). This number of individuals in each pool has been shown to be a good prerequisite to identify causative SNPs for a trait in *Drosophila* [20]. Mapping our data yielded a 12.32X mean coverage (0.24 SE) of unique mapped reads per river and SW return age (Supplementary Figure 1 and table S1). This coverage is similar to other studies in vertebrates including pig and chicken that have successfully identified signatures of selection [21, 22]. About 34% of the current genome has not been assigned to chromosomes, these contigs showed in or data to harbor only 1% of our uniquely mapped reads (Supplementary Figure 1). SNP calling revealed 4326591 SNPs in all sea ages and rivers using the chosen analytical criteria. Comparing 1SW and 3SW river traits using the Mantel test with a high threshold of significance (FDR<0.001) revealed 155 SNPs significantly associated with sea water age (Figure 2A, Supplementary table S1). 76 (49%) of these SNPs were found in a limited area Chr 25 (Figure 2A), covering ~230kb (Figure 2B). Additionally, a spread of single significant SNPs were found in chromosomes 1-7, 9-24 and 27-29, although only three of them showed a significance above –log (10) in Chr 12, Chr 17 and Chr 21 (Figure 1A). In a previous QTL study for precocious parr maturation the trait was shown to be linked to Chr 12 [23]. Chr 12 has also been associated with sea age at maturation in another study [24]. In a genome wide association study (GWAS) in Atlantic salmon, using a 6.5 kb SNP chip, grilsing was found to be weakly linked to both Chr 12 and Chr 25 [15]. From our data we conclude that in Western Norway there seems to be one single selective sweep in Chr 25 for age of maturation while other regions in the genome might contribute to a lesser degree. This is in contrast to earlier reports showing a polygenic nature of this trait, with contributions from several genomic regions [15, 16, 24]. However, one previous model based study have also suggested that time of maturation could be regulated by a stable genetic polymorphism which is in accordance with our finding[25].

To identify individual genotypes associated with the age at maturity revealed by the sequencing pools, we applied sequenom analysis for 11 of the most significant SNPs in the selective sweep found in Chr 25 (Supplementary table S2). Sequenom data revealed allele frequencies which to a large extent explained 1SW and 3SW traits for all of the SNPs (Supplementary Figure 2). To identify haplotypes in the 11 assayed SNPs we performed a genotype assay on all samples used in our GWAS [26]. This analysis revealed two dissimilar haplogroups comprising 11 haplotypes in one block (Figure 3A). One and five of these haplotypes showed significant association with either maturing early or late respectively. This data confirmed our finding in the pool sequencing and further supported the strength of this haplotype in determining the trait.

In samples from the pool sequencing we identified the genetic sea-age trait in the same yearclass (2008). These fish could possibly have been exposed to similar environmental conditions in these years, therefore showing a selection for those conditions as postulated by several previous studies in salmon [27-29]. To investigate this theory we assayed genotypes in other year classes: 1999 for Eidselva and 2004 for Suldalslågen. Allele frequencies for the 11

assayed SNPs showed correlation to either 1SW or 3SW trait (Supplemental Figure 3). Haplotype analysis of other year classes again revealed two haplogroups for the 11 SNPs assayed in Chr25 (Figure 3B). One of the haplotypes showed significant correlation to the 1SW trait. This specific variant was also found in the original year class of 2009 (Figure 3A), where it was associated to the 1SW trait. The 3SW trait was significantly represented by one haplotype, which was also found to be the most significant haplotype identified among the significant haplotypes found in the 2009 year class. This experiment clearly showed a strong single genetic predisposition for sea age at maturation independent of year class in wild fish in Western Norway.

Age at maturity can be significantly altered in salmon by modulating both light and temperature [1, 30, 31]. As a consequence, current aquaculture production methods include the use of constant light to inhibit maturation. However, in turn, this means that selection for a genetic predisposition to mature late in life has been relaxed since the early 90s, when light regimes became part of the standard rearing procedure. We were interested to see how much the identified genetic trait was contributing to the age of maturation in captive fish since wild fish live in a more diverse environment that may trigger time of puberty in another way. To assay the linkage between age at maturity in an aquaculture strain, we sampled DNA from maturing fish from four different families of the MOWI strain. The MOWI strain has been in aquaculture for at least 10 generations, has been selected for a variety of traits, and displayed significantly increased growth rates compared to wild salmon strains [32-34]. This strain was obtained from a combination of large salmon from Western Norway in 1969 and has been bred using a four year life cycle. The breeding company has thereby probably increased the allele frequency for the late maturity trait. In this experiment, fish were grown under natural light conditions in marine cages where males were naturally maturing after 1, 2 or 3 or more years in sea water. Haplotype analysis of these fish (n=97) revealed a shorter haplotype, consisting only of the first four SNPs assayed, covering only 2386 bp in the 5’ end of the region assayed (Figure 3C, Supplemental Figure 4). This analysis revealed two significant haplogroups which could explain late maturity or early maturity, respectively. The data clearly demonstrated that time of puberty can be explained by SNPs in this region also in an aquaculture strain.

The above mentioned experiments clearly show that the selective sweep in Chr25 significantly contributes to the age at maturity both in wild and domesticated salmon. Gene prediction in this area revealed three genes; *charged multivesicular protein 2B* (*chmp2B*), *vestigial like protein 3* (*vgll3*) and *a-kinase anchor protein 11* (*akap11*, Figure 2A). From the analysis of cultivated salmon we could decrease the area of selection to a 2.4 kb region covering only *vgll3*. This region contained two missense mutations in *vgll3;* at aa 54 and aa 323. In all fish having the 3SW haplotype these missense mutations led to amino acid changes Thr and Lys in 3SW while in 1SW fish they were represented by a Met and an Asp respectively. Our analysis could not conclude whether these missense mutations are causative for the time of maturity trait, but since they occurred consistently together in the material it is likely that they are. It is also known from other studies that co-occurring amino acid changes can confer a phenotype under selection [35]. The Vgll3 protein function as a cofactor for the

TEAD family of transcription factors [36] (Figure 2B). The transcription factor binding region (tondu) in aa105-aa134 in Vgll3 does not cover any of the aa changes discovered which means that any direct binding changes are less likely to occur in either 1SW or 3SW fish. It is difficult to predict how these amino acid changes affect the protein. To this point we cannot elucidate whether it is these missense mutations or other SNPs outside coding regions which confer the trait.

