Harnessing genomics to fast-track genetic improvement in aquaculture

Ross D. Houston1✉, Tim P. Bean1, Daniel J. Macqueen1, Manu Kumar Gundappa1, Ye Hwa Jin1, Tom L. Jenkins3, Sarah Louise C. Selly2, Samuel A. M. Martin4, Jamie R. Stevens2, Eduarda M. Santos3, Andrew Davie5 and Diego Robledo1

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 crop and livestock production, aquaculture 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.

Aquaculture has a crucial and rapidly increasing role in food security and economic stability worldwide. More than 90% of global aquaculture occurs 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. Global production of finfish and shellfish reached 172.6 million tons in 2017, approximately half of which is currently derived from aquaculture. 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. By contrast, aquaculture is the fastest-growing food production sector globally. With major limitations on wild capture and terrestrial farmland exploitation, 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 and infectious disease outbreaks, which have a negative impact on the health and welfare of farmed (and potentially wild) populations and may be exacerbated by climate change.

Selective breeding for genetic improvement of production traits has great potential to increase 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 (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). 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. Despite their diversity, aquaculture species tend to share two key features that enhance their potential for genetic improvement. Firstly, they remain under domestication and are therefore genetically improved animals for production. Therefore, there is a pressing opportunity to use domestication and selective breeding programmes to harness the as-yet largely untapped genetic potential of farmed aquatic species, as highlighted in a recent landmark report by the FAO. This potential for cumulative and permanent improvement of production traits is evident from the typically high genetic gains in aquaculture breeding programmes;
Fig. 1 | Summary of global aquaculture diversity and production. a | Phylogenetic tree showing farmed species with an annual production value higher than US$1 billion per annum (Supplementary Table 6). Estimated divergence times are from Refs 194–200. 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. The 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 (Supplementary Table 7).
for example, an average 13% growth increase per generation in Atlantic salmon (*Salmo salar*), which is substantially higher than the growth observed in breeding programmes for terrestrial livestock species. 

Genomic tools are hugely valuable to inform sustainable genetic improvement, 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 through to advanced genomic selection in closed commercial breeding nuclei (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.

**Box 1 | A road map for genomic 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 and the subsequent emergence of ‘crowd’ diseases in farmed populations, 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. Given recent and rapid technological developments, together with improved accessibility and increased 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 production began in 2009 and 2011, respectively, with closure of the life cycles in captivity in 2018 and 2016 and reference genomes released by 2016 and 2018. Both domestication processes have combined animal biology, health management and nutritional requirements with the development of genomic tools for genetic management and enhancement. 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 (see the figure).

When moving towards more advanced selective breeding programmes, bespoke tools such as single-nucleotide polymorphism (SNP) arrays can be applied, but their cost-effectiveness needs to be considered and contrasted with that of 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 the 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.

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 technologies, to promote step improvements in aquaculture breeding and production.

**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 to completion of the full life cycle in captivity, and use of modern selective breeding
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.

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

Behavioural plasticity
The ability of an organism to change its behaviour 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.

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.

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

Genotyping by sequencing (GBS)
A method using high-throughput sequencing to discover and genotype genome-wide single-nucleotide polymorphisms within a population.

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

for genetic improvement of production traits, such as growth and disease resistance. Historically, the selection of species amenable to reproduction in farmed environments was pivotal to defining which livestock and aquaculture species were farmed. For example, domesticated species tend to display behavioural plasticity that enables them to adapt to a range of captive environments.

A key difference between livestock and aquaculture species is that domestication of terrestrial livestock occurred in tandem with global human migration several millennia before the informed management of breeding populations, and modern livestock lines have typically undergone multiple major genetic bottlenecks. 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 (members of the family Cyprinidae) and tilapia (members of the family Cichlidae), aquaculture and domestication have been ongoing in some form for millennia, but selective breeding programmes to enable genetic improvement are much more recent (Fig. 13). Currently, only a minority of aquaculture production is derived from selectively bred stocks, estimated at approximately 10% in 2012 (Ref. 13). However, this proportion is increasing rapidly, particularly for species with high production volume and value, with approximately 75% of the top 10 fish, crustacean and mollusc species (by production volume) benefiting from some form of modern selective breeding programme (Supplementary Tables 3,4). The use of genetic technologies also varies dramatically by continent, with more than 80% of European aquaculture production derived from selective breeding programmes. The availability and application of selective breeding depends on the local environmental, social, political and economic landscapes, all of which can present major hurdles, especially in low- and middle-income countries. These programmes enable cumulative, permanent and sustainable genetic gain for target production traits, and are fundamental to scale up aquaculture production in the context of finite resources.

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

The burgeoning genomic toolbox
Genomic resources for aquaculture generally lag behind those for 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 widespread presence of duplicated loci due to genome duplication events, for example, in salmonids, cyprinids and sturgeons, and the exceptionally high heterozygosity observed, for example, in bivalve species and crustaceans. 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, and might contribute to the genetic regulation of production-relevant traits.

