Accepted for publication by *Nature Reviews Genetics* published by SpringerNature: https://doi.org/10.1038/s41576-020-0227-y

Harnessing genomics to fast-track genetic improvement in aquaculture

Ross D. Houston[1†], Tim P. Bean[1], Daniel J. Macqueen[1], Manu Kumar Gundappa[1], Ye Hwa Jin[1], Tom L. Jenkins[2], Sarah-Louise C. Selly[3], Samuel A.M. Martin[4], Jamie R. Stevens[2], Eduarda M. Santos[2], Andrew Davie[3] and Diego Robledo[1]

[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, UK.

[2]Sustainable Aquaculture Futures, Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter, UK.

[3]Institute of Aquaculture, University of Stirling, Stirling, UK.

[4]School of Biological Sciences, University of Aberdeen, Aberdeen, UK.

†e-mail: ross.houston@roslin.ed.ac.uk

Abstract

Aquaculture is the fastest growing farmed food sector and will soon become the primary source of fish and shellfish for human diets. In contrast to crops and livestock, production is derived from numerous, exceptionally diverse species that are typically in the early stages of domestication. Genetic improvement of production traits via well-designed, managed breeding programmes has great potential to help meet the rising seafood demand driven by human population growth. Supported by continuous advances in sequencing and bioinformatics, genomics is increasingly being applied across the broad range of aquaculture species and at all stages of the domestication process to optimize selective breeding. In the future, combining genomic selection with biotechnological innovations, such as genome editing and surrogate broodstock technologies, may further expedite genetic improvement in aquaculture.

[H1] Introduction

Aquaculture **[G]** has a crucial and rapidly increasing role in food security and economic stability worldwide. Over 90% of global aquaculture takes place in low- and middle-income countries, where it provides major contributions to the Sustainable Development Goals of the United Nations, either directly through human consumption or indirectly through economic growth[1]. Global production of finfish and shellfish reached 172.6 million tons in

2017, approximately half of which is currently derived from aquaculture[2]. Capture fisheries, which harvest organisms in naturally occurring marine and freshwater environments for commercial purposes, are placing serious pressures on wild stocks, with minimal scope for sustainable expansion[3]. By contrast, aquaculture is the fastest growing food production sector globally[1]. With major limitations on wild capture and terrestrial farmland exploitation[4], its future importance as a source of affordable and nutritious animal protein for human diets is evident. However, intensification of aquaculture production poses environmental concerns, such as habitat destruction[5] and infectious disease outbreaks, which have a negative impact on the health and welfare of farmed (and potentially wild) populations[6] and may be exacerbated by climate change[7].

Selective breeding for genetic improvement of production traits has great potential to improve the efficiency and reduce the environmental footprint of aquaculture. However, in contrast to the terrestrial livestock and crop sectors, aquaculture is based on a hugely diverse group of finfish and shellfish species (Fig. 1), comprising an estimated 543 different animal species, including 362 finfish, 104 molluscs, 62 crustaceans, 9 other aquatic invertebrates, and 6 frogs and reptiles[2] (although aquatic plants and algae are also cultured for human use and consumption, the aquaculture of these organisms is beyond the scope of this review and is covered elsewhere, for example, Ref.[8,9]). Farming of approximately 70 of these species underpins 80% of the global aquaculture production volume, compared with just three livestock species (pig, chicken and cow), which make up 80% of global meat production (Fig. 1b, Supplementary Tables 1-2), and four plant species (rice, wheat, maize and potatoes), which underlie two-thirds of worldwide crop production[10]. Despite their diversity, aquaculture species tend to share two key features that enhance their potential for genetic improvement. Firstly, they remain in the early stages of the domestication process[11] (Fig. 1), which is linked to higher within-species genetic diversity. Secondly, they are highly fecund, with typically external fertilization. This feature of their reproductive biology enables flexibility in breeding programme design and widespread dissemination of selectively bred strains to producers, often without the need for several tiers to multiply and disseminate sufficient numbers of genetically improved animals for production[12]. Therefore, there is a pressing opportunity to utilize domestication and selective breeding programmes to harness the as-yet largely untapped genetic potential of farmed aquatic species[13], as highlighted in a recent landmark report by the FAO[13]. This potential for cumulative and permanent improvement of production

traits is evident from the typically high genetic gains **[G]** in aquaculture breeding programmes, for example, an average of 13% growth increase per generation in Atlantic salmon (*Salmo salar*)[14], which is substantially higher than the growth observed in terrestrial livestock species breeding programmes[12,15].

Genomic tools are hugely valuable to inform sustainable genetic improvement[16], and their affordability and accessibility mean they can now be applied at all stages of the domestication and genetic improvement continuum, from informing the choice of base populations **[G]** through to advanced genomic selection **[G]** in closed commercial breeding nuclei **[G]** (Box 1). Furthermore, they can be applied to characterize, utilize and conserve wild aquatic genetic resources, and inform the management of interaction between farmed and wild aquatic animals throughout this continuum.

This Review provides an overview of the status of domestication and selective breeding in aquaculture species, highlights how tailored application of genomic tools can expedite sustainable genetic improvement in diverse species and environments, and explores the potential of emerging genomic and biotechnology techniques, such as genome editing or surrogate broodstock **[G]** technologies, to promote step-improvements in aquaculture breeding and production.

[H1] The domestication of aquaculture species

Domestication in the context of this Review is considered to be the process of moving from an exclusive reliance on wild broodstock **[G]**, to completion of the full lifecycle in captivity, and use of modern selective breeding for genetic improvement of production traits, such as growth and disease resistance. Historically, the selection of species amenable to reproduction in farmed environments has been pivotal to defining which livestock and aquaculture species were farmed. For example, domesticated species tend to display behavioural plasticity **[G]** that enables them to adapt to a range of captive environments[17,18]. A key difference between livestock and aquaculture species is that domestication of terrestrial livestock occurred in tandem with global human migration several millennia prior to the informed management of breeding populations, and modern livestock lines have typically undergone multiple major genetic bottlenecks **[G]**[11]. By contrast, the time lag between domestication and selective breeding is considerably shorter in aquaculture species, with both occurring in tandem in many cases. Consequently, genomic tools can be

used from the outset to inform, optimize and expedite the two processes (Box 1), providing a more detailed understanding of their impact on species' genomes and physiology.

For certain major aquaculture species, such as carp (*Cyprinidae* spp.) and tilapia (*Cichlidae* spp.), aquaculture and domestication have been ongoing in some form for millennia[19], but selective breeding programmes to enable genetic improvement are much more recent[20] (Fig. 1b). Currently, only a minority of aquaculture production is derived from selectively bred stocks, estimated at approximately 10% in 2012[21]. However, this number is increasing rapidly, particularly for species with high production volume and value, with approximately 75% of the top 10 finfish, crustacean and mollusc species (by production volume) benefitting from some form of modern selective breeding programme (Supplementary tables 3, 4). The use of genetic technologies also varies dramatically by continent, with >80% of European aquaculture production derived from selective breeding programmes[22]. The availability and application of selective breeding depends on the local environmental, social, political and economic landscape, all of which can present major hurdles, especially in low- and middle-income countries[23]. These programmes enable cumulative, permanent and sustainable genetic gain for target production traits[15,24], and are fundamental to scale up aquaculture production in the context of finite resources[13].

Moving towards genetic improvement via selective breeding requires progression along the 'levels of domestication' scale[25], which reflects our ability to control the lifecycle of the farmed species in captivity. While the number and diversity of aquaculture species present challenges to this process, new husbandry techniques linked to improved understanding of reproductive biology and larval rearing will help overcome these challenges.

[H1] The burgeoning genomic toolbox

Genomic resources for aquaculture generally lag behind terrestrial livestock, in particular for sequencing and assembly of reference genomes (Table 1). Several high-value species remain without a publicly available high-quality reference genome and have limited genomic resources. In part, this reflects the traditionally challenging nature of genome assembly in non-mammalian and non-avian species, particularly for aquatic species with complex genomic features. These include the widespead presence of duplicated loci due to genome duplication events, for example, in salmonids[26], cyprinids[27] and sturgeons[28], and the

exceptionally high heterozygosity observed, for example, in bivalve species[29,30] and crustaceans[31]. Such features seriously hinder assembly algorithms using short-read sequence data; as a result, many existing assemblies are very fragmented. However, these genomic features can underlie adaptive capacity and phenotypic plasticity in production environments[26,32], and might contribute to the genetic regulation of production-relevant traits[26,32].

The latest sequencing technologies, including platforms that generate long reads, for example, single-molecule real-time sequencing (Pacific Biosciences) and nanopore sequencing (Oxford Nanopore), and linked reads **[G]** (10X Genomics) are increasingly applied to aquaculture species to improve assemblies (Supplementary table 5). When combined with long-range scaffolding **[G]** technologies such as high-throughput chromatin conformation capture approaches (Hi-C; for example, Dovetail Genomics) and/or optical mapping (for example, Bionano Genomics), high-quality contiguous assemblies are possible even for challenging genomes[33]. For example, a recent genome assembly of the yellow perch (*Perca flavescens*) resulted in 24 (2n = 24) chromosome-size scaffolds covering 99% of the complete assembly, with an N50 of 37.4 Mb[34]. All major aquaculture species are likely to benefit from such high-quality assemblies in the near future.

With genome sequencing of a target species coming within reach of individual laboratories, it no longer requires the degree of coordinated effort and funding that led to the first farmed animal species' reference genome assemblies, including Atlantic salmon[26]. However, standardization and coordination of multiple assemblies, including population- or 'breed'-specific assemblies, and their functional annotation remain a challenge for which international coordination and community-led initiatives are required.

A key component of the genomic toolbox to inform domestication and selective breeding is genotyping. Single nucleotide polymorphism (SNP) array platforms have been created for many high-value aquaculture species (Table 1), and genotyping by sequencing **[G]** (GBS) techniques including restriction site-associated DNA sequencing (RAD-Seq[35]) and derivatives have been applied in many species to obtain population-level SNP data without major prior investment or the immediate need for a reference genome[36,37].

[H1] Genomics applied to domestication

The establishment and management of genetically diverse base populations is essential to domestication and the formation of breeding programmes, as it underlies the future genetic potential to be exploited via selective breeding[38]. Poor broodstock management and hatchery practices that lead to inbreeding depression **[G]** have been hypothesized to result in reduced population fitness, increased susceptibility to stress and disease and, ultimately, 'boom-and-bust' production cycles[39,40]. Tailored use of genomic tools can be applied at each stage of the domestication and selective breeding continuum to inform and optimize the process (Box 1).

An example of genomics-enabled domestication of a new target species is the Australasian snapper (*Pagrus auratus*) in New Zealand. Rapid generation of *de novo* genome maps[41], transcriptomes[42], GBS methods[41,43] and estimation of genetic diversity and genetic parameters[43] were applied to inform the selection of base populations, retention of genetic diversity during domestication, and investigations into the biology of production traits. Similarly, the recent widespread use of cleanerfish (for example, Ballan wrasse, *Labrus bergylta*, and lumpfish, *Cyclopterus lumpus*) for co-culture with Atlantic salmon to help tackle sea lice (*Lepeophtheirus salmonis* and *Caligus rogercresseyi*) has led to expedited genomics-enabled domestication and breeding of lumpfish (Box 1). These cases are early examples of how genomics technology has rapidly become accessible and should be applied from the outset to inform domestication and subsequent genetic improvement.

Moreover, genomics tools are valuable to tackle species-specific breeding and production issues related to the highly diverse biology of aquaculture species. For example, a key component of the domestication–genetic improvement continuum in aquaculture species is an early understanding of sex determination, where a diverse array of genetic and non-genetic systems have been described[44]. These can vary within genus and even within species, and sequential hermaphroditism **[G]** presents an additional challenge in several commercially important aquaculture species[45]. GBS techniques have been widely applied to assess the genetic basis of sex determination[46], for example, in Nile tilapia (*Oreochromis niloticus*)[47], Atlantic halibut (*Hippoglossus hippoglossus*)[48], European seabass (*Dicentrarchus labrax*)[49] and mud crabs (Scylla sp.)[50]. The genetic markers identified in these studies can be applied to predict the sex of juveniles and to control the sex ratio in both broodstock and production animals. An additional species-specific reproductive challenge is mass spawning **[G]**, which is a feature of several marine aquaculture species, such as

gilthead sea bream and barramundi. Mass spawning causes practical challenges such as uneven parental contribution and difficulty in tracking individual pedigrees, which can result in inbreeding[51]. Although multiple interventions are possible to enable pedigree tracking (for example, pair spawning or stripping using hormonal induction)[52], genetic markers are frequently applied to track stock relatedness to minimize loss of genetic diversity within a closed breeding nucleus[51].

Of note, the reliability of genomic data alone to predict adaptive potential of populations is questionable[53], and genomic tools should be used as a complement to phenotypic evaluations of stocks. These evaluations may include trial diallelic crosses between strains in multiple environments, which can inform on additive genetic and heterotic effect on traits of interest, in addition to genotype by environment (GxE) interactions (discussed in more detail below[54]). Such information can be used to optimize selection of the base population, ensuring it has substantial genetic variation to be utilised for effective directional selection[38]. However, while hybrid vigour resulting from strain crosses can result in notable one-off gains in production, and genomic tools can provide insight to the underlying molecular mechanisms of this heterosis[55], exploiting additive genetic variation via within-strain breeding programmes is likely to result in superior performance after a small number of generations of selection[54].

[H1] Genomics applied to selective breeding

The establishment of well-managed selective breeding programmes for aquaculture based on recording of pedigree and routine measurements of traits has been successful in increasing the production of several species[14]. Just as genomic tools are applied to inform and optimize domestication, they can improve selective breeding in several ways, including by maximizing genetic gain and minimizing inbreeding[16].

[H2] Major-effect loci in recently domesticated populations

A key factor in defining the optimal use of genomic tools is the genetic architecture of production traits in the breeding goal, that is, whether genetic variation in target traits is underpinned by few major-effect loci or (as is typically the case in farmed animal populations[12]) many loci of minor effect. Farmed aquatic populations face selection pressures that are vastly different to their wild counterparts. Due to the recent and ongoing

domestication process, previously neutral alleles in wild populations may be beneficial for production phenotypes, and these will remain amongst the standing genetic variation in aquaculture populations. During the millenia of domestication of terrestrial livestock, such loci are likely to already be fixed via soft sweeps **[G]** , but in aquaculture species they may present a one-off opportunity for rapid genetic improvement via marker-assisted selection **[G]** (MAS) based on the use of targeted quantitative trait locus **[G]** (QTL)-linked markers to augment breeding decisions.

