



# Atlantic salmon (*Salmo salar* L.) genetics in the 21st century: taking leaps forward in aquaculture and biological understanding

R. D. Houston\* and D. J. Macqueen\*<sup>†</sup>

\*The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Midlothian EH25 9RG, UK. <sup>†</sup>School of Biological Sciences, University of Aberdeen, Aberdeen AB24 2TZ, UK.

## Summary

Atlantic salmon (*Salmo salar* L.) is among the most iconic and economically important fish species and was the first member of Salmonidae to have a high-quality reference genome assembly published. Advances in genomics have become increasingly central to the genetic improvement of farmed Atlantic salmon as well as conservation of wild salmon stocks. The salmon genome has also been pivotal in shaping our understanding of the evolutionary and functional consequences arising from an ancestral whole-genome duplication event characterising all Salmonidae members. Here, we provide a review of the current status of Atlantic salmon genetics and genomics, focussed on progress made from genome-wide research aimed at improving aquaculture production and enhancing understanding of salmonid ecology, physiology and evolution. We present our views on the future direction of salmon genomics, including the role of emerging technologies (e.g. genome editing) in elucidating genetic features that underpin functional variation in traits of commercial and evolutionary importance.

**Keywords** evolutionary genetics, genome editing, genomics, salmonid, selective breeding, sequencing technology, whole genome duplication

## Introduction

Atlantic salmon (*Salmo salar* L.) (hereafter 'salmon') is among the most famous and economically important fish species globally. In addition to being a prized sport fish with a fascinating life cycle, major ecological importance and high conservation value, salmon is a nutritious food product farmed for human consumption. Salmon aquaculture is worth approximately 8.5 billion GBP (~9.7 billion Euro) annually (FAO 2017) and contributes significantly to food, economic and employment security in many nations, especially Norway, Chile, Canada and the United Kingdom. Genetics and genomics have key roles in the current and future management of farmed and wild salmon stocks. Consequently, huge research investment, often supported by industry, is driving the field forward at a remarkable

pace. Fuelled by the recent publication of a high-quality reference genome for salmon (Lien *et al.* 2016) and related species from the Salmonidae family (e.g. Christensen *et al.* 2018a,b; Narum *et al.* 2018), there is currently a major interest in applying genome-wide tools to enhance selective breeding for aquaculture and improve knowledge of genome biology, physiology, ecology and evolution (Macqueen *et al.* 2017). The goal of this article is to provide an overview of Atlantic salmon and its key genetic features before reviewing the current and future research landscape in genetics and genomics.

## Evolutionary history and key genetic features

### Phylogeny and macroevolution

Atlantic salmon is one of two recognized *Salmo* species, the other being brown trout (*Salmo trutta*). *Salmo* sits within the Salmoninae subfamily, which also includes *Oncorhynchus* (Pacific salmon), *Salvelinus* (charrs), *Parahucho* (Sakhalin taimen), *Hucho* (huchens/taimens) and *Brachymystax* (lenoks). The position of *Salmo* within Salmoninae has been long-debated, but a recent study used genome-wide markers to affiliate *Salmo* and *Parahucho* as a sister group (Lecaudey

### Address for correspondence

R. D. Houston and D. J. Macqueen, The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Midlothian EH25 9RG, UK.  
E-mails: ross.houston@roslin.ed.ac.uk;  
daniel.macqueen@roslin.ed.ac.uk

Accepted for publication 15 October 2018

*et al.* 2018). Evidently, *Salmo* and *Parahucho* split approximately 22 million years ago, whereas their ancestor diverged from a group containing *Oncorhynchus* and *Salvelinus* approximately 10 million years earlier (Lecaudey *et al.* 2018). The clade that includes *Salmo*, *Parahucho*, *Oncorhynchus* and *Salvelinus* shares a capacity for anadromy (Alexandrou *et al.* 2013)—the ability to migrate into seawater after spending early life in streams and rivers. This trait likely evolved after divergence from the *Hucho-Brachymystax* lineage, for which the full life-cycle is spent within freshwater, a feature present in more distantly related lineages including graylings (*Thymallinae*), along with *Eosalmo* (extinct), the earliest known salmonid in the fossil record (Wilson & Li 1999). Consequently, the famous ability of salmon to transform their juvenile physiology, migrate to and exploit oceanic feeding grounds—sometimes thousands of miles from their birthplace—has ancient evolutionary origins. This life-history strategy was also proposed to have driven species diversification (Macqueen & Johnston 2014), and its evolution may be linked to genetic features distinguishing salmonids from other fishes (see the later sub-section ‘The “Ss4R” WGD event’).

#### Intra-specific diversity and microevolution

Salmon have a broad distribution in the Northern hemisphere and diverged into North American and European lineages at least 0.6–0.7 million years ago (King *et al.* 2007), with an even deeper divergence 1.56–1.76 million years ago suggested recently (Rougemont & Bernatchez 2018). These lineages are characterized by notable differences in chromosomal organization (Hartley 1987) and mating incompatibilities (Cauwelier *et al.* 2012), so they can be reasonably classified as sub-species (King *et al.* 2007; Rougemont & Bernatchez 2018). Substantial structure exists within each lineage, including three differentiated European clades and several North American groups (Bourret *et al.* 2013; Moore *et al.* 2014). There is evidence of substantial recent gene flow between and within these major lineages and sub-populations (Rougemont & Bernatchez 2018).

Salmon encounter diverse environments across their range. Coupled with a strong tendency to reproduce in the streams of their birth, populations show significant genetic differentiation and adaptation at small spatial scales (Garcia de Leaniz *et al.* 2007; Fraser *et al.* 2011). Although most salmon maintain anadromous life-history strategies, populations on both continents have become trapped in post-glacial freshwater systems. These ‘landlocked’ fish have experienced rapid genetic and phenotypic differentiation owing to drift coupled with selection on a distinct set of traits, for example a loss/reduction in selection on the systems that prepare anadromous populations for seawater entry (e.g. Nilsen *et al.* 2008). Although rapid phenotypic change and plasticity in response to new environments is a

highly recognized feature of many salmonids (e.g. Klemetsen *et al.* 2003), the mechanisms involved remain poorly understood. However, genomic plasticity provided by an ancestral whole-genome duplication (WGD) event (see the next subsection) has been linked to the salmon’s high capacity for adaptation (Kjærner-Semb *et al.* 2016).

#### The ‘Ss4R’ WGD event

The salmon’s nuclear material sums to approximately 3 billion bases and contains 50–60% repeat content (Lien *et al.* 2016). Although this closely matches the human genome, the basis for these features differs markedly. The ancestor to salmonids experienced an autotetraploidization WGD (Allendorf & Thorgaard 1984) 88–103 million years ago (Macqueen & Johnston 2014), an event coined ‘Ss4R’ (salmonid-specific 4th round of WGD) to account for additional rounds of WGD at the base of teleost fishes (‘teleost-specific’ Ts3R) (Glasauer & Neuhauss 2014) and the vertebrate lineage (1R/2R) (Dehal & Boore 2005).

Autotetraploidization involves spontaneous doubling of all chromosomes, distinct from the other major WGD class, allotetraploidization, which involves hybridisation of distinct species. After the latter, the two genomes within a cell are usually different enough to segregate into two sets of bivalents during meiosis, which rescues pairing incompatibilities among hybridizing species prior to WGD (Otto 2007). Conversely, autotetraploidization leads to four chromosome sets that initially pair randomly during meiosis after WGD; preferential bivalent pairing must be re-established before duplicated genes created by WGD can diverge beyond an allelic state (Martin & Holland 2014; Lien *et al.* 2016; Robertson *et al.* 2017). This represents one of the key outcomes of rediploidization, the process whereby a tetraploid genome returns to diploidy. The re-establishment of bivalent pairing in salmonids involved large structural reorganizations (e.g. inversions) associated with bursts of transposable element proliferation, suggesting that Ss4R resulted in relaxed ‘policing’ of deleterious transposable element propagation (Lien *et al.* 2016). Remarkably, this process was delayed by tens of millions of years in around a quarter of the genome, which has had a pervasive impact on lineage-specific genome evolution and adaptive potential (Robertson *et al.* 2017).

