Advancing fish breeding in aquaculture through genome functional annotation

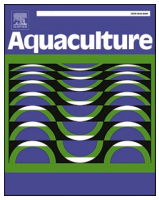
Ian A Johnston, Matthew P Kent, Pierre Boudinot, Mark Looseley, Luca Bargelloni, Sara Faggion, Gabriela A Merino, Garth R Ilsley, Julien Bobe, Costas S Tsigenopoulos, et al.

Aquaculture 583 (2024) 740589

Contents lists available at ScienceDirect

Aquaculture

journal homepage: www.elsevier.com/locate/aquaculture



Review

Advancing fish breeding in aquaculture through genome functional annotation

Ian A. Johnston [a,*], Matthew P. Kent [b], Pierre Boudinot [c], Mark Looseley [a], Luca Bargelloni [d], Sara Faggion [d], Gabriela A. Merino [e], Garth R. Ilsley [e], Julien Bobe [f], Costas S. Tsigenopoulos [g], Joseph Robertson [b], Peter W. Harrison [e], Paulino Martinez [h], Diego Robledo [i], Daniel J. Macqueen [i,*,1], Sigbjørn Lien [b,1]

[a] Xelect Ltd, Horizon House, St Andrews, Scotland KY16 9LB, UK
[b] Centre for Integrative Genetics, Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, Ås, Norway
[c] Université Paris-Saclay, National Institute for Agricultural Research (INRAE), Jouy-en-Josas, France
[d] Department of Comparative Biomedicine and Food Science, University of Padova, Italy
[e] European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, Cambridgeshire CB10 1SD, UK
[f] INRAE, LPGP, Fish Physiology and Genomics, Rennes F-35000, France
[g] Institute of Marine Biology, Biotechnology and Aquaculture (IMBBC), Hellenic Centre for Marine Research (HCMR), Heraklion, Greece
[h] Department of Zoology, Genetics and Physical Anthropology, University of Santiago de Compostela, Lugo, Spain
[i] The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, UK

ARTICLE INFO

Keywords:
Selective breeding
Fish genomes
Functional annotation
Genomic selection
Genome editing
Genetic variation
Teleost fish
Aquaculture

ABSTRACT

Genomics is increasingly applied in breeding programmes for farmed fish and shellfish species around the world. However, current applications do not include information on genome functional activity, which can enhance opportunities to predict relationships between genotypes and phenotypes and hence increase the accuracy of selection. Here, we review prospects for improving aquaculture breeding practises through the uptake of functional genomics data in light of the EU Horizon 2020 project AQUA-FAANG: '*Advancing European Aquaculture by Genome Functional Annotation*'. This consortium targeted the six major farmed fish species in European aquaculture, producing thousands of functional genomic datasets from samples representing embryos to mature adults of both sexes, and following immunological stimulation. This data was used to catalogue functional activity across the genome of each species, revealing transcribed regions, distinct chromatin states and regulatory elements impacting gene expression. These functional annotations were shared as open data through the Ensembl genome browser using the latest reference genomes for each species. AQUA-FAANG data offers novel opportunities to identify and prioritize causative genetic variants responsible for diverse traits including disease resistance, which can be exploited to enhance selective breeding. Such knowledge and associated resources have the potential to improve sustainability and boost production in aquaculture by accelerating genetic gain for health and robustness to infection, whilst reducing the requirement for animal testing. We further outline directions to advance and leverage genome functional annotation beyond the AQUA-FAANG project. Given the diversity of aquaculture sectors and businesses, the incorporation of functional genomic information into breeding decisions will depend on technological readiness level and scale of operation, with cost-benefit analysis necessary to determine the most profitable approach for each species and production system.

1. Introduction

A major challenge to the global food supply is to reconcile the growing human population, expected to reach 10 billion by 2050, with the United Nations sustainable development goals including the eradication of hunger. There is an urgent need to produce sufficient healthy food with fewer natural resources in order to reduce hunger and malnutrition, while reducing the environmental impact of animal

* Corresponding authors.
E-mail addresses: ian.johnston@xelect.co.uk (I.A. Johnston), daniel.macqueen@roslin.ed.ac.uk (D.J. Macqueen).
[1] Equal contribution.

https://doi.org/10.1016/j.aquaculture.2024.740589
Received 28 October 2023; Received in revised form 17 December 2023; Accepted 15 January 2024
Available online 17 January 2024

farming and conserving biodiversity. Aquaculture represents the fastest growing food production sector globally, with a crucial and rapidly increasing role in food and economic security (Anderson et al., 2017). As aquatic animals have become domesticated and farming industrialised, genetic selection has delivered impressive gains in many traits, including growth, survival, and muscle/fillet quality, whilst reducing feed conversion ratio (de Verdal et al., 2018; Houston et al., 2020).

Nevertheless, a major biological constraint to the expansion and long-term sustainability of aquaculture production remains the occurrence and severity of diseases, which seriously affects the productivity, profitability and welfare of all species produced (Stentiford et al., 2017). Resistance to many diseases shows moderate narrow sense heritability (0.2 to 0.4), which is within the range needed to improve the trait through selective breeding (Yáñez et al., 2015). Breeding for improved disease resistance typically requires measuring resistance phenotypes in controlled trials involving exposure to the pathogen of interest through immersion, injection, or cohabitation with infected individuals, and is usually performed on juvenile or sub-adult sibs of fish that are the candidates for selection. Since animal usage in disease challenge experiments is orders of magnitude lower than production losses caused by disease, there are major net animal welfare benefits associated with selective breeding for resistance. However, as current selective breeding approaches rely on phenotyping in the selection candidates or closely related animals (i.e., full sibs), such challenges need to be performed routinely for every generation of selection. These experiments are expensive, laborious and there is a strong industry and societal drive to reduce their frequency and minimise animal usage, to which the incorporation of functional genomic information in breeding decisions could potentially contribute.

Genetic variants located in non-coding regulatory regions of the genome play a major role in the genetic architecture of complex traits (Emilsson et al., 2008; Albert and Kruglyak, 2015). Selection methods that prioritize such 'functional variants', on the basis of their probability to affect gene expression, therefore have potential to increase selection accuracy (e.g., Xiang et al., 2021) and hence genetic gain. However, the reference genome sequences of aquaculture species have lacked sufficient functional annotation to explore this hypothesis until now.

This review explores how improved genome functional annotations capturing regulatory element activity can contribute to advancements in aquaculture selective breeding, framed in light of the recently completed EU Horizon 2020 project AQUA-FAANG. The AQUA-FAANG project aimed to address the major challenge of infectious diseases, while generating datasets and approaches relevant to understanding and exploiting the functional genetic architecture of diverse traits of importance to aquaculture, including growth, embryonic development, flesh quality, reproduction, and sexual maturation and dimorphism, across the targeted farmed fish species.

2. Functional annotation and the AQUA-FAANG project

Functional annotation seeks to catalogue the activity of sequences across the genome, revealing regions transcribed into RNA, and DNA elements involved in transcriptional regulation (**Box 1**). The scale and complexity of achieving comprehensive functional annotation requires a community approach using standardized methodologies to facilitate the amalgamation of data from different research groups. This was recognized in 2003 with the formation of the Encyclopaedia of DNA Elements (ENCODE) consortium, which had the aim of cataloguing and annotating functional elements in human and mice genomes, as well as developing search and visualisation tools (https://www.encodeproject.org/). Genetic variants associated with human diseases were highly enriched in non-coding annotations produced by ENCODE, which also captured at least half of all known human single nucleotide polymorphisms (SNPs; The ENCODE Project Consortium, 2012).

Following in many ways the model established by the ENCODE consortium, the Functional Annotation of Animal Genomes (FAANG) consortium was later established to characterise functional elements in farm animal genomes (https://www.faang.org/) (Andersson et al., 2015). In addition to producing extensive functional genomics datasets for several livestock species (Giuffra et al., 2019), FAANG has developed infrastructures to make data and metadata 'Findable, Accessible, Interoperable and Reusable' (FAIR), i.e., to be of most future value to stakeholders (Harrison et al., 2021). A "FAANG to fork" strategy was recently published, which describes a list of research priorities to improve farmed animal production by exploiting strategies that enhance understanding of genotype to phenotype relationships (Clark et al., 2020). This crucially mirrors the Farm to Fork strategy, for a fair, healthy and environmentally friendly food system, which lies at the heart of the European Commission Green Deal. Implementation of the "FAANG to fork" strategy within Europe was facilitated by establishment of the EuroFAANG consortium comprising six linked projects, AQUA-FAANG, BovReg, GENE-SWitCH, GEroNIMO, RUMIGEN and HoloRuminant, funded under the European Union's Horizon 2020 programme and covering the main farmed animal species in Europe (https://eurofaang.eu/). The continuation of EuroFAANG was recently formalized through a funded Horizon pan-European Research Infrastructure for genotype-to-phenotype research in terrestrial and aquatic farmed animals (https://cordis.europa.eu/project/id/101094718).

AQUA-FAANG was the only EuroFAANG project focussed on aquaculture, with thirteen academic institutions and ten companies participating (https://www.aqua-faang.eu/partners.html). The target species, Atlantic salmon (*Salmo salar*), rainbow trout (*Oncorhynchus mykiss*), common carp (*Cyprinus carpio*), gilthead sea bream (*Sparus aurata*), European sea bass (*Dicentrarcus labrax*) and turbot (*Scophthalmus maximus*) account for about 90% of commercial finfish production in Europe. In addition to producing comprehensive genome functional annotations for each species (see Section 3), a key focus was on revealing changes in genome activity dictating the host immune response to pathogen challenge.

The present article provides a synthesis of key results from AQUA-FAANG framed against current practises and the potential for using functional genomic information to advance breeding approaches beyond the current state-of-the-art. We also highlight future research and development priorities emerging from the AQUA-FAANG project and discuss industry requirements, particularly the need to quantify the likely financial returns of implementing any proposed new technology in breeding practise, which is very sector dependent.

Box 1
Genome functional annotation

Genome functional annotation seeks to assign functionality to DNA across the entire genome sequence under different conditions, for example in different tissues and life stages or in response to perturbations such as disease. Functional annotation moves beyond predictions of protein-coding genes to reveal DNA regulatory elements where nuclear proteins such as RNA polymerase and transcription factors (TFs) bind to initiate or influence RNA transcription. This includes promoters, which lie upstream of genes to activate transcription, and enhancers or silencers, which modulate the transcription of target genes by influencing the binding of TFs. Functional annotation encompasses both coding and non-coding RNA expression, the latter including molecules with regulatory roles, e.g., microRNAs and long non-coding RNAs. A range of functional genomics assays are available to quantitatively measure RNA expression, chromatin epigenetic state, and DNA element activity (The ENCODE Project Consortium, 2012). As detailed in Table 1, AQUA-FAANG applied 1) RNA Sequencing (RNA-Seq) to capture the expression of coding and non-coding genes, 2) ATAC Sequencing (ATAC-Seq; 'assay for transposase-accessible chromatin with sequencing') to quantify chromatin accessibility (Buenrostro et al., 2013) as a marker of DNA activity and promoter and enhancer elements, and 3) ChIP Sequencing (ChIP-Seq; Chromatin Immunoprecipitation with sequencing) to capture four histone modifications (H3K27ac, H3K4me1, H3K4me3, H3K27me3) that in combination can segregate the genome into distinct chromatin states, distinguishing various classes of regulatory elements (Ernst and Kellis, 2017). The usefulness of functional annotation increases with the diversity of sample types studied (both different tissues and developmental time points) and the quality of sample metadata recorded.