A well-known example is the major QTL affecting resistance to Infectious Pancreatic Necrosis (IPN) virus in Atlantic salmon, for which rapid uptake of MAS by the industry had a major role in preventing outbreaks of this disease (Box 2). Other applications of QTL for disease resistance include breeding of a Japanese flounder (*Paralichthys olivaceus*) strain with resistance to the viral disease lymphocystis[56], based on a major QTL for lymphocystis resistance[57], and use of MAS based on QTL affecting resistance to bacterial cold water disease in rainbow trout (*Oncorhynchus mykiss*)[58]. Other noteworty examples of major effect loci in salmon include *vgll3*, which controls the timing of sexual maturation and explains 30–40% of the phenotypic variation in age at maturity[59,60], as well as loci for resistance to pancreas disease[61], and cardiomyopathy syndrome[62,63]. Similarly, in Nile tilapia, a locus explaining 79% of the phenotypic variation in salinity tolerance was detected[64], although validation of the size of effect in independent populations is required to make generalized conclusions about this trait.

As genomics is increasingly used to study traits of interest to aquaculture in additional species and populations the number of loci of major effect will presumably rise. While MAS has had limited success in terrestrial livestock, its use within aquaculture populations at the early stages of domestication can provide rare but striking examples which help to highlight the value of genetic improvement to the industry.

[H2] Genomic selection to accelerate trait improvement

Genome-wide association studies (GWAS) in aquaculture species have highlighted that most traits of relevance to production are polygenic in nature[65,66], that is, under the control of many loci, typically of small effect. For genetic improvement of such traits, routine trait measurement and tracking of relationships between individual animals in a breeding population is required[67]. The availability of large full-sibling families gives both power and

flexibility to a breeding programme design, for example, allowing the routine testing of full-siblings of the selection candidates (sib-testing) for traits that are practically challenging or impossible to measure on the selection candidates themselves, such as disease resistance (Fig. 2, Box 2). However, for these sib-testing traits, selection candidates from a given family have the same estimated breeding value, placing limitations on genetic gain that can be achieved while maintaining genetic diversity. Genetic marker data are required to accurately capture the within-family (or Mendelian sampling **[G]**) component of genetic variation for such traits.

Genomic selection[68] was first tested in Atlantic salmon breeding, enabled by development of the first high-density SNP arrays **[G]** [69,70] and demonstration of their utility to accurately predict breeding values in a typical salmon breeding programme setting[70,71]. Genomic selection in aquaculture breeding is based on the same concept as for terrestrial livestock, with genome-wide genotype and phenotype measurements taken on a reference population **[G]** used to train a prediction model, which is then applied to genotyped selection candidates to predict genomic estimated breeding values (gEBVs)[12,68]. Importantly, the high fecundity and large family sizes in aquaculture species offer two major advantages. Firstly, the close relationship between the reference population and the selection candidates enables high selection accuracy **[G]** , even at low marker density, which is likely to be due to long genomic segments shared between close relatives. Secondly, routine phenotyping **[G]** can be performed on these close relatives for different traits and in diverse environments, including 'field' testing in commercial farm settings (Fig. 2). In the past 5 years, the majority of advanced breeding programmes for major aquaculture species have routinely employed genomic selection[66,72], and developments in low-cost genotyping technologies are enabling technology transfer to smaller and more fragmented sectors.

The availability of large full-sibling families can be exploited using within-family genomic selection, with very low-density markers used to estimated gEBVs within families with known pedigree-based EBVs[73]. The increased accuracy of genomic prediction compared to pedigree prediction is evident in a range of aquaculture species, with a median increase in prediction accuracy of 24% for growth-related traits and 22% for disease resistance traits (Table 2). These increases in prediction accuracy are fairly consistent across species and genotyping platforms, with SNP arrays primarily used in high-value species, but GBS giving equivalent findings in several other finfish, crustacean and shellfish

species (Table 2). The majority of studies of genomic selection in aquaculture species use genomic best linear unbiased prediction **[G]** (GBLUP) approaches, which harness genomic relationships to estimate genetic merit of individuals[66]. A range of Bayesian models **[G]** have been tested in several species, but without consistent differences in prediction accuracy compared to the simpler GBLUP approach[66]. Adequate sample size for the genotyped and phenotyped population is key to fully assess the efficacy of genomic selection (for example, >1,000 individuals), but the population structure is equally important, as prediction accuracy is very dependent on the proximity of relationships between animals in the training and validation sets[74]. While several thousand genome-wide markers are also required, it is noteworthy that a reduction in SNP density down to only 1,000 or 2,000 SNPs tends to be sufficient to achieve the asymptote of prediction accuracy where these close relationships exist[66,75]. However, the accuracy drops drastically as the relationship between the reference and test populations becomes more distant, as demonstrated in Atlantic salmon[76] and common carp (*Cyprinus carpio*)[77]; therefore, routine trait measurement and genotyping are required each generation to retrain the genomic prediction models.

[H2] Low-cost solutions for democratizing genomic selection

Capitalizing on the advantages offered by high fecundity in aquaculture breeding programmes requires genotyping of thousands of animals per generation, which can be prohibitively expensive. While genomic selection has become commonplace in a few highly developed aquaculture sectors (for example, salmonids, tilapia and shrimp), genomic tools are yet to be routinely incorporated into breeding programmes for many species (Table 1; Supplementary Table 5). Hence, to translate the benefits of genomic selection to most aquaculture species, there is a clear need to develop cost-effective and species-specific tools, together with effective knowledge transfer to help democratize the technologies. Lower-density SNP panels, potentially typed using targeted GBS techniques (for example, GT-Seq[78]) or fluorescence-based assays, tend to be cheaper than SNP arrays. Low-density genotyping can be integrated with genotype imputation **[G]** to increase the accuracy of genomic selection to levels approaching those obtained with high-density genotyping[79–81].

Imputation relies on genotyping only a subset of the animals at high density (in an aquaculture breeding scheme, typically the parents of the reference population and selection candidates), defining the set of haplotypes in this subset, followed by genotyping offspring

at low density and imputing to high density based on those haplotypes[79]. Considering that breeding programmes for many aquaculture species routinely use low-density SNP panels for parentage assignment[51], combined purpose low-density panels can offer the benefit of genomic selection at little added cost (and may reduce the need for physical tagging). The addition of selected functional markers linked to major QTL would add further value to combined purpose panels to enable concurrent parentage assignment, MAS and imputation-based genomic selection. Further research to develop cost-effective and pragmatic genomic selection approaches is essential to translate its benefits to aquaculture sectors with smaller margins, including in many low- and middle-income countries.

[H2] From sequence to consequence: identifying causative variants for target traits

Mapping and understanding the causative or functional variants that have an impact on complex traits is a fundamental goal of biology but also has potential additional benefits for improving rates of genetic gain in breeding either through improved selection accuracy or as targets for genome editing (Fig. 3). The reduction in prediction accuracy with more distant relationships between reference and validation sets[82] is partly due to the fact that QTL are captured via linked markers rather than causative genetic variants. Research from terrestrial livestock breeding hints at the potential of harnessing whole-genome sequencing data[83], and incorporating weighting of putative functional genomic variants (for example, Bayes RC[84]) into genomic selection models to improve accuracy, although improvements in prediction accuracy have been rather minor in most cases. The use of whole-genome sequencing of key selected individuals (for example, parents) combined with imputation to whole-genome sequences based on genome-wide SNP genotypes will result in population-scale sequence data for aquaculture species and allow testing of such approaches in the near future. However, the cost of whole-genome sequencing and the effectiveness of low-density SNP panels described above mean that substantial improvements in selection accuracy would be necessary to justify its routine use in breeding programmes.

The high fecundity harnessed for sib-testing is also advantageous for high-resolution genetic mapping experiments, and GWAS are used to highlight genomic regions associated with traits of interest. However, such regions often contain hundreds to thousands of candidate causative variants **[G]** and dozens of genes, and most of these variants are in non-coding regions potentially affecting transcriptional regulation. Shortlisting those variants

and genes that are more likely to be causal can be facilitated by employing a pipeline of functional genomics techniques, together with knowledge of the biology of the trait in question (Fig. 3).

Improvements to the annotation of reference genomes of aquaculture species is integral to the process of identification of causative genetic variants. RNA sequencing (RNA-seq) combined with advances in software for read alignment and quantification have facilitated genome-wide prediction of coding and non-coding genes in many aquaculture species[32], replacing microarrays as the standard for global quantification of gene expression. Single-cell RNA-seq is yet to be widely applied to aquaculture species, but offers opportunities to understand complex and rare cell populations and uncover regulatory relationships between genes, thereby improving genome annotation and detection of putative causative variants t[85].

Discovery and exploitation of epigenetic marks in aquaculture species also represents a crucial step to help bridge the genotype–phenotype gap **[G]**[86], and priortise variants for downstream functional testing. Emerging genomic technologies are enabling elucidation of genome-scale patterns of cytosine methylation, chromatin accessibility, histone modifications, transcriptional start sites and transcript variants[87]. These tools enhance the scope to identify putative causative variants within regulatory sequences (for example, enhancers) active under specific environmental conditions (for example, during disease outbreaks). In addition, aquaculture species also benefit from existence of extant and recently diverged wild counterparts, and use of comparative genomics and orthology analysis can help predict functional variants based on sequence conservation[88]. The Functional Annotation of Animal Genomes (FAANG) initiative[89] is a concerted effort to map such features in livestock, with the Functional Annotation of All Salmonid Genomes (FAASG) being an equivalent community initiative for salmonid fish[32], and equivalent initiatives are likely to follow for other major aquaculture species.

Ultimately, the identification of functional variants will require functional studies such as genome editing of a specific allele to assess consequences for the trait of interest in cell culture and/or whole animal systems (see section '*Genome editing to accelerate genetic improvement*' below).

[H2] Towards accurate high-throughput phenotyping

Obtaining accurate phenotypes *en masse* is critical for any breeding programme since the accuracy of trait measurement directly affects genetic gain per generation. Phenotype measurements can be particularly challenging for aquaculture species, because manual measurements prior to harvest typically require handling large numbers of animals outside the water, presenting a logistical and financial challenge. Therefore, the ability to collect such data both directly on the selection candidates in the breeding nucleus, and on relatives of those candidate in test or production environments, can present a limitation to genetic progress in breeding programmes. Computer vision technologies are being widely applied to automate plant and terrestrial livestock phenotyping, and its utility to accurately predict traits of interest has been demonstrated in several aquaculture species[66,90]. Optical sensors and machine vision systems can be used to monitor behavioural and health traits in tank or cage environments, while hyperspectral imaging approaches can inform on fillet content and characteristics[90]. For instance, the use of underwater cameras for real-time *in situ* data collection is being exploited for tasks such as sea lice monitoring in Atlantic salmon farms[91], and their use is likely to expand for more widespread data collection and phenotyping[90]. Connected mobile devices for affordable on-farm monitoring and automation of aquaculture environments (that is, sensor technologies and the 'internet of things **[G]** ') have major potential for monitoring individual traits such as behaviour and feed intake, in parallel to enabling the collection of huge volumes of environmental data. Transforming such data into meaningful phenotypes for breeding is a substantial challenge, and consequently data interpretation and decision tools such as machine learning and artificial intelligence will assume greater importance in aquaculture[92]. Together with routine genomic evaluations, the effective combination of increasingly high-resolution and high-volume phenotyping in breeding nuclei, production environments and post harvest will lead to more precise and effective genetic improvement of aquaculture species.

[H1] Genetics and the environment

[H2] Tackling genotype by environment interactions in aquaculture breeding

The performance and robustness of a farmed animal is dependent on the interaction between its genotype and the environment, which can vary substantially in aquaculture both within and across farms. For example, water quality presents a key challenge with limited

environmental control, resulting in substantial within- and across-farm variation in partial pressure of CO_2, temperature and other parameters. The transition to on-land recirculating aquaculture systems or floating closed-containment systems with close control of environmental conditions is plausible for certain species such as Atlantic salmon[93], but the level of investment required to establish and maintain these systems is substantial and is unlikely to be feasible for the majority of situations. As such, genetic improvement in a breeding programme is intrinsically linked to the environment in which traits are recorded, and G×E interactions commonly result in genotype re-ranking such that the best-performing genotypes in one environment may not be the best in another, placing a limitation to realizing genetic gains in breeding programmes[94,95]. The extent and nature of G×E interactions depend on the trait in question and can be quantified by measuring the genetic correlation between the trait in different environments. Studies across multiple aquaculture species have highlighted that such correlations tend to be positive, albeit only moderate in magnitude for growth and survival traits[94], highlighting the need to account for G×E interactions in aquaculture breeding programmes.

The domestication and genetic improvement of local strains and species, which may be better adapted to the local environment, is one route to reducing the impact of G×E interactions. However, well-managed breeding programmes are expensive, and as such the current trend is consolidation into large and high-tech programmes that harness high fecundity (often including multiplication layers) to disseminate single lines into production facilities worldwide. In this scenario, breeding programmes need to account for G×E interactions to maximize the benefits of genetic improvement[96]. The possibility of disseminating many closely related animals to diverse geographical locations and environmental conditions (Fig. 2) can be coupled with phenotyping technologies for routine data collection to feed back information on performance under diverse settings. This may facilitate production of differentiated strains tailored to specific environments, or inclusion of robustness as a target trait such that a single strain has phenotypic plasticity within and across diverse environments[97]. An example of a robust strain that performs well in multiple environments is the genetically improved farmed tilapia (GIFT) strain. In the late 1970s, inadequate tilapia stocks were hampering the development of aquaculture in Asia. To develop a strain with robust performance in high- and low-input systems across diverse environments, a base population including wild and farmed strains from eight African and

Asian countries was established. The breeding programme focused primarily on improving growth rate, but involved multiple farmers in different countries in evaluations to account for G×E interactions. The GIFT strain is now farmed in 16 countries across Asia, Africa and Latin America and grows 85% faster than the base population[98].

Genomic selection can facilitate the breeding of more robust strains in aquaculture species where reference populations (including close relatives of selection candidates) are tested in diverse environments[94,99]. The performance of a genotype along an environmental gradient for any measurable trait can be used to calculate the response curve, or reaction norm, of that genotype[94]. This reaction norm can be used as a target trait for genomic selection to reduce sensitivity to environmental variation, with notably superior results to sibling testing schemes alone[99]. The variation within and between production environments is typically larger for aquaculture in low- and middle-income countries; as breeding programmes in such settings increase in sophistication, low-cost genomic selection methods should be applied to help improve resilience of stock performance within and across environments to maximize the benefits of genetic gain for producers.