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 (10x Genomics), are increasingly being applied to aquaculture species to improve assemblies (Supplementary Table 5). When combined with long-range scaffolding 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. 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 N50 of 37.4 Mb. All major aquaculture species are likely to benefit from such high-quality assemblies soon.

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. 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 (GBS) techniques, including restriction site-associated DNA sequencing (RAD-seq) 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.

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. Poor broodstock management and hatchery practices that lead to inbreeding depression have been hypothesized to result in reduced population fitness, increased susceptibility to stress and
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, transcriptomes, GBS methods and estimation of genetic diversity and genetic parameters 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 cleaner fish (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. 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, genomic 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

| Species                  | Production value (US$ billion) | Genome size (Gb) | Scaffold N50 (Mb) | Coding genes | Published SNP arrays (number of SNPs) | Resequenced genomes |
|--------------------------|--------------------------------|------------------|-------------------|--------------|--------------------------------------|--------------------|
| **Finfish**              |                                 |                  |                   |              |                                      |                    |
| Atlantic salmon (Salmo salar) | 16.69                          | 2.96             | 1.36              | 48,775       | 7 (15,000–286,000)                   | 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 (50,000–58,000)                   | 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 (6,000)                           | –                  |
| 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 (6,000)                           | 2                  |
| Oriental river prawn (Macrobrachium nipponense) | 2.09                           | –                | –                 | –            | –                                    | –                  |
| **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 (27,000–190,000)                  | 516                |
| Blood cockle (Tegillarca granosa) | 1.02                           | –                | –                 | –            | –                                    | –                  |
| **Echinoderms**          |                                 |                  |                   |              |                                      |                    |
| Japanese sea cucumber (Apostichopus japonicus) | 1.40                           | 0.8              | 0.48              | 30,350       | –                                    | 1                  |

Full data are provided for the top 20 species per taxonomic group in Supplementary Table 5. Gb, gigabase; Mb, megabase; SNP, single-nucleotide polymorphism.
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. These can vary within a genus and even within a species, and sequential hermaphroditism presents an additional challenge in several commercially important aquaculture species. GBS techniques have been widely applied to assess the genetic basis of sex determination, for example, in Nile tilapia (Oreochromis niloticus), Atlantic halibut (Hippoglossus hippoglossus), European seabass (Dicentrarchus labrax) and mud crabs (Scylla sp.). 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, 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. Although multiple interventions are possible to enable pedigree tracking (for example, pair spawning or stripping using hormonal induction), genetic markers are frequently applied to track stock relatedness to minimize loss of genetic diversity within a closed breeding nucleus.

Of note, the reliability of genomic data alone to predict adaptive potential of populations is questionable, 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 effects on traits of interest, in addition to genotype and environment (G × E) interactions (discussed in more detail below). Such information can be used to optimize selection of the base population, ensuring it has substantial genetic variation to be utilized for effective directional selection. However, while hybrid vigour resulting from strain crosses can result in notable one-off gains in production, and genomic tools can provide insight into the underlying molecular mechanisms of this heterosis, exploiting additive genetic variation via within-strain breeding programmes is likely to result in superior performance after a small number of generations of selection.

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. 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.

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) many loci of minor effect. Farmed aquatic populations face selection pressures that are vastly different from those faced by 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 among the standing genetic variation in aquaculture populations. During the millennia of domestication of terrestrial livestock, such loci are likely to already be fixed via soft sweeps. However, in aquaculture species, they may present a one-off opportunity for rapid genetic improvement via marker-assisted selection (MAS) based on the use of targeted quantitative trait locus (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 IPN. Other applications of QTLs for disease resistance include breeding of a Japanese flounder (Paralichthys olivaceus) strain with resistance to the viral disease lymphocystis based on a major QTL for lymphocystis resistance, and use of MAS based on a QTL affecting resistance to bacterial cold water disease in rainbow trout (Oncorhynchus mykiss). Other noteworthy 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, as well as loci for resistance to pancreas disease and cardiomyopathy syndrome. Similarly, in Nile tilapia, a locus explaining 79% of the phenotypic variation in salinity tolerance was detected, although validation of the size of the 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 that highlight the value of genetic improvement to the industry.

Genomic selection to accelerate trait improvement
Genome-wide association studies in aquaculture species have highlighted that most traits of relevance to production are polygenic in nature (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. 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. However, for these sib-testing traits, selection candidates from a given family have the same estimated breeding value, placing limitations on the genetic gain that can be achieved while maintaining genetic diversity. Genetic marker data are required to accurately capture
Infectious disease outbreaks are a major and ongoing threat to the economic and environmental sustainability of aquaculture. 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 (REF. 219), and annual losses of shrimp equating to ~10% of the global industry due to white spot syndrome virus 220. Options to fully mitigate such diseases via vaccination (in finish fish 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 vaccine options 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 221–224, and sib-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 225, whereas in terrestrial livestock this is limited by logistical and financial challenges relating to routine measurement of disease resistance traits 226.