3. AQUA-FAANG achievements

Table 1 summarizes the key resources and datasets produced by the AQUA-FAANG project relevant to aquaculture breeding. At the start of the project, while genome assemblies were available for all six target species (Tine et al., 2014; Xu et al., 2014; Figueras et al., 2016; Lien et al., 2016; Pauletto et al., 2018; Pearse et al., 2019), these were fragmented by modern standards, limiting their genetic exploitation. An early project achievement was to incorporate highly contiguous, and in most cases, chromosome-level assemblies for each species into the Ensembl genome browser (https://projects.ensembl.org/aqua-faang/) (Martin et al., 2023), which were either produced by consortium members (Atlantic salmon: Stenløkk et al., 2022; common carp: Blasweiler et al., 2023; turbot: Martinez et al., 2021; European seabass: unpublished accession: GCA_905237075.1) or the wider research community (rainbow trout: Goa et al., 2021; gilthead seabream: unpublished accession: GCA_900880675.1). These improved assemblies were based on long-read sequencing data and provide a dramatically more accurate and complete depiction of the genome, with a large reduction in regions that are incorrectly assembled and/or fragmented. In turn, these updated assemblies provide an improved pathway for the development of genotyping tools that can be applied in commercial breeding. Improved assemblies allow variants to be more effectively identified and prioritized, increasing the success rate of each assay, and reducing development time and cost. A high-quality reference genome also increases the accuracy of the imputation of whole-genome sequencing (WGS) data. A further advantage is that highly contiguous assemblies allow precise mapping of genes and variants within quantitative trait loci (QTL) regions revealed by genome wide association studies (GWAS). For example, the availability of improved genomes for salmonid fish has enabled AQUA-FAANG to annotate complex T-receptor loci critical for adaptive immune responses (Boudinot et al., 2023).

AQUA-FAANG generated large-scale RNA-Seq (both mRNA-Seq and small RNA-Seq), ATAC-Seq and ChIP-Seq datasets (see Box 1 for information on these methods) covering diverse sample types across all six species as the basis for functional annotation. Much of the data produced contributed towards three functional annotation 'maps' using samples matched across species to support comparative analysis. These maps capture different stages of ontogeny, including developmental landmarks during embryogenesis (hereafter: DevMaps) and a panel of tissues from both sexes, selected for their relevance to production and fish health traits, and representing both sexually immature and mature fish (hereafter: BodyMaps). AQUA-FAANG also developed standardized protocols for immune stimulation and applied them across the six species, with the aim of activating innate immunity mechanisms against viruses and bacteria. Samples from these challenge experiments were used to generate the ImmunoMaps (Table 1). The ATAC-Seq and ChIP-Seq protocols required optimization across the partnering research centres due to the large diversity of tissues and species sampled. The same genomics/sequencing service provider was used to generate all raw sequence data, derived from over 4500 functional annotation libraries for the DevMap, BodyMap and ImmunoMap datasets, all of which are publicly available, with comprehensive descriptions of metadata and protocols used, through the FAANG data portal for AQUA-FAANG (Table 1; https://data.faang.org/projects/AQUA-FAANG), providing a novel resource for exploitation by the research community and industry.

Using the RNA-Seq data, the reference genomes of all six species have benefited from improved gene predictions through the Ensembl genebuild pipeline (https://www.ensembl.org/) (Martin et al., 2023). As Ensembl does not have a pipeline to predict microRNAs (miRNAs) using sequencing data, AQUA-FAANG collaborated with the FishmiRNA database (Desvignes et al., 2022) to annotate miRNAs using the small RNA-Seq datasets based on the evolution of miRNA gene repertoires in teleost fish. The ChIP-Seq and ATAC-Seq datasets produced by AQUA-FAANG were used in Ensembl's regulatory build (Martin et al., 2023)

Table 1
Key resources and datasets produced by the AQUA-FAANG project that empower selective breeding.

Achievement	Advancement	Data availability
High quality genomes for six farmed fish species shared via the Ensembl genome browser.	Essential foundation for accurate functional annotation, breeding programme design and implementation, and future genomic research.	Ensembl genome browser: Atlantic salmon [1], rainbow trout [2], European seabass [3], common carp [4], turbot [5], gilthead seabream [6]
689 functional annotation datasets from the six species representing key stages during embryonic development (DevMaps).	Reveals genes and regulatory elements driving key developmental processes. Genetic variants overlapping these regions expected to have major phenotypic effects.	Data and metadata available through FAANG data portal [7]
2183 functional annotation datasets from the six species, representing tissues (liver, brain, gill, intestine, fast-twitch myotomal muscle, head kidney, ovary and testis) in sexually immature and mature stages for both sexes (BodyMaps).	Allows identification of genes and regulatory elements associated with diverse traits relevant to fish health and production, including metabolism and growth, digestion, fillet quality, reproduction and sex, behaviour and neural function, haematopoiesis and immunity. Genetic variants in these regions may be relevant to enhancing production and health.	Data and metadata available through FAANG data portal [7]
1018 functional datasets from the six species representing immune-challenged head kidney tissue or primary cells, sampled after exposure to saline (control), poly I:C (stimulating antiviral response), or inactive *Vibrio* bacteria (stimulating inflammatory response) (ImmunoMaps).	Reveals genes and regulatory elements involved in early immune responses to simulated pathogens. Genetic variants overlapping these regions are expected to regulate disease resistance and health traits linked to immune functions.	Data and metadata available through FAANG data portal [7]
Improved gene annotations for the six species in the Ensembl genome browser.	More accurate representation of genes essential to prioritize potential causative genetic variants across the genome	Ensembl genome browser [1–6]
Regulatory annotations for five farmed fish species (Atlantic salmon, rainbow trout, turbot, European seabass, common carp)	Novel resource to identify and prioritize candidate causative genetic variants in non-coding regulatory regions (Fig. 1)	Ensembl genome browser from release 111 (BodyMaps and DevMaps data; ImmunoMaps data will be released in future) [1–6]
432 microRNA datasets from six farmed fish species representing selected DevMap, BodyMap and ImmunoMap samples.	Novel resource to identify non-coding RNAs influencing gene expression traits and their overlap with existing genetic variation datasets.	FishmiRNA database (in process) [8]

Notes: [1]https://www.ensembl.org/Salmo_salar/Info/Index; [2]https://www.ensembl.org/Oncorhynchus_mykiss/Info/Index; [3]https://www.ensembl.org/Dicentrarchus_labrax/Info/Index; [4]https://www.ensembl.org/Cyprinus_carpio_carpio/Info/Index; [5]https://www.ensembl.org/Scophthalmus_maximus/Info/Index; [6]https://www.ensembl.org/Sparus_aurata/Info/Index; [7]https://data.faang.org/projects/AQUA-FAANG; [8]https://www.fishmirna.org/

to predict DNA regulatory elements including promoters and enhancers, in addition to open chromatin regions. Providing the key datasets from AQUA-FAANG freely on the Ensembl genome browser gives access to a range of useful tools including Ensembl Variant Effect Predictor (VEP) (McLaren et al., 2016). Ensembl VEP allows a bespoke set of genetic variants to be classified by their predicted impacts on functional features within the genome. Using this approach, AQUA-FAANG functional annotation maps enable genetic variants overlapping non-coding regulatory elements to be prioritized across the genome (e.g., Fig. 1). The Ensembl genome browser provides tracks and annotations that allow functional annotation datasets to be navigated and visualized, providing insights into genomic regulation and gene expression across sample types and for candidate genes and genomic regions of interest, for instance QTL regions defined by GWAS.

Finally, an important legacy of the project is that a large number of aquaculture-focussed research centres in Europe have developed new competencies in producing and analysing functional annotation datasets. In this respect, beyond the bioinformatics work performed by Ensembl, the different participating groups developed expertise in analysing functional annotation datasets through nf-core pipelines (Ewels et al., 2020). Thus, as a consequence of AQUA-FAANG, many future research projects focussed on aquaculture species will benefit from the inclusion of advanced functional genomics datasets.

4. Aquaculture breeding

Box 2 provides a short overview of the diversity of selective breeding strategies currently employed in commercial aquaculture, which varies from species to species and with production system and scale of the farming operation. Currently, there are >500 cultivated aquatic species including >70 fish species, representing an enormous wealth of diversity with respect to taxon, genomic architecture, and reproductive systems (Naylor, 2021). There are three main categories of company operating selective breeding programmes in aquaculture. Firstly, germplasm providers which operate centralised breeding programmes for species such as Atlantic salmon, rainbow trout and Nile tilapia; secondly, large, and often multinational, fully integrated producers; and, thirdly, SME producers. The introduction of genetic technology is often facilitated by specialised breeding management companies or academic consultants with strong links to the scientific community. The large number of independent breeding programmes preserves genetic variation and allows breeding objectives to be highly focussed on specific production conditions.

5. Genomic technologies in aquaculture breeding

The incorporation of genome sequencing and other genomic technologies have brought advances in almost every aspect of the breeding of aquaculture species, in line with the state of the art for terrestrial farm animals (reviewed by Houston et al., 2020).

5.1. Parentage assignment

Traditional family selection programmes require offspring to be reared separately until large enough to be PIT-tagged at around 20 g, involving large numbers of tanks and significant infrastructure and labour costs. The advent of affordable genotyping technologies has allowed parentage to be retrospectively assigned at near 100% accuracy, enabling the mixing of families at very early life-stages well before tagging is possible. This not only saves on infrastructure and husbandry costs, but it is also critical to the accurate estimation of genetic effects without the confounding influence of tank effects. Thus, it is now common for family groups to be mixed shortly after production, with tagging and genotyping at a later developmental stage for parentage assignment and pedigree reconstruction.