[H2] Epigenetic programming to improve performance and environmental adaptation

Epigenetic mechanisms or 'marks' (for example, cytosine methylation, histone modifications, chromatin accessibility state) can be influenced by the environment to result in substantial phenotypic variation from the same genomic DNA blueprint[86]. Recent domestication can profoundly alter the epigenome of hatchery-reared animals via alterations to the DNA methylation profile[100], highlighting the potential for rapid epigenetic reprogramming. This potential can be harnessed by intentional environmental manipulation during crucial life stages, in particular larvae and broodstock, to improve production traits later in life and/or in subsequent generations[86,101,102]. For example, early-life use of plant-based diets improved the acceptance and utilization of these diets in later life in rainbow trout[103], and early-life stress can modulate future stress or immune responses in Atlantic salmon, which may have implications for robustness in adult stages[104,105]. Multigenerational epigenetic effects are of most relevance to selective breeding and have been proposed to play a role in the fitness of the Manila clam (*Ruditapes philippinarum*), where adults exposed to low pH levels during gonadal maturation had faster-growing offspring compared with controls[106], and in the Sydney rock oyster (*Saccostrea glomerata*), where larvae of parents incubated under low-

pH conditions grew and developed faster in low-pH conditions and had higher fitness as adults[107]. The development of assays to assess genome-wide cytosine modification, chromatin structure and accessibility across multiple aquaculture species will help elucidate the mechanisms underpinning these epigenetic phenomena, and the availability of isogenic finfish lines is a useful resource to help distinguish genetic and epigenetic effects[108].

For heritable epigenetic marks that affect production traits, it is highly likely that their impact will be directly captured and utilized by conventional sib-testing and genomic selection, which are both based on phenotypic similarity between relatives[109]. However, distinguishing additive genetic and epigenetic components of phenotypic variation may facilitate improvement in genetic parameter estimation and prediction of response to selection[102]. Furthermore, an interesting intersection between epigenetic programming and genetic improvement via selective breeding may be related to optimizing of robust performance of improved stocks in multiple environments. The use of genomics to support breeding of 'robust' strains for multiple environments described above can be augmented with tailored epigenetic programming to improve the performance of these strains in specific farmed environments. Furthermore, there is likely to be genetic variation in the response to targeted environmental manipulation, and genomic prediction using large full-sibling families each split into groups tested with targeted environmental treatments can be used to assess this (Fig. 2). Therefore, selection for improved response to epigenetic programming could be a route to realizing genetic improvement for impact across diverse production environments.

[H2] The microbiome as a predictor of performance

The microbiome is a critical component of the interaction between animals and their environment, and contributes to the health and performance of farmed animals[110,111]. Colonization and development of bacterial communities are essential to immune function and influenced by host physiology and immune response. Host microbial composition is heritable to some extent in marine species[112,113], and differences have also been observed between farmed and wild strains of Atlantic salmon[104] and Pacific whiteleg shrimp (*Litopenaeus vannamei*)[114]. Microbiome research in aquaculture species is currently primarily focused on gaining an understanding of its composition in various species[111,115]. Developments in DNA sequencing technologies have provided drastic improvements in microbiome analyses, in particular metagenomics approaches to sequencing all genomes

within a sample. Microbiome sequencing may have potential when paired with host genotyping for the prediction of production traits, with a potential example trait being the ability of salmonids to tolerate increasingly vegetarian diets[116]. In terrestrial livestock, microbiome similarity matrices have been used to replace or complement the host genomic relationship matrix **[G]** , with an improved predictive ability for feed conversion efficiency in Holstein Friesian dairy cattle[12]. In this context, microbiome composition can be considered as an 'intermediate phenotype' resulting from both host genetic and environmental influences, and has potential value in prediction of trait performance in later life, rather than prediction of offspring performance. The latter may depend in part on the heritable component of the microbiome, but is likely to be captured within additive genetic variation and breeding values for production traits.

[H2] Interaction between farmed and wild animals

The recent domestication of aquaculture species means that farmed species often co-exist in close proximity to wild counterparts, with frequent interaction and interbreeding possible between the two groups. As species move along the domestication scale towards closed selective breeding populations, the genetic divergence between farmed and wild populations widens. The genomes of farmed species are significantly altered by domestication and genetic improvement programmes, which exert intense selection pressures[117]. As domestication progresses, high-density genotyping or sequencing of multiple populations of farmed and wild populations, and comparison of genetic diversity across the genome to identify common signatures of selection can be applied to gauge these effects[118,119].

Divergence of wild and farmed populations results in notable differences in growth, morphology, life history, behaviour and physiology[120]. The impact of domestication on animal physiology has been demonstrated via studies of gene expression and genome methylation, which show marked differences after a few generations of hatchery breeding in salmonids[121]. Introgression **[G]** of potentially maladapted alleles into wild populations can lead to undesirable changes in life history traits, reduced population productivity and decreased resilience[122]. Many species of marine fish and invertebrates are characterized by high connectivity, with associated high gene flow, and high effective population size **[G]** [123], such that the effects of introgression from farm-reared animals is rapidly diluted. Such introgression may even be beneficial in some species, for example, bivalve shellfish, by

contributing to natural recruitment and adding genetic variation to wild populations[124,125]. By contrast, freshwater and anadromous species are characterized by fairly small effective population sizes[126], and gene flow can be heavily modified (or blocked)[127,128]. Consequently, inflow of genes from farmed animals can result in rapid and substantial alterations to the gene pool in populations of these species[126]. Therefore, methods of preventing escapees and interbreeding of farmed and wild animals are important for the sustainability of aquaculture and its long-term coexistence with extant wild populations[126,129,130]. Engineering and management solutions are unlikely to completely prevent escapees, and genetic technologies to prevent such introgression include triploidy, currently used in a range of species including salmonids and oysters[131,132], or other means of inducing sterility in production stocks such as germ cell ablation via genome editing[133] (see section '*Genome editing to accelerate genetic improvement'* below).

In addition to protecting wild stocks, it is important to maintain genetic resources for farmed strains as they undergo domestication. Biobanking is applied for conservation of germplasm **[G]** of aquatic animals, both for vulnerable wild species and farmed strains, to avoid losing genetic diversity. There are established repositories and gene banks for finfish and shellfish, and technologies for preservation of gametes, tissues and cell lines are developing rapidly, with detailed reviews available[134,135]. However, the field remains at a fairly early stage compared with equivalent efforts in crops and terrestrial livestock. Whereas cryopreservation of sperm is routine for several fish and shellfish species, the cryopreservation of oocytes is much more challenging to achieve. Cryopreservation of ovarian tissues is a promising alternative but would require research into the in vitro culture of these tissues[135]. Surrogate broodstock (discussed below) hold promise to preserve genetic resources through transplant of primordial germ cells **[G]**[136]. As these methods develop, preservation of aquatic genetic resources will benefit from more centralized efforts, akin to seedbanks for crops, together with associated FAO standards and procedures for biobanking[137].

[H1] Biotechnology in aquaculture breeding

Biotechnological innovations hold promise to tackle production barriers in aquaculture. These advances include the use of genome editing technologies to make targeted changes to the genomes of aquaculture species', resulting in improved health and performance, use

of reproductive biotechnologies such as surrogate broodstock to expedite genetic gain, and combinations of both approaches.

[H2] Genome editing to accelerate genetic gain

Genome editing tools such as engineered CRISPR–Cas9 systems[138,139] are invaluable to understanding genetic regulation of economically-important traits and have potential to accelerate genetic gain in aquaculture breeding programmes (Fig. 3). The Cas9 enzyme makes a DNA double-strand cut at a genomic site corresponding to a guide RNA, which results in either small insertions or deletions that can lead to loss-of-function mutations (non-homologous end joining) or user-defined edits to the genome based on a provided DNA template (homology-directed repair). Since the first demonstration of effective genome editing in Atlantic salmon[140], CRISPR–Cas9 has been successfully applied in various farmed finfish and mollusc species, primarily for gene knockout and as proof of principle[141]. Microinjection into early-stage embryos is the most commonly used delivery method but can be inefficient, and alternative delivery methods, such as electroporation of sperm, hold promise[142]. Genome editing can be used as a component of pipelines to identify putative causative genes and variants, for example, by assessing the effect of gene knockouts on traits of interest. Exploitation of genome-wide loss of function CRISPR screens such as GeCKO (Genome-scale CRISPR Knock-Out)[143] in aquaculture species offers a powerful tool to explore the genetic basis for resistance to certain pathogens; successful editing of a salmonid fish cell line using a lentivirus delivery system suggests that this approach is technically viable[144]. However, cell line resources for many aquaculture species, in particular invertebrate species, are limited, and targeted development of suitable cell lines for important aquaculture species is required. As an alternative, *in vivo* GeCKO may be plausible in species with external fertilization, abundance of embryos and feasible early-life screens[141]. This approach is likely to require the development of Cas9-stable broodstock and a method of delivering guide RNA libraries *en masse* to early-stage embryos. Combining such genome-wide screening approaches with mapping, and shortlisting causative functional variants in QTL regions, will create opportunities for targeted experiments testing candidate causative alleles, followed by assessment of the consequences on the trait (Fig. 3).

Several potential applications of genome editing could expedite genetic improvement. Firstly, it could enable the rapid fixation of favourable alleles at QTL segregating within

breeding populations[145]. Secondly, genome editing could facilitate introgression-by-editing of favourable alleles from other populations, strains or species, potentially including wild stocks, into a breeding population[141]. Finally, it is possible to create *de novo* alleles based on knowledge of the biology of the trait in question, or utilizing targets from GeCKO screens. For example, removal of an exon of the *CD163* gene in pigs (*Sus scrofa*) resulted in complete resistance to the porcine reproductive and respiratory syndrome virus[146].

Although disease resistance is likely to be the primary focus for genome editing in aquaculture, other traits, such as adaptation of stocks to plant-based diets or sterility to prevent introgression and unwanted effects of precocious maturity[147,148], are additional key objectives. For example, knockout of the germline-specific genes *dnd* in Atlantic salmon[133] and *nanos2* or *nanos3* in Nile tilapia[149] resulted in sterility. For practical applications, genome editing needs to be integrated into well-managed breeding programmes to ensure maintenance of genetic diversity. Genome editing *en masse* in production animals is unlikely to be feasible and, therefore, editing of the germline of broodstock animals is highly likely to be the most effective approach. Sterility requires special consideration because it is by definition non-heritable, and inducible transgenic targets may be required. However, sterility may be a useful trait to include with other genome editing targets to negate the risk of edited alleles being transferred to wild stocks (for example, via escapees).

Refinement of genome editing methods are occurring constantly, and use of modified CRISPR–Cas systems such as CRISPR activation (CRISPRa) or CRISPR interference (CRISPRi) can induce differences in expression levels of target genes instead of complete knockout[150–152]. Such tools will be valuable in elucidating the functional genetic basis of production traits, for fundamental understanding of genome function and for future application in aquaculture breeding programmes. However, it is critical that edited stocks are carefully studied to detect and avoid off-target editing and rigorously monitored to discount deleterious pleiotropic effects **[G]**; aquaculture can follow procedures used in terrestrial livestock to achieve these goals[153]. Furthermore, any practical application for aquaculture depends entirely on an acceptable regulatory and public approval landscape[154], and the approval of the genetically-modified AquaAdvantage salmon (Aquabounty) as fit for human consumption by the US FDA and the Canadian Food Inspection Agency was a recent landmark[155]. Target traits that have concurrent production and animal welfare or

environmental benefits should be a focus for genome editing in aquaculture, and public and policy-maker engagement on the technology, its benefits and its risks is absolutely vital.

[H2] Surrogate broodstock to reduce generation intervals

A key factor in the rate of genetic gain in a breeding programme is the length of the generation interval. Consider the breeder's equation;

$$\Delta G = \frac{i\, r\, \sigma_A}{y}$$

where ΔG is genetic gain over time, i is selection intensity **[G]** , r is selection accuracy, σ_A is additive genetic variance, and y is generation time. Genomic selection has resulted in a step increase in selection accuracy, and much research is now devoted to achieving further minor increases[66]. However, decreasing generation time has potential for more drastic changes to genetic gain, especially considering that many of the major aquaculture species have relatively long generation intervals (for example, up to 20 years in sturgeon, family *Acipenseridae*). Surrogate broodstock technologies are based on the concept of isolation of the primordial germ cells of selected broodstock animals at an early life stage, and transplantation of these cells into the surrogate, that is, a germ cell-ablated specimen of a species with a shorter generation time (Fig. 4). When combined with genomic selection to predict breeding values of embryos or juveniles, surrogate broodstock technology could potentially reduce the generation interval without substantial loss of selection accuracy. Germ cell isolation, transplantation and successful gamete production in surrogate broodstock have been demonstrated across species within a genus, and even across genera[156], for example, rainbow trout offspring were produced when spermatogonia from rainbow trout were injected into newly-hatched sterile masu salmon (*Oncorhynchus masou*)[157]. The same technology has other potential applications, for example, to produce offspring from a species which is challenging to rear in captivity using surrogates, such as production of Atlantic bluefin tuna (*Thunnus thynnus*) gametes from chub mackerel (*Scomber japonicus*) as a surrogate[156]. In addition, surrogate technology can be coupled with genome editing of primordial germ cells to create germline-edited animals, as successfully demonstrated in chickens[158]. This approach is a route to genome editing for aquaculture species where access to the newly fertilized embryos is challenging, such as certain crustaceans[159] or ovoviviparous **[G]** species such as rockfish (*Sebastes* spp.)[160].

While clearly a long-term and high-risk research goal, the combination of surrogate broodstock, genome editing and genomic selection has potential to drastically increase the rate of genetic gain in breeding programmes via the reduction of generation interval. Extensive effort and resources have been put into the use of functional genomic data to improve selection accuracy in breeding, and reproductive technologies require equivalent attention.

[H1] Conclusions

In contrast to terrestrial livestock and crop production, most aquaculture production derives from species for which domestication and breeding is at an early stage. Genetic improvement and dissemination of germplasm originating from a well-managed breeding programme enables cumulative increases in production traits, and facilitates adaptation to emerging challenges, such as climate change or infectious disease outbreaks. Due to recent growth and improved availability, genomics should be utilized from the outset of domestication and breeding programme design to inform base population composition, maintain genetic diversity and understand sex determination and differentiation. Genomic selection has revolutionized terrestrial livestock breeding and is commonplace in advanced aquaculture sectors such as salmon, but judicious application of multi-purpose cost-effective marker panels may be necessary to translate those benefits to most aquaculture species for which the industries are smaller and more fragmented.