In humans the VGLL3 protein has been linked to the age at maturity by a SNP in close proximity of the gene [37], this strengthens our notion that the salmon Vgll3 protein is also involved in the age of puberty in fish. Regarding the function of this protein in controlling age of maturation, it is known that Vgll3 is involved in the inhibition of adipocyte differentiation in mouse [38]. Likewise changes in fat metabolism may be causative for changes in the age of maturation, since increased adiposity has previously been linked to maturation in salmon [39-42]. In other studies in rodent testis, *Vgll3* transcipts have been associated to regulated expression during early stages of steroidogenesis in the embryonal testis [43], which may link this proteins function to time of testis maturation. Further functional studies of this protein and regions around will confirm if the previous study in humans and our study actually links this protein to time of puberty in vertebrates.

The most significant SNPs identified in this study covered the Vgll3 protein but also many of the other significant SNPs covered the two neighboring genes in the salmon genome; *chmp2B* and *akap11* (Figure 2B). Akap11 also contained a missense mutation which translates to a Val in 1SW and a Met in 3SW at aa 214. AKAP11 is involved in compartmentalization of cyclic AMP-dependent protein kinase (PKA). The missense mutation in AKAP11 is not located to any of the known functional domains related to PKA [44]. AKAP 11 has been shown to be highly expressed in elongating spermatocytes and mature sperm in human testis. This protein is also believed to contribute to cell cycle control in both germ cells and somatic cells. There are no reports clearly linking this protein to age of maturation but future functional studies will reveal if that is the case. Chmp2B did not contain any missense mutations but many significant SNPs were in proximity of the gene. This is a protein belonging to a complex molecule which is involved in endocytosis of proteins that need to be broken down [45] (Figure 2B). In humans this protein is known to be essential for the survival of nerve cells and is linked to both dementia and ALS [46-48]. Whether, this protein is involved in the regulation of puberty remains to be elucidated. However, it is well known that the neural system works as a gate keeper in controlling age of puberty also in fish [49].

Could there be a balancing selection process ongoing, in which under some conditions makes it beneficial to mature late or early? This could possibly depend on both the conditions in the sea and in the river for example temperature and food accessibility at feeding grounds. A previous study has shown that males maturing late grow faster in their first year at sea [50]. Interestingly there is an inverse relationship between Vgll3 protein expression in fat and body weight in mouse [38], possibly linking this region to metabolic functions of the fish. It is also known that there have been dramatic shifts in fractions of 1SW, 2SW and 3SW in later years in Western Norway[51] which further implies that changing environmental conditions in the

sea, influences growth rates which therefore influences fat deposition and therefore triggering the maturation differences. This implies that in some years it may be very beneficial to have one trait in the sea but in other years a different trait is favored.

Conclusions

In this study we have performed a GWAS by genome re-sequencing, with the aim to find one region in the genome which regulates age of maturation in male Atlantic salmon. By investigating male fish which mature late or early from 6 rivers in Western Norway we demonstrated that the genetic trait for maturity was strongly associated with one locus in the salmon genome at Chr25. This haplotype can be used in selective breeding to identify individuals with the late maturation trait, thereby possibly reducing the incidence of negative phenotypes associated with early maturation of males in salmon aquaculture. This study also shows that both 1SW and 3SW trait haplotypes are under strong selection in Atlantic salmon in Western Norway, and may also be implemented in surveillance of wild salmon populations in the face of changing environmental conditions such as increased sea temperatures. Interestingly, this study also demonstrates, together with a previous study in humans, that the *vgll3* gene or regulatory regions surrounding this is involved in regulating age of maturation in vertebrates.

Materials and Methods

Samples and sampling

The private environmental surveillance company in Bergen Rådgivende Biologer AS (http://www.radgivende-biologer.no/default.aspx?pageId=81) collected scale samples from game fishing of adult wild Atlantic salmon for the rivers assayed in Western Norway. These rivers were Eidselva, Gloppenelva, Flekke, Årdalselva, Suldalslågen and Vorma (see Figure 1). To minimize the potential influence of environmental variation on the age of maturation in this material, we used fish from the same year class and therefore 1 sea winter (SW) fish were collected as adults in these rivers in 2009, and 3SW as adults in rivers in 2011. Each river was represented by 20 1SW and 20 3SW males. For the sequenom SNP assay, we included two other year classes of fish, one from Eidselva and one from Suldalslågen. From Eidselva we got scales from 20 1SW males from year 2000 and 8 3SW males from 2002. From Suldalslågan we got male scales from 14 1SW fish from year 2005 and 14 3SW males from 2007.

In addition to samples of wild salmon captured in rivers, we investigated age of maturation in four full sibling families of domesticated salmon from the Norwegian Mowi strain maturing at either 1SW, 2SW or older. These fish were obtained from an ongoing study at the Matre Aquaculture station where they were reared in marine cages without the use of continuous light. We sampled fin clips from a total of 97 fish maturing at different ages in the sea. The four families consisted of 36, 24, 13 and 24 fish per family (Supplementary Figure S4).

DNA extraction and PCR-based sdY test

DNA from selected individuals was purified from 2 to 3 scales using Qiagen DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer´s protocols. Following Eisbrenner *et al*. (2014), sex of all samples used herein was validated by a PCR-based methodology aimed to detect the presence of the sdY gene [52, 53]. Briefly, any individual showing exon 2 and exon 4 amplicons was designated as male. As a PCR quality control, we used the presence of the 5S rRNA that provides species identification capabilities based on the the total lenght of the amplicon [54]. PCR amplifications were performed using reaction mixtures containing approximately 50 ng of extracted AS DNAtemplate, 10 nM Tris–HCl pH 8.8, 1.5 mM MgCl2,50 mM KCl, 0.1% Triton X-100, 0.35 uM of each primers, 0.5 Units of DNA Taq Polymerase (Promega, Madison, WI, USA) and 250 uM of each dNTP in a final volume of 20 uL. PCR products were visualized in 3% agarose gels. Prior to deep sequencing of DNA samples, all fish were screened for their genetic sex [52, 53]. Genetic analysis revealed that many of the 3SW fish had been wrongly assessed as males when caught in the river. This is due to that fish are not opened upon capture, and that distinct sexual characteristics in Atlantic salmon arrive later in the reproductive season than when fish were caught by angling (May-Sept).