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, outcomes such as survival or pathogen burden to quantify the resistance traits 227. 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 228. 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.

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%) in stocks both in freshwater hatcheries and following transfer to sea cages. Resistance to IPN was shown to be moderately to highly heritable 229, and breeding companies began to implement family-based selection. In parallel, teams from the UK and Norway identified a single major quantitative trait locus on chromosome 26 that could explain 80–100% of genetic variation in resistance to IPN virus in seawater field trials 230 and experimental freshwater trials 231–233. 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 233,234. The practical outcome of these experiments was extensive use of marker-assisted selection for the favourable allele in all major salmon breeding programmes, assisted by the fact that the resistance allele is dominant 235,236. The results were striking, with a sustained decrease in the incidence of IPN outbreaks to near zero 237 (see the figure). Follow-up functional studies highlighted marked differences in gene expression response to infection between resistant and susceptible salmon fry 234 and suggested that epithelial cadherin may be part of the mechanism underlying the quantitative trait locus 238. However, the exact causative mutations and the nature of their effect remain at least partly elusive.

Figure adapted from Ref. 225, Elsevier.

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

SNP arrays
A type of DNA microarray that are used to genotype genome-wide polymorphisms within a population.

Reference population
In genomic selection, the population of animals that 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.

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

Phenotyping
Collection of measurements relating to traits of interest in the goals of a breeding programme.
The availability of large full-sibling families can be exploited with use of within-family genomic selection, with very low-density markers used to estimate genomic breeding values within families with known pedigree-based estimated breeding values. The increased accuracy of genomic prediction compared with 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). Most studies of genomic selection in aquaculture species use genomic best linear unbiased prediction (GBLUP) approaches, which harness genomic relationships to estimate the genetic merit of individuals. A range of Bayesian models have been tested in several species but without consistent differences in prediction accuracy compared with the simpler GBLUP approach. Adequate sample size for the genotyped and phenotyped population is key to fully assess the efficacy of genomic selection (for example, more than 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. While several thousand genome-wide markers are also required, it is noteworthy that a reduction in SNP density to only 1,000 or 2,000 SNPs tends to be sufficient to achieve the asymptote of prediction accuracy where these close relationships exist. However, the accuracy drops drastically as the relationship between the reference and test populations becomes more distant, as demonstrated in Atlantic salmon and common carp (Cyprinus carpio); therefore, routine trait measurement and genotyping are required each generation to retrain the genomic prediction models.

**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, aquaculture of salmonids, tilapia and shrimp), genomic tools are yet to be routinely incorporated into breeding programmes for these species.

---

**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 (genomic best linear unbiased prediction) or Bayesian models. gEBV, genomic estimated breeding value.

**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.
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) or fluorescence-based assays, tend to be cheaper than SNP arrays. Low-density genotyping can be integrated with genotype imputation to increase the accuracy of genomic selection to levels approaching those obtained with high-density genotyping. 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, genotyping offspring at low density and imputing genotypes to high density on the basis of those haplotypes. Considering that breeding programmes for many aquaculture species routinely use low-density SNP panels for parentage assignment, 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.

From sequence to consequence: identifying causative variants for target traits. Mapping and understanding the causative variants or functional variants that have an impact on complex traits is a fundamental goal of biology but also has potential additional benefits for increasing rates of genetic gain in breeding either through improved selection accuracy or as targets for genome editing (Fig. 5). The reduction in prediction accuracy with more distant relationships between reference and validation sets is partly because 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, and incorporating weighting of putative functional genomic variants (for example, BayesRC) into genomic selection models to increase 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 soon. However, the cost of whole-genome sequencing and the effectiveness of low-density SNP panels described earlier mean that substantial increases 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 genome-wide association studies are used to highlight genomic regions associated with traits of interest. However, such regions often contain hundreds to thousands of candidate causative variants 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 using 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, 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.

Discovery and exploitation of epigenetic marks in aquaculture species also represents a crucial step to help bridge the genotype-phenotype gap and prioritize variants for downstream functional testing. Emerging genomic technologies are enabling the elucidation of genome-scale patterns of cytosine methylation, chromatin accessibility, histone modifications, transcriptional start sites and transcript variants. These tools enhance the scope to identify putative causative variants within regulatory sequences (for example, enhancers) that are active under specific environmental conditions (for example, during disease outbreaks). In addition, aquaculture species also benefit from the existence of extant and recently diverged wild counterparts, and use of comparative genomics and orthology analysis can help predict functional variants on the basis of sequence conservation. The Functional Annotation of Animal Genomes (FAANG) initiative 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, and comparable 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 the section ‘Genome editing to accelerate genetic improvement’).