The ability to disseminate closely related individuals to diverse testing and production environments, combined with genomic selection, should be applied to tackle G×E interactions and improve robustness. Genomic tools can also inform on the potential of the microbiome and epigenome as useful intermediate phenotypes, and as conduits to improve capacity for adaptation of stocks to environmental challenges. For the more advanced aquaculture sectors, the immediate future will include mapping and understanding functional genomic variants, harnessing the species' high fecundity to perform high-resolution genetics and genomics experiments paired with highly contiguous and well-annotated genome assemblies. Genome editing is key to this process and as such requires species-specific optimization both *in vivo* and in cell culture, with the development of suitable cell lines for aquaculture species being an important focus, for example, to assist with genome-wide CRISPR screens for disease resistance. The widespread commercial application of genome

editing in aquaculture seems to be several years away, but it has clear potential for step-changes in trait improvement to help address production barriers. In the longer term, developments in surrogate broodstock technology combined with genomic selection have the potential for shortening generation intervals to expedite genetic gain.

Underpinning many of these advances is an improved knowledge of the genetics and biology of key production traits, which is particularly pertinent for the many aquaculture species from understudied taxa with major knowledge gaps relating to fundamental inheritance and genome biology. Overall, there is now an unprecedented opportunity to harness genomics to fast-track the domestication and genetic improvement of farmed aquatic species, which will be necessary to secure the sustainable growth of aquaculture as one of the most promising solutions to the current global food security challenge.

Table 1. Genomic resources for aquaculture species with highest production value

Species	**Production value** (US$Bn)	**Genome size** (Gbp)	**Scaffold N50** (Mbp)	**Coding genes**	**Published SNP arrays**	**Re-sequenced genomes**
Finfish						
Atlantic salmon (*Salmo salar*)	16.69	2.96	1.36	48,775	7 (15–286K)	165
Grass carp (Ctenopharyngodon idella)	12.64	0.90	6.45	27,263	-	1
Silver carp (*Hypophthalmichthys molitrix*)	10.26	1.10	0.31	-	-	-
Nile tilapia (*Oreochromis niloticus*)	7.61	1.00	38.8	29,550	2 (50–58K)	65
Bighead carp (*Hypophthalmichthys nobilis*)	7.31	1.01	0.08	-	-	-
Crustaceans						
Whiteleg shrimp (*Litopenaeus vannamei*)	26.74	1.63	0.6	24,987	1 (6K)	-
Red swamp crawfish (*Procambarus clarkii*)	10.00	2.07	0.001	136,962	-	-
Chinese mitten crab (*Eriocheir sinensis*)	9.54	1.54	0.49	-	-	-
Giant tiger prawn (*Penaeus monodon*)	5.59	1.44	0.007	18,115	1 (6K)	2
Oriental river prawn	2.09	-	-	-	-	-

(*Macrobrachium nipponense*)						
Molluscs						
Japanese carpet shell (*Ruditapes philippinarum*)	6.95	2.56	0.048	108,034	-	15
Chilean mussel (*Mytilus platensis*)	2.50	-	-	-	-	-
Constricted tagelus (*Sinonovacula constricta*)	1.41	-	-	-	-	-
Pacific cupped oyster (*Crassostrea gigas*)	1.24	0.55	0.4	28,398	2 (27–190K)	516
Blood cockle (*Tegillarca granosa*)	1.02	-	-	-	-	-
Echinoderms						
Japanese sea cucumber (*Apostichopus japonicus*)	1.40	0.8	0.48	30,350	-	1

Full data provided for the top 20 species per each taxonomic group in Supplementary Table 5.

Table 2 | Summary of studies testing genomic prediction for production traits in aquaculture species.

Species	Trait	Measurement	Heritability (pedigree)	Accuracy (pedigree)	Relative increase	Genotyping technology	Ref
Atlantic salmon (*Salmo salar*)	Growth	Weight	0.60 (0.48)	0.70 (0.58)	21%	SNP array (132K, 112K post-filtering)	[161]
		Length	0.61 (0.51)	0.66 (0.56)	18%		[161]
	Resistance to sea lice	Lice count	0.33 (0.27)	0.60 (0.48)	25%	SNP array (132K, 33K post-filtering)	[162]
		Lice count	0.22(0.27)	0.46 (0.43)	7%		[162]
		Lice count	0.11 (0.10)	0.50 (0.41)	22%	SNP array (50K, 37K post-filtering)	[163]
		Log lice density	(0.14)	0.52 (0.34)	52%	SNP array (220K)	[70]
	Resistance to amoebic gill disease	Gill score	0.24 (0.25)	0.62 (0.51)	22%	Two species SNP array (17K, 7K post-filtering)	[164]
		Amoebic load	0.25 (0.36)	0.70 (0.60)	17%		[164]
		Gill score	0.28 (0.32)	0.72 (0.61)	18%	SNP array (55K, 53K post-filtering)	[165]

	Resistance to salmon rickettsial syndrome	Time to death	0.27 (0.18)	0.41[a] (0.34)	21%	SNP array (50K, 50K post-filtering)	[166]
		Binary survival	0.39 (0.26)	0.26 (0.20)	30%		[166]
	Fillet pigmentation	-	(0.43)	0.44 (0.36)	22%	SNP array (220K)	[70]
	Muscle fat	-	0.25 (0.36)	0.56 (0.60)	-7%	SNP array (57K, 50K post-filtering)	[167]
	Omega-3 fatty acid content	DHA	0.20 (0.21)	0.41 (0.33)	24%		[167]
		EPA	0.04 (0.06)	0.32 (0.37)	-14%		[167]
Rainbow trout (*Oncorhynchus mykiss*)	Resistance to bacterial cold water disease	Binary survival	–	0.68[a] (0.36)	89%	SNP array (57K, 45K post-filtering)	[168]
		Time to death	0.33 (0.37)	0.67[a] (0.34)	97%	SNP array (57K, 36K post-filtering)	[169]
		Binary survival	0.35 (0.35)	0.70[a] (0.36)	94%		[169]
		Time to death	0.29 (0.31)	0.49 (0.50)	-2%	SNP array (57K, 41K post-filtering)	[170]
		Binary survival	0.45 (0.48)	0.46 (0.41)	12%		[170]
	Resistance to infectious pancreatic necrosis virus	Time to death	0.25 (0.40)	0.53 (0.49)	8%	SNP array (57K, 38K post-filtering)	[171]
		Binary survival	0.24 (0.35)	0.56 (0.50)	12%		[171]
	Resistance to salmon rickettsial syndrome	Time to death	0.45 (0.38)	0.78[a] (0.61)	28%	SNP array (57K, 27K post-filtering)	[172]
		Binary survival	0.55 (0.54)	0.60[a] (0.47)	28%		[172]
	Resistance to Infectious haematopoietic necrosis virus	Time to death	0.23 (0.33)	0.33 (0.13)	154%	SNP array (57K, 35K post-filtering)	[173]
		Binary survival	0.25 (0.28)	0.39 (0.24)	63%		[173]
	Resistance to columnaris disease	Binary survival	0.32 (–)	0.11 (-0.02)	-650%	SNP array (57K, 36K post-filtering)	[174]
		Binary survival	0.51 (–)	0.22 (0.06)	267%	SNP array (57K, 34K post-filtering)	[174]
Coho salmon (*Oncorhynchus kisutch*)	Resistance to salmon rickettsial syndrome	Time to death	– (0.14)	0.52 (0.27)	93%	ddRAD (9K)	[175]
		Binary survival	– (0.27)	0.81 (0.31)	161%		[175]
	Growth	Length	0.33 (0.33)	0.71 (0.60)	18%	RAD-seq (20K)	[176]

Carp (*Cyprinus carpio*)	Resistance to koi herpesvirus	Binary survival	0.50 (0.61)	0.53[a] (0.49)	8%	RAD-seq (16K)	[77]
Nile tilapia (*Oreochromis niloticus*)	Growth	Harvest weight	0.36 (0.31)	0.60 (0.48)	25%	SNP array (43K, 32K post-filtering)	[177]
		Fillet yield	0.21 (0.21)	0.62 (0.54)	15%		[177]
		Harvest weight	0.17 (0.22)	0.29 (0.19)	53%	SNP array (59K, 48K post-filtering)	[178]
		Fillet weight	0.16 (0.24)	0.34 (0.18)	89%		[178].
		Fillet yield	0.23 (0.33)	0.54 (0.46)	17%		[178]
European sea bass (*Dicentrarchus labrax*)	Resistance to viral nervous necrosis	Binary survival	0.43 (0.27)	0.62[a] (0.67)	-7%	RAD-seq (9K)	[179]
Gilthead sea bream (*Sparus aurata*)	Resistance to pasteurellosis	Time to death	0.28 (0.22)	0.44[a] (0.30)	47%	2b-RAD (22K)	[180]
	Resistance to pasteurellosis	Time to death	0.32 (0.32)	0.54[a] (0.45)	20%	2b-RAD (28K)	[181]
		Binary survival	0.33 (0.31)	0.56[a] (0.46)	22%		[181]
Turbot (*Scophthalmus maximus*)	Resistance to Scuticociliatosis	Resilience	0.15 (–)	0.46 (0.41)	12%	2b-RAD (18K)	[182]
		Resistance	0.26 (–)	–	–		[182]
		Endurance	0.12 (–)	–	–		[182]
Japanese flounder (*Paralichthys olivaceus*)	Resistance to Edwardsiella tarda	Binary survival	– (–)	0.603 (–)	–	WGS (1.9M)	[183]
Channel catfish (*Ictalurus punctatus*)	Growth	Harvest weight	0.27 (–)	0.37 (0.29)	28%	SNP array (660K, 55K post-filtering)	[184]
		Residual carcass weight	0.34 (–)	0.31 (0.24)	29%		[184]
Large yellow croaker (*Larimichthys crocea*)	Growth	Body weight	0.60 (–)	0.41 (–)	–	ddRAD (30K)	[185]
		Body length	0.59 (–)	0.40 (–)	–		[185]
	n-3HUFA	–	0.44 (–)	0.30 (–)	–	ddRAD (32K)	[185]
Yellowtail kingfish (*Seriola lalandi*)	Growth	Weight	0.47 (0.42)	0.69 (–)	–	DArT-Seq (14K)	[186]
		Length	0.43 (0.42)	0.67 (–)	–		[186]
		Condition index	0.21 (0.11)	0.44 (–)	–		[186]
Yellow drum (*Nibea albiflora*)	Growth	Body length	– (–)	0.38[a] (–)	–	GBS (54K)	[187]
		Swimming bladder index	– (–)	0.17[a] (–)	–		[187]

		Swimming bladder weight	– (–)	0.22[a] (–)	–		[187]
		Body thickness	– (–)	0.24[a] (–)	–		[187]
		Body height	– (–)	0.30[a] (–)	–		[187]
		Body length / body height ratio	– (–)	0.36[a] (–)	–		[187]
		Gonad weight index	– (–)	0.37[a] (–)	–		[187]
Oyster (*Crassostrea gigas*)	Growth	Shell length	0.26 (0.23)	0.54 (0.44)	23%	Two species SNP array (38K, 23K post-filtering)	[188]
		Shell height	0.23 (0.20)	0.60 (0.47)	28%		[188]
		Wet weight	0.35 (0.31)	0.67 (0.54)	24%		[188]
	Resistance to Osterid Herpesvirus	Binary survival	0.37 (0.25)	0.76 (0.64)	19%		[189]
Yesso scallop (*Patinopecten yessoensis*)	Growth	Shell height	0.48 (–)	0.53 (–)	–	2b-RAD (2K)	[190]
		Shell length	0.48 (–)	0.46 (–)	–		[190]
		Shell width	0.36 (–)	0.55 (–)	–		[190]
Zhikong scallop (*Chlamys farreri*)	Growth	Shell length	0.42 (–)	0.65[a] (–)	–	2b-RAD (31K)	[191]
		Shell height	0.47 (–)	0.70[a] (–)	–		[191]
		Shell width	0.54 (–)	0.63[a] (–)	–		[191]
		Whole weight	0.28 (–)	0.64[a] (–)	–		[191]
Whiteleg shrimp (*Litopenaeus vannamei*)	Growth	Body weight	0.32 (–)	0.62 (–)	–	2b-RAD (23K)	[192]
		Body length	0.45 (–)	0.61 (–)	–		[192]
		Body length	– (–)	0.30[a] (–)	–	SLAF-seq (6K)	[193]
		Body weight	– (–)	0.41[a] (–)	–		[193]
	Resistance to AHPND	Time to death	0.26 (0.24)	0.50 (0.47)	6%	2b-RAD (23K)	[194]
		Binary survival	0.16 (0.15)	0.21 (0.20)	5%		[194]
Banana shrimp (*Fenneropenaeus merguiensis*)	Growth	Body weight	0.55	0.76 (0.65)	17%	DArT-Seq (9K)	[195]
		Body length	0.49	0.73 (0.60)	22%		[195]
		Head length	0.39	0.42 (0.32)	31%		[195]
		Body width	0.61	0.72 (0.60)	20%		[195]
		Tail weight	0.45	0.77 (0.66)	17%		[195]
		Meat yield	0.10	–	–		[195]
	Colour	Dark (raw shrimp)	0.18	0.59 (0.53)	11%		[195]
		Red (cooked shrimp)	0	NA	–		[195]
	'Flesh streaks'	–	0	NA	–		[195]

	Yellow hepatopancreas	–	0.03	NA	–		[195]
	Resistance to HPV	Viral load	0.35	0.60 (0.09)	567%		[195]

[a]Alternative statistical models to GBLUP were used, for example, Bayesian models or Ridge Regression Best Linear Unbiased Predictor.

Fig. 1 | A summary of global aquaculture diversity and production. a| Phylogenetic tree showing farmed species with an annual production value higher than US$1,000M per annum (Supplementary Table 6). Estimated divergence times are from refs [196–202]. **b|** The time at which species were first farmed or domesticated including species which account for 80% of all farmed seafood production and 95% of all meat globally. Arrow in the bar denotes the point at which the first scientifically-driven selective breeding studies were undertaken for each species (note, this could not be identified precisely for chickens or goats). Fading of timelines denotes uncertainty (Supplementary Tables 1,2,4.) **c|** Seafood production globally by sector and continent[2] (Supplementary Table 7).