Deep sequencing and mapping

After fluorometric quantification of DNA, equal amounts of DNA from ten males from each replicate were combined into a pool for sequencing. Paired-end libraries were generated using the Genomic DNA Sample Preparation Kit (Illumina, CA, USA) and then sequenced on the Illumina HiSeq2000 platform (Illumina, CA, USA). Sequencing was performed at the Norwegian Sequencing center (https://www.sequencing.uio.no/, Oslo, Norway). In each lane sequenced we used pools of 10 fish from each sea age and river which made a total of 24 lanes sequenced in the whole experiment (6 rivers, 2 replicates per sea age). The sequence data has been deposited at SRA with bioproject number PRJNA293012**.** Raw sequencing was obtained in FastQ files, this data were quality trimmed with Cutadapt (https://pypi.python.org/pypi/cutadapt/, http://dx.doi.org/10.14806/ej.17.1.200). Before quality trimming we peformed FASTQc quality control on raw sequencing was performed, none of the sequenced libraries showed a quality score below 20 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). All libraries were therefore approved for further analysis. All libraries were mapped to the most recent salmon genome release (Acc. No. AGKD0000000.4) using Bowtie 2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml, [55]).

SNP calling and statistical analysis of significant SNPs

Replicate datasets were merged using samtools merge (v. 0.1.19-96b5f2294a) so that each river was represented by a single bam file containing 1SW reads, and a another bam file containing 3SW reads [56]. Each file had ~12.3 x coverage (Supplementary Figure S1). A mpileup file was generated using samtools mpileup and was converted to a sync file for use with the PoPoolation2 package (build 201, 24. feb 2015) [57]. Minimum mapping quality and base quality was set to 20 for both criteria to remove reads with ambiguous mapping and to only analyze high quality sequence positions. SNPs with different allele frequencies between 1SW and 3SW samples were identified using the CMH-test (cmh-test.pl) as implemented in the PoPoolation2 package. The parameters min-count, min-coverage and max-coverage were set to 10, 5 and 100, respectively. A false discovery rate filter as described in [58] was implemented and using R custom scripts. False discovery rate was set to alpha = 0.001 with a cutoff value of 7.27. We only considered SNPs having coverage between 7 and 42 in each merged sample of 20 fish to avoid bias due to the assembly and mapping (95% percentile).

To provide additional support for the identified SNPs, F_{ST} values were estimated by merging all 1SW and all 3SW bam files to produce single 1SW and 3SW bam files, both with ~120 x coverage. The samtools mpilup command was used followed by the conversion from mpileup to sync format. The F_{ST} calculations were made using the F_{ST} software of the PoPoolation2 package (fst-sliding.pl) using every nucleotide position, with min-count, min-coverage, max-coverage and pool-size set to 20, 5, 400 and 120, respectively.

Genome annotation

Augustus gene prediction software was trained using PASA gene candidates, by mapping salmon ESTs from NCBI to the salmon genome assembly with PASA [59, 60]. The Augustus

de novo gene prediction contained coding sequences without UTRs. The genes were validated by RNASeq mapping and annotated with Swissprot.

Sequenom assays

The sequenome assay was designed and run on a Sequenom MassARRAY analyser (San Diego, CA, USA). The 11 most significant SNPs in the 230 kb region identified in Chr25 (indicated in Figure 2B, listed in Supplementary table S1). Primers and extension primers are found in Supplementary table S2.

Competing interests

There are no competing interests regarding this paper

Authors' contribution

AW, GLT, KG, VW and RBE designed the study. VW, HS and KU provided the wild salmon samples from Western Norway. KG and MS provided the cultivated salmon samples. FA and AW prepared the samples. FA ran the sex determination test for all samples assayed. FA, EK, TF, MS, CR, AW and RBE analysed the resequenced material. GD, FY and EK designed and performed the sequenom assays. FY and EK designed all figures for the paper. FA, EK, TF, RBE and AW wrote the manuscript with the contribution of all authors. All authors approved the final manuscript.

Acknowledgements

We would like to acknowledge the expert advice of Eva Andersson, Tom Hansen, Ketil Malde, Sigbjørn Lien and Per Gunnar Fjelldal during the project. We are also grateful to Anne-Grethe Sørvik, Stig Mæhle, Sara Olausson, Sven Leininger, Amèlie Juanchich and Lene Kleppe for providing laboratory support. This project was financed by the Norwegian research council (NFR) and their HAVBRUK-BIOTEK 2021 program (project number 226221- SALMAT). The farmed salmon strains analyzed here were produced under the NFR financed project INTERACT.