5.2. Monosex fish production

Sexual maturation diverts energetic resources from growth to gonad formation and results in an associated loss of flesh quality and value. In salmonids, which typically have an XX/XY sex determining system (Yano et al., 2012), all-female populations are often used for production because they become sexually mature after the fish reach harvest size. Whilst monosex fish are usually produced by hormone treatment of fry, in many countries use of hormones is prohibited. Consequently, genetic strategies have been developed to eliminate hormone use in the generation of fish destined for consumption e.g., in Nile tilapia feminised XY males are crossed with YY males to produce offspring that are 25% YY, 50% XY and 25% XX. YY neofemales are then crossed with XY males to produce all male fish for production (Mair et al., 1997). Discovery of reliable markers to confirm the genetic sex of progeny before maturation is central to the commercial production of monosex populations. Production of all-female populations in Atlantic turbot (ZZ/ZW) is more complex (Martinez et al., 2009) and requires two generations to obtain superfemales (WW), which are then crossed to normal males (ZZ) to produce all-female offspring. The AQUA-FAANG project used updated whole genome assemblies to discover sex markers for two flatfish species (*sox2* in turbot, Martinez et al., 2021; and *fshr* in Senegalese sole, de la Herrán et al., 2023), which are now being used in commercial progeny

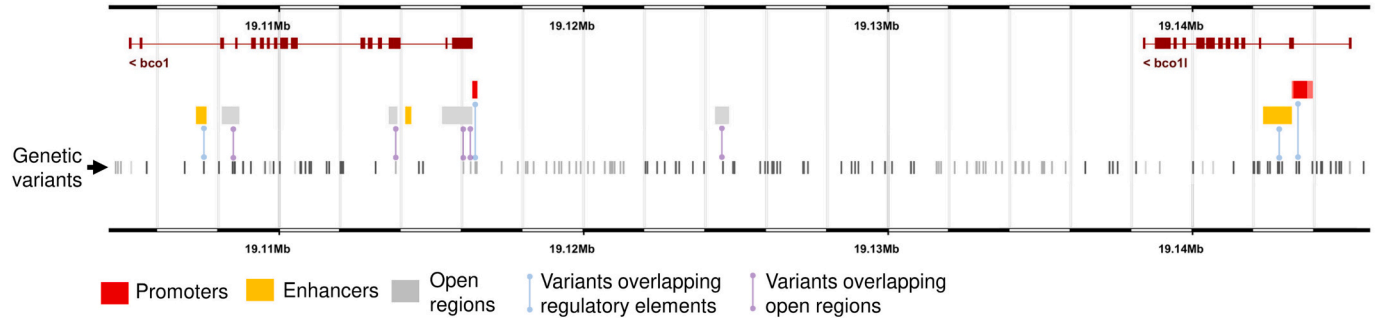


Fig. 1. The novel value of AQUA-FAANG datasets to prioritize genetic variants within the Ensembl genome browser. Shown is an approximately 40 kb region on Chromosome 26 of the Atlantic salmon genome within a QTL for flesh colour (Helgeland et al., 2019). This QTL region harbours two paralogous genes called *bco1* (ENSSSAG00000115144) and *bco1l* (ENSSSAG00000101643), each encoding beta-carotene oxygenases, with the latter considered the more likely causal gene. Also shown are regulatory elements and open chromatin regions predicted by the Ensembl regulatory build (Martin et al., 2023), integrating ATAC-Seq and ChIP-Seq datasets for the DevMaps and BodyMaps (Table 1). The location of SNPs derived from a WGS experiment (ENA accession: PRJEB34225) and their overlap with regulatory elements and open chromatin regions is shown. The SNPs overlapping these regions represent functional variants with greater potential to influence the expression of nearby genes compared to the much larger set of SNPs covering this genomic region. The data is derived from Ensembl 111 (released 11 January 2024). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

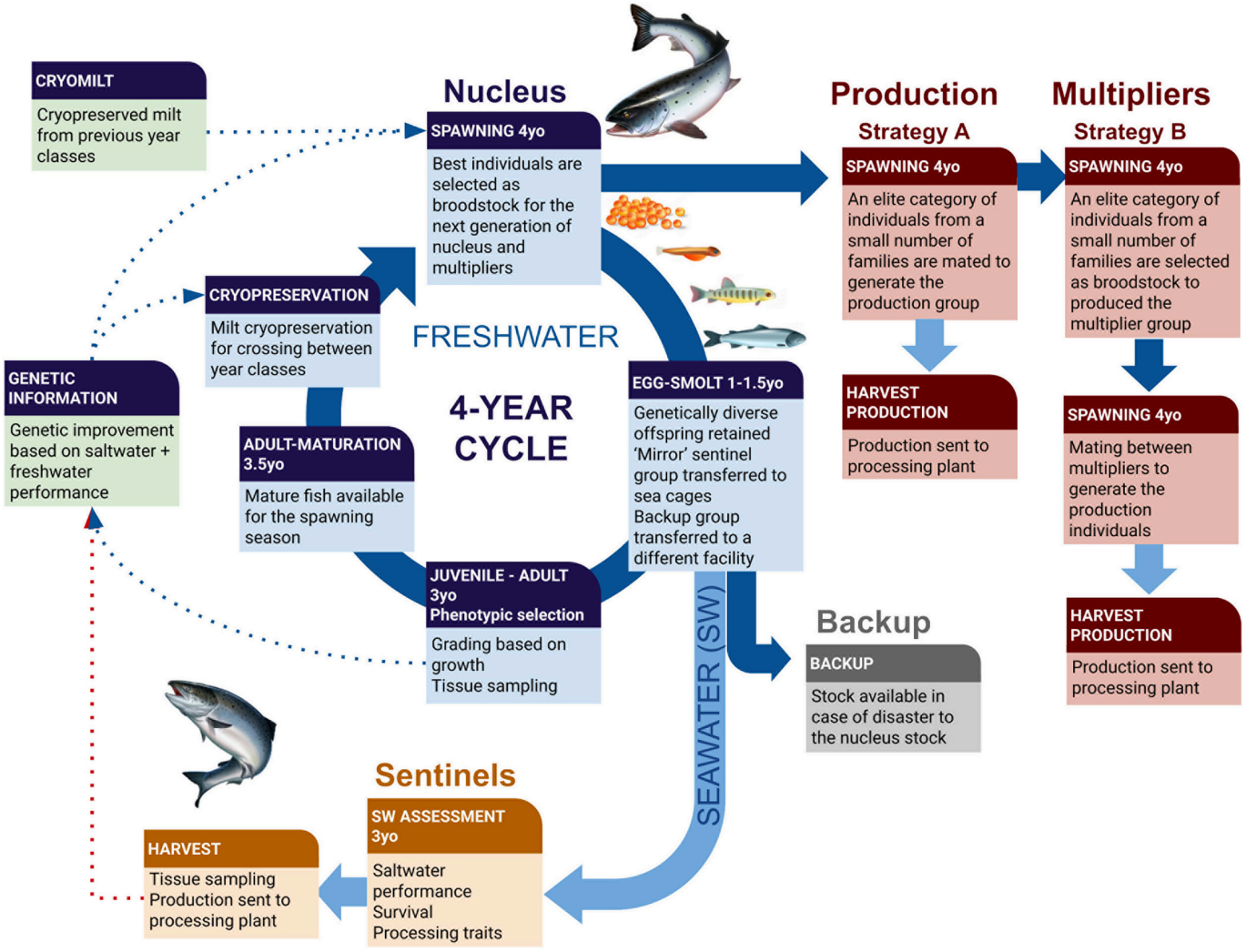


Fig. 2. Shows a typical genetic selection scheme for Atlantic salmon.

testing.

5.3. Marker assisted selection (MAS)

GWAS involving high density genotyping is widely used to discover genomic regions with a large effect on the trait of interest, i.e., QTLs. Markers with a significant association with the trait can be used to directly select fish that are also likely to carry favourable alleles at the QTL, with the potential for rapid trait improvement when QTL effects are large. An important proviso is that such SNPs will rarely affect the trait directly, rather significant associations arise because they were inherited on the same segment of chromosome as one or more yet unknown causal variants. Such associations arise through linkage disequilibrium (LD) and break down following the recombination of chromosome segments within each generation. As a result, statistically significant associations between favourable markers in one population will become weaker or absent altogether in more distantly related populations. As such, QTL selection using MAS requires regular validation in the population and environment of interest, to be sure that selection effort will be effective. Consequently, the identification of genetic variants that are causal (or located in causal genes) are much more likely to retain effectiveness across populations.

For some diseases a single or very few strong QTLs can explain over 50% of the genetic variance in the resistance trait including for cardiomyopathy syndrome in Atlantic salmon (Boison et al., 2019), *Vibrio anguillarum* (Karami et al., 2020), Rabdovirus (VHSV; Verrier et al., 2013) in rainbow trout and Tilapia Lake virus in Nile tilapia (Barria et al., 2021). Multiplexed marker panels for QTL resistance traits for multiple diseases have been successfully used by Atlantic salmon breeding companies to improve robustness and survival. The identification and application of a large QTL for IPN resistance in Atlantic salmon breeding programmes has contributed to an 80% decrease in mortality in the freshwater stages over the last few decades (Moen et al., 2015; Norris, 2017). WGS and functional annotation indicated that the NEDD-8 activating enzyme (encoded by *nae1*) underlies the QTL, supported by CRISPR-Cas 9 knockout and chemical inhibition studies in cell lines (Pavelin et al., 2021). This represents a rare example of a strong candidate gene discovered for a QTL region in an aquaculture species.

5.4. Proof-of-concept studies for functional marker discovery in AQUA-FAANG

One of the most complete proof-of-concept studies in AQUA-FAANG focussed on viral nervous necrosis (VNN), a viral disease that causes periodic and sometimes massive mortalities in farmed fish including European sea bass, Asian sea bass, grouper, Senegalese sole, gilthead sea bream, and turbot (Toffan et al., 2017). Recent studies have identified a major QTL on linkage group 12 that confers resistance to VNN in different seabass populations (Griot et al., 2021; Vela-Avitúa et al., 2022) and this region was subsequently fine-mapped using WGS

Box 2
The principles of aquaculture breeding.

The starting point for any breeding programme is the establishment of a closed genetic nucleus with sufficient numbers of genetically distinct broodstock to sustainably support genetic gain in commercially important traits, whilst at the same time maintaining high levels of genetic diversity and low levels of inbreeding and co-ancestry (Box Fig. 2). In species where artificial fertilisation can be used to make single crosses (e.g., salmonids, European seabass) the nucleus typically comprises 100–300 families with an approximately equal representation of family members. Broodstock for such species are usually sacrificed to obtain the gonads and therefore, if the generation time is n years, then n genetic nucleus groups are required, referred to as year classes. Cryopreserved milt can be used to overlap the year classes to produce more uniform performance between year classes. However, reproduction in species that live in established social groups (e.g., gilthead sea bream, amberjacks) is more complicated and requires natural spawning. In this case the nucleus comprises a series of tanks containing mature individuals of known sex ratio and with a low level of relatedness. Control of crosses is at the level of tank composition and highly dependent on the spawning dynamics of each species. Fish spawn for multiple generations and genetic gain is achieved through the annual replacement of a proportion of tank members with genetically selected individuals. In some cases, the genetic nucleus is maintained in a biosecure hatchery whilst in others future candidate breeders are brought from the production environment following a period of quarantine prior to their introduction. For larger companies, in case of accident it is common to maintain a back-up nucleus at a separate location to safeguard the genetic progress made in the breeding programme.

The next generation of breeding candidates is derived from representative groups of offspring from the nucleus that are all the same age, called sentinel or evaluation groups in classical selection (Fig. 2). Phenotypes are measured on a few thousand individuals either after on-growing in the commercial production environment, or after exposure to a pathogen in a tank-based disease challenge trail. Since commercial cages can contain hundreds of thousands of individuals, only a subsection of a few thousand is usually sampled to collect phenotype and genotype records.