Fig. 2 | Genomic selection within an aquaculture breeding programme. Full-siblings from a number of families are split into selection candidates and animals for phenotypic evaluation. These full-siblings of the selection candidates can be grown in different environmental conditions and phenotyped for different traits, for example using pathogen challenges to estimate resistance to different diseases or measuring performance traits in diverse production environments. The selection candidates and their phenotyped full-siblings are all genotyped, and a genomic relationship matrix reflecting the genetic similarity between each pair of animals is built. This relationship matrix and the collected phenotypes enable the estimation of breeding values for the selection candidates through the use of genomic selection models such as GBLUP or Bayesian models [12].

Fig. 3 | Discovering functional variants using genomics and genome editing. Three complementary strategies to discover causative variants affecting traits of interest for aquaculture breeding are represented. The first is ‘Mapping and understanding QTL’ which harnesses GWAS and within-family QTL mapping approaches to detect genomic regions associated with these traits, followed by functional genomic comparison of animals carrying alternate genotypes at the identified QTL. Identified SNPs within the region of candidate genes are then annotated according to their position in the genome to prioritise them as targets for validation using CRISPR–Cas9 genome editing. The second is ‘Comparative genomics’ where two closely related species that differ for a high priority trait (for example, resistance to sea lice) are compared using comparative and functional genomics, again leading to potential genome editing targets for validation. The third is ‘Reverse genetics’

where pooled, genome-wide CRISPR screens can be applied in cell culture, followed by screening based on markers of infection or resistance to infection to identify key genes involved in disease resistance. The high fecundity of aquaculture species may allow analogous approaches *in vivo* using Cas9 transgenic broodstock followed by screening of embryos or juveniles. The three categories of functional variants identified in the inner circle all have potential for genetic improvement, either via marker-assisted or functionally-enriched genomic selection, or directly via genome editing of broodstock after a further testing and validation phase of research.

Fig. 4 | Potential application of surrogate broodstock technology to accelerate genetic gain. This approach involves the transplantation of germ cells from a donor species (target) to a recipient species (surrogate), which then produces gametes of the donor. The main interest for aquaculture is to transfer the germ cells of the selected breeders of the farmed species to a surrogate that is easier to maintain in captivity and has a shorter generation time, reducing the time between two successive rounds of selection. This approach ensures the success of production and accelerates the rate of genetic gain of the breeding programme. The germ cells of the surrogate must be ablated before transplantation. In this respect, germ cell free animals can be obtained through chromosome set manipulation (i.e. triploidy [157]) or the functional manipulation of genes fundamental for germ cell survival (for example, through genome editing[133]).

Box 1 | **A roadmap for genomics tools matched to different stages of the domestication process**

Historically, the mismanagement of genetic resources and diversity during the domestication process has led to reduced genetic resilience[39] and the subsequent emergence of ‘crowd’ diseases in farmed populations[203], which can be catastrophic for emerging industries. Targeted use of appropriate genomic tools throughout the domestication process can help to retain genetic diversity in both wild and farmed populations, which is likely to contribute to mitigation or prevention of these issues.

Genomic tools have already made substantial contributions to the optimization of scientific breeding programmes and to proactive species conservation strategies for both farmed and wild populations of target species[204,205]. Given recent and rapid technological

developments, together with improved accessibility and cost-efficiency, optimal genomic tools can be applied at each stage of the progression along the domestication and selective breeding continuum (see the figure). For example, cleaner fish, such as Ballan wrasse (*Labrus bergylta*) and lumpfish (*Cyclopterus lumpus*), are used in commercial salmon production to eat sea lice from the skin of the salmon, and are a key aspect of integrated pest management. Wrasse and lumpfish[206] production began in 2009 and 2011[207], respectively, with life cycles in captivity closed in 2018 and 2016[208] and reference genomes released by 2016[209] and 2018[210]. Both domestication processes have combined animal biology, health management and nutritional requirements with the development of genomic tools for genetic management and enhancement[208]. The aforementioned trial crosses, which are crucial when establishing base populations for breeding, can be performed in combination with the cost-effective genotyping by sequencing (GBS), and both phenotype and genomic information can be used to optimize broodstock selection. This process should run concurrently with evaluation of wild stock population structure, using the same genomic tools to inform management strategies for species conservation and rapid diagnostics of genetic introgression[204] (see the figure).

When moving towards more advanced selective breeding programmes, bespoke tools such as SNP arrays can be applied, but their cost-effectiveness needs to be considered and contrasted to GBS. Both of these tools can then be applied to understand the genetic architecture of production traits, and to support genomic selection to maximize genetic gain and minimize inbreeding. SNP discovery and high-density genotyping also pave the way for generation of targeted low-density SNP panels, which can have concurrent uses to support parentage assignment, stock management, traceability and low-cost genomic selection. Finally, due to the relative ease of generating reference genome assemblies, they should be created at the outset of the domestication of a new species for aquaculture, as they inform the choice of marker panels for genotyping and subsequent studies to understand the biology of production traits.

Box 2 | **Genetic solutions to major diseases in aquaculture**

Infectious disease outbreaks represent a major and ongoing threat to economic and environmental sustainability of aquaculture[211]. Most farming occurs in open water environments, providing frequent contact with pathogens (including wild reservoirs of

infection), and at high stocking densities conducive to the rapid spread of infection. Outbreaks of single pathogens can destroy national aquaculture industries, as highlighted by outbreaks of Infectious Salmon Anaemia Virus in Chile in 2007–2010[212], and annual losses of shrimp equating to ~10% of the global industry due to White Spot Syndrome Virus[213]. Options to fully mitigate such diseases via vaccination (in finfish only), biosecurity and pharmaceutical interventions are limited in aquaculture systems for several reasons. Firstly, physical handling is logistically and financially challenging; secondly, the open-water nature of many farming systems makes outbreaks difficult to contain; and thirdly, the early-stage of research in many species means there is a paucity of vaccination and/or treatment options for diseases.

The power of genetic and breeding technologies to prevent or mitigate infectious diseases is increasingly recognized. Encouragingly, host resistance to most aquaculture diseases is heritable[214–216], and sibling testing schemes together with genomic selection provide an effective route to breeding more resistant stocks without compromising the biosecurity of the breeding nucleus (Fig. 2). Indeed, disease resistance has become a major component of advanced aquaculture breeding programmes[22], whereas in terrestrial livestock this is limited by logistical and financial challenges relating to routine measurement of disease resistance traits[217].

Refining and optimizing the collection of disease resistance data in both experimental and production environments is an important goal. Disease resistance is typically measured using laboratory-based pathogen challenges of pedigreed populations of animals, using outcomes such as survival or pathogen burden to quantify the resistance traits[214]. However, disease outcomes in an outbreak depend on several epidemiological factors, and new traits such as the propensity of an infected individual to transmit disease have been suggested to have a genetic basis in farmed fish[218]. Benchmarking disease resistance traits measured in experimental settings with respect to outcomes in production environments is key to achieving disease prevention and control via improved genetics.

[b1] The example of IPN in salmon

Infectious pancreatic necrosis (IPN) is a viral disease that was one of the primary concerns for salmon farming, particularly around the turn of the 21st century, with frequent outbreaks causing high levels of mortality (up to 90%) to stocks both in freshwater hatcheries and

following transfer to sea cages. Resistance to IPN was shown to be moderately to highly heritable[219], and breeding companies began to implement family-based selection. In parallel, teams from the UK and Norway identified a single major quantitative trait locus (QTL) on chromosome 26 that could explain 80–100% of genetic variation in resistance to IPNV in sea water field trials[220] and experimental freshwater trials[221–223]. High-throughput sequencing subsequently enabled the development of SNP-based genetic tests to predict IPN resistance of salmon without the need for regular disease challenge experiments[224,225]. The practical outcome of these experiments was extensive use of marker-assisted selection (MAS) for the favourable allele in all major salmon breeding programmes, assisted by the fact that the resistance allele is dominant[222,225]. The results were striking, with a sustained decrease in the incidence of IPN outbreaks to near zero (see figure[72]). Follow-up functional studies highlighted marked differences in gene expression response to infection between resistant and susceptible salmon fry[226] and suggested that epithelial cadherin may be part of the mechanism underlying the QTL[225]. However, the exact causative mutations and nature of their effect remain at least partly elusive.

Figure adapted from Ref.[72]

References

1. Anderson, J. L., Asche, F., Garlock, T. & Chu, J. Aquaculture: Its role in the future of food. *Front. Econ. Glob.* **17**, 159–173 (2017).
2. FAO. *FAO yearbook. Fishery and Aquaculture Statistics 2017*. (FAO, 2019).
3. Longo, S. B., Clark, B., York, R. & Jorgenson, A. K. Aquaculture and the displacement of fisheries captures. *Conserv. Biol.* **33**, cobi.13295 (2019).
4. Froehlich, H. E., Runge, C. A., Gentry, R. R., Gaines, S. D. & Halpern, B. S. Comparative terrestrial feed and land use of an aquaculture-dominant world. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 5295–5300 (2018).
5. Ahmed, N., Thompson, S. & Glaser, M. Global Aquaculture Productivity, Environmental Sustainability, and Climate Change Adaptability. *Environ. Manage.* **63**, 159–172 (2019).
6. Jennings, S. *et al.* Aquatic food security: insights into challenges and solutions from an analysis of interactions between fisheries, aquaculture, food safety, human health, fish and human welfare, economy and environment. *Fish Fish.* **17**, 893–938 (2016).
7. Handisyde, N., Telfer, T. C. & Ross, L. G. Vulnerability of aquaculture-related livelihoods to changing climate at the global scale. *Fish Fish.* **18**, 466–488 (2017).
8. Charrier, B., Rolland, E., Gupta, V. & Reddy, C. R. K. Production of genetically and developmentally modified seaweeds: exploiting the potential of artificial selection techniques. *Front. Plant Sci.* **6**, 127 (2015).
9. Kim, J. K., Yarish, C., Hwang, E. K., Park, M. & Kim, Y. Seaweed aquaculture: Cultivation technologies, challenges and its ecosystem services. *Algae* vol. 32 1–13 (2017).
10. Troell, M. *et al.* Does aquaculture add resilience to the global food system? *Proc. Natl. Acad. Sci. U. S. A.* **111**, 13257–63 (2014).
11. Teletchea, F. Animal Domestication: A Brief Overview. in *Animal Domestication* (IntechOpen, 2019). doi:10.5772/intechopen.86783.
12. Georges, M., Charlier, C. & Hayes, B. Harnessing genomic information for livestock improvement. *Nat. Rev. Genet.* **20**, 135–156 (2019).
13. FAO. *The State of the World's Aquatic Genetic Resources for Food and Agriculture.* (FAO, 2019).

This report highlights the value of genetic resources and their potential to enhance

the contributions of aquaculture to food security.

14. Gjedrem, T. & Rye, M. Selection response in fish and shellfish: a review. *Rev. Aquac.* **10**, 168–179 (2018).
15. Hill, W. G. Is continued genetic improvement of livestock sustainable? *Genetics* **202**, 877–81 (2016).
16. Abdelrahman, H. *et al.* Aquaculture genomics, genetics and breeding in the United States: current status, challenges, and priorities for future research. *BMC Genomics* **18**, 191 (2017).
17. Mignon-Grasteau, S. *et al.* Genetics of adaptation and domestication in livestock. *Livest. Prod. Sci.* **93**, 3–14 (2005).
18. Driscoll, C. A., Macdonald, D. W. & O'Brien, S. J. From wild animals to domestic pets, an evolutionary view of domestication. *Proc. Natl. Acad. Sci. U. S. A.* **106 Suppl**, 9971–8 (2009).
19. Harland, J. The origins of aquaculture. *Nat. Ecol. Evol.* 1–2 (2019) doi:10.1038/s41559-019-0966-3.
20. Nguyen, N. H. Genetic improvement for important farmed aquaculture species with a reference to carp, tilapia and prawns in Asia: achievements, lessons and challenges. *Fish Fish.* **17**, 483–506 (2016).
21. Gjedrem, T., Robinson, N. & Rye, M. The importance of selective breeding in aquaculture to meet future demands for animal protein: A review. *Aquaculture* **350–353**, 117–129 (2012).

This review suggests that only 10 % of aquaculture production is derived from selective breeding programmes.

22. Janssen, K., Chavanne, H., Berentsen, P. & Komen, H. Impact of selective breeding on European aquaculture. *Aquaculture* **472**, 8–16 (2017).
23. Kumar, G., Engle, C. & Tucker, C. Factors driving aquaculture technology adoption. *J. World Aquac. Soc.* **49**, 447–476 (2018).
24. Janssen, K., Saatkamp, H. & Komen, H. Cost-benefit analysis of aquaculture breeding programs. *Genet. Sel. Evol.* **50**, 2 (2018).
25. Teletchea, F. & Fontaine, P. Levels of domestication in fish: implications for the sustainable future of aquaculture. *Fish Fish.* **15**, 181–195 (2014).
26. Lien, S. *et al.* The Atlantic salmon genome provides insights into rediploidization.

Nature **533**, 200–205 (2016).

27. Xu, P. *et al.* Genome sequence and genetic diversity of the common carp, Cyprinus carpio. *Nat. Genet.* **46**, 1212–1219 (2014).
28. Ludwig, A., Belfiore, N. M., Pitra, C., Svirsky, V. & Jenneckens, I. Genome duplication events and functional reduction of ploidy levels in sturgeon (Acipenser, Huso and Scaphirhynchus). *Genetics* **158**, 1203–15 (2001).
29. Plough, L. V. Genetic load in marine animals: a review. *Curr. Zool.* **62**, 567–579 (2016).
30. Hollenbeck, C. M. & Johnston, I. A. Genomic tools and selective breeding in molluscs. *Front. Genet.* **9**, 253 (2018).
31. Zhang, X. *et al.* Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. *Nat. Commun.* **10**, 356 (2019).
32. Macqueen, D. J. *et al.* Functional Annotation of All Salmonid Genomes (FAASG): an international initiative supporting future salmonid research, conservation and aquaculture. *BMC Genomics* **18**, 484 (2017).

The FAASG white paper, which followed on from the FAANG equivalent (Andersson et al.) described an initiative to improve annotation of all salmonid genomes, and similar initiatives are likely to follow for other major aquaculture species.