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 before harvest typically require the handling of large numbers of animals outside the water, presenting a logistical and financial challenge. Therefore,
| Species | Trait | Measurement | Heritability (pedigree) | Accuracy (pedigree) | Relative increase (%) | Genotyping technology (number of SNPs) | Ref. |
|---------|-------|-------------|-------------------------|---------------------|-----------------------|----------------------------------------|------|
| Atlantic salmon (Salmo salar) | Growth | Weight | 0.60 (0.48) | 0.70 (0.58) | 21 | SNP array (132,000, 112,000 postfiltering) | 159 |
| | | Length | 0.61 (0.51) | 0.66 (0.56) | 18 | | | 159 |
| | Resistance to sea lice | Lice count | 0.33 (0.27) | 0.60 (0.48) | 25 | SNP array (132,000, 33,000 postfiltering) | 160 |
| | | Lice count | 0.22 (0.27) | 0.46 (0.43) | 7 | | | 160 |
| | | Lice count | 0.11 (0.10) | 0.50 (0.41) | 22 | SNP array (50,000, 37,000 postfiltering) | 161 |
| | Resistance to amoebic gill disease | Gill score | 0.24 (0.25) | 0.62 (0.51) | 22 | Two-species SNP array (17,000, 7,000 postfiltering) | 162 |
| | | Amoebic load | 0.25 (0.36) | 0.70 (0.60) | 17 | | | 162 |
| | Resistance to salmon rickettsial syndrome | Time to death | 0.27 (0.18) | 0.41* (0.34) | 21 | SNP array (50,000, 50,000 postfiltering) | 164 |
| | | Binary survival | 0.39 (0.26) | 0.26 (0.20) | 30 | | | 164 |
| Rainbow trout (Oncorhynchus mykiss) | Resistance to bacterial cold water disease | Binary survival | – | 0.68* (0.36) | 89 | SNP array (57,000, 45,000 postfiltering) | 166 |
| | | Time to death | 0.33 (0.37) | 0.67* (0.34) | 97 | SNP array (57,000, 36,000 postfiltering) | 167 |
| | | Binary survival | 0.35 (0.35) | 0.70* (0.36) | 94 | | | 167 |
| | | Time to death | 0.29 (0.31) | 0.49 (0.50) | 2 | SNP array (57,000, 41,000 postfiltering) | 168 |
| | | Binary survival | 0.45 (0.48) | 0.46 (0.41) | 12 | | | 168 |
| | Resistance to infectious pancreatic necrosis virus | Time to death | 0.25 (0.40) | 0.53 (0.49) | 8 | SNP array (57,000, 38,000 postfiltering) | 169 |
| | | Binary survival | 0.24 (0.35) | 0.56 (0.50) | 12 | | | 169 |
| | Resistance to salmon rickettsial syndrome | Time to death | 0.45 (0.38) | 0.78* (0.61) | 28 | SNP array (57,000, 27,000 postfiltering) | 170 |
| | | Binary survival | 0.55 (0.54) | 0.60* (0.47) | 28 | | | 170 |
| | Resistance to infectious haematopoietic necrosis virus | Time to death | 0.23 (0.33) | 0.33 (0.13) | 154 | SNP array (57,000, 35,000 postfiltering) | 171 |
| | | Binary survival | 0.25 (0.28) | 0.39 (0.24) | 63 | | | 171 |
| | Resistance to columnaris disease | Binary survival | 0.32 (–) | 0.11 (–0.02) | –650 | SNP array (57,000, 36,000 postfiltering) | 172 |
| | | Binary survival | 0.51 (–) | 0.22 (0.06) | 267 | | | 172 |
| Coho salmon (Oncorhynchus kisutch) | Resistance to salmon rickettsial syndrome | Time to death | – (0.14) | 0.52 (0.27) | 93 | ddRAD (9,000) | 173 |
| | | Binary survival | – (0.27) | 0.81 (0.31) | 161 | | | 173 |
| Carp (Cyprinus carpio) | Growth | Length | 0.33 (0.33) | 0.71 (0.60) | 18 | RAD-seq (20,000) | 174 |
| | Resistance to koi herpesvirus | Binary survival | 0.50 (0.61) | 0.53* (0.49) | 8 | RAD-seq (16,000) | 77 |
| Nile tilapia (Oreochromis niloticus) | Growth | Harvest weight | 0.36 (0.31) | 0.60 (0.48) | 25 | SNP array (43,000, 32,000 postfiltering) | 175 |
| | | Fillet yield | 0.21 (0.21) | 0.62 (0.54) | 15 | | | 175 |
| | | Harvest weight | 0.17 (0.22) | 0.29 (0.19) | 53 | SNP array (59,000, 48,000 postfiltering) | 176 |
| | | Fillet weight | 0.16 (0.24) | 0.34 (0.18) | 89 | | | 176 |
| | | Fillet yield | 0.23 (0.33) | 0.54 (0.46) | 17 | | | 176 |
| European sea bass (Dicentrarchus labrax) | Resistance to viral nervous necrosis | Binary survival | 0.43 (0.27) | 0.62* (0.67) | 7 | RAD-seq (9,000) | 177 |
| Species                          | Trait                                  | Measurement                  | Heritability (pedigree) | Accuracy (pedigree) | Relative increase (%) | Genotyping technology (number of SNPs) | Ref. |
|---------------------------------|----------------------------------------|------------------------------|-------------------------|---------------------|-----------------------|----------------------------------------|------|
| Gilthead sea bream (*Sparus aurata*) | Resistance to pasteurellosis           | Time to death                | 0.28 (0.22)             | 0.44* (0.30)        | 47                    | 2b-RAD (22,000)                       | 178  |
|                                 |                                        | Time to death                | 0.32 (0.32)             | 0.54* (0.45)        | 20                    | 2b-RAD (28,000)                       | 179  |
|                                 |                                        | Binary survival              | 0.33 (0.31)             | 0.56* (0.46)        | 22                    |                                        |      |
| Turbot (*Scophthalmus maximus*) | Resistance to scuticociliatosis        | Resilience                   | 0.15 (–)                | 0.46 (0.41)         | 12                    | 2b-RAD (18,000)                       | 180  |
|                                 |                                        | Resistance                  | 0.26 (–)                | –                   | –                     |                                        | 180  |
|                                 |                                        | Endurance                    | 0.12 (–)                | –                   | –                     |                                        | 180  |
| Japanese flounder (*Paralichthys olivaceus*) | Resistance to Edwardsiella tarda | Binary survival              | – (–)                   | 0.603 (–)           | –                     | WGS (1.9 × 10⁶)                      | 181  |
| Channel catfish (*Ictalurus punctatus*) | Growth                              | Harvest weight               | 0.27 (–)                | 0.37 (0.29)         | 28                    | SNP array (660,000, 55,000 postfiltering) | 182  |
|                                 |                                        | Residual carcass weight      | 0.34 (–)                | 0.31 (0.24)         | 29                    |                                        |      |
| Large yellow croaker (*Larimichthys crocea*) | Growth                              | Body weight                  | 0.60 (–)                | 0.41 (–)            | –                     | ddRAD (30,000)                        | 183  |
|                                 |                                        | Body length                  | 0.59 (–)                | 0.40 (–)            | –                     |                                        |      |
|                                 |                                        | Omega-3 HUFA                 | –                       | 0.44 (–)            | 0.30 (–)              | ddRAD (32,000)                        | 183  |