The genetic component of a trait is estimated as a breeding value (EBV) in the selection candidates or their siblings by inferring additive genetic relationships based on individual phenotype records across one or more generations. EBVs for multiple traits can be combined into a single selection index by weighting each trait according to commercial priorities. To produce a mating plan, it is necessary to consider the EBV, average relatedness between selected candidates, the specific relatedness of parents within matings, and the number of families that should be produced. Sustainable genetic gain in the nucleus is achieved through a trade-off between the maximum achievable genetic gain and the requirement to keep rates of inbreeding and co-ancestry at low levels.

For small farms the surplus offspring from the nucleus may be sufficient for production. However, for large-scale production the nucleus can be used to generate multiplier broodstock, which are randomly mated to generate the production fish (Fig. 2). Since the offspring of the multiplier broodstock are not used for reproduction, genetic gain can be maximised by relaxing inbreeding and coancestry constraints. Furthermore, since requirements for production can likely be satisfied by a relatively small number of elite families, this allows faster genetic progress than directly obtained from the nucleus.

Measured phenotypes have a heritable genetic component, and a non-heritable environmental component (e.g., temperature, daylength, feed quality and quantity, stocking density, water current velocity etc.). Some traits have a significant genetic x environment interaction, such that the same genetics produces different outcomes depending on the farming location and conditions. For successful selective breeding the additive genetic component should ideally represent at least 20% of the total variation i.e., have a narrow sense heritability >0.20 (Gjedrem and Baranski, 2009).

The genetic gain is described by the so-called Breeder's equation, $R = i.r.\sigma_A$, which describes how the response to selection (R) varies with the selection intensity (i), selection accuracy (r) (equivalent to the narrow sense heritability) and σ_A, the square root of the additive genetic variance. The expected genetic gain over time is R/generation time and is reduced by the inbreeding constraints applied to ensure sustainability of the programme. Genetic gain is favoured by keeping the generation time as short as possible, often at the expense of a reduction in fecundity (Gjedrem and Baranski, 2009).

The requirement to maintain the effective population size of the nucleus demands a higher number of candidate breeders as selection intensity increases. In addition, when the number of traits included in the Selection Index increases, the genetic gain for each trait decreases; this can be offset to some degree by increasing the number of candidate breeders. The availability of a large number of candidate breeders and the control of matings optimise the genetic gain by facilitating the control of inbreeding rates (Gjedrem and Baranski, 2009).

Fig. 2 *Shows a typical genetic selection scheme for Atlantic salmon Colour figure required.*

(Delpuech et al., 2023). In AQUA-FAANG, resistance to VNN was studied in a tank-based disease challenge, and the combination of WGS, imputation, RNA-Seq and genome functional annotation led to the identification of two putative causative SNPs in the promoter of an interferon-stimulated immune gene underlying a major QTL for resistance to VNN, explaining >37% of the genetic variation in the resistance trait (Mukiibi et al., 2023). The incorporation of functional annotation was key to pull apart those putative causative SNPs from the hundreds found in the same region, thereby identifying targets for CRISPR-Cas9 gene editing to confirm the causative gene and causative mutation.

AQUA-FAANG also carried out an in-depth study on infection of rainbow trout by Viral Haemorrhagic Septicaemia Virus (VHSV) and by the bacterium *Flavobacterium psychrophilum* (Fp), responsible for two important diseases of rainbow trout in freshwater that cause significant mortality and economic losses. Using double-haploid rainbow trout isogenic lines with opposing resistance/susceptibility to VHSV and Fp, the immune pathways associated with resistance were found to be different for the two pathogens, and therefore the resistance traits are unlikely to share a common genetic basis (Fraslin et al., 2020); meaning selection will need to be conducted for each disease independently (Verrier et al., 2012, 2018). The mapping of transcriptome and epigenetic responses to infection will provide a better understanding of the genomic basis for differential resistance mechanism to support more targeted selection that minimally affect immune responses to other pathogens.

The use of common viral and bacterial mimics delivered using standardized methods for the AQUA-FAANG ImmunoMaps (Table 1) enables reliable comparative studies of immune responses across species. For example, the antiviral response of Atlantic salmon and rainbow trout includes both a highly conserved component, as well as a non-conserved, species-specific component, explained by distinct gene families that likely underlie both common and unique genetic architectures influencing disease resistance (Clark et al., 2023).

5.5. Genomic selection

Genomic Selection (GS) is widely used in larger breeding programmes because of potential gains to the accuracy of selection relative to classical methods (reviewed in Houston et al., 2020; Song et al., 2023), particularly for traits that cannot be measured in candidate breeders. GS requires genotyping at much higher densities (thousands to tens/hundreds of thousands of SNPs) than parentage assignment. Standard GS methods rely on neutral markers, commonly SNPs in incomplete LD with causative genetic variants, and selection of animals for breeding is based on EBVs calculated from the joint effects of markers covering the whole genome. In most breeding programmes, a reference or training population (usually full or half-sibs of the selection candidates) is genotyped and phenotyped, and various statistical approaches including genomic BLUP and Bayesian models are applied to estimate the EBVs of selection candidates, which are only genotyped (Song et al., 2023). GS is therefore particularly useful for destructive traits such as disease resistance or fillet yield, which cannot be directly measured on selection candidates. The increase in prediction accuracy of EBVs achieved with GS depends on numerous factors, including heritability, genotyping density, genetic relatedness between the training population and selection candidates, the statistical model used, and the number of generations of records available for models that utilize pedigrees (Bangera et al., 2017; Dufflocq et al., 2019). Single-step genomic BLUP methods incorporating phenotype information from non-genotyped

candidates offer an additional route to maximizing selection accuracy (Lagarra et al., 2014).

Relatedness is a key factor affecting the precision of selection. GS is most effective when applied to selection candidates that are closely related to the reference population but becomes less accurate at higher genetic distances between training and test populations, or when prediction is performed across generations (Fraslin et al., 2022). It has been shown for several aquaculture species that when the training population and selection candidates are closely related, prediction accuracies plateau at relatively low marker densities (e.g., 1-5 K SNPs), but may increase somewhat at higher SNP densities across different populations (Robledo et al., 2018; Kriaridou et al., 2020; Boudry et al., 2021). Experimental and methodological differences between studies preclude direct comparisons of the likely gain in prediction accuracy using GS across populations, but most fall within 5–35% (Boudry et al., 2021). Academic studies typically evaluate GS accuracy using random k-fold cross validations, where individuals belonging to the reference population are randomly split into equally sized groups; 80% of the individuals are normally used to train the model and estimate SNP effects; the obtained prediction equation is then used to predict the genomic EBV or phenotypes of the remaining 20% of the individuals (test set). This is repeated as many times as the number of groups, changing the test set at each permutation. Prediction accuracy is measured as the correlation between observed and predicted values. However, due to a random partition of the experimental population into a training set and a test set, the average genetic relatedness between the two sets is likely to be inflated, which might not apply in a commercial breeding programme. In addition, it is ultimately the response to selection that is important in breeding and some of the apparent gains in EBVs may not be realized in practice because of constraints applied when developing the mating plan, which needs to consider the relatedness of potential crosses.

A major consideration regarding the use of GS is the cost of genotyping, which can be considerably more expensive than parentage assignment, unless sample numbers are high enough to produce significant economies of scale. Until the price difference between genotyping at high- and low-density narrows, it may be more cost effective for small producers to invest in other levers to increase genetic gain e.g., improvements in phenotyping accuracy and/or genotyping higher numbers of candidates to increase selection intensity.

Genotype imputation is a strategy to reduce genotyping costs whilst sacrificing few of the benefits arising from increased selection accuracy (Tsai et al., 2017). Closely related individuals share longer haplotype blocks and have higher levels of LD than more distantly related individuals. The genetic nucleus of fish breeding programmes is typically closed with only 100 to 300 families represented per year class and inbreeding rates maintained at around 0.5% per generation. The resulting high rates of LD result in up to 95% accuracies with pedigree-based imputation methods providing most of the benefits obtainable from GS (Kriaridou et al., 2023). With this approach, higher density genotyping is usually restricted to the parents and the much more numerous offspring are genotyped with a cheaper low-density panel comprising hundreds of carefully selected SNPs from the high-density panel.

5.6. The development of multifunctional low density SNP panels

The AQUA-FAANG consortium has developed multifunctional amplicon sequencing panels, comprising 350–400 SNPs to be used for parentage assignment, genotype imputation, and marker assisted selection. The favourable marker score can be regarded as a trait with pedigree-based selection making the approach suitable to many hundreds of small and medium sized producers. As a first step, AQUA-FAANG examined the frequency of favourable alleles for VNN resistance in populations sampled from six sea bass farms across Europe and North Africa. The frequency of the favourable allele for the top 5 markers varied between 5 and 40% which is promising for the future commercial use of this type of relatively cost-efficient and low-density multifunctional panel (Wilbourn, R., unpublished results).

6. Functional annotation to accelerate genetic improvements in aquaculture

Whilst GS can offer improved prediction accuracy over classical selection approaches, a fundamental assumption is that every region of the genome has an equal influence on the trait, such that the overall genetic relationship between pairs of individuals determines their similarity in performance. Under the infinitesimal model, where the trait is controlled by an infinite number of causal variants, this assumption is valid. However, when a small number of QTLs have a substantial effect, relationships based on whole genome similarity may not accurately reflect similarities in performance. Methods to address this constraint have been proposed, whereby information from each marker is weighted according to its estimated effect on trait performance (Ren et al., 2021). The success of such approaches is specific to the genetic architecture of the trait, and the degree of genetic relatedness of the breeding population, but has potential to significantly increase prediction accuracy (Song et al., 2023). In particular, improvements in selection accuracy may be modest where a closed genetic nucleus comprises closely related individuals showing a high degree of LD and in cases where a major QTL explains a major proportion of trait variation. In contrast, more substantial improvements in accuracy would be expected on theoretical grounds across populations and/or generations and for highly polygenic traits. Experimental studies are required to quantify the benefits of incorporating functional and causative markers in different contexts and this remains a priority for future research.