33. van Dijk, E. L., Jaszczyszyn, Y., Naquin, D. & Thermes, C. The third revolution in sequencing technology. *Trends Genet.* **34**, 666–681 (2018).
34. Feron, R. *et al.* Characterization of a Y-specific duplication/insertion of the anti-Mullerian hormone type II receptor gene based on a chromosome-scale genome assembly of yellow perch, Perca flavescens. *bioRxiv* doi:10.1101/717397 (2019).
35. Baird, N. A. *et al.* Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**, e3376 (2008).

The discovery of RAD sequencing and its applications have been instrumental for application of genomics to genetic improvement of aquaculture species.

36. Campbell, C. R., Poelstra, J. W. & Yoder, A. D. What is Speciation Genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. *Biol. J. Linn. Soc.* **124**, 561–583 (2018).
37. Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P. & Houston, R. Applications of genotyping by sequencing in aquaculture breeding and genetics. *Rev. Aquac.* **10**,

670–682 (2018).

38. Fernández, J., Toro, M. Á., Sonesson, A. K. & Villanueva, B. Optimizing the creation of base populations for aquaculture breeding programs using phenotypic and genomic data and its consequences on genetic progress. *Front. Genet.* **5**, 414 (2014).
39. You, W. & Hedgecock, D. Boom-and-bust production cycles in animal seafood aquaculture. *Rev. Aquac.* doi:10.1111/raq.12278 (2018)
40. Doyle, R. W. Inbreeding and disease in tropical shrimp aquaculture: a reappraisal and caution. *Aquac. Res.* **47**, 21–35 (2016).
41. Ashton, D. T., Ritchie, P. A. & Wellenreuther, M. High-density linkage map and QTLs for growth in snapper (Chrysophrys auratus). *G3 (Bethesda)* **9**, 1027–1035 (2019).
42. Wellenreuther, M., Le Luyer, J., Cook, D., Ritchie, P. A. & Bernatchez, L. Domestication and temperature modulate gene expression signatures and growth in the Australasian snapper Chrysophrys auratus. *G3 (Bethesda).* **9**, 105–116 (2019).
43. Ashton, D. T., Hilario, E., Jaksons, P., Ritchie, P. A. & Wellenreuther, M. Genetic diversity and heritability of economically important traits in captive Australasian snapper (Chrysophrys auratus). *Aquaculture* **505**, 190–198 (2019).
44. Devlin, R. H. & Nagahama, Y. Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture* **208**, 191–364 (2002).
45. Kobayashi, Y., Nagahama, Y. & Nakamura, M. Diversity and plasticity of sex determination and differentiation in fishes. *Sex Dev.* **7**, 115–25 (2013).
46. Martínez, P. *et al.* Genetic architecture of sex determination in fish: applications to sex ratio control in aquaculture. *Front. Genet.* **5**, 340 (2014).
47. Palaiokostas, C. *et al.* Mapping and validation of the major sex-determining region in Nile tilapia (Oreochromis niloticus L.) Using RAD sequencing. *PLoS One* **8**, e68389 (2013).
48. Palaiokostas, C. *et al.* Mapping the sex determination locus in the Atlantic halibut (Hippoglossus hippoglossus) using RAD sequencing. *BMC Genomics* **14**, 566 (2013).
49. Palaiokostas, C. *et al.* A new SNP-based vision of the genetics of sex determination in European sea bass (Dicentrarchus labrax). *Genet. Sel. Evol.* **47**, 68 (2015).

50. Shi, X. *et al.* Female-specific SNP markers provide insights into a WZ/ZZ sex determination system for mud crabs Scylla paramamosain, S. tranquebarica and S. serrata with a rapid method for genetic sex identification. *BMC Genomics* **19**, 981 (2018).
51. Vandeputte, M. & Haffray, P. Parentage assignment with genomic markers: a major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. *Front. Genet.* **5**, 432 (2014).
52. Khang, P. Van, Phuong, T. H., Dat, N. K., Knibb, W. & Nguyen, N. H. An 8-Year Breeding Program for Asian Seabass Lates calcarifer: Genetic Evaluation, Experiences, and Challenges. *Front. Genet.* **9**, 191 (2018).
53. Reed, D. H. & Frankham, R. How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution (N. Y).* **55**, 1095–1103 (2001).
54. Bentsen, H. B. *et al.* Genetic improvement of farmed tilapias: growth performance in a complete diallel cross experiment with eight strains of Oreochromis niloticus. *Aquaculture* **160**, 145–173 (1998).
55. Govindaraju, D. R. An elucidation of over a century old enigma in genetics—Heterosis. *PLOS Biol.* **17**, e3000215 (2019).
56. Fuji, K. *et al.* Marker-assisted breeding of a lymphocystis disease-resistant Japanese flounder (Paralichthys olivaceus). *Aquaculture* **272**, 291–295 (2007).
57. Fuji, K. *et al.* Identification of a single major genetic locus controlling the resistance to lymphocystis disease in Japanese flounder (Paralichthys olivaceus). *Aquaculture* **254**, 203–210 (2006).
58. Liu, S. *et al.* Retrospective evaluation of marker-assisted selection for resistance to bacterial cold water disease in three generations of a commercial rainbow trout breeding population. *Front. Genet.* **9**, 286 (2018).
59. Ayllon, F. *et al.* The vgll3 locus controls age at maturity in wild and domesticated Atlantic salmon (Salmo salar L.) males. *PLoS Genet.* **11**, e1005628 (2015).
60. Barson, N. J. *et al.* Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature* **528**, 405–408 (2015).

Together with Ayllon et al., Barson et al. showed the impact of a major effect locus on an important life history and production trait in salmon.

61. Gonen, S. *et al.* Mapping and validation of a major QTL affecting resistance to pancreas disease (salmonid alphavirus) in Atlantic salmon (Salmo salar). *Heredity (Edinb).* **115**, 405–414 (2015).
62. Boison, S. *et al.* QTLs associated with resistance to cardiomyopathy syndrome in Atlantic salmon. *J. Hered.* (2019) doi:10.1093/jhered/esz042.
63. Hillestad, B. & Moghadam, H. K. Genome-wide association study of piscine myocarditis virus (PMCV) resistance in Atlantic salmon (Salmo salar). *J. Hered.* (2019) doi:10.1093/jhered/esz040.
64. Gu, X. H. *et al.* Identifying a major QTL associated with salinity tolerance in Nile Tilapia using QTL-Seq. *Mar. Biotechnol.* **20**, 98–107 (2018).
65. Houston, R. D. Future directions in breeding for disease resistance in aquaculture species. *Rev. Bras. Zootec.* **46**, 545–551 (2017).
66. Zenger, K. R. *et al.* Genomic selection in aquaculture: application, limitations and opportunities with special reference to marine shrimp and pearl oysters. *Front. Genet.* **9**, 693 (2019).
67. Gjedrem, T. & Baranski, M. *Selective breeding in aquaculture : an introduction*. (Springer Science & Business Media, 2010).
68. Meuwissen, T. H., Hayes, B. J. & Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–29 (2001).

This study highlighted the potential of use of genome-wide markers for prediction of breeding values, a technique now widely applied in advanced aquaculture breeding programmes.

69. Houston, R. D. *et al.* Development and validation of a high density SNP genotyping array for Atlantic salmon (Salmo salar). *BMC Genomics* **15**, 90 (2014).
70. Odegård, J. *et al.* Genomic prediction in an admixed population of Atlantic salmon (Salmo salar). *Front. Genet.* **5**, 402 (2014).

This was the first empirical study of genomic selection in aquaculture, highlighting the benefit compared to pedigree approaches for predicting breeding values.

71. Tsai, H. Y. *et al.* The genetic architecture of growth and fillet traits in farmed Atlantic salmon (Salmo salar). *BMC Genet.* **16**, 51 (2015).
72. Norris, A. 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 (2017).

73. Lillehammer, M., Meuwissen, T. H. E. & Sonesson, A. K. A low-marker density implementation of genomic selection in aquaculture using within-family genomic breeding values. *Genet. Sel. Evol.* **45**, 39 (2013).
74. Daetwyler, H. D., Calus, M. P. L., Pong-Wong, R., de los Campos, G. & Hickey, J. M. Genomic prediction in animals and plants: Simulation of data, validation, reporting, and benchmarking. *Genetics* vol. 193 347–365 (2013).
75. Kriaridou, C., Tsairidou, S., Houston, R. D. & Robledo, D. Genomic Prediction Using Low Density Marker Panels in Aquaculture: Performance Across Species, Traits, and Genotyping Platforms. *Front. Genet.* **11**, 124 (2020).
76. Tsai, H. Y. *et al.* Construction and annotation of a high density SNP linkage map of the Atlantic salmon (Salmo salar) genome. *G3 GENES, GENOMES, Genet.* **6**, 2173–2179 (2016).
77. Palaiokostas, C. *et al.* Optimizing genomic prediction of host resistance to koi herpesvirus disease in carp. *Front. Genet.* **10**, 543 (2019).
78. Campbell, N. R., Harmon, S. A. & Narum, S. R. Genotyping-in-Thousands by sequencing (GT-seq): a cost effective SNP genotyping method based on custom amplicon sequencing. *Mol. Ecol. Resour.* **15**, 855–867 (2015).
79. Tsai, H.-Y. *et al.* Genotype imputation to improve the cost-efficiency of genomic selection in farmed Atlantic salmon. *G3 (Bethesda).* **7**, 1377–1383 (2017).

This study highlighted a potentially cost-efficient approach to genomic selection in aquaculture that could help democratize the use of the technology to smaller aquaculture sectors.

80. Yoshida, G. M. *et al.* 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**, 147–154 (2018).
81. Tsairidou, S., Hamilton, A., Robledo, D., Bron, J. E. & Houston, R. D. Optimizing Low-Cost Genotyping and Imputation Strategies for Genomic Selection in Atlantic Salmon. *G3 (Bethesda).* **10**, 581–590 (2020).
82. Daetwyler, H. D., Calus, M. P. L., Pong-Wong, R., de los Campos, G. & Hickey, J. M. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics* **193**, 347–365 (2013).

83. Ni, G., Cavero, D., Fangmann, A., Erbe, M. & Simianer, H. Whole-genome sequence-based genomic prediction in laying chickens with different genomic relationship matrices to account for genetic architecture. *Genet. Sel. Evol.* **49**, 8 (2017).
84. MacLeod, I. M. *et al.* Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics* **17**, 144 (2016).
85. Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* **50**, 96 (2018).
86. Gavery, M. R. & Roberts, S. B. Epigenetic considerations in aquaculture. *PeerJ* **5**, e4147 (2017).
87. Giuffra, E., Tuggle, C. K. & FAANG Consortium. Functional Annotation of Animal Genomes (FAANG): current achievements and roadmap. *Annu. Rev. Anim. Biosci.* **7**, 65–88 (2019).
88. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
89. Andersson, L. *et al.* Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biol.* **16**, 57 (2015).
90. Saberioon, M., Gholizadeh, A., Cisar, P., Pautsina, A. & Urban, J. Application of machine vision systems in aquaculture with emphasis on fish: state-of-the-art and key issues. *Rev. Aquac.* **9**, 369–387 (2017).
91. Føre, M. *et al.* Precision fish farming: a new framework to improve production in aquaculture. *Biosyst. Eng.* **173**, 176–193 (2018).
92. Liakos, K. *et al.* Machine learning in agriculture: a Review. *Sensors* **18**, 2674 (2018).
93. Badiola, M., Basurko, O. C., Piedrahita, R., Hundley, P. & Mendiola, D. Energy use in Recirculating Aquaculture Systems (RAS): a review. *Aquac. Eng.* **81**, 57–70 (2018).
94. Sae-Lim, P., Gjerde, B., Nielsen, H. M., Mulder, H. & Kause, A. A review of genotype-by-environment interaction and micro-environmental sensitivity in aquaculture species. *Rev. Aquac.* **8**, 369–393 (2016).
95. Nguyen, N. H. Genetic improvement for important farmed aquaculture species with a

reference to carp, tilapia and prawns in Asia: achievements, lessons and challenges. *Fish Fish.* **17**, 483–506 (2016).

96. Sae-Lim, P. *et al.* Genetic (co)variance of rainbow trout (Oncorhynchus mykiss) body weight and its uniformity across production environments. *Genet. Sel. Evol.* **47**, 46 (2015).
97. Saltz, J. B. *et al.* Why does the magnitude of genotype-by-environment interaction vary? *Ecol. Evol.* **8**, 6342–6353 (2018).
98. Eknath, A. E. & Acosta, B. O. *Genetic Improvement of Farmed Tilapias (GIFT) project: final report, March 1988–December 1997*. (1998).
99. Mulder, H. A. Genomic Selection Improves Response to Selection in Resilience by Exploiting Genotype by Environment Interactions. *Front. Genet.* **7**, 178 (2016).
100. Luyer, J. Le *et al.* Parallel epigenetic modifications induced by hatchery rearing in a Pacific salmon. *Proc. Natl. Acad. Sci.* **114**, 12964–12969 (2017).
101. Jonsson, B. & Jonsson, N. Early environment influences later performance in fishes. *J. Fish Biol.* **85**, 151–188 (2014).
102. Moghadam, H., Mørkøre, T. & Robinson, N. Epigenetics—potential for programming fish for aquaculture? *J. Mar. Sci. Eng.* **3**, 175–192 (2015).
103. Geurden, I. *et al.* Early-feeding exposure to a plant-based diet improves its future acceptance and utilization in rainbow trout. *Commun. Agric. Appl. Biol. Sci.* **78**, 157–60 (2013).
104. Uren Webster, T. M., Consuegra, S., Hitchings, M. & Garcia de Leaniz, C. Interpopulation variation in the Atlantic salmon microbiome reflects environmental and genetic diversity. *Appl. Environ. Microbiol.* **84**, e00691-18 (2018).
105. Robinson, N. A., Johnsen, H., Moghadam, H., Andersen, Ø. & Tveiten, H. Early developmental stress affects subsequent gene expression response to an acute stress in Atlantic salmon: an approach for creating robust fish for aquaculture? *G3 (Bethesda).* **9**, 1597–1611 (2019).
106. Zhao, L. *et al.* Transgenerational acclimation to seawater acidification in the Manila clam Ruditapes philippinarum: Preferential uptake of metabolic carbon. *Sci. Total Environ.* **627**, 95–103 (2018).
107. Parker, L. M., O'Connor, W. A., Raftos, D. A., Pörtner, H.-O. & Ross, P. M. Persistence of positive carryover effects in the oyster, Saccostrea glomerata,

following transgenerational exposure to ocean acidification. *PLoS One* **10**, e0132276 (2015).