A significant percentage of the genome (10–20%) in salmonids has yet to complete the rediploidization process and maintains tetraploid genetic characteristics including potential for tetrasomic inheritance (Allendorf *et al.* 2015). Although such regions are long-recognized (e.g. Allendorf & Thorgaard 1984), their significance is now becoming better appreciated through application of modern genomics in wild populations (e.g. Waples *et al.* 2016). However, the role played by such regions in influencing commercially relevant trait variation remains unknown, because they are preferentially filtered and removed during genomic analysis

(e.g. Limborg *et al.* 2016) and remain challenging to incorporate into standard experimental designs. This rediploidization process is also thought to be the primary cause of a major disparity in recombination rate between males and females; males have very limited recombination over large parts of the genome, but with recombination 'hotspots' near the telomeres, which tend to be regions showing residual tetraploidy (Allendorf *et al.* 2015). Finally, a key outcome of rediploidization is the retention of at least half of all salmonid genes in duplicated pairs from Ss4R (Berthelot *et al.* 2014; Lien *et al.* 2016); for some gene families, the retention rate of Ss4R gene duplicates is 100% (e.g. Garcia de la Serrana & Macqueen 2018). Additionally, one in five salmon genes belongs to a pair of more ancient gene duplicates retained from Ts3R, leading to highly expanded gene families compared to most non-teleost vertebrates.

Though fascinating, the complexity of salmonid genomes brings challenges, firstly by adding uncertainty to the quality of reference genome sequences in regions where rediploidization was delayed. In such regions, distinguishing duplicated regions during bioinformatic sequence assembly remains challenging, particularly when using short-read data (see the later sub-section 'Improvements in genome assemblies'). Moreover, the global presence of duplicated regions can reduce confidence when mapping short-read sequence data to reference genomes, with potential impacts on RNASeq, SNP calling and population genetic analyses (see the section 'Growing toolbox for genome-wide investigations'). Interpreting functional signals, especially gene expression, in the face of gene family expansions is likewise challenging, as salmon often retain multiple co-orthologues of single genes found in model taxa like zebrafish or human. Such duplicated copies are often differentially expressed (e.g. Lien *et al.* 2016; Robertson *et al.* 2017) and can have divergent protein sequences, making it important to interpret their functions and expression as a 'sum of parts' when establishing the roles of candidate genetic systems under investigation.

## Domestication and selective breeding

Commercial-scale salmon farming began in Norway in the 1960s, expedited by trials in the early 1970s that demonstrated the huge potential of family-based breeding programmes (Gjedrem 2012). In these trials, gametes from salmon taken from approximately 40 Norwegian rivers were collected and formed the basis of robust estimations of genetic parameters and the first commercial breeding programme (Gjøen & Bentsen 1997). Other similar breeding programme initiatives were instigated, including the establishment of the Mowi, Rauma, Jakta and Bolaks strains in Norway (Glover *et al.* 2017). Together, following various crossing and international export events, these strains underpin the vast majority of global salmon aquaculture. The consolidation of breeding companies over recent years

has resulted in very few but large international players that supply eggs to all the major salmon-producing countries. These include AquaGen (Norway), Benchmark (UK; owners of both SalmoBreed and StofnFiskr), Hendrix Genetics (Netherlands; owners of Landcatch) and AquaInnovo (Chile), with further consolidation underway via a joint venture between Benchmark and AquaInnovo.

The Norwegian family-based breeding programmes successfully focussed on increasing growth rate, with estimates of genetic gain per generation of approximately 15% (Gjedrem & Rye 2016). This is vastly superior to terrestrial livestock, albeit the generation interval of salmon is relatively long, typically 3–4 years. This high level of genetic gain may be due in part to the selection intensity associated with the high fecundity of salmon (several thousand offspring per female) and in part to a very recent domestication history, providing high levels of genetic variability influencing traits of importance for farming. In contrast, terrestrial livestock species have been domesticated and selected (directly or indirectly) for favourable traits for approximately 10 000 years (Mignon-Grasteau *et al.* 2005).

Subsequently, from the 1990s onwards, as breeding programmes became more advanced and needs of producers changed, the breeding goals broadened to include traits such as disease resistance, rate of sexual maturation and fillet characteristics (Gjedrem & Rye 2016). The typical structure of a breeding programme developed to take advantage of the amenable features of salmon biology, in particular external fertilisation and high fecundity. As a result, it was possible to maintain breeding nuclei of approximately 100–300 families, retaining a proportion of juveniles from each family within the nucleus while setting aside their full siblings for production and performance testing. This process is known as 'sib testing' (short for sibling testing) and enables recording of traits impossible or impractical to measure directly on selection candidates in the nucleus (e.g. resistance to specific pathogens or invasive fillet traits). In addition, technology advances began enabling genetic markers to be applied to capitalise on the within-family component of genetic variation in addition to the between-family component. The first example of this was the extensive use of marker-assisted selection for favourable alleles at a major QTL explaining the vast majority of variation in host resistance to infectious pancreatic necrosis virus (IPNV) (Houston *et al.* 2008, 2010; Moen *et al.* 2009; Gheyas *et al.* 2010). The result was a sustained decrease in the incidence of IPN outbreaks to near zero and widespread recognition of the potential of (molecular) genetics in selective breeding to tackle infectious disease (Norris 2017). Subsequent studies have demonstrated that most other traits of importance for salmon production are heritable but highly polygenic (for reviews, see Yáñez *et al.* 2014; Houston 2017), and

therefore genomic selection (GS) is considered the state of the art for application of genomics to genetic improvement (see the sub-section 'Mapping QTL and genomic selection').

Due to the outcomes of domestication and selective breeding, there are both genetic and phenotypic differences between wild and farmed salmon populations. Escapees from salmon farms are thought to have resulted in significant introgression into wild stocks, which may impact life-history traits and the subsequent fitness of natural populations (e.g. Glover *et al.* 2017). As such, approaches to prevent interbreeding of wild and farmed fish are being developed, including mass generation of triploids (Benfey 2001) and gene editing to induce sterility in farmed stocks (see the sub-section 'Genome editing for understanding and improving traits'). Comparisons of farmed and wild stocks are useful for detecting genetic signatures of domestication. Salmon present an interesting model due to the passage of relatively few generations since organised farming began, perhaps around 13 generations. Comparisons between the genomes of farmed and wild populations have revealed selection signals related to various domestication-related traits, affecting genes associated with growth, early sexual maturation and immune response (Gutierrez *et al.* 2016; Liu *et al.* 2017b).

### Growing toolbox for genome-wide investigations

High-throughput sequencing has transformed salmon genetics, in particular the ease of generating genome-wide genetic marker datasets. A major step forward came with the arrival of restriction site-associated DNA sequencing (RAD-seq; Baird *et al.* 2008) and subsequent variations. The cost-effective discovery and concurrent genotyping of multiple, multiplexed samples in a single Illumina sequencing lane has been widely applied in many salmonid species (reviewed by Robledo *et al.* 2017). RAD-seq and similar genotyping-by-sequencing techniques were applied in salmon even before the availability of a reference genome and have been used for QTL mapping, linkage mapping, genome-wide association (GWA) studies, population genetics and SNP discovery for creating genotyping tools, including SNP arrays (Robledo *et al.* 2017). Subsequently, high-density SNP arrays were published for salmon (Houston *et al.* 2014; Yáñez *et al.* 2016), in addition to multiple unpublished custom arrays used in research and development projects by individual breeding companies. These arrays have enabled many high-resolution genetic association and population genetic studies (e.g. see the next section 'Modern applications of genomics'), in addition to the first tests of GS in salmon breeding programmes (Ødegård *et al.* 2014; Tsai *et al.* 2015b). Whole-genome (re)sequencing (WGS) methods have also been applied for

variant detection and calling in salmon but remain expensive, and population-scale genotyping by WGS requires further research (see the sub-section 'Moving towards WGS for population analysis'). Genetic marker resources have been utilized to develop linkage maps of the salmon genome, including high-density SNP linkage maps created using SNP arrays (Lien *et al.* 2011, 2016; Tsai *et al.* 2016a) and RAD-seq (Gonen *et al.* 2014).

