

Applications of genotyping by sequencing in aquaculture breeding and genetics

Diego Robledo¹ , Christos Palaiokostas¹, Luca Bargelloni², Paulino Martínez³ and Ross Houston¹ 

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

² Department of Comparative Biomedicine and Food Science, University of Padova, Legnaro, Padova, Italy

³ Department of Zoology, Genetics and Physical Anthropology, Faculty of Veterinary, University of Santiago de Compostela, Lugo, Spain

Correspondence

Ross Houston, The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Midlothian, EH25 9RG UK. Email: ross.houston@roslin.ed.ac.uk

Received 12 October 2016; accepted 27 December 2016.

Abstract

Selective breeding is increasingly recognized as a key component of sustainable production of aquaculture species. The uptake of genomic technology in aquaculture breeding has traditionally lagged behind terrestrial farmed animals. However, the rapid development and application of sequencing technologies has allowed aquaculture to narrow the gap, leading to substantial genomic resources for all major aquaculture species. While high-density single-nucleotide polymorphism (SNP) arrays for some species have been developed recently, direct genotyping by sequencing (GBS) techniques have underpinned many of the advances in aquaculture genetics and breeding to date. In particular, restriction-site associated DNA sequencing (RAD-Seq) and subsequent variations have been extensively applied to generate population-level SNP genotype data. These GBS techniques are not dependent on prior genomic information such as a reference genome assembly for the species of interest. As such, they have been widely utilized by researchers and companies focussing on nonmodel aquaculture species with relatively small research communities. Applications of RAD-Seq techniques have included generation of genetic linkage maps, performing genome-wide association studies, improvements of reference genome assemblies and, more recently, genomic selection for traits of interest to aquaculture like growth, sex determination or disease resistance. In this review, we briefly discuss the history of GBS, the nuances of the various GBS techniques, bioinformatics approaches and application of these techniques to various aquaculture species.

Key words: aquaculture, genotyping, next-generation sequencing, restriction-site associated DNA, selective breeding, single nucleotide polymorphism.

Background

Despite the critical role for aquaculture in global food security, the vast majority of world fish and shellfish production is based on stocks without advanced selective breeding programmes (Gjedrem *et al.* 2012; Janssen *et al.* 2016). Aquaculture breeding schemes tend to lag behind their terrestrial livestock counterparts in terms of the uptake of genomic technologies, and for many aquaculture species, molecular genetic tools are only applied for pedigree reconstruction (Chavanne *et al.* 2016). In comparison, most modern breeding programmes in livestock are now underpinned by genomic selection (GS, Meuwissen *et al.* 2001), the benefits of which are well-illustrated in dairy cattle (Hayes *et al.* 2009). GS typically requires genome-wide genetic marker

data for a large number of individual animals. Up until a few years ago, obtaining genetic markers was costly and laborious; hence, large numbers of markers were only available for a handful of well-studied species. However, the recent advances in next-generation sequencing (NGS) have greatly reduced the cost of nucleic acid sequencing, and therefore also genetic marker discovery. This has opened the door for rapid generation of genome-wide genetic marker datasets, either via generation and application of SNP arrays, or directly via genotyping by sequencing (GBS) techniques (Davey *et al.* 2011). GBS techniques have revolutionized the field of evolutionary genomics (reviewed in Andrews *et al.* 2016) and have also led to several advances in genetics and breeding of aquaculture species, the subject of this review.