| Yellowtail kingfish (*Seriola lalandi*) | Growth                              | Weight                       | 0.47 (0.42)             | 0.69 (–)            | –                     | DArT-seq (14,000)                     | 184  |
|                                 |                                        | Length                       | 0.43 (0.42)             | 0.67 (–)            | –                     |                                        |      |
|                                 |                                        | Condition index              | 0.21 (0.11)             | 0.44 (–)            | –                     |                                        |      |
| Yellow drum (*Nibe a albiflora*)  | Growth                                | Body length                  | – (–)                   | 0.38* (–)           | –                     | GBS (54,000)                          | 185  |
|                                 |                                        | Swimming bladder index       | – (–)                   | 0.17* (–)           | –                     |                                        |      |
|                                 |                                        | Swimming bladder weight      | – (–)                   | 0.22* (–)           | –                     |                                        |      |
|                                 |                                        | Body thickness               | – (–)                   | 0.24* (–)           | –                     |                                        |      |
|                                 |                                        | Body height                  | – (–)                   | 0.30* (–)           | –                     |                                        |      |
|                                 |                                        | Body length/body height ratio| – (–)                   | 0.36* (–)           | –                     |                                        |      |
|                                 |                                        | Gonad weight index           | – (–)                   | 0.37* (–)           | –                     |                                        |      |
| Pacific oyster (*Crassostrea gigas*) | Growth                              | Shell length                 | 0.26 (0.23)             | 0.54 (0.44)         | 23                    | Two-species SNP array (38,000, 23,000 postfiltering) | 186  |
|                                 |                                        | Shell height                 | 0.23 (0.20)             | 0.60 (0.47)         | 28                    |                                        |      |
|                                 |                                        | Wet weight                   | 0.35 (0.31)             | 0.67 (0.54)         | 24                    |                                        |      |
|                                 | Resistance to ostreid herpesvirus     | Binary survival              | 0.37 (0.25)             | 0.76 (0.64)         | 19                    |                                        |      |
| Yesso scallop (*Patinopecten yessoensis*) | Growth                              | Shell height                 | 0.48 (–)                | 0.53 (–)            | –                     | 2b-RAD (2,000)                        | 187  |
|                                 |                                        | Shell length                 | 0.48 (–)                | 0.46 (–)            | –                     |                                        |      |
|                                 |                                        | Shell width                  | 0.36 (–)                | 0.55 (–)            | –                     |                                        |      |
| Zhikong scallop (*Chlamys farreri*) | Growth                              | Shell length                 | 0.42 (–)                | 0.65* (–)           | –                     | 2b-RAD (31,000)                       | 188  |
|                                 |                                        | Shell height                 | 0.47 (–)                | 0.70* (–)           | –                     |                                        |      |
|                                 |                                        | Shell width                  | 0.54 (–)                | 0.63* (–)           | –                     |                                        |      |
|                                 |                                        | Whole weight                 | 0.28 (–)                | 0.64* (–)           | –                     |                                        |      |
| Whiteleg shrimp (*Litopenaeus vannamei*) | Growth                              | Body weight                  | 0.32 (–)                | 0.62 (–)            | –                     | 2b-RAD (23,000)                       | 189  |
|                                 |                                        | Body length                  | 0.45 (–)                | 0.61 (–)            | –                     |                                        |      |
|                                 |                                        | Body length                  | – (–)                   | 0.30* (–)           | –                     | SLAF-seq (6,000)                      | 190  |
|                                 |                                        | Body weight                  | – (–)                   | 0.41* (–)           | –                     |                                        |      |
|                                 | Resistance to AHPND                   | Time to death                | 0.26 (0.24)             | 0.50 (0.47)         | 6                     | 2b-RAD (23,000)                       | 191  |
|                                 |                                        | Binary survival              | 0.16 (0.15)             | 0.21 (0.20)         | 5                     |                                        | 192  |
the ability to collect such data both directly on the selection candidates in the breeding nucleus and on relatives of those candidates 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. 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. For instance, underwater cameras for real-time in situ data collection are being used for tasks such as sea lice monitoring in Atlantic salmon, and their use is likely to expand for more widespread data collection and phenotyping. Connected mobile devices for affordable on-farm monitoring and automation of aquaculture environments (that is, sensor technologies and the Internet of things) have major potential for monitoring individual traits such as behaviour and feed intake, while 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. Together with routine genomic evaluations, the effective combination of increasingly high-resolution and high-volume phenotyping in breeding nuclei, in production environments and after harvest will lead to more precise and effective genetic improvement of aquaculture species.