The landmark publication of the salmon genome (Lien *et al.* 2016) provided a reference assembly that advanced possibilities for high-resolution genomic analyses. Genome-wide gene expression profiling has traditionally been performed in salmon by microarrays, and these reliable tools are still widely applied (e.g. Król *et al.* 2016; Robledo *et al.* 2016; Vera *et al.* 2017). However, RNASeq performed against the reference genome is now routinely used for functional genomic investigations focussed on evolution (e.g. He *et al.* 2017; Robertson *et al.* 2017), aquaculture (e.g. Robledo *et al.* 2018a) and physiology (e.g. Gillard *et al.* 2018). Mapping against a reference genome, compared to a transcriptome assembly, also has the benefit that highly similar duplicated regions can be distinguished in the analysis, assuming such regions have been correctly assembled. Conversely, the assembly of transcript sequence data in species with recent WGD is prone to the collapse of contigs and generation of chimeric contigs (e.g. Krasileva *et al.* 2013), which makes RNAseq analyses and interpretation more challenging. Given the wide range of approaches available for RNAseq and other mapping-based genomic analyses, the field would benefit from a move towards standardizing pipelines and converging on best-practices to increase comparability across studies. This is one of the goals of the recently established 'Functional Annotation of All Salmonid Genomes' (FAASG) initiative (Macqueen *et al.* 2017), described fully in the sub-section 'Improved annotation and understanding of genome function and regulation'. As increasing quantities of genetic and functional genomic data are generated, a portal for interrogating and visualising these data is necessary for widespread community uptake beyond the standard public repositories, and the genome browser Salmobase.org (Samy *et al.* 2017) is currently serving this purpose.

High-quality annotations of protein products across a genome also enable investigations applying high-throughput proteomic approaches that couple liquid chromatography with mass spectrometry to identify huge numbers of putative peptides; data that can be used for quantitative comparisons of protein abundance and modifications (e.g. Breker & Schuldiner 2014). This represents a powerful new tool in the functional genomics armoury for salmonids and is consequently being up-taken rapidly for investigations of fish physiology and health (e.g. Liu *et al.* 2017a; Causey *et al.* 2018a,b; Kumar *et al.* 2018; Nuez-Ortín *et al.* 2018).



## Modern applications of genomics: sequence to consequence

### Mapping QTL and genomic selection

The genomic toolbox developed for salmon has enabled a wide range of applications in aquaculture breeding and genetics. The case of IPN resistance is one of the most famous examples of a major QTL impacting an economically important trait in a farmed animal species (Houston *et al.* 2008, 2010; Moen *et al.* 2009; Gheyas *et al.* 2010). High-throughput sequencing approaches have subsequently allowed development of SNP-based genetic tests to predict IPN resistance of salmon without the need for regular disease challenge experiments (Houston *et al.* 2012; Moen *et al.* 2015). Furthermore, functional studies have been undertaken to highlight the marked differences in gene expression response to infection between resistant and susceptible salmon fry (Robledo *et al.* 2016) and to suggest that epithelial cadherin may be part of the mechanism underlying the QTL (Moen *et al.* 2015). However, in subsequent QTL scans, including GWA studies, there has been little evidence for additional major QTL affecting disease resistance or any other economically important trait (except for maturation; described in the next sub-section 'Population genetics to discover the basis of life-history traits'). For example, significant QTL of relatively minor effect have been identified for salinity tolerance (Norman *et al.* 2012), body weight (Houston *et al.* 2009; Gutierrez *et al.* 2012, Tsai *et al.* 2015a; Yoshida *et al.* 2017) and resistance to several diseases and pathogens, namely pancreatic disease (Gonen *et al.* 2015), salmon rickettsial syndrome (Correa *et al.* 2015), amoebic gill disease (Robledo *et al.* 2018b) and sea lice (Correa *et al.* 2016; Tsai *et al.* 2016b). The percentage of genetic variation (heritability) explained by the identified QTL in all these studies was low (between 2 and 20%, compared to 80–100% for IPN resistance), and therefore marker-assisted selection is unlikely to be a fruitful strategy for improving these target traits.

Genomic selection was first described by Meuwissen *et al.* (2001) and involves the use of genome-wide genetic marker data to predict breeding values for selection candidates. The premise of GS is that marker effects are estimated in a 'training' population that has been measured for both phenotypes and genotypes, and the model developed is used to predict breeding values for individuals with genotype information only. GS has transformed the livestock breeding industry, generating substantially faster genetic gain for key economic traits compared to the traditional pedigree-based approach (Meuwissen *et al.* 2013). Applications of GS in aquaculture began with the development of the first high-density SNP arrays, containing hundreds of thousands of SNPs (Houston *et al.* 2014; Ødegård *et al.* 2014; Yáñez *et al.*

2016). The focus of GS in salmon has been on disease resistance due to its economic importance and the practical impossibility of trait measurement on the selection candidates themselves. In all published GS studies in salmon, the results have shown higher prediction accuracy of breeding values than with pedigree information alone (Ødegård *et al.* 2014; Tsai *et al.* 2015b, 2016a,b; Bangera *et al.* 2017; Robledo *et al.* 2018b). A major downside to GS is that high-density genotyping in large numbers of individuals can be prohibitively expensive. Approaches to reduce genotyping costs, such as the use of low-density marker panels, including with genotype imputation, have shown promising results (Tsai *et al.* 2017; Yoshida *et al.* 2018). GS has been shown to be effective in salmon for which the training and test populations are closely related (such as in a typical sib-testing scheme), but the ability to predict breeding values in animals more distantly related to the training population is rather limited (Tsai *et al.* 2016b) and may require new approaches including increased focus on potential functional variants, such as those identified under FAASG initiative studies (see the sub-section 'Improved annotation and understanding of genome function and regulation').

### Population genetics to discover the basis of life-history traits

The latest genomic tools have also been used to reveal the genomic basis of salmon traits with significance for adaptation in natural environments. As a prime example, which also has significance for aquaculture, two closely timed publications identified a major locus (harbouring the *vgll3* gene) explaining a large proportion (approximately one third) of individual variation in the age that salmon undergo maturation, which is under divergent selection in males and females. Both investigations used GWA methods to locate the same genomic region, either using WGS following pooling of individuals from multiple populations (Ayllon *et al.* 2015) or by applying a high-density SNP array on a large number of populations and subsequent WGS to interrogate potential functional variants (Barson *et al.* 2015). The latter study provides a classic example of how sexual conflict—when selection acts in different directions in males and females—can be partly resolved by balancing selection on a single autosomal gene (Mank 2017). Follow-up studies are providing insights into the mechanisms by which *vgll3* is operating in reproductive systems, providing evidence for distinct regulation between sexes (Kjærner-Semb *et al.* 2018). Another recent study applied WGS using pooling of individuals to identify highly differentiated regions of the genome that harboured genes with important immune functions, comparing Northern and Southern populations of salmon in Norway (Kjærner-Semb *et al.* 2016). Similarly, the same high-density SNP

array applied by Barson *et al.* (2015) was recently applied to identify candidate genomic regions and genes under divergent selective pressures in sub-populations of salmon inhabiting the Teno River in Finland (Pritchard *et al.* 2018).

## State of the art and perspectives

### Improvements in genome assemblies

As reviewed in the section 'Growing toolbox for genome-wide investigations', it is an exciting time for genome-enabled biology in salmon. This sentiment extends to other Salmonidae members, for which high-quality and draft genomes have been published, including for rainbow trout (*Oncorhynchus mykiss*; Berthelot *et al.* 2014), Chinook salmon (*Oncorhynchus tshawytscha*; Christensen *et al.* 2018a; Narum *et al.* 2018), European grayling (*Thymallus thymallus*; Varadharajan *et al.* 2018) and Arctic charr (*Salvelinus alpinus*; Christensen *et al.* 2018b), or are available yet currently unpublished, including for coho salmon (*Oncorhynchus kisutch*; NCBI accession no. GCA\_002021735) and Danube salmon (*Hucho hucho*; NCBI accession no. GCA\_003317085). Several of these assemblies have been anchored to chromosomes and are annotated to a high standard (Macqueen *et al.* 2017). Such resources provide a powerful framework that, when coupled with our understanding of phylogeny (e.g. Macqueen & Johnston 2014; Lecaudey *et al.* 2018), will enable salmon researchers to harness comparative approaches to reconstruct the evolutionary origins of traits of commercial and evolutionary relevance—interesting in the context of Ss4R and the diversity of ecological adaptations present among salmonid lineages (Robertson *et al.* 2017; Varadharajan *et al.* 2018). However, despite such substantial recent progress, improvements to existing genome assemblies and annotations will be vital to more fully exploit genomic information in salmon and related species.