Due to the high fecundity of aquaculture species, the majority of breeding programmes are based on collection of trait data on close relatives (e.g. full siblings) of the selection candidates, particularly where the trait of interest cannot be measured on the candidates themselves (e.g. fillet quality, disease resistance). Without genetic markers, this set-up enables family selection, whereby family-level estimated breeding values (EBVs) for selection candidates are calculated using the data collected on the relatives. However, to utilize the within-family genetic variation in these traits, genetic markers are necessary to distinguish between selection candidates. Implementation of markers in breeding can broadly be split into two categories; marker-assisted selection (MAS) and GS. MAS is based on the use of targeted markers linked to major quantitative traits loci (QTL) affecting the trait, and one of the first examples in aquaculture was host resistance to infectious pancreatic necrosis virus (IPNV) in Atlantic salmon (*Salmo salar*, Houston *et al.* 2008; Moen *et al.* 2009). For traits with a polygenic architecture, GS is a more appropriate approach, whereby the relatives of the selection candidates become the 'training' population with genotypes and phenotypes, and those data are used to calculate genomic breeding values (GEBVs) for selection candidates with genotype data only. This application of genomic selection in aquaculture breeding is at a formative stage, and most examples to date have focussed on improved breeding for resistance to infectious diseases (e.g. Ødegård *et al.* 2014; Tsai *et al.* 2015, 2016b; Vallejo *et al.* 2016; Dou *et al.* 2016; Palaiokostas *et al.* 2016). The majority of high-resolution genetic studies in aquaculture species, and applications of genomic selection, have been underpinned by GBS techniques, either by directly providing genotype data or by discovering markers for the design of SNP arrays, which are currently only available for a handful of aquaculture species (e.g. Atlantic salmon, Houston *et al.* 2014; Yáñez *et al.* 2016; Pacific oyster, *Crassostrea gigas*, and European flat oyster, *Ostrea edulis*, Lapègue *et al.* 2014; channel catfish, *Ictalurus punctatus*, Liu *et al.* 2014; common carp, *Cyprinus carpio*, Xu *et al.* 2014a; rainbow trout, *Oncorhynchus mykiss*, Palti *et al.* 2015a).

The most common GBS techniques involve library preparation steps that result in deep sequence data at a repeatable subset of sites dispersed throughout the genome, typically using one or two restriction enzymes (RE), although also new GBS techniques based on targeted sequencing have been recently developed (i.e. GT-Seq, discussed below). The reason behind this genome complexity reduction is that high-coverage sequencing of a typical aquaculture species' genome with enough depth to confidently call genotypes is still prohibitively expensive for the number of animals required for high-resolution genetic studies and breeding programme applications. Genome

complexity reduction via RE is fast and inexpensive. Indeed, RE-based techniques have been commonplace in genotyping for many years, with RFLP and AFLP being widely applied to generate genotyping assays for limited numbers of genetic markers. The marriage of these ideas with NGS has enabled a major breakthrough for genetic studies of complex traits in nonmodel organisms, and their application to improve aquaculture production.

RAD sequencing

Restriction-site associated DNA sequencing (RAD sequencing or RAD-Seq) covers a range of GBS techniques which combine the use of genome complexity reduction with REs and the high sequencing output of NGS technologies. RAD-Seq was first described by Baird *et al.* (2008), following on from a similar idea based on microarrays (Miller *et al.* 2007). Some of the main reasons for its instant success are that RAD-Seq does not require any prior genomic knowledge, it allows generation of population-specific genotype data (i.e. no ascertainment bias) and it offers flexibility in terms of desired marker density across the genome. The use of different REs or innovative modifications to the base technique allows a high level of control over the number of markers obtained for a specific study. RAD-Seq and similar techniques are also amenable tools for aquaculture breeding, where genetic markers have typically been used in family assignment and pedigree reconstruction (Vandeputte & Haffray 2014). Mass spawning species are common in aquaculture, where mixed rearing and unknown parental contribution necessitate the use of genotyping for family-based breeding. RAD-Seq potentially facilitates a single experiment whereby pedigrees are reconstructed, genetic diversity is quantified, QTL can be mapped and genomic breeding values calculated (Palaiokostas *et al.* 2016). Since the original RAD-Seq paper by Baird *et al.* (2008), several variants of this methodology have been described. Three of them have been extensively used in aquaculture genetics research: the original RAD-Seq (Baird *et al.* 2008), 2b-RAD (Wang *et al.* 2012) and ddRAD (Peterson *et al.* 2012). Other RAD-based techniques like ezRAD (Toonen *et al.* 2013) or SLAF-seq (Sun *et al.* 2013) introduced minor modifications, which do not confer a major advantage for aquaculture applications. All available RAD-based techniques have been recently reviewed in depth elsewhere (Andrews *et al.* 2016); therefore, here we have focused on those most relevant in aquaculture breeding. The main features of original RAD-Seq, 2b-RAD and ddRAD are shown in Table 1, and they are briefly described below.