6.1. Use of AQUA-FAANG data to prioritize functional genetic variants

A key achievement of the AQUA-FAANG project was to produce openly available functional annotation data to prioritize genetic variants across the genomes of the target species by association with functional regions (Table 1; Fig. 1), which have several novel applications (Fig. 3). Most variants exist within regions of the genome with no recognized function and only a small fraction of variants alter protein coding sequences. Evidence from model fishes (Albert and Kruglyak, 2015), as well as humans (The ENCODE Project Consortium, 2012), cattle (Prowse-Wilkins et al., 2021) and swine (Pan et al., 2021) suggests that most genetic variants underlying key phenotypes are in regulatory, non-coding genomic regions. However, it is extremely difficult to tell apart phenotype-causative regulatory variants from close-by linked SNPs, as their effect is not easily predicted based on just sequence information. To this end, the functional annotation data generated by AQUA-FAANG (Table 1), coupled with single-base resolution of genetic variation, dramatically increases the probability of identifying and prioritizing variants (Fig. 1). This is important because any species will typically have millions of genetic variants, most of which have limited or no impact on phenotypes. Using Ensembl VEP (McLaren et al., 2016) or related methods, it is straightforward for researchers and companies leading aquaculture breeding programmes to quickly identify functional genetic variants that overlap annotations produced by AQUA-FAANG, which have a much higher chance of impacting phenotypes and thus representing causal variants for QTLs. Once detected, predicted causal variants can be targeted on customized genotyping arrays with the aim of more accurately predicting genetic merit for traits across populations and generations. This approach has proven to work in practice in dairy cattle, where a biology-informed SNP-array containing a modest number of markers (50 k) outperformed the standard genotyping product for predicting genetic value of multiple traits across dairy cattle populations globally (Xiang et al., 2021). Similar evidence was obtained in humans, where prioritization of SNPs located in regulatory regions (after accounting for predicted effects on TF binding) significantly increased the accuracy of predicted risk scores for complex diseases using populations

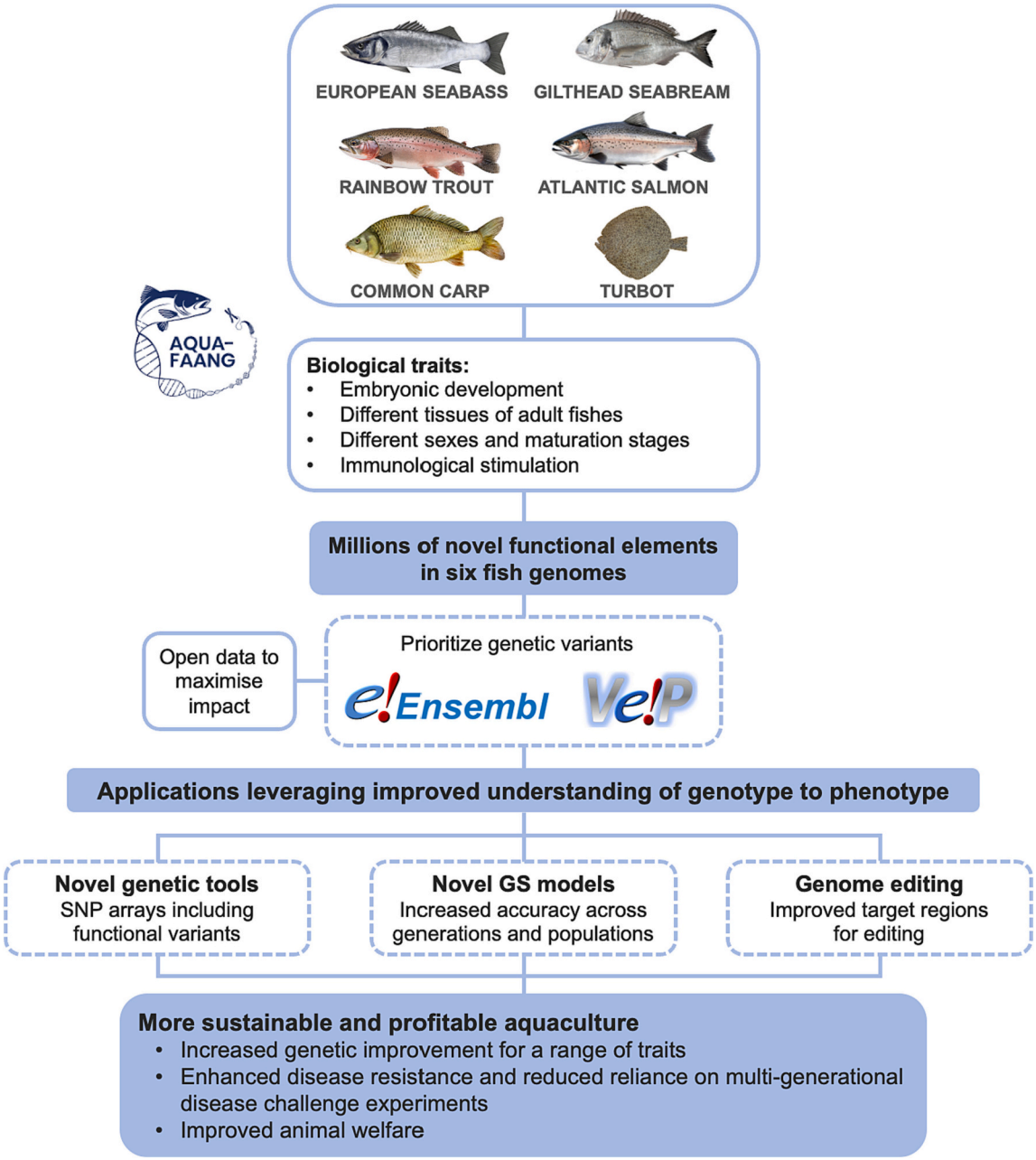


Fig. 3. A pipeline to exploit AQUA-FAANG datasets in novel breeding applications. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of different ancestry (Amariuta et al., 2020), a notoriously difficult task akin to predicting EBVs across breeds or distantly related genetic backgrounds in animal genetics.

6.2. *Genome editing to validate and create precision changes for trait improvement*

AQUA-FAANG annotations greatly enrich our understanding of functional regions within the investigated fish genomes, enabling the identification of regulatory elements and genes within QTLs and other candidate regions of interest. This enables previously overlooked putative functional associations to be explored, for example an intergenic genetic variant falling within a regulatory element showing activity in a sample type linked to a trait of interest (Fig. 1). Additional prioritization of functional variants in such regions can easily incorporate impacts on TF binding sites within regulatory elements. However, any such functional associations remain putative until experimentally tested. Genome editing (GE) by CRISPR-CAS9 is an enabling technology that can be used to test and validate such functional associations, but also has the potential to transfer or propagate desirable genetic variants into breeding programs with much greater precision than traditional crossings and in fewer generations (Tait-Burkard et al., 2018). For breeding purposes, GE represents a means to increase genetic gain by increasing the frequency of favourable variants in the genome through various means including fixing alleles at existing trait loci, creating de novo alleles, and by introducing desirable alleles from other closely related strains or species (Houston et al., 2020).

There are significant technological barriers that must be overcome before GE can be applied to breeding, including an incomplete understanding of the genes, pathways and causal variants underlying complex phenotypes such as disease resistance, the lack of high-throughput embryo editing techniques, complexities with generating (or accumulating)

multiple genome edits, and low efficiency resulting in cellular mosaicism are other issues that will need resolving prior to the practical application of the technology in aquaculture. Importantly, the use of GE in breeding would require both societal and regulatory approval and an accepted legal framework.

6.3. Can the use of functional variants in breeding reduce animal testing?

Selective breeding for disease resistance requires the repeated use of specific disease challenge experiments involving thousands of fish. Disease challenge experiments are not only expensive, but also raise ongoing ethical and animal welfare concerns, despite the fact that they provide a huge net benefit to fish welfare by enabling selective breeding for disease resistance. To date, genomic EBVs for disease traits have been estimated with little or no understanding of causative functional variants. Integrating results from studies of genomic regulatory regions with GWAS has shown promise for resolving causal variants underlying human diseases (Cano-Gamez and Trynka, 2020; Schaid et al., 2018). We anticipate the same will take place with AQUA-FAANG functional annotations, especially when it comes to prioritizing causative variants associated with disease that are difficult to rank merely by association mapping (i.e., variants in high LD).

GS and MAS deploying causal variants will not be subject to loss of effectiveness through the breakdown of LD with continued breeding across generations. This would reduce the need for routine disease challenge trials, thereby delivering significant gains in animal welfare compared to the current situation, while also reducing commercial costs. Nonetheless, the approach we are advocating still requires disease challenge trials to detect and validate the initial causal variant, and periodic testing will be required to ensure that novel pathogen strains or mutations escaping the resistance mechanism have not made the markers ineffective.

7. The next steps in genomics and their impact for breeding

The ability to predict traits in aquaculture has been hampered by limited understanding of the functional mechanisms linking genotype to phenotype. Combining AQUA-FAANG results with emerging methods that further improve the ability to discover genotype to phenotype relationships, paves the way for more accurate selection and increased genetic gain. In the following sections we discuss forward-looking approaches which, together with functional annotation, have large potential to further accelerate targeted genetic improvements.

7.1. Structural variation and pan-genomics to better capture genome diversity

A limitation of current aquaculture breeding programmes is that they do not capture structural variants (SVs), representing large insertions, duplications, deletions, and inversions (among others) ranging from tens of base pairs up to megabases in size. SVs together encompass more genomic sequence than SNPs and have a higher probability of affecting gene function and expression by disrupting coding regions or altering regulatory elements (Chiang et al., 2017; Alonge et al., 2020; Scott et al., 2021). SVs are therefore disproportionately likely to represent causal genetic variants in comparison to SNPs. WGS using short-read technologies has been used to generate accurate SV calls in Atlantic salmon (Bertolotti et al., 2020). However, SVs are diverse and often structurally complex, and as a consequence they remain undetected or incorrectly represented using short-read data (Ho et al., 2020). Importantly, SVs are not always in LD with nearby SNPs, and therefore, important variants may not be routinely captured in GWAS or GS. In this respect, recent papers have indicated that SVs account for considerable 'missing heritability' and are widely involved in disease resistance, local adaptation, and important agricultural traits (e.g., Zhou et al., 2022).

A challenge that limits scope to capture SVs in aquaculture genetics is the current reliance on single reference genomes per species, which relates to the cost and complexity involved in genome sequencing and annotation (Houston et al., 2020). Single reference genomes arbitrarily represent one individual from a population of interest, which becomes limiting when used with more distantly related individuals/populations carrying distinct genetic variation to the reference, which is hence ignored (e.g., Talenti et al., 2022). Fortunately, using increasingly competitively priced long-read sequencing data, it is now possible to build high-quality reference genomes from multiple individuals and combine these into species-specific pan-genomes, which naturally capture both SVs and SNPs (Eizenga et al., 2020; Talenti et al., 2022). A recent study in Atlantic salmon used Oxford Nanopore long-read sequencing to generate a pan-genome resource for Atlantic salmon, representing 11 high-quality assemblies from individuals sampled across the natural distribution of the species (Stenløkk, 2023). The study confidently detects more than one million SVs, accounting for >360 Mb of sequence (i.e. >10% of the genome) and representing a dramatic improvement compared to past work based on short-read WGS (Bertolotti et al., 2020). Integrating pan-genome resources with functional annotation maps from the AQUA-FAANG project, will contribute significantly to our understanding of genotype to phenotype relationships for disease resistance and other traits. Due to a lack of data, aquaculture breeding companies have completely overlooked SVs in their breeding programmes to date, but this may be about to change. We expect that functional genome annotation will soon be used to identify SVs directly impacting coding and regulatory features of importance for disease resistance.