108. Franěk, R. *et al.* Isogenic lines in fish – a critical review. *Rev. Aquac.* raq.12389 (2019) doi:10.1111/raq.12389.
109. Goddard, M. E. & Whitelaw, E. The use of epigenetic phenomena for the improvement of sheep and cattle. *Front. Genet.* **5**, 247 (2014).
110. Brugman, S. *et al.* A comparative review on microbiota manipulation: lessons from fish, plants, livestock, and human research. *Front. Nutr.* **5**, 80 (2018).
111. Derome, N. *Microbial communities in aquaculture ecosystems: improving productivity and sustainability*. (Springer International Publishing, 2019). doi:10.1007/978-3-030-16190-3.
112. Smith, C. C. R., Snowberg, L. K., Gregory Caporaso, J., Knight, R. & Bolnick, D. I. Dietary input of microbes and host genetic variation shape among-population differences in stickleback gut microbiota. *ISME J.* **9**, 2515–2526 (2015).
113. Li, W. *et al.* Genetic effects on the gut microbiota assemblages of hybrid fish from parents with different feeding habits. *Front. Microbiol.* **9**, 2972 (2018).
114. Cornejo-Granados, F. *et al.* Microbiome of Pacific Whiteleg shrimp reveals differential bacterial community composition between wild, aquacultured and AHPND/EMS outbreak conditions. *Sci. Rep.* **7**, 11783 (2017).
115. Llewellyn, M. S., Boutin, S., Hoseinifar, S. H. & Derome, N. Teleost microbiomes: the state of the art in their characterization, manipulation and importance in aquaculture and fisheries. *Front. Microbiol.* **5**, 207 (2014).
116. Naylor, R. L. *et al.* Feeding aquaculture in an era of finite resources. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 15103–10 (2009).
117. Wang, G.-D., Xie, H.-B., Peng, M.-S., Irwin, D. & Zhang, Y.-P. Domestication genomics: evidence from animals. *Annu. Rev. Anim. Biosci.* **2**, 65–84 (2014).
118. Lopez Dinamarca, M. E. *et al.* Multiple selection signatures in farmed Atlantic salmon adapted to different environments across Hemispheres. *Front. Genet.* **10**, 901 (2019).
119. López, M. E. *et al.* Comparing genomic signatures of domestication in two Atlantic salmon (Salmo salar L.) populations with different geographical origins. *Evol. Appl.* **12**, 137–156 (2019).

120. Glover, K. A. *et al.* Half a century of genetic interaction between farmed and wild Atlantic salmon: status of knowledge and unanswered questions. *Fish Fish.* **18**, 890–927 (2017).

This review highlights the impact of interaction between farmed salmon escapees and wild salmon, showing the need to avoid interbreeding as farmed fish become more domesticated.

121. Roberge, C., Einum, S., Guderley, H. & Bernatchez, L. Rapid parallel evolutionary changes of gene transcription profiles in farmed Atlantic salmon. *Mol. Ecol.* **15**, 9–20 (2005).
122. Skaala, Ø. *et al.* An extensive common-garden study with domesticated and wild Atlantic salmon in the wild reveals impact on smolt production and shifts in fitness traits. *Evol. Appl.* **12**, 1001–1016 (2019).
123. Cowen, R. K. & Sponaugle, S. Larval dispersal and marine population connectivity. *Ann. Rev. Mar. Sci.* **1**, 443–466 (2009).
124. Varney, R. L., Watts, J. C. & Wilbur, A. E. Genetic impacts of a commercial aquaculture lease on adjacent oyster populations. *Aquaculture* **491**, 310–320 (2018).
125. Hornick, K. M. & Plough, L. V. Tracking genetic diversity in a large-scale oyster restoration program: effects of hatchery propagation and initial characterization of diversity on restored vs. wild reefs. *Heredity (Edinb).* **123**, 92–105 (2019).
126. Hindar, K., Fleming, I. A., McGinnity, P. & Diserud, O. Genetic and ecological effects of salmon farming on wild salmon: modelling from experimental results. *ICES J. Mar. Sci.* **63**, 1234–1247 (2006).
127. Hansen, M. M., Limborg, M. T., Ferchaud, A.-L. & Pujolar, J.-M. The effects of Medieval dams on genetic divergence and demographic history in brown trout populations. *BMC Evol. Biol.* **14**, 122 (2014).
128. Horreo, J. L. *et al.* Long-term effects of stock transfers: synergistic introgression of allochthonous genomes in salmonids. *J. Fish Biol.* **85**, 292–306 (2014).
129. Heggberget, T. G. *et al.* Interactions between wild and cultured Atlantic salmon: a review of the Norwegian experience. *Fish. Res.* **18**, 123–146 (1993).
130. Naylor, R. *et al.* Fugitive salmon: assessing the risks of escaped fish from net-pen aquaculture. *Bioscience* **55**, 427–437 (2005).
131. O'Flynn, F., McGeachy, S. A., Friars, G. W., Benfey, T. J. & Bailey, J. K.

Comparisons of cultured triploid and diploid Atlantic salmon (Salmo salar L.). *ICES J. Mar. Sci.* **54**, 1160–1165 (1997).

132. Piferrer, F. *et al.* Polyploid fish and shellfish: production, biology and applications to aquaculture for performance improvement and genetic containment. *Aquaculture* **293**, 125–156 (2009).
133. Wargelius, A. *et al.* Dnd knockout ablates germ cells and demonstrates germ cell independent sex differentiation in Atlantic salmon. *Sci. Rep.* **6**, 21284 (2016).

This study showed the potential of genome editing using CRISPR to induce sterility in farmed fish.

134. Adams, S. L., Smith, J. F., Taylor, J., McGowan, L. T. & Tervit, H. R. Cryopreservation of Greenshell™ Mussel (Perna canaliculus) Sperm. in 329–336 (Springer, New York, NY, 2015). doi:10.1007/978-1-4939-2193-5_14.
135. Goswami, M., Mishra, A., Ninawe, A., Trudeau, V. & Lakra, W. Bio-banking: an emerging approach for conservation of fish germplasm. *Poultry, Fish. Wildl. Sci.* **4**, (2016).
136. Robles, V. *et al.* Biology of teleost primordial germ cells (PGCs) and spermatogonia: Biotechnological applications. *Aquaculture* **472**, 4–20 (2017).
137. FAO. *Genebank Standards for Plant Genetic Resources for Food and Agriculture*. (2014).
138. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–23 (2013).
139. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–6 (2013).
140. Edvardsen, R. B., Leininger, S., Kleppe, L., Skaftnesmo, K. O. & Wargelius, A. 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 (2014).

This study was the first to apply CRISPR genome editing to an aquaculture species, and highlighted the possibility of creating double allele knockout individuals in the F0 generation.

141. Gratacap, R. L., Wargelius, A., Edvardsen, R. B. & Houston, R. D. Potential of genome editing to improve aquaculture breeding and production. *Trends Genet.* **35**,

672–684 (2019).

142. Sato, M. & Nakamura, S. Possible production of genome-edited animals using gene-engineered sperm. in *Gene Editing - Technologies and Applications* (IntechOpen, 2019). doi:10.5772/intechopen.84859.
143. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
144. Gratacap, R. L. *et al.* Efficient CRISPR/Cas9 genome editing in a salmonid fish cell line using a lentivirus delivery system. *bioRxiv* 734442 (2019) doi:10.1101/734442.
145. Jenko, J. *et al.* Potential of promotion of alleles by genome editing to improve quantitative traits in livestock breeding programs. *Genet. Sel. Evol.* **47**, 55 (2015).
146. Burkard, C. *et al.* Pigs lacking the scavenger receptor cysteine-rich domain 5 of CD163 are resistant to porcine reproductive and respiratory syndrome virus 1 infection. *J. Virol.* **92**, (2018).
147. Zohar, Y. Endocrinology and fish farming: Aspects in reproduction, growth, and smoltification. *Fish Physiol. Biochem.* **7**, 395–405 (1989).
148. Wong, T.-T. & Zohar, Y. Production of reproductively sterile fish: a mini-review of germ cell elimination technologies. *Gen. Comp. Endocrinol.* **221**, 3–8 (2015).
149. Li, M. *et al.* Efficient and heritable gene targeting in tilapia by CRISPR/Cas9. *Genetics* **197**, 591–599 (2014).
150. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–51 (2013).
151. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–83 (2013).
152. Sanson, K. R. *et al.* Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat. Commun.* **9**, (2018).
153. Tait-Burkard, C. *et al.* Livestock 2.0 – genome editing for fitter, healthier, and more productive farmed animals. *Genome Biol.* **19**, 204 (2018).
154. Bruce, A. Genome edited animals: Learning from GM crops? *Transgenic Res.* **26**, 385–398 (2017).
155. Waltz, E. First genetically engineered salmon sold in Canada. *Nature* (2017) doi:10.1038/nature.2017.22116.
156. Yoshizaki, G. & Yazawa, R. Application of surrogate broodstock technology in

aquaculture. *Fish. Sci.* **85**, 429–437 (2019).

157. Okutsu, T., Shikina, S., Kanno, M., Takeuchi, Y. & Yoshizaki, G. Production of trout offspring from triploid salmon parents. *Science* **317**, 1517 (2007).
158. Sid, H. & Schusser, B. Applications of gene editing in chickens: a new era is on the horizon. *Front. Genet.* **9**, 456 (2018).
159. Shelley, C. & Lovatelli, A. *Mud crab aquaculture –A practical manual*. *FAO Fisheries and Aquaculture Technical Paper No. 567* (FAO, 2011).
160. Xu, S., Zhao, L., Xiao, S. & Gao, T. Whole genome resequencing data for three rockfish species of Sebastes. *Sci. Data* **6**, 97 (2019).
161. Tsai, H.-Y. *et al.* Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics* **16**, 969 (2015).
162. Tsai, H.-Y. *et al.* Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. *Genet. Sel. Evol.* **48**, 47 (2016).
163. Correa, K., Bangera, R., Figueroa, R., Lhorente, J. P. & Yáñez, J. M. The use of genomic information increases the accuracy of breeding value predictions for sea louse (Caligus rogercresseyi) resistance in Atlantic salmon (Salmo salar). *Genet. Sel. Evol.* **49**, 15 (2017).
164. Robledo, D., Matika, O., Hamilton, A. & Houston, R. D. Genome-Wide Association and Genomic Selection for Resistance to Amoebic Gill Disease in Atlantic Salmon. *G3 (Bethesda).* **8**, 1195–1203 (2018).
165. Boison, S. A., Gjerde, B., Hillestad, B., Makvandi-Nejad, S. & Moghadam, H. K. Genomic and Transcriptomic Analysis of Amoebic Gill Disease Resistance in Atlantic Salmon (Salmo salar L.). *Front. Genet.* **10**, 68 (2019).
166. Bangera, R., Correa, K., Lhorente, J. P., Figueroa, R. & Yáñez, J. M. Genomic predictions can accelerate selection for resistance against Piscirickettsia salmonis in Atlantic salmon (Salmo salar). *BMC Genomics* **18**, 121 (2017).
167. Horn, S. S., Meuwissen, T. H. E., Moghadam, H., Hillestad, B. & Sonesson, A. K. Accuracy of selection for omega-3 fatty acid content in Atlantic salmon fillets. *Aquaculture* 734767 (2019) doi:10.1016/j.aquaculture.2019.734767.
168. Vallejo, R. L. *et al.* Accurate genomic predictions for BCWD resistance in rainbow trout are achieved using low-density SNP panels: Evidence that long-range LD is a

major contributing factor. *J. Anim. Breed. Genet.* **135**, 263–274 (2018).

169. Vallejo, R. L. *et al.* Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. *Genet. Sel. Evol.* **49**, 17 (2017).
170. Vallejo, R. L. *et al.* Evaluation of genome-enabled selection for bacterial cold water disease resistance using progeny performance data in rainbow trout: insights on genotyping methods and genomic prediction models. *Front. Genet.* **7**, 96 (2016).
171. Yoshida, G. M., Carvalheiro, R., Rodríguez, F. H., Lhorente, J. P. & Yáñez, J. M. Single-step genomic evaluation improves accuracy of breeding value predictions for resistance to infectious pancreatic necrosis virus in rainbow trout. *Genomics* **111**, 127–132 (2019).
172. Yoshida, G. M. *et al.* Genomic Prediction Accuracy for Resistance Against Piscirickettsia salmonis in Farmed Rainbow Trout. *G3 (Bethesda).* **8**, 719–726 (2018).
173. Vallejo, R. L. *et al.* Genome-wide association analysis and accuracy of genome-enabled breeding value predictions for resistance to infectious hematopoietic necrosis virus in a commercial rainbow trout breeding population. *Genet. Sel. Evol.* **51**, 47 (2019).
174. Silva, R. M. O. *et al.* Whole-genome mapping of quantitative trait loci and accuracy of genomic predictions for resistance to columnaris disease in two rainbow trout breeding populations. *Genet. Sel. Evol.* **51**, 42 (2019).
175. Barría, A. *et al.* Genomic predictions and genome-wide association study of resistance against piscirickettsia salmonis in coho salmon (Oncorhynchus kisutch) using ddRAD sequencing. *G3 (Bethesda).* **8**, 1183–1194 (2018).
176. Palaiokostas, C., Kocour, M., Prchal, M. & Houston, R. D. Accuracy of genomic evaluations of juvenile growth rate in common carp (Cyprinus carpio) using genotyping by sequencing. *Front. Genet.* **9**, 82 (2018).
177. Yoshida, G. M. *et al.* Genome-wide association study and cost-efficient genomic predictions for growth and fillet yield in Nile tilapia (Oreochromis niloticus). *G3 Genes, Genomes, Genet.* **9**, 2597–2607 (2019).
178. Joshi, R., Skaarud, A., de Vera, M., Alvarez, A. T. & Ødegård, J. Genomic prediction for commercial traits using univariate and multivariate approaches in Nile tilapia

(Oreochromis niloticus). *Aquaculture* **516**, 734641 (2020).