References

1. Taranger GL, Carrillo M, Schulz RW, Fontaine P, Zanuy S, Felip A, Weltzien FA, Dufour S, Karlsen O, Norberg B, et al: **Control of puberty in farmed fish.** *Gen Comp Endocrinol* 2010, **165:**483-515.
2. Glover KA, Quintela M, Wennevik V, Besnier F, Sorvik AGE, Skaala O: **Three Decades of Farmed Escapees in the Wild: A Spatio-Temporal Analysis of Atlantic Salmon Population Genetic Structure throughout Norway.** *Plos One* 2012, **7**.
3. Glover KA, Pertoldi C, Besnier F, Wennevik V, Kent M, Skaala O: **Atlantic salmon populations invaded by farmed escapees: quantifying genetic introgression with a Bayesian approach and SNPs.** *Bmc Genetics* 2013, **14**.
4. Olsen RE, Skilbrei OT: **Feeding preference of recaptured Atlantic salmon Salmo salar following simulated escape from fish pens during autumn.** *Aquaculture Environment Interactions* 2010, **1:**167-174.
5. Lorentzen T: **A statistical analysis of sea temperature data.** *Theoretical and Applied Climatology* 2015, **119:**585-610.
6. Fjelldal PG, Hansen T, Huang TS: **Continuous light and elevated temperature can trigger maturation both during and immediately after smoltification in male Atlantic salmon (Salmo salar).** *Aquaculture* 2011, **321:**93-100.
7. Wild V, Simianer H, Gjoen HM, Gjerde B: **Genetic-Parameters and Genotype X Environment Interaction for Early Sexual Maturity in Atlantic Salmon (Salmo-Salar).** *Aquaculture* 1994, **128:**51-65.
8. Gjerde B, Simianer H, Refstie T: **Estimates of Genetic and Phenotypic Parameters for Body-Weight, Growth-Rate and Sexual Maturity in Atlantic Salmon.** *Livestock Production Science* 1994, **38:**133-143.
9. Gjerde B, Gjedrem T: **Estimates of Phenotypic and Genetic-Parameters for Carcass Traits in Atlantic Salmon and Rainbow-Trout.** *Aquaculture* 1984, **36:**97-110.
10. Gjerde B: **Response to Individual Selection for Age at Sexual Maturity in Atlantic Salmon.** *Aquaculture* 1984, **38:**229-240.
11. Gjedrem T: **Genetic improvement of cold-water fish species.** *Aquaculture Research* 2000, **31:**25-33.
12. Moghadam HK, Poissant J, Fotherby H, Haidle L, Ferguson MM, Danzmann RG: **Quantitative trait loci for body weight, condition factor and age at sexual maturation in Arctic charr (Salvelinus alpinus): comparative analysis with rainbow trout (Oncorhynchus mykiss) and Atlantic salmon (Salmo salar).** *Mol Genet Genomics* 2007, **277:**647-661.
13. Davidson WS, Koop BF, Jones SJ, Iturra P, Vidal R, Maass A, Jonassen I, Lien S, Omholt SW: **Sequencing the genome of the Atlantic salmon (Salmo salar).** *Genome Biol* 2010, **11:**403.
14. Johnston SE, Orell P, Pritchard VL, Kent MP, Lien S, Niemela E, Erkinaro J, Primmer CR: **Genome-wide SNP analysis reveals a genetic basis for sea-age variation in a wild population of Atlantic salmon (Salmo salar).** *Mol Ecol* 2014, **23:**3452-3468.
15. Gutierrez AP, Yanez JM, Fukui S, Swift B, Davidson WS: **Genome-Wide Association Study (GWAS) for Growth Rate and Age at Sexual Maturation in Atlantic Salmon (Salmo salar).** *PLoS One* 2015, **10:**e0119730.
16. Gutierrez AP, Lubieniecki KP, Fukui S, Withler RE, Swift B, Davidson WS: **Detection of quantitative trait loci (QTL) related to grilsing and late sexual maturation in Atlantic salmon (Salmo salar).** *Mar Biotechnol (NY)* 2014, **16:**103-110.

17. Lien S, Gidskehaug L, Moen T, Hayes BJ, Berg PR, Davidson WS, Omholt SW, Kent MP: **A dense SNP-based linkage map for Atlantic salmon (Salmo salar) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns.** *BMC Genomics* 2011, **12:**615.
18. Rubin CJ, Zody MC, Eriksson J, Meadows JR, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S, et al: **Whole-genome resequencing reveals loci under selection during chicken domestication.** *Nature* 2010, **464:**587-591.
19. Schulz RW, de Franca LR, Lareyre JJ, Le Gac F, Chiarini-Garcia H, Nobrega RH, Miura T: **Spermatogenesis in fish.** *Gen Comp Endocrinol* 2010, **165:**390-411.
20. Ferretti L, Ramos-Onsins SE, Perez-Enciso M: **Population genomics from pool sequencing.** *Mol Ecol* 2013, **22:**5561-5576.
21. Rubin CJ, Megens HJ, Barrio AM, Maqbool K, Sayyab S, Schwochow D, Wang C, Carlborg O, Jern P, Jorgensen CB, et al: **Strong signatures of selection in the domestic pig genome.** *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109:**19529-19536.
22. Rubin CJ, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S, et al: **Whole-genome resequencing reveals loci under selection during chicken domestication.** *Nature* 2010, **464:**587-U145.
23. Pedersen S, Berg PR, Culling M, Danzmann RG, Glebe B, Leadbeater S, Lien S, Moen T, Vandersteen W, Boulding EG: **Quantitative trait loci for precocious parr maturation, early smoltification, and adult maturation in double-backcrossed trans-Atlantic salmon (Salmo salar).** *Aquaculture* 2013, **410:**164-171.
24. Johnston SE, Orell P, Pritchard VL, Kent MP, Lien S, Niemela E, Erkinaro J, Primmer CR: **Genome-wide SNP analysis reveals a genetic basis for sea-age variation in a wild population of Atlantic salmon (Salmo salar).** *Mol Ecol* 2014.
25. Gurney WS, Bacon PJ, Speirs DC, McGinnity P, Verspoor E: **Sea-age variation in maiden Atlantic salmon spawners: phenotypic plasticity or genetic polymorphism?** *Bull Math Biol* 2012, **74:**615-640.
26. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC: **PLINK: A tool set for whole-genome association and population-based linkage analyses.** *American Journal of Human Genetics* 2007, **81:**559-575.
27. Miller AS, Sheehan TF, Renkawitz MD, Meister AL, Miller TJ: **Revisiting the marine migration of US Atlantic salmon using historical Carlin tag data.** *Ices Journal of Marine Science* 2012, **69:**1609-1615.
28. Bacon PJ, Palmer SCF, MacLean JC, Smith GW, Whyte BDM, Gurney WSC, Youngson AF: **Empirical analyses of the length, weight, and condition of adult Atlantic salmon on return to the Scottish coast between 1963 and 2006.** *Ices Journal of Marine Science* 2009, **66:**844-859.
29. Hansen LP, Quinn TR: **The marine phase of the Atlantic salmon (Salmo salar) life cycle, with comparisons to Pacific salmon.** *Canadian Journal of Fisheries and Aquatic Sciences* 1998, **55:**104-118.
30. Taranger GL, Haux C, Hansen T, Stefansson SO, Bjornsson BT, Walther BT, Kryvi H: **Mechanisms underlying photoperiodic effects on age at sexual maturity in Atlantic salmon, Salmo salar.** *Aquaculture* 1999, **177:**47-60.
31. Taranger GL, Vikingstad E, Klenke U, Mayer I, Stefansson SO, Norberg B, Hansen T, Zohar Y, Andersson E: **Effects of photoperiod, temperature and GnRHa treatment on the reproductive physiology of Atlantic salmon (Salmo salar L.) broodstock.** *Fish Physiology and Biochemistry* 2003, **28:**403-406.