Original RAD-Seq

In original RAD-Seq (Baird *et al.* 2008), genomic DNA samples from several animals are individually digested with

Table 1 Summary of the different genotyping by sequencing (GBS) techniques

Technique	Key features	Advantages	Disadvantages
RAD-Seq	Digestion with one RE	<ul style="list-style-type: none"> ● Paired-end contigs ● PCR duplicate removal 	<ul style="list-style-type: none"> ● Complex library preparation
2bRAD	Digestion with type IIB REs	<ul style="list-style-type: none"> ● No size-selection step ● High reproducibility ● Easy library preparation ● Strand bias detection 	<ul style="list-style-type: none"> ● Short fragments ● Removal of PCR duplicates not possible
ddRAD	Digestion with two different REs	<ul style="list-style-type: none"> ● Can multiplex many samples ● Easy library preparation ● Flexibility over SNP density 	<ul style="list-style-type: none"> ● Repeatability dependent on size-selection step

a RE of choice. The digested DNA is then randomly sheared and pooled after ligation of adaptors with nucleotide barcodes for unique identification of each sample. The resulting restriction fragments are selected for suitable size range (i.e. for Illumina sequencing, typically 300–600 bp), and after a subsequent polymerase chain reaction (PCR) step, the fragments are sequenced. The result is high-coverage sequence data for flanking regions of the RE cut sites, which are typically dispersed quite evenly throughout the genome. As such, a genome-wide genetic marker dataset can be produced across a population of individuals at a fraction of the cost of whole genome resequencing. Illumina sequencing of short fragments either involves sequencing one (one read, single end) or both (two reads, paired end) ends of each fragment and currently gives reads of up to 300 bp in length. Each flanking sequence of the RE cut site is referred to as a RAD locus (or RAD-tag), and the high coverage of RAD tags facilitates simultaneous SNP detection and genotyping. The number of RAD tags, and therefore SNPs, generated in the experiment is tuneable via the choice of rarer or more frequent cutting RE. The most commonly used enzyme to date is *SbfI* which has an eight base recognition site and therefore cuts relatively infrequently throughout the genome. Online tools are available to guide the choice of the most appropriate RE according to the requirements and budget of the study (Lepais & Weir 2014). In addition to sequencing and genotyping individuals, the approach is also amenable to genotyping pooled populations for bulk-segregant analysis (Baird *et al.* 2008; Hohenlohe *et al.* 2010). One of the main drawbacks of the original technique is that shearing by sonication is random and variable, potentially hindering the efficiency and the reproducibility of RAD-Seq (Davey *et al.* 2013). However, this random shearing step can also be a

benefit, as the variable size of the genomic fragments anchored at the RE cut site facilitates the assembly of a contig based on the paired-end reads. This augments annotation of the RAD loci when there is no reference genome available, and also the design of specific primers for re-genotyping of targeted SNPs. In addition, the paired-end data from RAD-Seq allow identification and removal of putative PCR duplicates (reads originated from the same original DNA fragment, therefore presenting identical sequences), which can hinder analysis and interpretation of Illumina sequencing data (Schweyen *et al.* 2014). While there are several sources of potential bias and error in RAD-Seq techniques (see review by Andrews *et al.* 2016), several theoretical and empirical studies have demonstrated that RAD-Seq does render reproducible genotyping data across different laboratories, populations and even species (e.g. DaCosta & Sorenson 2014; Gonen *et al.* 2015).