Central to improvement of genome assemblies are technological advancements, which salmonid researchers have been quick to capitalise on, for example by incorporating long-read data generated on Pacific Biosciences (PACBIO) platforms (e.g. Lien *et al.* 2016; Christensen *et al.* 2018a,b). This trend will continue, ensuring improvement in the annotation of poorly represented regions in salmon genomes, notably repetitive and tetrasomic regions. Although PACBIO and some classes of short-read data (e.g. mate-pair libraries) provide long-range information that facilitates resolution of complex regions in genome assemblies, emerging approaches hold greater advantages in the same respect. Nanopore sequencing on Oxford Nanopore Technologies (ONT) platforms, including the portable MinION sequencer, generates ultra-long reads that reach lengths beyond PACBIO's capabilities and is being successfully applied to assemble large and complex eukaryotic genomes (e.g. Jain *et al.* 2018; Michael *et al.* 2018). The

ONT approach is currently being developed by several salmon research groups and is considered to hold great promise going forward.

Additional tools that provide the long-range information necessary to improve reference genome assemblies include high-throughput chromatin conformation capture (Hi-C), which generates genome-wide data on chromatin interactions that can be applied to scaffold existing assemblies to a high level (e.g. Burton *et al.* 2013; Putnam *et al.* 2016), an approach applied in an improved assembly of the rainbow trout genome (accession no. GCA\_002163495; unpublished). Optical mapping similarly generates very long-range genomic information that can be used to improve complex genome assemblies (reviewed by Howe & Wood, 2015), though as far as we are understand, is yet to be applied in salmonids in published studies. A highly promising tool for salmon research is 'linked-read' sequencing, using the 10× genomics microfluidic platform to partition fragmented genomic DNA into large molecules that are subsequently sequenced as short reads that retain a unique barcode matching the original fragment (Zheng *et al.* 2016). This approach can be used to generate assemblies that distinguish both chromosome sets (i.e. a 'diploid assembly') (Weisenfeld *et al.* 2017). Generation of such long-range haplotype information would have major applications in salmon population genomics and in theory could be used to distinguish tetrasomic regions directly during sequence assembly. It is also crucial to note that the merging of data gathered across the range of established and emerging sequencing platforms is essential for fully exploiting the unique advantages of different approaches while offsetting their varying limitations (e.g. using highly accurate short-read data to clean up ultra-long sequence data that currently suffer from high error rates).

### Improved annotation and understanding of genome function and regulation

A further step advance in understanding of how variation in the blueprint of the salmon genome leads to trait variation will require improved knowledge of genome function and the complex regulation of gene expression. Following in the footsteps of the FAANG initiative for terrestrial farmed animals (Andersson *et al.* 2015), the FAASG initiative was established to improve knowledge of genome function for salmonids (Macqueen *et al.* 2017). FAASG is a community-led initiative that will harness modern experimental molecular biology and sequencing technologies to identify and characterise functional elements in the genome. This will include studies of polymorphic variation within species, fixed variation across species, gene expression phenotypes covering multiple RNA classes and their variants, epigenetics and gene expression regulation, along with protein-level variation.

The epigenetic molecular component of phenotypic variability in salmonids is relatively poorly understood but holds promise for translational research relevant to stock enhancement in aquaculture (for recent reviews, see Gavery & Roberts 2017; Best *et al.* 2018). In this respect, FAASG aims to exploit a range of well-established technologies that enable profiling of DNA methylation, repressive and permissive histone modifications, chromatin accessibility and higher chromatin structure (Macqueen *et al.* 2017). The salmonid research community is already applying several of these approaches. For example, a recent study integrated transcriptomics with genome-wide epigenetic analyses to demonstrate remodelling of methylation status due to stress (Moghadam *et al.* 2017). A role for global changes in methylation in shaping phenotypic variation in response to the environment was also proposed recently with respect to the reduced fitness observed in hatchery-reared salmon used to re-stock wild populations (Le Luyer *et al.* 2017). Moreover, the possible role of histone modifications for the thermal dependence of salmon immune responses was recently reported (Boltana *et al.* 2018). In addition, an increased understanding of variation in the salmon microbiome will be important for improving our understanding of its role in complex traits, and this is of increasing interest for salmon biologists (e.g. Gajardo *et al.* 2016; Dehler *et al.* 2017; Uren Webster *et al.* 2018). Microbiome composition is almost certain to influence genome functional and epigenetic responses, with resulting impacts on phenotype, but there remains much left for fish biologists to learn in this area, with many promising avenues for genomic investigations (e.g. Llewellyn *et al.* 2014; Ghanbari *et al.* 2015).

A unique feature of the FAASG initiative is that functional annotation will facilitate an improved understanding of genome functional evolution after the Ss4R event. At the population scale, an improved understanding of genome function will allow prioritisation of polymorphisms that may be expected to have direct effects on traits of interest, rather than simply as genetic markers. Further, it will enable shortlisting of candidate variants for use with gene-editing technologies to demonstrate function and potentially improve traits for aquaculture (see the subsection 'Genome editing for understanding and improving traits').

#### Moving towards WGS for population analysis

Population-scale WGS has the potential to significantly enhance understanding of the genetic basis of traits of evolutionary and economic interest. Although genotyping-by-sequencing techniques, such as RAD-seq, have been widely applied (Andrews *et al.* 2016; Robledo *et al.* 2017), the ongoing reduction in sequencing and high-power computing costs is expected to make WGS routine in the future. Studies using pooled WGS have been applied to investigate signatures of selection (Kjærner-Semb *et al.*

2016) and to map a major QTL affecting maturation (Ayllon *et al.* 2015). However, individual-level population-scale WGS can offer insights including the role of different types of polymorphic variation in trait architecture (e.g. structural variants including copy number variation, inversions, etc.) and would enable the study of rare and *de novo* variants that are unlikely to be detected using SNP arrays due to ascertainment bias. To be affordable in the short term for population datasets, WGS can be performed at low individual coverage. This raises issues with potentially erroneous variant calling due to sequence errors and/or heterozygous sites being called as homozygous due to sequencing of just one allele (Bilton *et al.* 2018). Harnessing pedigree information together with imputation approaches within a breeding programme may be an effective route for improving the quality of low-coverage WGS data and may have downstream benefits for genomic prediction accuracy (Hickey 2013). When combined with GWA approaches and the functional annotation described above, WGS can provide the means to discover and characterise candidate causative variants within QTL regions that can be selected for functional testing.

#### Genome editing for understanding and improving traits

Genome editing technologies allow targeted changes to the genomic DNA at a specific location, and engineered CRISPR/Cas9 systems (Cong *et al.* 2013; Mali *et al.* 2013) are widely applied for this purpose. The Cas9 enzyme makes a double-stranded cut at a specific target site enabled by the guide RNAs. The resulting DNA changes are the result of two major categories of DNA repair mechanisms. The first of these is non-homologous end joining (NHEJ), whereby the repair mechanism does not require a homologous template and will result in small insertions or deletions at the cut site that can result in loss-of-function mutations. The second is homology-directed repair (HDR), whereby a DNA template is provided that is similar to the flanking sequence of the cut site (but may contain a user-targeted change in sequence), and the cell uses the template to repair at the cut. The successful use of CRISPR/Cas9 with NHEJ to generate *slc45a2* knockout salmon in the F0 generation via microinjection into one-cell stage embryos demonstrated the efficacy of the technology in salmon (Edvardsen *et al.* 2014). Subsequent studies have successfully applied CRISPR/Cas9 to generate sterile salmon via ablation of germ cells caused by *dnd* knockout (Wargelius *et al.* 2016). In addition to these *in vivo* successes, CRISPR/Cas9 has been successfully applied for gene knockout in a salmonid cell line (CHSE-214, Dehler *et al.* 2016). Evidence for targeted changes made via incorporation of a template DNA using HDR has not yet been published for salmon, though such work is currently underway in several groups. 'Base editing' is another emerging gene editing approach that can make specific targeted changes in the genome without the need to

cut the genomic DNA or utilize a template DNA (Komor *et al.* 2016) and has been successfully applied in zebrafish (Zhang *et al.* 2017) but is yet to be trialled in salmon as far as we are aware.