32. Glover KA, Ottera H, Olsen RE, Slinde E, Taranger GL, Skaala O: **A comparison of farmed, wild and hybrid Atlantic salmon (Salmo salar L.) reared under farming conditions.** *Aquaculture* 2009, **286:**203-210.
33. Solberg MF, Zhang ZW, Nilsen F, Glover KA: **Growth reaction norms of domesticated, wild and hybrid Atlantic salmon families in response to differing social and physical environments.** *Bmc Evolutionary Biology* 2013, **13**.
34. Solberg MF, Skaala O, Nilsen F, Glover KA: **Does Domestication Cause Changes in Growth Reaction Norms? A Study of Farmed, Wild and Hybrid Atlantic Salmon Families Exposed to Environmental Stress.** *Plos One* 2013, **8**.
35. de Juan D, Pazos F, Valencia A: **Emerging methods in protein co-evolution.** *Nat Rev Genet* 2013, **14:**249-261.
36. Helias-Rodzewicz Z, Perot G, Chibon F, Ferreira C, Lagarde P, Terrier P, Coindre JM, Aurias A: **YAP1 and VGLL3, encoding two cofactors of TEAD transcription factors, are amplified and overexpressed in a subset of soft tissue sarcomas.** *Genes Chromosomes Cancer* 2010, **49:**1161-1171.
37. Cousminer DL, Berry DJ, Timpson NJ, Ang W, Thiering E, Byrne EM, Taal HR, Huikari V, Bradfield JP, Kerkhof M, et al: **Genome-wide association and longitudinal analyses reveal genetic loci linking pubertal height growth, pubertal timing and childhood adiposity.** *Hum Mol Genet* 2013, **22:**2735-2747.
38. Halperin DS, Pan C, Lusis AJ, Tontonoz P: **Vestigial-like 3 is an inhibitor of adipocyte differentiation.** *J Lipid Res* 2013, **54:**473-481.
39. Trombley S, Mustafa A, Schmitz M: **Regulation of the seasonal leptin and leptin receptor expression profile during early sexual maturation and feed restriction in male Atlantic salmon, Salmo salar L., parr.** *General and Comparative Endocrinology* 2014, **204:**60-70.
40. Larsen DA, Beckman BR, Strom CR, Parkins PJ, Cooper KA, Fast DE, Dickhoff WW: **Growth modulation alters the incidence of early male maturation and physiological development of hatchery-reared spring Chinook salmon: A comparison with wild fish.** *Transactions of the American Fisheries Society* 2006, **135:**1017-1032.
41. Silverstein JT, Shearer KD, Dickhoff WW, Plisetskaya EM: **Regulation of nutrient intake and energy balance in salmon.** *Aquaculture* 1999, **177:**161-169.
42. Silverstein JT, Shearer KD, Dickhoff WW, Plisetskaya EM: **Effects of growth and fatness on sexual development of chinook salmon (Oncorhynchus tshawytscha) parr.** *Canadian Journal of Fisheries and Aquatic Sciences* 1998, **55:**2376-2382.
43. McDowell EN, Kisielewski AE, Pike JW, Franco HL, Yao HHC, Johnson KJ: **A Transcriptome-Wide Screen for mRNAs Enriched in Fetal Leydig Cells: CRHR1 Agonism Stimulates Rat and Mouse Fetal Testis Steroidogenesis.** *Plos One* 2012, **7**.
44. Reinton N, Collas P, Haugen TB, Skalhegg BS, Hansson V, Jahnsen T, Tasken K: **Localization of a novel human A-kinase-anchoring protein, hAKAP220, during spermatogenesis.** *Dev Biol* 2000, **223:**194-204.
45. Bodon G, Chassefeyre R, Pernet-Gallay K, Martinelli N, Effantin G, Hulsik DL, Belly A, Goldberg Y, Chatellard-Causse C, Blot B, et al: **Charged multivesicular body protein 2B (CHMP2B) of the endosomal sorting complex required for transport-III (ESCRT-III) polymerizes into helical structures deforming the plasma membrane.** *J Biol Chem* 2011, **286:**40276-40286.
46. Ferrari R, Kapogiannis D, Huey ED, Grafman J, Hardy J, Momeni P: **Novel missense mutation in charged multivesicular body protein 2B in a patient with frontotemporal dementia.** *Alzheimer Dis Assoc Disord* 2010, **24:**397-401.

47. Ghazi-Noori S, Froud KE, Mizielinska S, Powell C, Smidak M, Fernandez de Marco M, O'Malley C, Farmer M, Parkinson N, Fisher EM, et al: **Progressive neuronal inclusion formation and axonal degeneration in CHMP2B mutant transgenic mice.** *Brain* 2012, **135:**819-832.
48. Parkinson N, Ince PG, Smith MO, Highley R, Skibinski G, Andersen PM, Morrison KE, Pall HS, Hardiman O, Collinge J, et al: **ALS phenotypes with mutations in CHMP2B (charged multivesicular body protein 2B).** *Neurology* 2006, **67:**1074-1077.
49. Zohar Y, Munoz-Cueto JA, Elizur A, Kah O: **Neuroendocrinology of reproduction in teleost fish.** *Gen Comp Endocrinol* 2010, **165:**438-455.
50. Jonsson N, Jonsson B: **Sea growth, smolt age and age at sexual maturation in Atlantic salmon.** *Journal of Fish Biology* 2007, **71:**245-252.
51. Urdal K, Sægrov H: **Analysar av skjelprøvar frå elvar på Vestlandet 1999-2012. Rådgivende Biologer AS, rapport 1797, 28 sider.**
ISBN 978-82-8308-016-2.
52. Eisbrenner WS, Botwright N, Cook M, Davidson EA, Dominik S, Elliott NG, Henshall J, Jones SL, Kube PD, Lubieniecki KP, et al: **Evidence for multiple sex-determining loci in Tasmanian Atlantic salmon (Salmo salar).** *Heredity* 2014, **113:**86-92.
53. Yano A, Guyomard R, Nicol B, Jouanno E, Quillet E, Klopp C, Cabau C, Bouchez O, Fostier A, Guiguen Y: **An immune-related gene evolved into the master sex-determining gene in rainbow trout, Oncorhynchus mykiss.** *Curr Biol* 2012, **22:**1423-1428.
54. Pendas AM, Moran P, Martinez JL, Garciavazquez E: **Applications of 5s-Rdna in Atlantic Salmon, Brown Trout, and in Atlantic Salmon X Brown Trout Hybrid Identification.** *Molecular Ecology* 1995, **4:**275-276.
55. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2.** *Nat Methods* 2012, **9:**357-359.
56. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25:**2078-2079.
57. Kofler R, Pandey RV, Schlotterer C: **PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq).** *Bioinformatics* 2011, **27:**3435-3436.
58. Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing.** *Journal of the Royal Statistical Society Series B-Methodological* 1995, **57:**289-300.
59. Stanke M, Morgenstern B: **AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints.** *Nucleic Acids Res* 2005, **33:**W465-467.
60. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK, Jr., Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, et al: **Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies.** *Nucleic Acids Res* 2003, **31:**5654-5666.