2b-RAD

The first major modification of the original RAD technique was termed 2b-RAD (Wang *et al.* 2012). The main innovation in 2b-RAD is the use of type IIB REs, which share the feature of cutting the genomic DNA at both sides of the recognition site at a fixed distance, resulting in protruding noncohesive ends. The result is short genomic DNA fragments of identical size at each IIB RE site in the genome. Library construction in the 2b-RAD protocol is simple. Following DNA digestion, adaptors are ligated to the fragments, and specific barcodes are added to each sample through PCR amplification using degenerated linkers. Samples are then pooled and sequenced typically using Illumina technology, but allowing for runs of shorter read length due to the smaller size of the fragments in comparison to original RAD (2b-RAD fragments are 33–36 bp). The use of type IIB REs theoretically facilitates the sampling and sequencing of identical sites across individuals, circumventing the potential bias of RAD-Seq caused by the random shearing step. It also avoids the time-consuming and potentially error-prone size-selection step, which characterizes the majority of other RAD methods. Additionally, 2b-RAD is currently the only member of the RAD family that allows removal of loci exhibiting strand bias (Puritz *et al.* 2014a). The possibility to produce individually barcoded libraries allows targeted adjustment before pooling to obtain more equal representation of individual samples. The main caveat of this method is that it produces short sequencing reads (33–36 bp), which are less amenable for alignment to reference genome assemblies, and hinders follow-up applications such as the design of individual SNP assays (due to lack of SNP flanking sequence). However, this is not an issue if a draft genome sequence is available for the species, as is becoming the case in many aquaculture fish species.

ddRAD

Peterson *et al.* (2012) developed a new RAD-Seq platform using a double digestion of genomic DNA with two REs (ddRAD), thus eliminating the shearing step of original RAD. The ddRAD protocol is more flexible than RAD-Seq or 2b-RAD in terms of targeted marker density; the number of fragments and SNPs can be readily tailored by combining different RE pairs. Due to the typical use of a rare and a common cutting enzyme, ddRAD results in fewer sequenced sites than RAD-Seq, facilitating higher sequence coverage and/or more individuals multiplexed within a single sequencing lane. Higher multiplexing is possible due to combinatorial multiplex indexing, whereby a first barcode is introduced in the ligation step and a second during the PCR. Therefore, a larger number of samples can potentially be sequenced in a single lane than with the other RAD techniques. Compared to the RAD-Seq protocol, the workflow of preparation of ddRAD libraries is simpler, quicker and also substantially cheaper. However, the workflow is still more complex than the 2b-RAD protocol and requires a size-selection step. To ensure repeatability of sampled ddRAD loci across samples and libraries, consistency of size selection is paramount (Andrews *et al.* 2016). A simplified variation of the initial ddRAD protocol, where both P1 and P2 adaptors with individual barcodes are ligated prior to size selection (Palaiokostas *et al.* 2015a), further reduces hands-on time for library preparation.

RAD bioinformatic analyses

The advent of NGS posed important challenges in terms of data storage, transfer and analysis, which necessitated the development of specialized hardware and software. Consequently, the improvement of NGS-based sequencing platforms occurred in tandem with continuous development and improvement of suitable bioinformatics tools to analyse the large datasets. A wealth of software is available for analysing data originating from the RAD family of techniques. In the current review, a general framework for data analysis will be described, rather than attempting to provide a comprehensive overview of all available tools. Accordingly, the most popular, straightforward to use and regularly updated of the available tools are highlighted in terms of a suggested order of usage that might form a complete RAD analysis pipeline.

Experimental design and simulation

Sequencing and library construction typically account for the bulk of the cost of any experiment utilizing NGS. This leads to a balancing exercise, whereby researchers strive to include as many samples as possible per sequencing lane

(multiplexing), without compromising the read coverage required for accurate SNP genotype calling. Therefore, two key variables for a RAD experiment are the choice of the RE (affecting how many sites are sequenced), and the desired read coverage per locus. *In silico* simulation is a valuable tool for any well-designed RAD experiment. The R-based package SimRAD (Lepais & Weir 2014) can be utilized for simulation-based prediction of the expected number of loci for each RE (or their combination) and the genome of study. Although simulation estimates are likely to differ from the empirical data, valuable information can be gained to optimize experimental design before committing to the high cost associated with library construction and sequencing.