There are a number of potential future applications of genome editing for increasing understanding of salmon biology and improving traits of importance for salmon production and welfare. Genome-wide screening approaches, including the use of the genome-scale CRISPR knockout (GeCKO) technique (Shalem *et al.* 2014), may facilitate identification of genes involved in traits of importance, particularly traits that can be measured in cell cultures (e.g. resistance to viral disease). GeCKO involves lentiviral delivery of a library of tens of thousands of unique guide RNAs into cell cultures for genome-wide gene knockout followed by negative or positive selection screening (Shalem *et al.* 2014). There are technical hurdles to overcome before GeCKO screens could be applied in salmon, in particular relating to delivery of guide RNAs, as lentiviruses are not considered an effective delivery method in salmon cells. CRISPR/Cas9 is also likely to be used to test hypotheses relating to causative variants underlying QTL. Ideally, HDR or base-editing approaches could be applied to 'swap' one version of the allele at the candidate variant for the alternate version before assessing the impact on the trait of interest. For all editing approaches, it is important to consider, and if possible exclude, potential off-target effects, which remains a contentious issue in medical research (Nutter *et al.* 2018). However, there are several exciting potential applications of genome editing in salmon breeding programmes (subject to public and regulatory acceptance; see the following subsection) which could include (i) fixing of favourable alleles at QTL affecting traits of economic interest; (ii) rapid 'introgression' of favourable alleles from other populations, strains or species into a salmon breeding population; and (iii) creation of 'de novo' alleles based on knowledge of the biology of the trait in question. For the latter application, an example from terrestrial livestock is the removal of an exon of the *CD163* gene in pigs, which results in complete resistance to the porcine reproductive and respiratory syndrome virus (PRRSV) (Burkard *et al.* 2017).

### Regulatory and public perception landscape

Finally, it is important to briefly consider on-going changes surrounding the regulation and uptake of genetically modified (GM) or gene-edited (GE) salmon for production in aquaculture. These methods have the potential to rapidly introduce favourable traits (as described above) and to provide solutions to major challenges faced by the salmon aquaculture sector. However, there clearly are regulatory and perception issues to consider, and these include the definition of what constitutes GM and the extent to which gene editing should be considered separately and/or split into different categories according to the nature of the induced

change to the genome. These decisions will need to involve a wide variety of stakeholders, including in the aquaculture and retail industries, policymakers, consumers and other members of the public. At one end of the scale, it is now possible to generate GE animals with single base changes in the genome that are already segregating in wild populations. At the other end of the scale are more radical changes in the genome that are absent (or perhaps rare) in nature (e.g. the PRRSV example in pigs). Clearly, there are many scenarios in between, and the challenge is to find a balance that allows the revolutionary potential of gene editing to be realized in an objective (i.e. scientifically informed) manner with appropriate regulatory frameworks. Although arguments have been presented that gene editing for alleles that occur naturally in agricultural populations should not be considered gene modification even under strict legal frameworks (Custers 2017), the recent ruling by the European Court of Justice that GE crops should be considered GM organisms is a major setback (Callaway 2018). However, a landmark was set recently by the approval of a GM salmon strain as fit for human consumption by the US Food and Drug Administration and the Canadian Food Inspection Agency (and for sale by the latter) after a long period of regulatory limbo (Waltz 2017). The AquaAdvantage (AquaBounty Technologies) strain shows enhanced growth due to the integration of a growth hormone (GH) gene from Chinook salmon linked to a promoter from another fish species that drives high GH expression. Ultimately, research and development relating to potential uses of gene modification and gene editing in aquaculture will continue to develop rapidly and should do so in parallel with an extensive dialogue between the various stakeholder groups described above to help establish a knowledge-driven regulatory framework for future applications.

### Acknowledgements

RDH acknowledges funding for Atlantic salmon genetics research from the Biotechnology and Biological Sciences Research Council (grants BB/N024044/1, BB/R008612/1, BB/S004343/1, BB/S004432/1), including Institute Strategic Programme Grants (BBS/E/D/20002172, BBS/E/D/30002275) and funding from the Scottish Aquaculture Innovation Centre (Ref: SL\_2017\_09). DJM acknowledges funding for Atlantic salmon genomics research from the Biotechnology and Biological Sciences Research Council (grants BB/M016455/1, BB/P02582X/1, BB/R008442/1, BB/S004181/1), the Natural Environment Research Council (grant NBAF1089) and the Norwegian Research Council (grant refs: 241016 and 275310).

### References

Alexandrou M.A., Swartz B.A., Matzke N.J. & Oakley T.H. (2013) Genome duplication and multiple evolutionary origins of