7.2. CRISPR screening to reveal genotype to phenotype relationships

Combining GWAS data with AQUA-FAANG annotations represents a strategy to identify which of the relatively small number of variants overlapping functional elements within a QTL might have a causal effect (Fig. 1). In many cases there are potentially hundreds of causal mutations contributing to complex trait variation. While GE is most often deployed to specifically test single loci, it can also be used to test vast numbers of loci without pre-existing knowledge of their roles. High-quality annotated reference genomes can be used as a basis to develop CRISPR screening libraries to explore the functional role of any sequence including genes, regulatory elements, and non-coding RNAs, for traits of interest (Bock et al., 2022). CRISPR screening libraries are complex oligo-nucleotide pools containing typically >100 K guide RNA (gRNA) molecules. The large size of these libraries is a result of using multiple independent gRNAs to target genomic features defined by the researcher. The library is introduced to a pool of cells in vitro with the goal of perturbing thousands of genes and/or regulatory elements such that each cell contains no more than one edit (Bock et al., 2022). Cells containing edits can be selected using antibiotic resistance, before being exposed to selective pressure treatments, such as pathogen challenges, to identity and rank genes that either provide cytopathic resistance (e.g., revealing genes involved in viral entry) or promote cytopathic susceptibility (e.g., revealing genes involved in immunity).

Combining results from CRISPR screens with information on known genomic variation, can further highlight causative (and eliminate non-functional) variants. Furthermore, GWAS may not always reveal QTLs, either because the investigated population lacks segregating alleles or because SNP-array tools do not have sufficient resolution to interrogate the appropriate variants. CRISPR screening libraries can overcome this by creating genome-wide, novel genetic variations and revealing their effects.

Not all traits can be usefully assessed in vitro with the available cells, stressors, and measurable phenotypes. For example, while viral infection pressures can likely be simulated in cell culture, the complex host-pathogen interactions associated with extracellular parasites cannot be easily recaptured. There is therefore a need to expand our repertoire of cell lines and pathogenic organisms, and address fundamental

challenges related to the efficient introduction of a screening library into cells using viral transduction. Nevertheless, CRISPR screening is a powerful and popular approach to biological discovery used in humans (Li et al., 2023) and model organisms (Chen et al., 2015), as well as livestock (Yu et al., 2022).

7.3. Single cell resolved genome functional annotation

AQUA-FAANG has delivered unprecedented high-quality functional genomics data for the six major farmed fish species in European aquaculture, leading to a step-change in the quality of genome functional annotation. We are therefore now in a much better position to consider the role of specific genetic variants in the gene regulatory landscape. However, all data produced was based on bulk sequencing methods that fail to account for the fact that tissue samples are a heterogeneous mixture of different cell types. Thus, after homogenization and sequencing, it becomes impossible to link specific genomic features e.g., regulatory elements or gene expression, to particular cell types that may be responsible for phenotypes. For example, it is well established that diverse cells of the immune system and their subpopulations underlie specific immune functions responsible for disease resistance phenotypes (e.g. Ota et al., 2021). Recently developed single cell sequencing technologies offer transformative opportunities to understand the poorly characterised cellular biology of aquaculture species and are already being widely applied in diverse taxa (Ruiz Daniels et al., 2023). Single cell sequencing technologies provide novel opportunities to generate cell-type resolved functional annotations of fish genomes. This approach will further improve the scope to pre-select functional genetic variants by their association with cell-type resolved genomic features, including cell-specific expression and epigenetic responses, and will support the definition of target genes and variants for GE impacting specific cell types.

7.4. Genotype–tissue expression atlases

The AQUA-FAANG datasets provide a novel foundation for prioritizing functional genetic variants within regulatory elements in each targeted fish species. However, this approach falls short of revealing formal links between genetic variation and molecular phenotypes influencing higher level traits of relevance to aquaculture production and sustainability. In humans (GTEx Consortium, 2020) and farmed terrestrial animals (e.g. Liu et al., 2022; see https://www.farmgtex.org/; last accessed 15/12/2023), large-scale Genotype-Tissue Expression (GTEx) projects have been established to identify and characterise QTLs and genetic variants influencing gene expression, mRNA splicing, protein abundance and other molecular traits, both in tissues and specific cell types, the latter leveraging single cell sequencing (Section 7.3). A GTEx initiative in farmed fishes is a crucial next step for the field to more fully uncover and exploit the regulatory genetic basis for aquaculture traits and it has become more feasible in the light of falling sequencing costs. For some species, such as Atlantic salmon, sufficient raw data may be available in public databases, particularly for bulk RNA sequencing, but for many other species and most other molecular traits, new datasets need to be generated. This would allow the standardisation of conditions and assays, facilitating statistical associations between molecular traits and genetic variants. However, standardization will carry the risk of missing potentially relevant condition-specific associations, and special attention should be put on disease states, given the relevance of infectious agents in aquaculture. Ideally, efforts in this field would be coordinated through an initiative similar to AQUA-FAANG, so that comparable datasets are generated and made publicly available to the whole community. GTEx atlases can be directly overlaid with AQUA-FAANG functional annotations to reveal causal variants within active regulatory elements, which can then be recruited into our proposed pipeline for novel breeding applications (Fig. 3).

8. Industry perspectives

Fish farming is mostly conducted by commercial companies that need to provide a financial return to shareholders, whilst at the same time generating the societal benefits of a sustainable supply of healthy food. The best method of aquaculture breeding for individual companies is the one that produces the most economic return on investment, and this will vary with the scale of operation, production system and the species. All the classical and genomic methods described in this article have something to offer and knowledge about associated, functional, or causal genetic markers are integral to many of them, e.g., sex markers for monosex fish production or weighted single-step genomic BLUP methods, which also incorporate phenotype information. From a commercial viewpoint, it is essential to express genetic gain as an economic value to support investment decisions on the most appropriate breeding strategy to adopt for each business. Quantitative studies to determine which breeding strategies deliver the highest cost-benefit in different contexts are largely missing and would be a priority for future research from an industry perspective, particularly given the pace at which genomic science is advancing.

However, it should be noted that germplasm suppliers with the highest resources and motivation to maintain internal R&D projects may derive immediate gains from AQUA-FAANG data due to their in-house expertise and existing datasets (e.g., WGS for their stocks), which enable the updating of SNP panels with prioritized variants (e.g., Xiang et al., 2021) and will allow trials comparing the performance of functionally prioritized to randomly selected SNPs in GS, while also providing novel targets for GE research programs.

Funding

This review was funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 817923 (AQUA-FAANG).

CRediT authorship contribution statement

Ian A. Johnston: Conceptualization, Writing – original draft. **Matthew P. Kent:** Writing – original draft. **Pierre Boudinot:** Writing – review & editing. **Mark Looseley:** Writing – review & editing. **Luca Bargelloni:** Writing – original draft. **Sara Faggion:** Writing – review & editing. **Gabriela A. Merino:** Writing – review & editing. **Garth R. Ilsley:** Writing – review & editing. **Julien Bobe:** Conceptualization, Writing – review & editing. **Costas S. Tsigenopoulos:** Conceptualization, Writing – review & editing. **Joseph Robertson:** Writing – review & editing. **Peter W. Harrison:** Writing – review & editing. **Paulino Martinez:** Writing – review & editing. **Diego Robledo:** Writing – review & editing. **Daniel J. Macqueen:** Conceptualization, Writing – original draft. **Sigbjørn Lien:** Conceptualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare no conflict of interest.

Data availability

Data will be made available on request.

Acknowledgements

The authors are grateful to members of the AQUA-FAANG Exploitation Working Group for useful comments that helped shape the manuscript, as well as Dr. Lidia de Los Rios Pérez for assistance in preparing the figures. The authors acknowledge the efforts of a large number of AQUA-FAANG partners and individuals that contributed to

the generation and/or analysis of the functional annotation datasets mentioned in Table 1. AQUA-FAANG datasets mentioned in this review will be fully reported in peer-reviewed articles, including additional papers that will be published in the future. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

References

Albert, F.W., Kruglyak, L., 2015. The role of regulatory variation in complex traits and disease. Nat. Rev. Genet. 16 (4), 197–212. https://doi.org/10.1038/nrg3891. Apr.

Alonge, M., Wang, X., Benoit, M., Soyk, S., Pereira, L., Zhang, L., Suresh, H., Ramakrishnan, S., Maumus, F., Ciren, D., et al., 2020. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. Cell 182 (1), 145–161.e23.

Amariuta, T., Ishigaki, K., Sugishita, H., Ohta, T., et al., 2020. Improving the trans-ancestry portability of polygenic risk scores by prioritizing variants in predicted cell-type-specific regulatory elements. Nat. Genet. 52, 1346–1354. https://doi.org/10.1038/s41588-020-00740-8.

Anderson, J.L., Asche, F., Garlock, T., Chu, J., 2017. Aquaculture: its role in the future of food. Front. Econ. Glob. 17, 159–173. https://doi.org/10.1108/S1574-871520170000017011.

Andersson, L., Archibald, A.L., Bottema, C.D., et al., 2015. Co-ordinated international action to accelerate genome-to-phenome withing FAAG, the functional annotation of animal genome project. Genome Biol. 15. Art. No. 57 (2015). https://genomebiology.biomedcentral.com/articles/10.1186/s13059-015-0622-4.

Bangera, R., Correa, K., Lhorente, J.P., Figueroa, R., Yáñez, J.M., 2017. Genomic predictions can accelerate selections for resistance against *Piscirickettsia salmonis* in Atlantic salmon. BMC Genomics 18. https://doi.org/10.1186/s12864-017-3487-y. Article No. 121.

Barria, A., Trinh, T.Q., Mahmuddin, M., et al., 2021. A major quantitative trait locus affecting resistance to Tilapia Lake virus in farmed Nile tilapia (*Oreochromis niloticus*). Heredity 127 (3), 334–343. https://doi.org/10.1038/s41437-021-00447-4.

Bertolotti, A.C., Layer, R.M., Gundappa, M.K., et al., 2020. The structural variation landscape in 492 Atlantic salmon genomes. Nat. Commun. 11, 2020. https://doi.org/10.1038/s41467-020-18972-x. Article No. 5176.

Blasweiler, A., Megens, H.L., Goldman, M.R.G., et al., 2023. Symmetric expression of ohnologs encoding conserved antiviral responses in tetraploid common carp suggest absence of subgenome dominance after whole genome duplication. Genomics 115 (6), 110723. https://pubmed.ncbi.nlm.nih.gov/37804957/.

Bock, C., Datlinger, P., Chordon, F., Coelho, M.A., et al., 2022. High-content CRISPR screening. Nat. Rev. Meth. Prim. 2 https://doi.org/10.1038/s43586-021-00093-4. Article No 8.

Boison, S., Ding, J., Leder, E., Gjerde, B., et al., 2019. QTLs associated with resistance to cardiomyopathy syndrome in Atlantic Salmon. J. Hered. 110 (6), 727–737. https://doi.org/10.1093/jhered/esz042.

Boudinot, P., Novas, S., Jouneay, L., Monodot, S., et al., 2023. Evolution of T cell receptor beta loci in salmonids. Front. Immunol. Sec. Comparat. Immunol. 15 https://doi.org/10.3389/fimmu.2023.1238321, 15 August 2023.

Boudry, P., Allal, F., Aslam, M.C., et al., 2021. Current status and potential of genomic selection to improve selective breeding in the main aquaculture species of International Council for the Exploration of the sea (ICES) member countries. Aquacult. Report. 20, 100700 https://doi.org/10.1016/j.aqrep.2021.100700.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., Greanleaf, W.J., 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA binding proteins and nucleosome position. Nat. Methods 10, 1213–1218. https://doi.org/10.1038/nmeth.2688.