Demultiplexing libraries

The files that are generated by the sequencer (typically FastQ files) require demultiplexing into individual samples based on nucleotide barcodes. The most popular packages for this task include *Stacks* (Catchen *et al.* 2011) and *pyRAD* (Eaton 2014). Standard quality control procedure is to discard sequence reads below user-defined acceptable quality scores, erroneous barcodes and reads missing the characteristic sequence pattern obtained from the RE. Following demultiplexing, sequence files corresponding to each individual are generated for downstream analyses, including SNP calling and genotyping.

SNP identification

One of the key advantages of RAD-Seq approaches for non-model organisms (including many aquaculture species) is the ability to identify and genotype SNPs without requiring a reference genome for the organism under study. This approach, commonly defined in the literature as *de novo* assembly, can be performed using either *Stacks* (Catchen *et al.* 2011), *pyRAD* (Eaton 2014) or *dDocent* (Puritz *et al.* 2014b); however, the latter is limited to ddRAD or ezRAD data. The *de novo* approach involves identification and assembling of RAD loci in each individual, based on user-defined parameters related to read coverage required per locus, and sequence divergence between loci (Catchen *et al.* 2011). Identification of SNPs and inference of alleles within RAD loci is performed using a maximum-likelihood-based algorithm (Hohenlohe *et al.* 2010), which undertakes statistical tests at each nucleotide position to assess the likelihood of a particular diploid genotype. In doing so, the model implicitly estimates and accounts for sequencing error rate (Catchen *et al.* 2011). The *Stacks* software does not currently support SNP identification and genotyping in the paired-end (P2) read, unless anchored to a second RE (e.g. in ddRAD). Therefore, in original RAD experiments

using Stacks, the P2 read is typically used for quality control (e.g. removal of PCR duplicates), and for constructing paired-end ‘mini-contigs’ which facilitate BLAST alignment and genotyping assay design (Etter *et al.* 2011). The simultaneous use of P1 and P2 reads in the case of *dDocent*, and the application of an alignment-clustering algorithm in the case of *pyRAD*, allow the identification of insertion/deletion polymorphisms (indels) and identification of SNPs in the P2 reads.

Due to the decreasing cost of NGS, reference genome sequences are becoming available for many important aquaculture species. The number of species with reference genome assemblies is rapidly increasing (Atlantic cod, *Gadus morhua*, Star *et al.* 2011; Pacific oyster, Zhang *et al.* 2012; European sea bass, *Dicentrarchus labrax*, Tine *et al.* 2014; rainbow trout, Berthelot *et al.* 2014; Japanese eel, *Anguilla japonica*, Kai *et al.* 2014; half-smooth tongue sole, *Cynoglossus semilaevis*, Chen *et al.* 2014; common carp, Xu *et al.* 2014b; Northern pike, *Esox lucius*, Rondeau *et al.* 2014; Nile tilapia, *Oreochromis niloticus*, Brawand *et al.* 2015; Asian sea bass, *Lates calcarifer*, Vij *et al.* 2016; Mediterranean mussel, *Mytilus galloprovincialis*, Murgarella *et al.* 2016; turbot, *Scophthalmus maximus*, Figueras *et al.* 2016; Atlantic salmon, Lien *et al.* 2016; channel catfish, Chen *et al.* 2016), and new sequencing data will improve genome quality and annotation. Therefore, reference-guided RAD-Seq approaches are likely to be increasingly utilized. Both *Stacks* and *dDocent* can utilize reference genome information, using standard alignment tools followed by similar SNP calling algorithms to the *de novo* approach described above.