- complex migratory behavior in Salmonidae. *Molecular Phylogenetics and Evolution* **69**, 514–23.
- Allendorf F.W. & Thorgaard G.H. (1984) Tetraploidy and the evolution of salmonid fishes. In: *Evolutionary Genetics of Fishes* (Ed. by B.J. Turner), pp. 1–53. Plenum Press, New York, NY.
- Allendorf F.W., Bassham S., Cresko W.A., Limborg M.T., Seeb L.W. & Seeb J.E. (2015) Effects of crossovers between homeologs on inheritance and population genomics in polyploid-derived Salmonid fishes. *Journal of Heredity* **106**, 217–27.
- Andersson L., Archibald A.L., Bottema C.D. *et al.* (2015) Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biology* **16**, 57.
- Andrews K.R., Good J.M., Miller M.R., Luikart G. & Hohenlohe P.A. (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* **17**, 81–92.
- Ayllon F., Kjærner-Semb E., Furmanek T. *et al.* (2015) The *vgl3* locus controls age at maturity in wild and domesticated Atlantic salmon (*Salmo salar* L.) males. *PLoS Genetics* **11**, e1005628.
- Baird N.A., Etter P.D., Atwood T.S., Currey M.C., Shiver A.L., Lewis Z.A., Selker E.U., Cresko W.A. & Johnson E.A. (2008) Rapid SNP discovery and genetic mapping using sequenced rad markers. *PLoS ONE* **3**, e3376.
- Bangera R., Correa K., Lhorente J.P., Figueroa R. & Yáñez J.M. (2017) Genomic predictions can accelerate selection for resistance against *Piscirickettsia salmonis* in Atlantic salmon (*Salmo salar*). *BMC Genomics* **18**, 121.
- Barson N.J., Aykanat T., Hindar K. *et al.* (2015) Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature* **528**, 405–8.
- Benfey T.J. (2001) Use of sterile triploid Atlantic salmon (*Salmo salar* L.) for aquaculture in New Brunswick, Canada. *ICES Journal of Marine Science* **58**, 525–9.
- Berthelot C., Brunet F. & Chalopin D. (2014) The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature Communications* **5**, 3657.
- Best C., Ikert H., Kostyniuk D.J., Craig P.M., Navarro-Martin L., Marandel L. & Mennigen J.A. (2018) Epigenetics in teleost fish: from molecular mechanisms to physiological phenotypes. *Comparative Biochemistry and Physiology, Part B, Biochemistry and Molecular Biology* **224**, 210–44.
- Bilton T.P., Schofield M.R., Black M.A., Chagné D., Wilcox P.L. & Dodds K.G. (2018) Accounting for errors in low coverage high-throughput sequencing data when constructing genetic maps using biparental outcrossed populations. *Genetics* **209**, 65–76.
- Boltana S., Aguilar A., Sanhueza N., Donoso A., Mercado L., Imarai M. & Mackenzie S. (2018) Behavioural fever drives epigenetic modulation of the immune response in fish. *Frontiers in Immunology* **9**, 1241.
- Bourret V., Kent M.P., Primmer C.R. *et al.* (2013) SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Molecular Ecology* **22**, 532–51.
- Breker M. & Schuldiner M. (2014) The emergence of proteome-wide technologies: systematic analysis of proteins comes of age. *Nature Reviews Molecular Cell Biology* **15**, 453–64.
- Burkard C., Lillico S.G., Reid E., Jackson B., Mileham A.J., Ait-Ali T., Whitelaw B.A. & Archibald A.L. (2017) Precision engineering for PRRSV resistance in pigs: macrophages from genome edited pigs lacking CD163 SRCR5 domain are fully resistant to both PRRSV genotypes while maintaining biological function. *PLoS Pathogens* **13**, e1006206.
- Burton J.N., Adey A., Patwardhan R.P., Qiu R., Kitzman J.O. & Shendure J. (2013) Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nature Biotechnology* **31**, 1119–25.
- Callaway E. (2018) CRISPR plants now subject to tough GM laws in European Union. *Nature* **560**, 16.
- Causey D.R., Kim J.-H., Stead D.A., Martin S.A.M., Devlin R.H. & Macqueen D.J. (2018a) Proteomic comparison of selective breeding and growth hormone transgenesis in fish: unique pathways to enhanced growth. *Journal of Proteomics*, <https://doi.org/10.1016/j.jpro.2018.08.013>.
- Causey D.R., Pohl M.A.N., Stead D.A., Martin S.A.M., Secombes C.J. & Macqueen D.J. (2018b) High-throughput proteomic profiling of the fish liver following bacterial infection. *BMC Genomics* **19**, 719.
- Cauwelier E., Gilbey J., Jones C.S., Noble L.R. & Verspoor E. (2012) Asymmetrical viability in backcrosses between highly divergent populations of Atlantic salmon (*Salmo salar*): implications for conservation. *Conservation Genetics* **13**, 1665–9.
- Christensen K.A., Leong J.S., Sakhrani D., Biagi C.A., Minkley D.R., Withler R.E., Rondeau E.B., Koop B.F. & Devlin R.H. (2018a) Chinook salmon (*Oncorhynchus tshawytscha*) genome and transcriptome. *PLoS ONE* **13**, e0195461.
- Christensen K.A., Rondeau E.B., Minkley D.R. *et al.* (2018b) The Arctic charr (*Salvelinus alpinus*) genome and transcriptome assembly. *PLoS ONE* **13**, e0204076.
- Cong L., Ran F.A., Cox D. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–23.
- Correa K., Lhorente J.P., López M.E. *et al.* (2015) Genome-wide association analysis reveals loci associated with resistance against *Piscirickettsia salmonis* in two Atlantic salmon (*Salmo salar* L.) chromosomes. *BMC Genomics* **16**, 854.
- Correa K., Lhorente J.P., Bassini L., López M.E., Di Genova A., Maass A., Davidson W.S. & Yáñez J.M. (2016) Genome wide association study for resistance to *Caligus rogercresseyi* in Atlantic salmon (*Salmo salar* L.) using a 50K SNP genotyping array. *Aquaculture* **472**(S1), 61–5.
- Custers R. (2017) The regulatory status of gene-edited agricultural products in the EU and beyond. *Emerging Topics in Life Sciences* **1**, 221–9.
- Dehal P. & Boore J.L. (2005) Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biology* **3**, e314.
- Dehler C.E., Boudinot P., Martin S.A.M. & Collet B. (2016) Development of an efficient genome editing method by CRISPR/Cas9 in a fish cell line. *Marine Biotechnology* **18**, 449–52.
- Dehler C.E., Secombes C.J. & Martin S.A. (2017) Environmental and physiological factors shape the gut microbiota of Atlantic salmon parr (*Salmo salar* L.). *Aquaculture* **467**, 149–57.
- Edvardsen R.B., Leininger S., Kleppe L., Skaftnesmo K.O. & Wargelius A. (2014) Targeted mutagenesis in Atlantic salmon (*Salmo salar* L.) using the CRISPR/Cas9 system induces complete knockout individuals in the F0 generation. *PLoS ONE* **9**, e108622.
- FAO (2017) *Fishery and Aquaculture Statistics 2015*. Food and Agriculture Organization of the United Nations, Rome, Italy.
- Fraser D.J., Weir L.K., Bernatchez L., Hansen M.M. & Taylor E.B. (2011) Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. *Heredity* **106**, 404–20.

- Gajardo K., Rodiles A., Kortner T.M., Krogdahl Å., Bakke A.M., Merrifield D.L. & Sorum H. (2016) A high-resolution map of the gut microbiota in Atlantic salmon (*Salmo salar*): a basis for comparative gut microbial research. *Scientific Reports* **6**, 30893.
- Garcia de la Serrana D. & Macqueen D.J. (2018) Insulin-like growth factor-binding proteins of teleost fishes. *Frontiers in Endocrinology* **9**, 80.
- Garcia de Leaniz C., Fleming I.A., Einum S. *et al.* (2007) A critical review of adaptive genetic variation in Atlantic salmon: implications for conservation. *Biological Reviews of the Cambridge Philosophical Society* **82**, 173–211.
- Gavery M.R. & Roberts S.B. (2017) Epigenetic considerations in aquaculture. *PeerJ* **5**, e4147.
- Ghanbari M., Kneifel W. & Domig K.J. (2015) A new view of the fish gut microbiome: advances from next-generation sequencing. *Aquaculture* **448**, 464–75.
- Gheyas A.A., Houston R.D., Mota-Velasco J.C., Guy D.R., Tinch A.E., Haley C.S. & Woolliams J.A. (2010) Segregation of infectious pancreatic necrosis resistance QTL in the early life cycle of Atlantic salmon (*Salmo salar*). *Animal Genetics* **41**, 531–6.
- Gillard G., Harvey T.N., Gjuvland A. *et al.* (2018) Life-stage-associated remodelling of lipid metabolism regulation in Atlantic salmon. *Molecular Ecology* **27**, 1200–13.
- Gjedrem T. (2012) Genetic improvement for the development of efficient global aquaculture: a personal opinion review. *Aquaculture* **344**, 12–22.
- Gjedrem T. & Rye M. (2016) Selection response in fish and shellfish: a review. *Reviews in Aquaculture* **10**, 168–79.
- Gjøen H.M. & Bentsen H.B. (1997) Past, present, and future of genetic improvement in salmon aquaculture. *ICES Journal of Marine Science* **54**, 1009–14.
- Gutierrez A.P., Lubieniecki K.P., Davidson E.A., Lien S., Kent M.P., Fukui S., Withler R.E., Swift B. & Davidson W.S. (2012) Genetic mapping of quantitative trait loci (QTL) for body-weight in Atlantic salmon (*Salmo salar*) using a 6.5K SNP array. *Aquaculture* **358–9**, 61–70.
- Glasauer S.M. & Neuhauss S.C. (2014) Whole-genome duplication in teleost fishes and its evolutionary consequences. *Molecular Genetics and Genomics* **289**, 1045–60.
- Glover K.A., Solberg M.F., McGinnity P., Hindar K., Verspoor E., Coulson M.W., Hansen M.M., Araki H., Skaala O. & Svåsand F. (2017) Half a century of genetic interaction between farmed and wild Atlantic salmon: status of knowledge and unanswered questions. *Fish and Fisheries* **18**, 890–927.
- Gonen S., Lowe N.R., Cezard T., Gharbi K., Bishop S.C. & Houston R.D. (2014) Linkage maps of the Atlantic salmon (*Salmo salar*) genome derived from RAD sequencing. *BMC Genomics* **15**, 166.
- Gonen S., Baranski M., Thorland I., Norris A., Grove H., Arnesen P., Bakke H., Lien S., Bishop S.C. & Houston R.D. (2015) Mapping and validation of a major QTL affecting resistance to pancreas disease (salmonid alphavirus) in Atlantic salmon (*Salmo salar*). *Heredity* **115**, 405–14.
- Gutierrez A.P., Yáñez J.M. & Davidson W.S. (2016) Evidence of recent signatures of selection during domestication in an Atlantic salmon population. *Marine Genomics* **26**, 41–50.
- He X., Houde A.L.S., Neff B.D. & Heath D.D. (2017) Transcriptome response of Atlantic salmon (*Salmo salar*) to competition with ecologically similar non-native species. *Ecology and Evolution* **8**, 1769–77.
- Hartley S.E. (1987) The chromosomes of salmonid fishes. *Biological Reviews* **62**, 197–214.
- Hickey J.M. (2013) Sequencing millions of animals for genomic selection 2.0. *Journal of Animal Breeding and Genetics* **130**, 331–2.
- Houston R.D. (2017) Future directions in breeding for disease resistance in aquaculture species. *Revista Brasileira de Zootecnia* **46**, 545–51.
- Houston R.D., Haley C.S., Hamilton A., Guy D.R., Tinch A.E., Taggart J.B., McAndrew B.J. & Bishop S.C. (2008) Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). *Genetics* **178**, 1109–15.
- Houston R.D., Bishop S.C., Hamilton A., Guy D.R., Tinch A.E., Taggart J.B., Derayat A., McAndrew B.J. & Haley C.S. (2009) Detection of QTL affecting harvest traits in a commercial Atlantic salmon population. *Animal Genetics* **40**, 753–5.
- Houston R.D., Haley C.S., Hamilton A. *et al.* (2010) The susceptibility of Atlantic salmon fry to freshwater infectious pancreatic necrosis is largely explained by a major QTL. *Heredity* **105**, 318–27.
- Houston R.D., Davey J.W., Bishop S.C. *et al.* (2012) Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC Genomics* **13**, 244.
- Houston R.D., Taggart J.B., Cézard T. *et al.* (2014) Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). *BMC Genomics* **15**, 90.
- Howe K. & Wood J.M.D. (2015) Using optical mapping data for the improvement of vertebrate genome assemblies. *Gigascience* **4**, 10.
- Jain M., Koren S., Miga K.H. *et al.* (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology* **36**, 338–45.
- King T.L., Verspoor E., Spidle A.P., Gross R., Phillips R.B., Koljonen M.-L., Sanchez J.A. & Morrison C.L. (2007) Biodiversity and population structure. In: *The Atlantic Salmon* (Ed. by E. Verspoor, L. Stradmeyer & J. Nielsen), pp. 117–66. Blackwell Publishing Ltd, Hoboken, NJ.
- Kjærner-Semb E., Ayllon F., Furmanek T. *et al.* (2016) Atlantic salmon populations reveal adaptive divergence of immune related genes – a duplicated genome under selection. *BMC Genomics* **17**, 610.
- Kjærner-Semb E., Ayllon F., Kleppe L. *et al.* (2018) Vgll3 and the Hippo pathway are regulated in Sertoli cells upon entry and during puberty in Atlantic salmon testis. *Scientific Reports* **8**, 1912.
- Klemetsen A., Amundsen P.-A., Dempson J.B., Jonsson B., Jonsson N., O'Connell M.F. & Mortensen E. (2003) Atlantic salmon *Salmo salar* L., brown trout *Salmo trutta* L. and Arctic charr *Salvelinus alpinus* (L.): a review of aspects of their life histories. *Ecology of Freshwater Fish* **12**, 1–59.
- Komor A.C., Kim K.B., Packer M.S., Zuris J.A. & Liu D.R. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–4.
- Krasileva K.V., Buffalo V., Bailey P. *et al.* (2013) Separating homeologs by phasing in the tetraploid wheat transcriptome. *Genome Biology* **14**, R66.
- Król E., Douglas A., Tocher D.R., Crampton V.O., Speakman J.R., Secombes C.J. & Martin S.A. (2016) Differential responses of the gut transcriptome to plant protein diets in farmed Atlantic salmon. *BMC Genomics* **17**, 156.