Cano-Gamez, E., Trynka, G., 2020. From GWAS to functional genomics to identify the mechanisms underlying complex diseases. Front. Genetics. Sec. Computat. Genom. 13 May 2020. https://www.frontiersin.org/articles/10.3389/fgene.2020.00424/full.

Chen, S., Sanjana, N.E., Zheng, K., et al., 2015. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. Cell. 160 (6), 1246–1260. https://doi.org/10.1016/j.cell.2015.02.038.

Chiang, C., Scott, A.J., Davis, J.R., Tsang, E.K., Li, X., Kim, Y., Hadzic, T., Damani, F.N., Ganel, L., Montgomery, S.B., et al., 2017. The impact of structural variation on human gene expression. Nat. Genet. 49 (5), 692–699. https://doi.org/10.1038/ng.3834.

Clark, E.L., Archibald, A.L., Daetwyler, H.D., Groenen, M.A.M., et al., 2020. From FAANG to fork: genome: application of highly annotated genomes to improve farm animal production. Genome Biol. 21, 285 (2020). https://doi.org/10.1186/s13059-020-02197-8.

Clark, T.C., Naseer, S., Gundappa, M.K., Laurent, A., et al., 2023. Conserved and divergent arms of the antiviral response in the duplicated genomes of salmonid fishes. Genomics 115 (4), 110663. July 2023. https://www.sciencedirect.com/science/article/pii/S0888754323001076.

de la Herrán, R., Hermida, M., Rubido, J.A., Gómez-Garrido, J., et al., 2023. A chromosome-level genome assembly enables the identification of the follicle stimulating hormone receptor as the master-sex determining gene in the flatfish *Solea senegalensis*. Mol. Ecol. Resour. 23 (4), 886–904. https://doi.org/10.1111/1755-0998.13750.

de Verdal, H., Komen, H., Quillet, E., Chatain, B., Allal, F., Benzie, J.A.H., et al., 2018. Improving feed efficiency in fish using selective breeding: a review. Rev. Aquac. 10, 833–851. https://doi.org/10.1111/raq.12202.

Delpuech, E., Vandeputte, M., Morvezen, R., et al., 2023. Whole-genome sequencing identifies interferon-induced protein IFI6/IFI27-like as a strong candidate gene for VNN resistance in European sea bass. Genet. Sel. Evol. 55 (1), 30. https://doi.org/10.1186/s12711-023-00805-2.

Desvignes, T., Bordou, P., Monfort, J., Sydes, J., et al., 2022. Fish miRNA : An evolutionary supported MicroRNA annotation and expression database for ray-finned fishes. Mol. Biol. Evol. 39 (2) https://doi.org/10.1093/molbev/msac004, 2022 Feb 3.

Dufflocq, R., Pérez-Encisco, M., Lhorente, J.P., Yāñez, J.M., 2019. Accuracy of genomic predictions using different imputation error rates in aquaculture breeding programs: a simulation study. Aquaculture 503, 225–230. https://doi.org/10.1016/j.aquaculture.2018.12.061.

Eizenga, J.M., Novak, A.M., Sibbesen, J.A., Heumos, S., Ghaffaari, A., Hickey, G., Chang, X., Seaman, J.D., Rounthwaite, R., Ebler, J., et al., 2020. Pangenome Graphs. Annu. Rev. Genomics Hum. Genet. 21 (1), 139–162. https://doi.org/10.1146/annurev-genom-120219-080406.

Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., et al., 2008. (2008) genetics of gene expression and its effect on disease. Nature 452, 423–428. https://doi.org/10.1038/nature06758.

Ernst, J., Kellis, M., 2017. Chromatin-state discovery and genome annotation with ChromHMM. Nat. Protoc. 12, 2478–2492. https://doi.org/10.1038/nprot.2017.124.

Ewels, P.A., Pettzer, A., Fillinger, S., Patel, H., et al., 2020. The nf-core framework for community-curated bioinformatics pipelines. Nat. Biotechnol. 38, 276–278. https://doi.org/10.1038/s41587-020-0439-x.

Figueras, A., Robledo, D., Corvelo, A., et al., 2016. Whole genome sequencing of turbot (*Scophthalmus maximus*. Pleuronectiformes) a fish adapted to demersal life. DNA Res. June 23 (3). https://pubmed.ncbi.nlm.nih.gov/26951068/.

Fraslin, C., Quillet, E., Rochat, T., Dechamp, N., Bernardet, J.-F., Collet, B., Lallias, D., Boudinot, P., 2020. Combining multiple approaches and models to dissect the genetic architecture of resistance to infections in fish. Front. Genet. 11, 677. https://doi.org/10.3389/fgene.2020.00677.

Fraslin, C., Yáñez, J.M., Robledo, D., Houston, R.D., 2022. The impact of genetic relationship between training and validation populations on genomic prediction accuracy in Atlantic salmon. Aquacult. Report. 23 April 2023, 101033 https://doi.org/10.1016/j.aqrep.2022.101033.

Giuffra, E., Tuggle, C.K., Archibald, A.L., Acloque, H., et al., 2019. Functional annotation of animal genomes (FAANG): Current achievements and road map. Ann. Rev. Animal Biosciences 7, 65–68. https://doi.org/10.1146/annurev-animal-020518-114913.

Gjedrem, T., Baranski, M., 2009. Selective Breeding in Aquaculture: An Introduction. Springer, Dordrecht, Heidelberg, London, New York, p. 217. https://doi.org/10.1007/978-90-481-2773-3.

Goa, G., Magaden, S., Walbiesser, G.C., et al., 2021. A long reads-based de-novo assembly of the genome of the Arlee homozygous line reveals chromosomal rearrangements in rainbow trout. G3 Genes|Genomes|Genet. 11 (4) https://doi.org/10.1093/g3journal/jkab052. April 2021,jkab052.

Griot, R., Allal, F., Phocos, F., Brard-Fudulea, S., et al., 2021. Genome-wide association studies for resistance to viral nervous necrosis in three populations of European sea bass using a novel 57k SNP array DlabChip. Aquaculture 530, 735930. https://doi.org/10.1016/j.aquaculture.2020.735930, 15th January. 2021.

GTEx Consortium, 2020. The GTEx consortium atlas of genetic regulatory effects across human tissues. Science 369 (6509), 1318–1330. https://doi.org/10.1126/science.aaz1776.

Harrison, P.W., Sokolov, A., Nayak, A., Fan, J., et al., 2021. The FAANG data portal: global, open access, "FAIR", and richly validated genotype to phenotype data for high quality functional annotation of animal genomes. Front. Genet. 12 https://doi.org/10.3389/fgene.2021.639238.

Helgeland, H., Sodeland, M., Zoric, N., et al., 2019. Genomic and functional studies suggest a key role of beta-carotene oxygenase 1like (bco1i) gene in salmon flesh colour. Sci. Rep. 9 https://doi.org/10.1038/s41598-019-56438-3. Article No. 20061.

Ho, S.S., Urban, A.E., Mills, R.E., 2020. Structural variation in the sequencing era. Nat. Rev. Genet. 21, 171–189. https://doi.org/10.1038/s41576-019-0180-9.

Houston, R.D., Bean, T.P., Macqueen, D.J., et al., 2020. Harnessing genomics to fast-track genetic improvement in aquaculture. Nat. Rev. Genet. 21, 389–409. https://doi.org/10.1038/s41576-020-0227-y.

Karami, A.M., Ødegård, J., Marana, M.H., et al., 2020. A major QTL for resistance to *Vibrio anguillarum* in rainbow trout. Front. Genet. 11–2020 https://doi.org/10.3389/fgene.2020.607558.

Kriaridou, C., Tsairidou, S., Houston, R.D., Robledo, D., 2020. Genomic prediction using low density panels in aquaculture: performance across species, traits and genotyping platforms. Front. Genet. 11. 27 Feb https://doi.org/10.3389/fgene.2020.00124.

Kriaridou, C., Tsairidou, S., Fraslin, C., et al., 2023. Evaluation of low-density SNP panels and imputation for cost-effective genomic selection in four aquaculture species. Front. Genet. Sec. Livestock Genomics. https://doi.org/10.3389/fgene.2023.1194266.

Lagarra, A., Christensen, O.F., Aguilar, I., Misztal, I., 2014. Single step, a general approach to genomic selection. Livest. Sci. 186, 54–65. https://doi.org/10.1016/j.livsci.2014.04.029.

Li, K., Ouyang, M., Zhan, J., Tian, R., 2023. CRISPR-based functional genomics screening in human-pluripotent-stem-cell-derived cell types. Cell Genom. 3, 100300 https://doi.org/10.1016/j.xgen.2023.100300.