Potential bias and sources of error

While the bioinformatic pipelines for the RAD-like approaches are becoming increasingly standardized, there remains potential intrinsic barriers that must be overcome to ensure the generation of accurate and repeatable SNP datasets. One example that is particularly relevant to the aquaculture research community is distinguishing between genuine allelic SNPs and paralogous variants resulting from ancestral whole genome duplication. This is particularly a challenge for salmonid species, and strategies to account for this include (i) assessing read coverage for patterns suggestive of paralogous variation, (ii) checking for excessive heterozygosity at loci and (iii) sequencing (double) haploid individuals as the basis for filtering out paralogous sequence variants (e.g. Everett & Seeb 2014; Houston *et al.* 2014; Palti *et al.* 2015a,b). Another potential source of error for all RAD-Seq studies is the problem of RAD allele dropout (Gautier *et al.* 2013), where mutations within the recognition sequence for the RE segregating in the population are a common

source of null alleles. The extent of the issue is related to the length of the RE recognition sequence, and it is therefore potentially more of a problem for ddRAD (which requires two REs) versus other methods (Gonen *et al.* 2015; Andrews *et al.* 2016). Both read coverage levels and assessment of segregation distortion in pedigreed crosses can assist in identifying and removing, or accounting for, these null alleles. Finally, the concept of PCR duplicates is raised above, and this is due to preferential amplification of certain clonal DNA fragments derived from the original genomic DNA fragments. PCR duplicates can give rise to the situation where one allele is overrepresented in the resulting sequence data and causes problems with differentiating homozygous and heterozygous individuals at that locus (Schweyen *et al.* 2014).

Applications of RAD sequencing in aquaculture

Since its first description by Baird *et al.* (2008), RAD-Seq has quickly spread through different fields of genetic research, and it has been used in different aquaculture species to construct genetic maps (e.g. Recknagel *et al.* 2013; Gonen *et al.* 2014), for comparative genomics (e.g. Kakioka *et al.* 2013; Manousaki *et al.* 2015), for mapping genes associated with production traits (e.g. Houston *et al.* 2012; Shao *et al.* 2015; Fu *et al.* 2016), mapping sex determining loci (e.g. Palaiokostas *et al.* 2013a,b), studying population dynamics (e.g. Bradic *et al.* 2013), for fisheries management (e.g. Ogden *et al.* 2013), assembling reference genomes (e.g. Tine *et al.* 2014) or generating SNP resources for future SNP array development (e.g. Houston *et al.* 2014; Palti *et al.* 2014). A summary of the studies performed directly relevant for aquaculture is detailed below and in Table 2.

Genetic marker discovery for SNP array development

Early studies using RAD-Seq typically focussed on simply generating a genetic marker resource for nonmodel organisms. When the genome size of the target species is large, then whole genome (re)sequencing is arguably not cost-effective for SNP discovery across many individuals, and genome complexity reduction is advantageous. As such, RAD-Seq and similar techniques enabled a step change in the number of genetic markers (SNPs) available for several species (e.g. sturgeon, *Acipenser* genus, Ogden *et al.* 2013; or rainbow trout, Palti *et al.* 2014), and these have subsequently been used for several high-resolution genetic studies. SNPs generated by RAD techniques have also been applied to produce SNP arrays for several aquaculture species, including Atlantic salmon (Houston *et al.* 2014), rainbow trout

Table 2 Summary of aquaculture-oriented studies using restriction-site associated DNA sequencing (RAD-Seq)