Genetics and the environment

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-farm and across-farm variation in partial pressure of CO₂, 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, but the level of investment required to establish and maintain these systems is substantial and is unlikely to be feasible for most 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 reranking such that the best-performing genotypes in one environment may not be the best in another, placing a limitation on realizing genetic gains in breeding programmes. 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, 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-technology 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. 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 environments⁸⁶. 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 population⁸⁹.

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 environments⁹²,⁹⁷. 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 genotype⁹⁸. This reaction norm can be used as a target trait for genomic selection to reduce sensitivity to environmental variation, with notably superior results to sib-testing schemes alone⁹⁸. 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.

**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 and result in substantial phenotypic variation from the same genomic DNA blueprint⁹⁹. Recent domestication can profoundly alter the epigenome of hatchery-reared animals via alterations to the DNA methylation profile⁹⁹, 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 generations⁹⁹,⁹⁶,¹⁰⁰. For example, early-life use of plant-based diets increased the acceptance and use of these diets in later life in rainbow trout¹⁰¹, and early-life stress can modulate future stress or immune responses in Atlantic salmon, which may have implications for robustness in adult stages¹⁰²,¹⁰³. 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 controls¹⁰⁴, 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 adults¹⁰⁵. 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 effects¹⁰⁶.

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 relatives¹⁰⁷. However, distinguishing additive genetic and epigenetic components of phenotypic variation may facilitate improvement in genetic parameter estimation and prediction of response to selection¹⁰⁸. Furthermore, an interesting intersection between epigenetic programming and genetic improvement via selective breeding may be related to optimizing robust performance of improved stocks in multiple environments. The use of genomics to support breeding of ‘robust’ strains for multiple environments described earlier 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.

**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 animals¹⁰⁹,¹¹⁰. Colonization and development of bacterial communities are essential for immune function and are influenced by host physiology and immune response. Host microbiota composition is heritable to some extent in marine species¹¹⁰,¹¹¹, and differences have also been observed between farmed and wild strains of Atlantic salmon¹¹² and Pacific whiteleg shrimp (*Litopenaeus vannamei*)¹¹². Microbiome research in aquaculture species is currently primarily focused on gaining an understanding of its composition in various species¹⁰⁹,¹¹³. 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 diets¹¹⁴. In terrestrial livestock, microbiome similarity matrices have been used to replace or complement the host genomic relationship matrix, with an improved predictive ability for feed conversion.

---

**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.
efficiency in Holstein Friesian dairy cattle\(^2\). 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.
**Interaction between farmed and wild animals.** The recent domestication of aquaculture species means that farmed species often coexist 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. 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.

Divergence of wild and farmed populations results in notable differences in growth, morphology, life history, behaviour and physiology. The impact of domestication on animal physiology has been demonstrated by studies of gene expression and genome methylation, which show marked differences after a few generations of hatchery breeding in salmonids. Introngression of potentially maladapted alleles into wild populations can lead to undesirable changes in life history traits, reduced population productivity and decreased resilience. Many species of marine fish and invertebrates are characterized by high connectivity, with associated high gene flow, and high effective population size, 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.