- Kumar G., Hummel K., Noebauer K., Welch T.J., Razzazi-Fazeli E. & El-Matbouli M. (2018) Proteome analysis reveals a role of rainbow trout lymphoid organs during *Yersinia ruckeri* infection process. *Scientific Reports* **8**, 13998.
- Le Luyer J., Laporte M., Beacham T.D., Kaukinen K.H., Withler R.E., Leong J.S., Rondeau E.B., Koop B.F. & Bernatchez L. (2017) Parallel epigenetic modifications induced by hatchery rearing in a Pacific salmon. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 12964–9.
- Lecaudey L.A., Schliwien U.K., Osinov A.G., Taylor E.B., Bernatchez L. & Weiss S.J. (2018) Inferring phylogenetic structure, hybridization and divergence times within Salmoninae (Teleostei: Salmonidae) using RAD-sequencing. *Molecular Phylogenetics and Evolution* **124**, 82–99.
- Lien S., Gidskehaug L., Moen T., Hayes B.J., Berg P.R., Davidson W.S., Omholt S.W. & Kent M.P. (2011) 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* **12**, 615.
- Lien S., Koop B.F., Sandve S.R. *et al.* (2016) The Atlantic salmon genome provides insights into rediploidization. *Nature* **533**, 200–5.
- Limborg M.T., Seeb L.W. & Seeb J.E. (2016) Sorting duplicated loci disentangles complexities of polyploid genomes masked by genotyping by sequencing. *Molecular Ecology* **25**, 2117–29.
- Liu L., Ang K.P., Elliott J.A.K., Kent M.P., Lien S., MacDonald D. & Boulding E.G. (2017a) A genome scan for selection signatures comparing farmed Atlantic salmon with two wild populations: testing colocalization among outlier markers, candidate genes, and quantitative trait loci for production traits. *Evolutionary Applications* **10**, 276–96.
- Liu P.F., Du Y., Meng L., Li X. & Liu Y. (2017b) Proteomic analysis in kidneys of Atlantic salmon infected with *Aeromonas salmonicida* by iTRAQ. *Developmental and Comparative Immunology* **72**, 140–53.
- Llewellyn M.S., Boutin S., Hoseinifir S.H. & Derome N. (2014) Teleost microbiomes: the state of the art in their characterization, manipulation and importance in aquaculture and fisheries. *Frontiers in Microbiology* **5**, 207.
- Macqueen D.J. & Johnston I.A. (2014) A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proceedings of the Royal Society B: Biological Sciences* **281**, 20132881.
- Macqueen D.J., Primmer C.R., Houston R.D. *et al.* (2017) Functional annotation of all salmonid genomes (FAASG): an international initiative supporting future salmonid research, conservation and aquaculture. *BMC Genomics* **18**, 1–9.
- Mali P., Yang L., Esvelt K.M., Aach J., Guell M., DiCarlo J.E., Norville J.E. & Church J.M. (2013) RNA-guided human genome engineering via Cas9. *Science* **339**, 823–6.
- Mank J.E. (2017) Population genetics of sexual conflict in the genomic era. *Nature Genetics* **18**, 721–30.
- Martin K.J. & Holland P.W. (2014) Enigmatic orthology relationships between Hox clusters of the African butterfly fish and other teleosts following ancient whole-genome duplication. *Molecular Biology and Evolution* **31**, 2592–611.
- Meuwissen T.H., Hayes B.J. & Goddard M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–29.
- Meuwissen T.H., Hayes B.J. & Goddard M.E. (2013) Accelerating improvement of livestock with genomic selection. *Annual Review of Animal Biosciences* **1**, 221–37.
- Michael T.P., Jupe F., Bemm F. *et al.* (2018) High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. *Nature Communications* **9**, 541.
- Mignon-Grasteau S., Boissy A., Bouix J. *et al.* (2005) Genetics of adaptation and domestication in livestock. *Livestock Production Science* **93**, 3–14.
- Moen T., Baranski M., Sonesson A.K. & Kjøglum S. (2009) Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): population-level associations between markers and trait. *BMC Genomics* **10**, 368.
- Moen T., Torgersen J., Santi N. *et al.* (2015) Epithelial cadherin determines resistance to infectious pancreatic necrosis virus in Atlantic salmon. *Genetics* **200**, 1313–26.
- Moghadam H.K., Johnsen H., Robinson N., Andersen Ø., H. Jørgensen E., Johnsen H.K., Bæhr V.J. & Tveiten H. (2017) Impacts of early life stress on the methylome and transcriptome of Atlantic salmon. *Scientific Reports* **7**, 5023.
- Moore J.S., Bourret V., Dionne M., Bradbury I., O'Reilly P., Kent M., Chaput G. & Bernatchez L. (2014) Conservation genomics of anadromous Atlantic salmon across its North American range: outlier loci identify the same patterns of population structure as neutral loci. *Molecular Ecology* **23**, 5680–97.
- Narum S.R., Di Genova A., Micheletti S.J. & Maass A. (2018) Genomic variation underlying complex life-history traits revealed by genome sequencing in Chinook salmon. *Proceedings of the Royal Society B: Biological Sciences* **285**, 20180935.
- Nilsen T.O., Ebbesson L.O., Küllerich P., Björnsson B.T., Madsen S.S., McCormick S.D. & Stefansson S.O. (2008) Endocrine systems in juvenile anadromous and landlocked Atlantic salmon (*Salmo salar*): seasonal development and seawater acclimation. *General and Comparative Endocrinology* **155**, 762–72.
- Norman J.D., Robinson M., Glebe B., Ferguson M.M. & Danzmann R.G. (2012) Genomic arrangement of salinity tolerance QTLs in salmonids: a comparative analysis of Atlantic salmon (*Salmo salar*) with Arctic charr (*Salvelinus alpinus*) and rainbow trout (*Oncorhynchus mykiss*). *BMC Genomics* **13**, 420.
- Norris A. (2017) Application of genomics in salmon aquaculture breeding programs by Ashie Norris: who knows where the genomic revolution will lead us? *Marine Genomics* **36**, 13–5.
- Nuez-Ortín W.G., Carter C.G., Nichols P.D. *et al.* (2018) Liver proteome response of pre-harvest Atlantic salmon following exposure to elevated temperature. *BMC Genomics* **19**, 133.
- Nutter L.M.J., Heaney J.D., Kent Lloyd K.C., Murray S.A., Seavitt J.R., Skarnes W.C., Teboul L., Brown S.D.M. & Moore M. (2018) Response to “Unexpected mutations after CRISPR–Cas9 editing *in vivo*”. *Nature Methods* **15**, 235–6.
- Ødegård J., Moen T., Santi N., Korsvoll S.A., Kjøglum S. & Meuwissen T.H.E. (2014) Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). *Frontiers in Genetics* **5**, 402.
- Otto S.P. (2007) The evolutionary consequences of polyploidy. *Cell* **31**, 452–62.
- Pritchard V.L., Mäkinen H., Vähä J.P., Erkinaro J., Orell P. & Primmer C.R. (2018) Genomic signatures of fine-scale local selection in Atlantic salmon suggest involvement of sexual