Figure Legends

Figure 1. *Geographical location of salmon rivers used.* Map of Norway and a magnification of Western Norway showing rivers used in the experiment. Rivers selected; three rivers in the North of Western Norway Eidselven, Gloppenelven and Flekkeelven and three rivers in the south Western Norway Suldalslågen, Vormo and Årdalselven.

Figure 2. I*dentification of a selective region conferring time of maturity in Atlantic salmon.* (A) Manhattan plot showing the genomic region under selection covering significant SNPs identified in the whole genome analysis. The X axis presents genomic coordinates along chromosome 1-29 in Atlantic salmon. On the Y axis the negative logarithm of the SNPs associated P-value is displayed. All SNPs found above the solid horizontal dotted line in the plot shows SNPs which are significantly associated with the trait (FDR<0.001). (B) Description of the 230 kb region (28550-28780 kb), plotted on the X axis covering the 76 significants SNPs identified in Chr 25. The SNPs are marked in black dots in the plot. The 8 larger black dots together with the 3 larger squared black dots demarcate those SNPs which were genotyped with the sequenom analysis. The squared dots also indicate those SNPs which confer a missense mutation in either *vgll3* or *akap11*. Below the plot, genomic organization of the three genes found in the region is illustrated. The dashed line within the plot indicates the threshold for significance for significant SNPs (FDR<0.001). On the left positive y-axis the negative logarithm of the SNPs associated P-value is displayed. Below on the negative y-axis depth of sequencing in identified regions is demarcated. The x-axis is showing the location of the region in the Chr 25 and covers 28550-28780 kb. The grey area around the *vgll3* gene demarcates the shorter selective region discovered in the cultivated strain.

Figure 3 *Haplotype frequencies in different yearclasses and I year class 2008*. (A) Haplotype frequency associated with either 1SW (black bars) or 3SW (dark grey bars) in male Atlantic salmon for 6 rivers in Western Norway from year class 2008. (B) Historical Haplotypes frequencies associated with either, 1SW (black bars) or 3SW (dark grey bars) in male Atlantic salmon in year classes 1999 from Eidselva and 2004 from Suldalslågen yearclass 2004. (C) Haplotypes frequencies linked to time of maturity in the cultivated MOWI strain maturing either after 1 (black bars), 2 (grey bars) or 3(dark grey bars) or more years in sea water. This analysis identified a shorter haplotype block than in previous studies, covering only 4 of the SNPs in the 5' region of the block. In all graphs, the X axis indicates frequency of that trait for the identified haplotype, while the Y axis presents the haplotype block obtained from the genotype assay. * Indicates that the haplotype was significantly linked to the trait. The bold base in the haplotype is indicating a missense mutation.

Supplemental Material

Sequenced pool material submitted to SRA (Bioproject number PRJNA293012)

Supplementary Figure 1. (A)Average number of uniquely mapped sequences (y-axis) on each chromosome (x-axis) with +/-SEM. Scaffolds refers to contigs not assigned to chromosomes in the current genome assembly. (B) Average genome coverage. The x axis is showing read depth in all pooled samples +/-SEM. The y axis is showing average percent in the genome covered.

Supplementary Figure 2. Individual SNP frequency in rivers assayed in yearclass 2008 for fish returning after 1SW (green bars) and 3SW (blue bars). These SNPs are indicated by the coordinate shown on the y-axis. This coordinate directly link them to the SNPs indicated in Figure 2B. Above each bar the number genotyped fish is indicated, in all these fish the allele frequency was measured.

Supplementary Figure 3. Individual SNP frequency in Eidselva yearclass 1999 and Suldalslågen yearclass 2004 for fish returning after 1SW (green bars) and 3SW (blue bars). These SNPs are indicated by the coordinate shown on the y-axis. This coordinate directly link them to the SNPs indicated in Figure 2B.Individual SNP frequency in Eidselva yearclass 1999 and Suldalslågen yearclass 2004 for fish returning after 1SW and 3SW. Above each bar the number genotyped fish is indicated, in all these fish the allele frequency was measured.

Supplementary Figure 4. Individual SNP frequency found in farmed salmon (MOWI strain) strains, maturing either after 1SW (green bars), 2SW (yellow bars) or after 3SW or more in sea water (blue bars). These SNPs are indicated by the coordinate shown on the y-axis. This coordinate directly link them to the SNPs indicated in Figure 2B. Above each bar the number genotyped fish is indicated, in all these fish the allele frequency was measured.