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

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 of aquatic animals, both for vulnerable wild species and for 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. 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. Cryopreservation of ovarian tissues is a promising alternative but would require research into the in vitro culture of these tissues. Surrogate broodstock (discussed later) hold promise to preserve genetic resources through transplant of primordial germ cells. 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.

### 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.

#### Genome editing to accelerate genetic gain.

Genome editing tools such as engineered CRISPR–Cas9 systems are invaluable for understanding genetic regulation of economically important traits and have potential to accelerate genetic gain in aquaculture breeding programmes (Fig. 3). The enzyme Cas9 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, CRISPR–Cas9 has been successfully applied in various farmed finfish and mollusc species, primarily for gene knockout and as proof of principle. Microinjection into early-stage embryos is the most commonly used delivery method but can be inefficient, and alternative delivery methods,
such as electroproportion of sperm, hold promise. 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. Use of genome-wide loss-of-function CRISPR screens such as genome-scale CRISPR knockout (GeCKO) 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. 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 pliable in species with external fertilization, abundance of embryos and feasible early-life screens. 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. 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. Finally, it is possible to create de novo alleles on the basis of knowledge of the biology of the trait in question or using targets from GeCKO screens. For example, removal of an exon of the CD163 gene in pigs (Sus scrofa) resulted in complete resistance to porcine reproductive and respiratory syndrome virus.

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, are additional key objectives. For example, knockout of the germ-line-specific genes dnd1 in Atlantic salmon and nanos2 or nanos3 in Nile tilapia 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 not 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 escapes).

Refinement of genome editing methods is occurring constantly, and use of modified CRISPR–Cas systems such as CRISPR activation or CRISPR interference can induce differences in expression levels of target genes instead of complete knockout. 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 are rigorously monitored to discount deleterious pleiotropic effects; aquaculture can follow procedures used for terrestrial livestock to achieve these goals. Furthermore, any practical application for aquaculture depends entirely on an acceptable regulatory and public approval landscape, and the approval of the genetically modified AquaAdvantage salmon (Aquabounty) as fit for human consumption by the US Food and Drug Administration and the Canadian Food Inspection Agency was a recent landmark. Target traits that have concurrent production and animal welfare or environmental benefits should be a focus for genome editing in aquaculture, and public and policymaker engagement with the technology, its benefits and its risks is absolutely vital.

**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:

\[
\Delta G = \frac{ir\sigma_y}{y}
\]

where \(\Delta G\) is the genetic gain over time, \(i\) is the selection intensity, \(r\) is the selection accuracy, \(\sigma_y\) is the additive genetic variance and \(y\) is the generation time. Genomic selection has resulted in a step increase in selection accuracy, and much research is now devoted to achieving further minor increases. 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; for example, rainbow trout offspring were produced when spermatogonia from rainbow trout were injected into newly hatched sterile masu salmon (Oncorhynchus masou). 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 club mackerel (Scomber japonicus) as
In addition, surrogate technology can be coupled with genome editing of primordial germ cells to create germ-line-edited animals, as successfully demonstrated in chickens. This approach is a route to genome editing for aquaculture species where access to the newly fertilized embryos is challenging, such as in certain crustaceans or ovoviviparous species such as rockfish (Sebastes spp.). 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 the generation interval. Extensive effort and resources have been put into the use of functional genomic data to increase selection accuracy in breeding, and reproductive technologies require equivalent attention.

**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 makes possible cumulative increases in production traits, and facilitates adaptation to emerging challenges, such as climate change or infectious disease outbreaks. Due to recent growth and increased availability, genomics should be used 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 the salmon sector, but judicious application of multipurpose cost-effective marker panels may be necessary to translate these 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 increase 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, and 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

---

**Ovoviviparous**

Producing offspring by means of eggs that are hatched within the body of the parent.
adoption of breeding programs. In the longer term, develop-
ments in surrogate broodstock technology combined with
genomic selection have the potential to shorten
ance intervals to expedite genetic gain.
Underpinning many of these advances is improved
knowledge of the genetics and biology of key produc-
tion traits, which is particularly pertinent for the many
aquaculture species from understudied taxa with major
knowledge gaps relating to fundamental inheritance
and genome evolution. Overall, there is now an unpre-
cedented 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.

Published online 16 April 2020
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 reference 59, this study shows the importance of a single locus on an important life history and production trait in salmon.

Gonen, S., et al. Mapping and validation of a major QTL for pancreas disease (salmonid alphonsea) in Atlantic salmon (Salmo salar). Heredity 115, 405–414 (2015).

Gu, X. H. et al. Associated with cardiomyopathy syndrome in Atlantic salmon. J. Hered. 110, 727–737 (2019).