- maturation, energy homeostasis and immune defence-related genes. *Molecular Ecology* **27**, 2560–75.
- Putnam N.H., O'Connell B.L., Stites J.C. *et al.* (2016) Chromosome-scale shotgun assembly using an *in vitro* method for long-range linkage. *Genome Research* **26**, 342–50.
- Robertson F.M., Gundappa M.K., Grammes F. *et al.* (2017) Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. *Genome Biology* **18**, 111.
- Robledo D., Taggart J.B., Ireland J.H. *et al.* (2016) Gene expression comparison of resistant and susceptible Atlantic salmon fry challenged with infectious pancreatic necrosis virus reveals a marked contrast in immune response. *BMC Genomics* **17**, 279.
- Robledo D., Palaiokostas C., Bargelloni L., Martínez P. & Houston R. (2017) Applications of genotyping by sequencing in aquaculture breeding and genetics. *Reviews in Aquaculture* **10**, 670–82.
- Robledo D., Gutierrez A.P., Barria A., Yáñez J.M. & Houston R.D. (2018a) Gene expression response to sea lice in Atlantic salmon skin: an RNA-Seq comparison between resistant and susceptible animals. *Frontiers in Genetics* **9**, 287.
- Robledo D., Matika O., Hamilton A. & Houston R.D. (2018b) Genome-wide association and genomic selection for resistance to amoebic gill disease in Atlantic salmon. *G3 (Bethesda)* **8**, 1195–1203.
- Rougemont Q. & Bernatchez L. (2018) The demographic history of Atlantic salmon (*Salmo salar*) across its distribution range reconstructed from approximate Bayesian computations. *Evolution* **72**, 1261–77.
- Samy J.K.A., Mulugeta T.D., Nome T., Sandve S.R., Grammes F., Kent M.P., Lien S. & Våge D.I. (2017) SalmoBase: an integrated molecular data resource for Salmonid species. *BMC Genomics* **18**, 482.
- Shalem O., Sanjana N.E., Hartenian E. *et al.* (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–7.
- Tsai H.Y., Hamilton A., Guy D.R., Tinch A.E., Bishop S.C. & Houston R.D. (2015a) The genetic architecture of growth and fillet traits in farmed Atlantic salmon (*Salmo salar*). *BMC Genetics* **16**, 51.
- Tsai H.Y., Hamilton A., Tinch A.E. *et al.* (2015b) Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics* **16**, 969.
- Tsai H.Y., Robledo D., Lowe N.R., Bekaert M., Taggart J.B., Bron J.E. & Houston R.D. (2016a) Construction and annotation of a high density SNP linkage map of the Atlantic salmon (*Salmo salar*) genome. *G3 (Bethesda)* **6**, 2173–9.
- Tsai H.Y., Hamilton A., Tinch A.E. *et al.* (2016b) Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. *Genetics Selection Evolution* **48**, 47.
- Tsai H.Y., Matika O., Edwards S.M. *et al.* (2017) Genotype imputation to improve the cost-efficiency of genomic selection in farmed Atlantic salmon. *G3 (Bethesda)* **7**, 1377–83.
- Uren Webster T.M., Consuegra S., Hitchings M. & Garcia de Leaniz C. (2018) Inter-population variation in the Atlantic salmon microbiome reflects environmental and genetic diversity. *Applied and Environmental Microbiology* **84**. <https://doi.org/10.1128/aem.00691-18>. in press
- Varadharajan S., Sandve S.R., Gillard G. *et al.* (2018) The grayling genome reveals selection on gene expression regulation after whole genome duplication. *Genome Biology and Evolution* **10**, 2785–800. <https://doi.org/10.1093/gbe/evy201>. in press.
- Vera L.M., Metochis C., Taylor J.F., Clarkson M., Skjærven K.H., Migaud H. & Tocher D.R. (2017) Early nutritional programming affects liver transcriptome in diploid and triploid Atlantic salmon, *Salmo salar*. *BMC Genomics* **18**, 886.
- Waltz E. (2017) First genetically engineered salmon sold in Canada. *Nature* **548**, 148.
- Waples R.K., Seeb L.W. & Seeb J.E. (2016) Linkage mapping with paralogs exposes regions of residual tetrasomic inheritance in chum salmon (*Oncorhynchus keta*). *Molecular Ecology Resources* **16**, 17–28.
- Wargelius A., Leininger S., Skaftnesmo K.O., Kleppe L., Andersson E., Taranger G.L., Schulz R.W. & Edvardsen R.B. (2016) *Dnd* knockout ablates germ cells and demonstrates germ cell independent sex differentiation in Atlantic salmon. *Scientific Reports* **6**, 21284.
- Weisenfeld N.I., Kumar V., Shah P., Church D.M. & Jaffe D.B. (2017) Direct determination of diploid genome sequences. *Genome Research* **27**, 757–67.
- Wilson M.V.H. & Li G.Q. (1999) Osteology and systematic position of the Eocene salmonid *Eosalmo driftwoodensis* Wilson from western North America. *Zoological Journal of the Linnean Society* **125**, 279–311.
- Yáñez J.M., Houston R.D. & Newman S. (2014) Genetics and genomics of disease resistance in salmonid species. *Frontiers in Genetics* **5**, 415.
- Yáñez J.M., Naswa S., López M.E. *et al.* (2016) Genomewide single nucleotide polymorphism discovery in Atlantic salmon (*Salmo salar*): validation in wild and farmed American and European populations. *Molecular Ecology Resources* **16**, 1002–11.
- Yoshida G.M., Lhorente J.P., Carvalheiro R. & Yáñez J.M. (2017) Bayesian genome-wide association analysis for body weight in farmed Atlantic salmon (*Salmo salar* L.). *Animal Genetics* **48**, 698–703.
- Yoshida G.M., Carvalheiro R., Lhorente J.P., Correa K., Figuero R., Houston R.D. & Yáñez J.M. (2018) Accuracy of genotype imputation and genomic predictions in a two-generation farmed Atlantic salmon population using high-density and low-density SNP panels. *Aquaculture* **491**, 147–54.
- Zhang Y., Qin W., Lu X., Xu J., Huang H., Bai H., Li S. & Lin S. (2017) Programmable base editing of zebrafish genome using a modified CRISPR-Cas9 system. *Nature Communications* **8**, 118.
- Zheng G.X., Lau B.T., Schnall-Levin M. *et al.* (2016) Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nature Biotechnology* **34**, 303–11.