179. Palaiokostas, C. *et al.* Genome-wide association and genomic prediction of resistance to viral nervous necrosis in European sea bass (Dicentrarchus labrax) using RAD sequencing. *Genet. Sel. Evol.* **50**, 30 (2018).
180. Palaiokostas, C., Ferraresso, S., Franch, R., Houston, R. D. & Bargelloni, L. Genomic prediction of resistance to pasteurellosis in gilthead sea bream (Sparus aurata) using 2b-RAD sequencing. *G3: Genes|Genomes|Genetics* **6**, 3693–3700 (2016).
181. Aslam, M. L. *et al.* Genetics of resistance to photobacteriosis in gilthead sea bream (Sparus aurata) using 2b-RAD sequencing. *BMC Genet.* **19**, 43 (2018).
182. Saura, M. *et al.* Disentangling genetic variation for resistance and endurance to scuticociliatosis in turbot using pedigree and genomic information. *Front. Genet.* **10**, 539 (2019).
183. Liu, Y. *et al.* Genomic Selection Using BayesCπ and GBLUP for Resistance Against Edwardsiella tarda in Japanese Flounder (Paralichthys olivaceus). *Mar. Biotechnol.* **20**, 559–565 (2018).
184. Garcia, A. L. S. *et al.* Development of genomic predictions for harvest and carcass weight in channel catfish. *Genet. Sel. Evol.* **50**, 66 (2018).
185. Dong, L., Xiao, S., Wang, Q. & Wang, Z. Comparative analysis of the GBLUP, emBayesB, and GWAS algorithms to predict genetic values in large yellow croaker (Larimichthys crocea). *BMC Genomics* **17**, 460 (2016).
186. Nguyen, N. H., Premachandra, H. K. A., Kilian, A. & Knibb, W. Genomic prediction using DArT-Seq technology for yellowtail kingfish Seriola lalandi. *BMC Genomics* **19**, 107 (2018).
187. Liu, G. *et al.* Evaluation of genomic selection for seven economic traits in yellow drum (Nibea albiflora). *Mar. Biotechnol.* **21**, 806–812 (2019).
188. Gutierrez, A. P., Matika, O., Bean, T. P. & Houston, R. D. Genomic Selection for Growth Traits in Pacific Oyster (Crassostrea gigas): Potential of Low-Density Marker Panels for Breeding Value Prediction. *Front. Genet.* **9**, (2018).
189. Gutierrez, A. P. *et al.* Potential of genomic selection for improvement of resistance to Ostreid Herpes virus in Pacific oyster (Crassostrea gigas). *bioRxiv* 754473 (2019) doi:10.1101/754473.

190. Dou, J. *et al.* Evaluation of the 2b-RAD method for genomic selection in scallop breeding. *Sci. Rep.* **6**, 19244 (2016).
191. Wang, Y. *et al.* Predicting growth traits with genomic selection methods in Zhikong scallop (Chlamys farreri). *Mar. Biotechnol.* **20**, 769–779 (2018).
192. Wang, Q. *et al.* Effects of marker density and population structure on the genomic prediction accuracy for growth trait in Pacific white shrimp Litopenaeus vannamei. *BMC Genet.* **18**, 45 (2017).
193. Wang, Q., Yu, Y., Li, F., Zhang, X. & Xiang, J. Predictive ability of genomic selection models for breeding value estimation on growth traits of Pacific white shrimp Litopenaeus vannamei. *Chinese J. Oceanol. Limnol.* **35**, 1221–1229 (2017).
194. Wang, Q. *et al.* Evaluation on the genomic selection in Litopenaeus vannamei for the resistance against Vibrio parahaemolyticus. *Aquaculture* **505**, 212–216 (2019).
195. Nguyen, N. H., Phuthaworn, C. & Knibb, W. Genomic prediction for disease resistance to Hepatopancreatic parvovirus and growth, carcass and quality traits in Banana shrimp Fenneropenaeus merguiensis. *Genomics* (2019) doi:10.1016/J.YGENO.2019.11.014.
196. Zhang, X. *et al.* The sea cucumber genome provides insights into morphological evolution and visceral regeneration. *PLOS Biol.* **15**, e2003790 (2017).
197. Hughes, L. C. *et al.* Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proc. Natl. Acad. Sci.* **115**, 6249–6254 (2018).
198. Wanninger, A. & Wollesen, T. The evolution of molluscs. *Biol. Rev. Camb. Philos. Soc.* **94**, 102 (2018).
199. Wolfe, J. M. *et al.* A phylogenomic framework, evolutionary timeline and genomic resources for comparative studies of decapod crustaceans. *Proc. R. Soc. B Biol. Sci.* **286**, 20190079 (2019).
200. Dohrmann, M. & Wörheide, G. Dating early animal evolution using phylogenomic data. *Sci. Rep.* **7**, 3599 (2017).
201. Plazzi, F. & Passamonti, M. Towards a molecular phylogeny of mollusks: bivalves' early evolution as revealed by mitochondrial genes. *Mol. Phylogenet. Evol.* **57**, 641–657 (2010).
202. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: a resource for

timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
203. Diamond, J. Evolution, consequences and future of plant and animal domestication. *Nature* vol. 418 700–707 (2002).
204. Bernatchez, L. *et al.* Harnessing the Power of Genomics to Secure the Future of Seafood. *Trends in Ecology and Evolution* vol. 32 665–680 (2017).
205. Yue, G. H. Recent advances of genome mapping and marker-assisted selection in aquaculture. *Fish Fish.* **15**, 376–396 (2014).
206. Brooker, A. J. *et al.* Sustainable production and use of cleaner fish for the biological control of sea lice: recent advances and current challenges. *Vet. Rec.* **183**, 383–383 (2018).
207. Sveier, H. & Breck, O. Cleaner fish application in Norway. in *Cleaner fish Biology and Aquaculture applications* (ed. Treasurer, J.) 370–385 (5m Publishing Ltd, 2018).
208. Treasurer, J. *Cleaner Fish Biology and Aquaculture Applications*. (5M Publishing Ltd, 2018).
209. Lie, K. K. *et al.* Loss of stomach, loss of appetite? Sequencing of the ballan wrasse (Labrus bergylta) genome and intestinal transcriptomic profiling illuminate the evolution of loss of stomach function in fish. *BMC Genomics* **19**, 186 (2018).
210. Knutsen, T. M. Lumpfish (Cyclopterus lumpus) draft genome assembly. *figshare* (2018) doi:10.6084/m9.figshare.7301546.v1.
211. Lafferty, K. D. *et al.* Infectious diseases affect marine fisheries and aquaculture economics. *Ann. Rev. Mar. Sci.* **7**, 471–496 (2015).
212. Asche, F., Hansen, H., Tveteras, R. & Tveteras, S. The salmon disease crisis in Chile. *Mar. Resour. Econ.* **24**, 405–411 (2009).
213. Verbruggen, B. *et al.* Molecular mechanisms of white spot syndrome virus infection and perspectives on treatments. *Viruses* **8**, 23 (2016).
214. Ødegård, J., Baranski, M., Gjerde, B. & Gjedrem, T. Methodology for genetic evaluation of disease resistance in aquaculture species: challenges and future prospects. *Aquac. Res.* **42**, 103–114 (2011).
215. Yáñez, J. M. *et al.* Genetic co-variation between resistance against both Caligus rogercresseyi and Piscirickettsia salmonis, and body weight in Atlantic salmon (Salmo salar). *Aquaculture* **433**, 295–298 (2014).
216. Gjedrem, T. Disease resistant fish and shellfish are within reach: a review. *J. Mar.*

Sci. Eng. **3**, 146–153 (2015).

217. Bishop, S. C. & Woolliams, J. A. Genomics and disease resistance studies in livestock. *Livest. Sci.* **166**, 190–198 (2014).
218. Anacleto, O. *et al.* Genetic differences in host infectivity affect disease spread and survival in epidemics. *Sci. Rep.* **9**, 4924 (2019).
219. Storset, A., Strand, C., Wetten, M., Kjøglum, S. & Ramstad, A. Response to selection for resistance against infectious pancreatic necrosis in Atlantic salmon (Salmo salar L.). *Aquaculture* **272**, (2007).
220. Houston, R. D. *et al.* Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (Salmo salar). *Genetics* **178**, 1109–15 (2008).
221. Moen, T., Baranski, M., Sonesson, A. K. & Kjøglum, S. 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 (2009).

Together with Houston et al. 2008, Moen et al. described a very large effect QTL for disease resistance in aquaculture populations, leading to widespread adoption of marker-assisted selection to help reduce disease incidence.

222. Houston, R. D. *et al.* The susceptibility of Atlantic salmon fry to freshwater infectious pancreatic necrosis is largely explained by a major QTL. *Heredity (Edinb).* **105**, 318–327 (2010).
223. Gheyas, A. A. *et al.* Segregation of infectious pancreatic necrosis resistance QTL in the early life cycle of Atlantic Salmon (Salmo salar). *Anim. Genet.* **41**, 531–536 (2010).
224. Houston, R. D. *et al.* Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC Genomics* **13**, 244 (2012).
225. Moen, T. *et al.* Epithelial cadherin determines resistance to infectious pancreatic necrosis virus in Atlantic salmon. *Genetics* **200**, 1313–26 (2015).
226. Robledo, D. *et al.* 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 (2016).

Acknowledgements

The authors acknowledge funding from the Biotechnology and Biological Sciences Research Council (BBSRC), the Natural Environment Research Council, and the Scottish Aquaculture Innovation Centre via the consortium project 'AquaLeap' (BB/S004343/1, BB/S004181/1, BB/S004416/1, BB/S004300/1), and BBSRC Institute Strategic Programme grants (BBS/E/D/20241866 & BBS/E/D/20002172 & BBS/E/D/20002174).

Author contributions

The authors contributed equally to all aspects of the article.

Competing interests

The authors declare no competing interests.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Peer review information

Nature Reviews Genetics thanks L. Bernatchez, D. Jerry, N. H. Nguyen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Supplementary information

Supplementary information is available for this paper at https://doi.org/10.1038/s415XX-XXX-XXXX-X

Glossary

AQUACULTURE

The farming of fish, crustaceans, molluscs, aquatic plants, algae in fresh water or saltwater environments, typically for human food.

Broodstock

A group of sexually mature individuals used in aquaculture for breeding purposes.

Genetic gains

Improvement in average genetic value, and therefore improved phenotypes, in a population due to selection over cycles of selective breeding.

OVOVIVIPAROUS

Producing offspring by means of eggs which are hatched within the body of the parent.

MASS SPAWNING

Release of high numbers of eggs and sperm into the water, where fertilization occurs externally. Also known as broadcast spawning.

SEQUENTIAL HERMAPHRODITISM

Where an individual in a species is born as one sex, but can later change into the opposite sex.

BEHAVIOURAL PLASTICITY

The ability of an organism to change its behavior following exposure to stimuli, such as changing environmental conditions.

GENETIC BOTTLENECKS

Sharp reductions in genetic diversity, typically due to large reductions in population size caused by environmental events or human activities.

BASE POPULATIONS

Populations of animals used to start a selective breeding programme.

INBREEDING DEPRESSION

The reduced biological fitness in a given population as a result of inbreeding, typically due to deleterious recessive alleles.

SNP ARRAYS

A type of DNA microarray which is used to genotype genome-wide polymorphisms within a population.

GENOTYPING BY SEQUENCING

A method using high-throughput sequencing to discover and genotype genome-wide SNPs within a population.

SCAFFOLDING

An approach during genome assembly where contigs (that is, continuous assembled sequences) are linked into larger continguous sequences including gaps of known length.

INTROGRESSION

The deliberate movement of a target locus from one species or strain (donor) into another (recipient) by the creation and repeated backcrossing of a hybrid with one of the donor species or strains.

EFFECTIVE POPULATION SIZE

The size of an idealised population which would give rise to the rate of inbreeding and the rate of change in variance of allele frequencies actually observed in the population under consideration. It is approximate to the number of individuals that contribute gametes to the next generation.

SOFT SWEEPS

Increases in frequency and/or fixation of a favourable allele at an existing polymorphic locus due to strong positive selection pressure.

MARKER-ASSISTED SELECTION

The selection of breeding individuals for genetic improvement of a trait of interest based on genetic markers linked to a quantitative trait locus affecting that trait.

Quantitative trait locus

A region of the genome which explains a significant component of variation in a trait of interest.

Mendelian sampling

The chance factor in the process of distributing half the genetic material from each parent to their offspring, which is the source of within-family genetic variation.

Genomic selection

The selection of breeding individuals for genetic improvement of a trait of interest based on the use of genome-wide genetic markers to estimate genomic breeding values. Genetic marker genotypes and phenotypes are measured in a reference population to predict breeding values of selection candidates that have genotypes only.

Reference population

In genomic selection, the population of animals which have both genotypes and phenotypes. These data are used to estimate genetic marker effects, which are then applied to predict breeding values for genotyped selection candidates.

Phenotyping

Collection of measurements relating to traits of interest to the goals of a breeding program.

Accuracy

In the context of genomic selection, accuracy is the correlation between the estimated genomic breeding values and the true breeding values.

Genotype imputation

The statistical inference of unobserved genotypes based on knowledge of haplotypes in a population, typically used to predict high density marker genotypes when most individuals are genotyped for low density marker genotypes.

CAUSATIVE VARIANT

A polymorphism within the genome of a population that has a direct effect on a trait of interest, as opposed to simply being a genetic marker associated with the trait.

INTERNET OF THINGS

A network of physical objects that use sensors and application program interfaces to connect and exchange data over the Internet.

GENOMIC RELATIONSHIP MATRIX

A matrix containing the estimation of the proportion of total genomic DNA shared by any two individuals based on genome-wide genetic marker data.

BREEDING NUCLEI

The elite broodstock animals that are maintained only for breeding, which is followed by multiplication and dissemination of the genetically improved animals for production.

SURROGATE BROODSTOCK

Sterile animals used for the production of gametes of another individual, strain, or species.

PLEIOTROPIC EFFECTS

In the context of genome editing, the unintended impacts on traits other than the target trait due to a specific edit.

PRIMORDIAL GERM CELLS

The stem cells specified during early development that will differentiate to form male and female gametes, therefore representing the precursors of the germline.

GERMPLASM

In the context of animal breeding, the genetic material of a breeding program.

LINKED READS

Linking together of short sequence reads to provide long range orientation, based on the addition of a unique DNA barcode to each read generated from an individual molecule.

GENOTYPE–PHENOTYPE GAP

The gap in knowledge of how variation at the level of the genome causes an effect on a phenotype of interest.

SELECTION INTENSITY

The number of phenotypic standard deviation units that selected parents are superior to the mean of a population.

Genomic best linear unbiased prediction

(GBLUP) A modification of the pedigree-based best linear unbiased prediction method that incorporates SNP information in the form of a genomic relationship matrix and defines the additive genetic covariance among individuals to predict breeding values.

BAYESIAN MODELS

In the context of genomic selection, the use of multiple-regression methods incorporating prior information on marker effects which are used widely for genomic prediction of breeding values.