Hillestad, B. & Moghadam, H. K. Genome-wide association study of major QTLs for viral resistance in Atlantic salmon (Salmo salar). Mar. Biotechnol. 21, 720–726 (2019).

Houston, R. D. Future directions in breeding for disease resistance in aquaculture species. Rev. Bras. Zootec. 46, 545–551 (2017).

Zenger, K. R. et al. Genomic selection in aquaculture: application, limitations and opportunities with special reference to shrimp and pearl oysters. Front. Genet. 9, 693 (2019).

Giedreo, T. & Barbosa, J. Selective Breeding in Aquaculture: an Introduction (Springer, 2010).

Meeuwissen, T. H., Hayes, B. J. & Goddard, M. E. Prediction of genetic merit using genome-wide dense marker maps. Genetics 157, 1819–1829 (2001).

This study highlights the potential of using genomic selection in aquaculture and validating the benefit compared with pedigree approaches for predicting breeding values.

Houle, D. et al. Development and validation of a high density SNP genotyping array for Atlantic salmon (Salmo salar). BMC Genomics 15, 90 (2014).

Odegård, J. et al. Genomic prediction in an admixed population of Atlantic salmon (Salmo salar) Front. Genet. 5, 402 (2014).

This is the first empirical study of genomic selection in an admixed population of Atlantic salmon, validating the benefit compared with pedigree approaches for predicting breeding values.

Tsai, Y. Y. et al. The genetic architecture of growth and fillet traits in farmed Atlantic salmon (Salmo salar). BMC Genet. 16, 51 (2015).

This study highlights the potential of using genomic selection in aquaculture and validating the benefit compared with pedigree approaches for predicting breeding values.

Norris, A. Application of genomics in aquaculture: breeding programs by Nicky Norris: who knows where the genomic revolution will lead us? Mar. Genomics 36, 15–17 (2015).

Linthorst, H. H. T. E. & Sonesson, A. K. A low-marker density implementation of genomic selection in aquaculture within within-family genomic breeding schemes. Front. Genet. 6, 359 (2015).

Daetwyler, H. D., Calus, M. P. L., Pong-Wong, R., T sai, H. Y. et al. Construction and annotation of a genome-wide SNP panel of Atlantic salmon (Salmo salar). Front. Genet. 5, 176–193 (2014).

Saltz, J. B. et al. Why does the magnitude of genotype-by-environment interaction vary? Ecol. Evol. 8, 6342–6553 (2018).

Elanat, A. E. & Acosta, B. O. Genetic Improvement of Farmed Tilapias (GIFT Project): Final Report, March 2015 to February 2016. Aquaculture 491, 207–219 (2016).

Moghadam, H., Mørkøre, T. & Robinson, N. A. Isogenic lines in fish: a critical review. Rev. Aquac. 7, 2173–2179 (2015).

Robinson, N. A., Johnsen, H., Moghadam, H., Kriaridou, C., T sairidou, S., Houston, R. D. & Odegård, J. et al. Genomic prediction in an admixed population of Atlantic salmon (Salmo salar) Front. Genet. 5, 402 (2014).

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

Roberge, C., Eirum, S., Gudler, H. & Bernatchez, L. Rapid parallel evolutionary changes of gene transcription profiles in farmed Atlantic salmon. Mol. Biol. Evol. 15, 8–20 (2005).

Skalsa, O. et al. An extensive common-garden study with domesticated and wild Atlantic salmon in the wild reveals impacts on important traits in fitness traits. Evol. Appl. 12, 157–156 (2019).

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, 893–927 (2017).

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

Hornick, K. M. & Plough, L. V. Tracking genetic diversity in a large-scale oyster restoration project: effects of hatchery propagation and initial characterization of diversity on restored vs. wild reefs. Heredity 123, 92–105 (2015).

Hindar, K., Fleming, I. A., McGinnity, P. D. & Diergard, O. Genetic and ecological effects of salmon farming on wild populations: modelling effects on marine mammals. Ices J. Mar. Sci. 63, 1234–1247 (2006).

Hansen, M. M., Limborg, M. T., Ferchau, A.-L. & Pujolar, J. M. The effect of genetic drift on genetic divergence and demographic history in brown trout populations. BMC Evol. Biol. 14, 122 (2014).

Horreo, J. L. et al. Long-term effects of stock transfers: synergistic introgression of allodomic genomes in salmonids. J. Fish. Biol. 85, 292–306 (2014).

Heggenstaller, T. et al. Breeding and selection in the wild and cultured Atlantic salmon: a review of the Norwegian experience. Fish. Res. 18, 125–146 (1993).

Naylor, R. R. et al. Fugitive stocks and the risks of escaped fish from net-pen aquaculture. Bioscience 55, 427–457 (2005).

O’Flynn, F. M., McGeachy, S. A., Frairs, G. W., Benley, T. J. & Bailey, J. K. Comparisons of cultured triploid and diploid Atlantic salmon (Salmo salar L.). Ices J. Mar. Sci. 54, 1160–1165 (1997).