Supplementary Table S1. Table of significant SNPs, alleles and genomic location
Supplemental Table S2. Primers and extension primers listed for the Sequenom assay

Figure 1

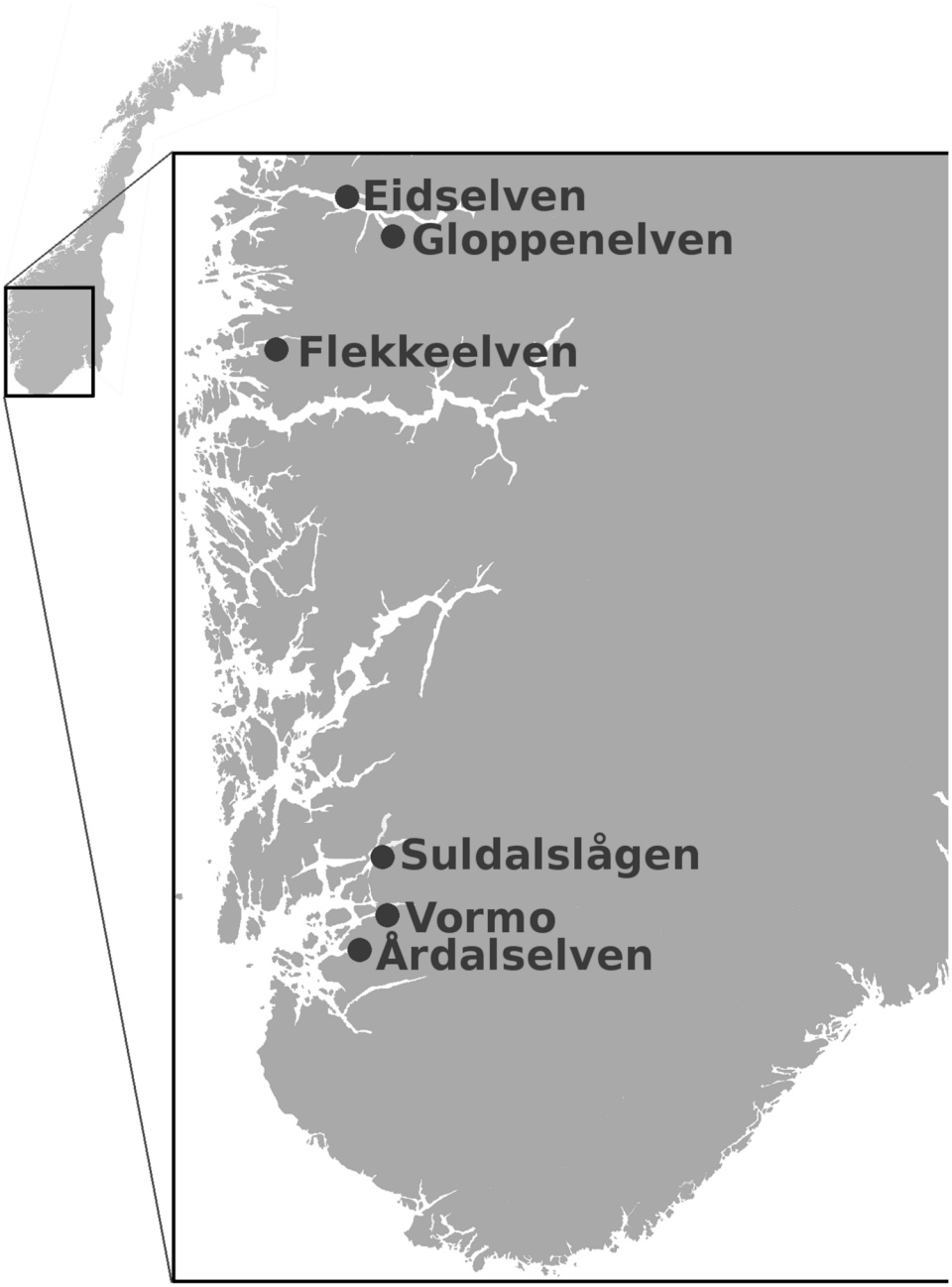

Eidselven
Gloppenelven
Flekkeelven
Suldalslågen
Vormo
Årdalselven

Figure 2

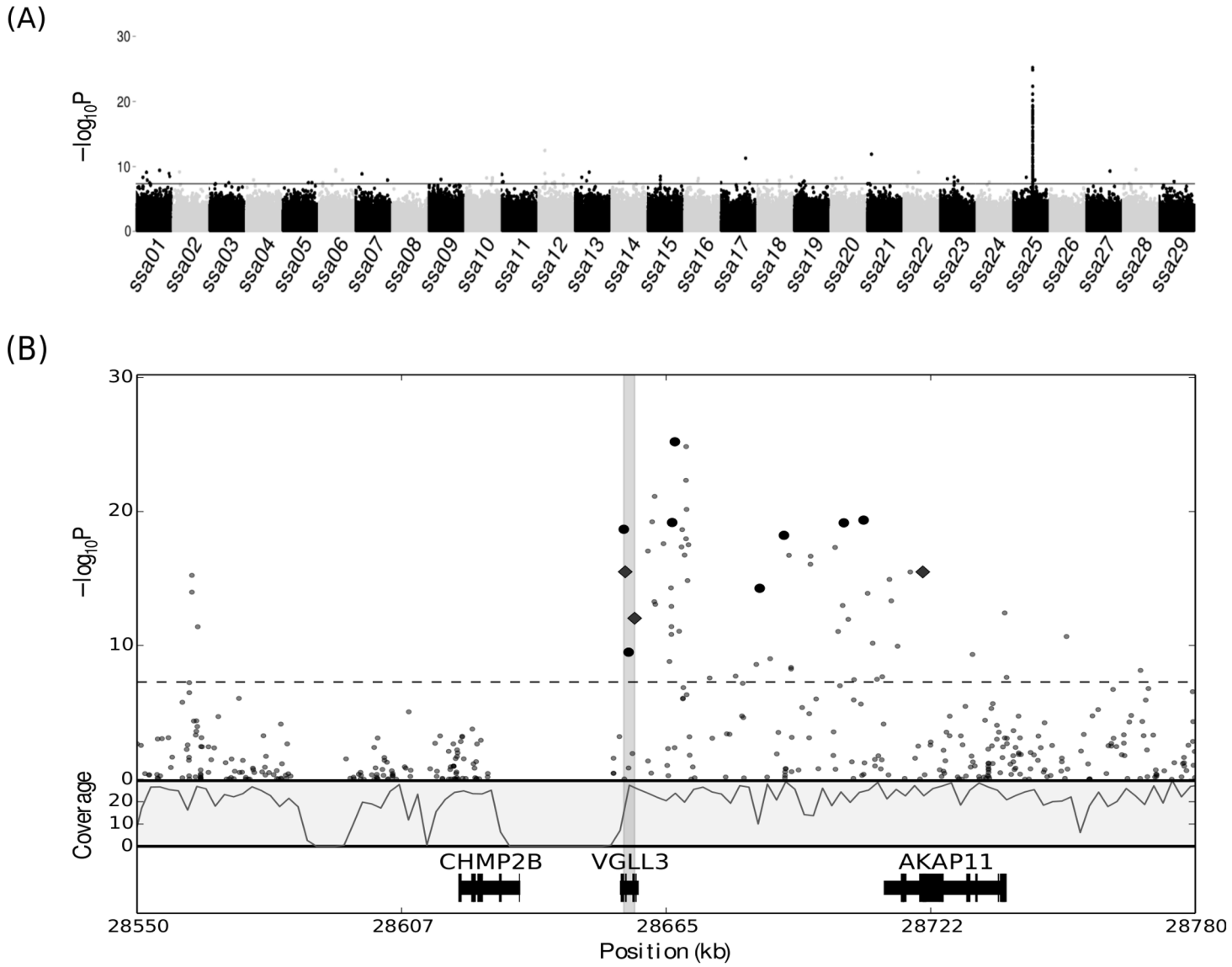
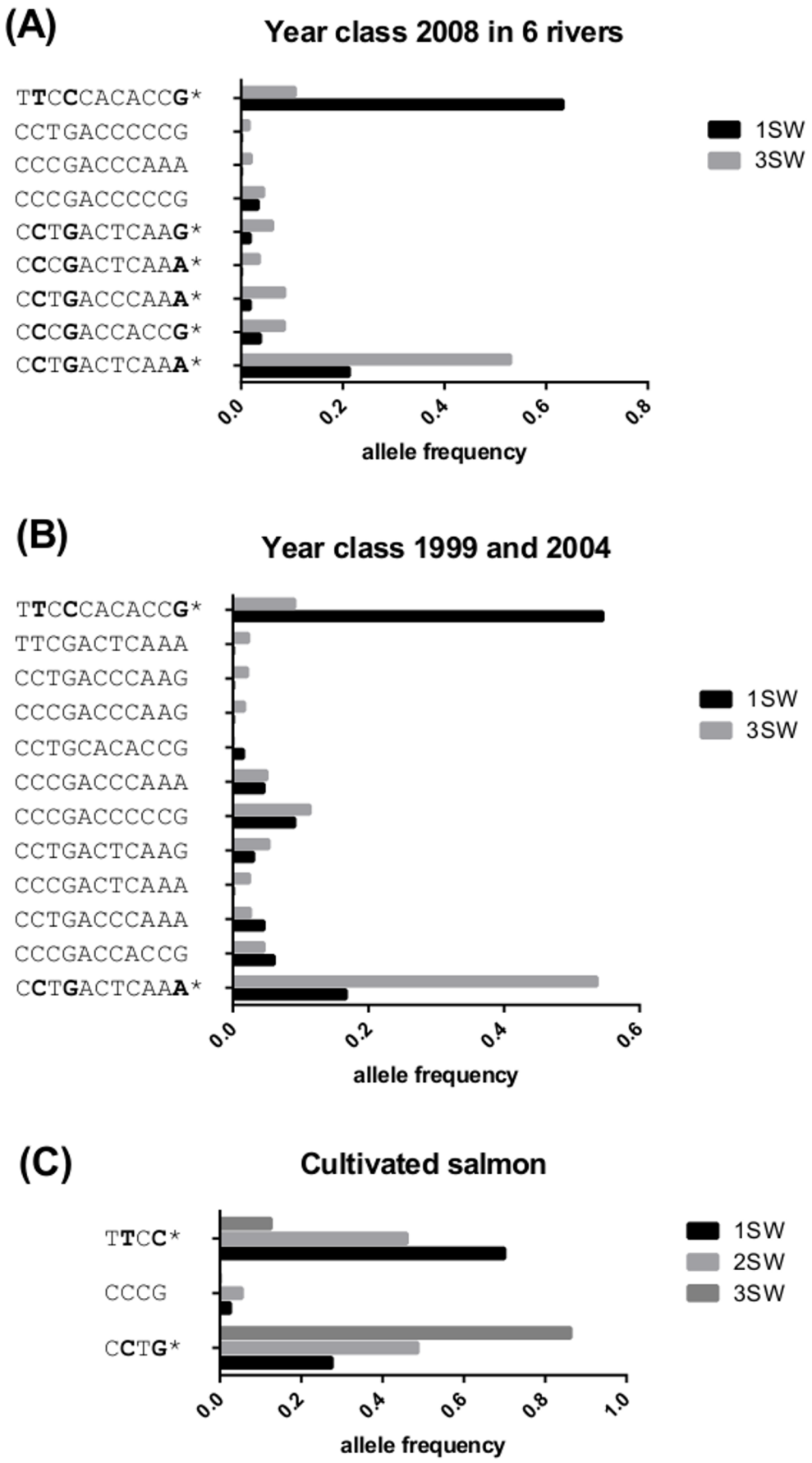

(A)
−log10P
30
20
10
0
ssa01
ssa02
ssa03
ssa04
ssa05
ssa06
ssa07
ssa08
ssa09
ssa10
ssa11
ssa12
ssa13
ssa14
ssa15
ssa16
ssa17
ssa18
ssa19
ssa20
ssa21
ssa22
ssa23
ssa24
ssa25
ssa26
ssa27
ssa28
ssa29
(B)
−log10P
Coverage
CHMP2B
VGLL3
AKAP11
28550
28607
28665
28722
28780
Position (kb)

Figure 3


(A)
Year class 2008 in 6 rivers
TTCCCACACCG*
CCTGACCCCCG
CCCGACCCAAA
CCCGACCCCCG
CCTGACTCAAG*
CCCGACTCAAA*
CCTGACCCAAA*
CCCGACCACCG*
CCTGACTCAAA*
0.0
0.2
0.4
0.6
0.8
allele frequency
1SW
3SW
(B)
Year class 1999 and 2004
TTCCCACACCG*
TTCGACTCAAA
CCTGACCCAAG
CCCGACCCAAG
CCTGCACACCG
CCCGACCCAAA
CCCGACCCCCG
CCTGACTCAAG
CCCGACTCAAA
CCTGACCCAAA
CCCGACCACCG
CCTGACTCAAA*
0.0
0.2
0.4
0.6
allele frequency
1SW
3SW
(C)
Cultivated salmon
TTCC*
CCCG
CCTG*
0.0
0.2
0.4
0.6
0.8
1.0
allele frequency
1SW
2SW
3SW