Lien, S., Koop, B.F., Sandve, S.R., et al., 2016. The salmon genome provides insights into rediploidization. Nature 533, 200–205. https://doi.org/10.1038/nature17164.
Liu, S., Gao, Y., Canela-Xandri, O., et al., 2022. A multi-tissue atlas of regulatory variants in cattle. Nat. Genet. 54 (9), 1438–1447. https://doi.org/10.1038/s41588-022-01153-5.
Mair, G.C., Abucay, J.S., Skibinsky, D.O.F., Abolia, T.A., Beardmore, J.A., 1997. Genetic manipulation of the sex ratio for the large-scale production of all-male tilapia *Oreochromis niloticus*. Can. J. Fish. Aquat. Sci. 54 (2), 396–404. https://doi.org/10.1139/f96-282.
Martin, F.J., Amode, M.R., Aneja, A., et al., 2023 Jan 6. (2023). Ensembl 2023. Nucleic Acids Res. 51 (D1), D933–D941. https://doi.org/10.1093/nar/gkac958.
Martinez, P., Bouza, C., Hermida, M., et al., 2009. Identification of the major sex-determining region of turbot (*Scophthalmus maximus*). Genetics 183 (4), 1443–1452. https://doi.org/10.1534/genetics.109.107979.
Martinez, P., Robledo, D., Taboada, X., et al., 2021. A genome-wide association study supported by a new chromosome level assembly, suggests sox2 as a main driver of the undifferentiated ZZ/ZW sex determination of turbot (*Scophthalmus maximus*). Genomics July 113 (4), 1705–1718. https://doi.org/10.1016/j.ygeno.2021.04.007.
McLaren, W., Gil, L., Hunt, S.E., et al., 2016. The Ensembl variant effect predictor. Genome Biol. 17 https://doi.org/10.1186/s13059-016-0974-4. Article number:122.
Moen, T., Torgersen, J., Santi, N., Davidson, W.S., et al., 2015. Epithelial cadherin determines resistance to infectious pancreatic necrosis virus in Atlantic salmon. Genetics 200 (4), 1313–1326. https://doi.org/10.1534/genetics.115.175406.
Mukiibi, R., Peruzza, L., Penaloza, C., Babbucci, M., et al., 2023. Multi-functional genomic analysis identifies causal gene and variants modulating viral nervous necrosis resistance. In: 39<ce:sup>th</ce:sup> international society for animal genetics, July 2023. Cape Town, South Africa. In: https://www.isag.us/Docs/Proceedings/ISAG_2023_Abstracts.pdf.
Naylor, et al., 2021. Composition and growth of global live-weight aquaculture production. Nature 591, 551–563. https://doi.org/10.1038/s41586-021-03308-6.
Norris, A., 2017. Application of genomics in salmon aquaculture breeding programs by Ashie Norris: who knows where the genomic revolution will lead us? Mar. Genomics 36, 13–15. https://doi.org/10.1016/j.margen.2017.11.013.
Ota, M., Nagafuchi, Y., Hatano, H., et al., 2021. Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. Cell 184 (11). https://doi.org/10.1016/j.cell.2021.03.056, 3006-3021. e17.
Pan, Z., Yao, Y., Yin, H., et al., 2021. Pig genome functional annotation enhances the biological interpretation of complex traits and human disease. Nat. Commun. 12 https://doi.org/10.1038/s41467-021-26153-7. Article No. 5848.
Pauletto, M., Manousaki, T., Ferraresso, S., Babbucci, M., Tsakogiannis, A., et al., 2018. Genomic analysis of Sparus aurata reveals the evolutionary dynamics of sex-biased genes in a sequential hermaphrodite fish. Commun. Biol. 17 (1), 119. https://doi.org/10.1038/s42003-018-0122-7, 2018 Aug.
Pavelin, J., Jin, Y.H., Gratacap, R.L., et al., 2021. The nedd-8 activating enzyme gene underlies genetic resistance to infectious pancreatic necrosis virus in Atlantic salmon. Genomics. 113, 3842–3850. https://doi.org/10.1016/j.ygeno.2021.09.012.
Pearse, D.E., Barson, N.J., Nome, T., Gao, G., et al., 2019. Sex-dependent dominance maintains migration supergene in rainbow trout. Nat. Ecol. Evol. 3 (12), 1731–1742. https://doi.org/10.1038/s41559-019-1044-6, 2019 Dec.
Prowse-Wilkins, C.P., Wang, J., Xiang, R., et al., 2021. Putative causal variants are enriched in annotated functional regions from six bovine tissues. Front. Genetics. Sec. Livestock genomics 12. https://doi.org/10.3389/fgene.2021.664379, 27th June.
Ren, D., An, L., Li, B., et al., 2021. Efficient weighting methods for genomic best linear-prediction (BLUP) adapted to the genetic architecture of quantitative traits. Heredity 126, 320–334. https://doi.org/10.1038/s41437-020-00372-y.
Robledo, D., Palaiokostas, C., Bargelloni, L., Martinez, P., Houston, R., 2018. Applications of genotyping by sequencing in aquaculture breeding and genetics. Rev. Aquac. 10 (3) https://doi.org/10.1111/raq.12193, 67-0 – 682.
Ruiz Daniels, R., Taylor, R.S., Robledo, R., Macqueen, D.J., 2023. Single cell genomics as a transformative approach for aquaculture research and innovation. Rev. Aquac. 07 March 2023 https://doi.org/10.1111/raq.12806.
Schaid, D.J., Chen, W., Larson, N.B., 2018. From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat. Rev. Genet. 19 (8), 491–504. https://doi.org/10.1038/s41576-018-0016-z.
Scott, A.J., Chiang, C., Hall, I.M., 2021. Structural variants are a major source of gene expression differences in humans and often effect multiple nearby genes. Genome Res. 31, 2249–2257. https://genome.cshlp.org/content/31/12/2249.full.
Song, H., Dong, T., Yan, X., Wang, W., et al., 2023. Genomic selection and its research progress in aquaculture breeding. Rev. Aquac. 15 (1). Jan 2023. https://onlinelibrary.wiley.com/doi/full/10.1111/raq.12716.
Stenløkk, K., 2023. Genomic Structural Variations as Drivers of Adaptation in Salmonid Fishes. PhD-thesis. Norwegian University of Life Sciences. ISBN 978-82-575-2056-4. https://nmbu.brage.unit.no/nmbu-xmlui/handle/11250/3093077.
Stenløkk, K., Saiton, M., Rud-Johansen, L., et al., 2022. The emergence of supergenes from inversions in Atlantic salmon. Philos. Trans. R. Soc. Lon. B Biol. SCo August 377 (1856). https://doi.org/10.1098/rstb.2021.0195, 20210195.
Stentiford, G.D., Stritunyalucjsana, K., Flegel, T.W., Williams, B.A.P., et al., 2017. New paradigms to help solve the global aquaculture disease crisis. PLoS Pathog. 13, e1006160 https://doi.org/10.1371/journal.ppat.1006160.
Tait-Burkard, C., Doeschl-Wilson, A., McGrew, M.J., Archibald, A.L., et al., 2018. Livestock 2.0 – genome editing for fitter, healthier, and more productive farm animals. Genome Biol. 19 https://doi.org/10.1186/s13059-018-1583-1. Article no. 204 (2018).
Talenti, A., Powell, J., Hemmink, J.D., Cook, E.A.J., et al., 2022. A cattle graph genome incorporating global breed diversity. Nat. Commun. 13. Article Number 910. https://www.nature.com/articles/s41467-022-28605-0.
The ENCODE Project Consortium, 2012. An integrated encyclopaedia of DNA elements in the human genome. Nature 489, 57–74. https://doi.org/10.1038/nature11247.
Tine, M., Kuhl, H., Gagnairem, P.-A., et al., 2014. European sea bass genome and its variation provides insights into adaptation and speciation. Nat. Commun. https://doi.org/10.1038/ncomms6770. Article No. 5770.
Toffan, A., Pascoli, F., Oretto, T., et al., 2017. Viral nervous necrosis in gilthead sea bream (*Sparus aurata*) caused by reassortant betanodavirus RGNNV/SJNNV: an emerging threat for Mediterranean aquaculture. Sci. Rep. 7 https://doi.org/10.1038/srep46755. Article number 46755.
Tsai, H.Y., Matika, O., Edwards, S.M.K., Antolín-Sánchez, R., Hamilton, A., et al., 2017. Genotype imputation to improve the cost-efficiency of genomic selection in farmed Atlantic salmon. G3 Genes Genomes, Genet. 7, 1377–1383. https://doi.org/10.1534/g3.117.040717.
Vela-Avitúa, S., Thorland, I., Bakopoulos, V., et al., 2022. Genetic basis for resistance against viral nervous necrosis: GWAS and potential of genomic prediction explored in European sea bass (*Dicentrarchus labrax*). Front. Genetics Sec. Livestock Genom. 13 https://doi.org/10.3389/fgene.2022.804584.
Verrier, E.R., Langevin, C., Tohry, C., Houel, A., Ducrocq, V., Benmansour, A., et al., 2012. Genetic resistance to rhabdovirus infection in teleost fish is paralleled to the derived cell resistance status. PLoS One 7, e33935. https://doi.org/10.1371/journal.pone.0033935.
Verrier, E.R., Dorson, M., Mauger, S., Torhy, C., Ciobotaru, C., et al., 2013. Resistance to a Rhabdovirus (VHSV) in rainbow trout: identification of a major QTL related to innate mechanisms. PLoS One 8 (2), e55302. https://doi.org/10.1371/journal.pone.0055302.
Verrier, E.R., Genet, C., Laloë, D., Jaffrezic, F., Rau, A., Esquerre, D., et al., 2018. Genetic and transcriptomic analyses provide new insights on the early antiviral response to VHSV in resistant and susceptible rainbow trout. BMC Genomics 19, 482. https://www.doi.org/10.1186-s12864-018-4860-1.
Xiang, R., MacLeod, I.M., Daetwyler, H.H., et al., 2021. Genome-wide fine mapping identifies pleiotropic and functional variants that predict many traits across global cattle populations. Nat. Commun. 12, Article No. 860, 2021. https://www.nature.com/articles/s41467-021-21001-0.
Xu, P., Zhang, X., Wang, X., et al., 2014. Genome sequence and genetic diversity of the common carp. *Cyprinus carpio*. Nat. Genet. 46, 1212–1219. https://doi.org/10.1038/ng.3098.
Yáñez, J.M., Houston, R.D., Newman, S., 2015. Genetics and genomics of disease resistance in salmonid species. Front. Genet. 5, 415, 2014 Nov 26. https://pubmed.ncbi.nlm.nih.gov/25505486/.
Yano, A., Nicol, B., Jouanno, E., Quillet, E., et al., 2012. The sexually dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. Evol. Appl. 6 (3), 486–496. https://doi.org/10.1111/eva.12032.
Yu, C., Zhong, H., Yang, X., Li, G., Wu, Z., Yang, H., 2022. Establishment of a pig CRISPR/Cas9 knockout library for functional gene screening in pig cells. Biotechnol. J. Jul;17 (7) https://doi.org/10.1002/biot.202100408 e2100408.
Zhou, Y., et al., 2022. Graph pangenome captures missing heritability and empowers tomato breeding. Nature 606, 527–534. https://doi.org/10.1038/s41586-022-04808-9.

Glossary

Causal genetic variant: is taken to mean a genetic variant that is directly affecting a trait of interest.
ChIP-Seq: is a technique that combines chromatin immunoprecipitation (ChIP) with high-throughput sequencing to identify genome-wide DNA binding sites for transcription factors and other nuclear proteins, including to assess histone modifications.
Combined Family Selection: is a methodology which exploits the additive genetic variation between and within families to identify the best selection candidates for breeding.
Effective Population Size: is the number of breeding individuals in a theoretical population that would maintain genetic diversity.
Estimated Breeding Value (EBV): is a measure of the breeding potential of an animal for a specific trait.
Family: represents the progeny of a single mating.
Family selection: is a breeding technique involving the selection of a mating pair on the basis of the average performance of their progeny.
Functional variant: is taken to mean a genetic variant that is prioritized as likely to influence a trait due to its association with functional regions in the genome.
Genome editing: is a group of technologies that enable targeted genetic and epigenetic modifications at specific locations in the genome.
Genotype: is the allele carried by an individual within a wider population at a particular polymorphic location in the genome.
Genomic selection: is a form of marker-assisted selection using genetic markers covering the whole genome so that all quantitative trait loci (QTL) are in linkage disequilibrium with at least one marker.
GWAS: stands for genome-wide association study, an approach where variation in a trait of interest is statistically associated with genetic variation to identify genomic regions that explain trait variation within a population or species.
Linkage disequilibrium: is the non-random association of alleles at different loci in a population, which increases with genetic relatedness.
Phenotype: refers to the observable traits or characters of an individual.
Quantitative Trait Loci (QTL): are regions of DNA associated with a specific trait that varies within a population.

RNA-Seq: refers to RNA sequencing using high-throughput sequencing technologies.

SNP: is short for single nucleotide polymorphism, the main type of genetic marker used in aquaculture breeding and genetics, representing a variant corresponding to a single DNA base pair in the genome.

SV: is short for structural variant, referring to large genetic polymorphisms including deletions, inversions, duplications, and insertions, which are widely overlooked in current aquaculture breeding, but more likely to represent causal genetic variants than SNPs.