Study	Species	Aim	Technique	Samples	SNPs	Families
Salmonids						
Houston <i>et al.</i> (2012)	<i>Salmo salar</i>	Disease resistance QTL (IPNV)	RAD	32	6712	Two families
Gonen <i>et al.</i> (2014)	<i>Salmo salar</i>	Linkage map	RAD	96	8257	Two families
Campbell <i>et al.</i> (2014)	<i>Oncorhynchus mykiss</i>	Disease resistance QTL (BCWD and IHN)	RAD	456	4661	40 families
Palti <i>et al.</i> (2014)	<i>Oncorhynchus mykiss</i>	SNP resource	RAD (×2)	19	145 168	19 genetic lines
Palti <i>et al.</i> (2015b)	<i>Oncorhynchus mykiss</i>	Disease resistance QTL (BCWD)	RAD	252	5612/4946	Two families
Liu <i>et al.</i> (2015b)	<i>Oncorhynchus mykiss</i>	Cortisol response to crowding QTL	RAD	234	4874	One family
Liu <i>et al.</i> (2015b)	<i>Oncorhynchus mykiss</i>	Disease resistance QTL (BCWD) and spleen size QTL	RAD	301	7849	Two half-sib families
Vallejo <i>et al.</i> (2016)	<i>Oncorhynchus mykiss</i>	Genomic selection (BCWD)	RAD	711	24 465	81 families
Everett and Seeb (2014)	<i>Oncorhynchus tshawytscha</i>	Thermotolerance and growth QTL	RAD	422	3534	Six families
Larson <i>et al.</i> (2016)	<i>Oncorhynchus nerka</i>	Thermotolerance and growth QTL	RAD	491	11 457	Five families
Nonsalmonid fish						
Palaiokostas <i>et al.</i> (2013b)	<i>Oreochromis niloticus</i>	Sex determination QTL	RAD	88	3904/4477	Two families
Palaiokostas <i>et al.</i> (2015a)	<i>Oreochromis niloticus</i>	Sex determination QTL	ddRAD	372	1279	Five families
Palaiokostas <i>et al.</i> (2013a)	<i>Hippoglossus hippoglossus</i>	Sex determination QTL	RAD	93	7572/5954	2 half-sib families
Palaiokostas <i>et al.</i> (2015b)	<i>Dicentrarchus labrax</i>	Sex determination QTL	RAD	187	6706	4 + 4 half-sib families
Wang <i>et al.</i> (2015a,b)	<i>Scophthalmus maximus</i>	Sex determination and growth QTL	RAD	151	6647	One family
Brown <i>et al.</i> (2016)	<i>Polyprion oxygeneios</i>	Sex determination and growth QTL	ddRAD	59	1609	One family
Manousaki <i>et al.</i> (2015)	<i>Pagellus erythrinus</i>	Linkage map	ddRAD	99	920	One family
Shao <i>et al.</i> (2015)	<i>Paralichthys olivaceus</i>	Disease resistance QTL (<i>Vibrio anguillarum</i>)	RAD	218	13 362	One family
Palaiokostas <i>et al.</i> (2016)	<i>Sparus aurata</i>	Disease resistance genomic selection	2b-RAD	777	12 085	75 families
Wang <i>et al.</i> (2015a,b)	<i>Lates calcarifer</i>	Growth QTL	ddRAD	144	3349	One family
Fu <i>et al.</i> (2016)	<i>Hypophthalmichthys nobilis</i>	Growth QTL	2b-RAD	119	3323	One family
Invertebrates						
Jiao <i>et al.</i> (2014)	<i>Chlamys farreri</i>	Sex determination and growth QTL	2b-RAD	98	7458	One family
Li and He (2014)	<i>Pinctada fucata</i>	Growth QTL	RAD	100	1381	One family
Shi <i>et al.</i> (2014)	<i>Pinctada fucata</i>	Growth QTL	2b-RAD	98	10 577	One family
Tian <i>et al.</i> (2015)	<i>Apostichopus japonicus</i>	Growth QTL	2b-RAD	102	11 306	One family
Lu <i>et al.</i> (2016)	<i>Marsupenaeus japonicus</i>	Thermotolerance and growth QTL	RAD	152	9829	One family
Dou <i>et al.</i> (2016)	<i>Patinopecten yessoensis</i>	Genomic selection (growth)	2b-RAD	349	2364	Five families
Ren <i>et al.</i> (2016)	<i>Haliotis diversicolor</i>	Growth QTL	RAD	142	3317	One family

(Palti *et al.* 2015a) and Pacific oyster (Lapègue *et al.* 2014). With the reduction in sequencing costs over recent years, whole genome (re)sequencing (i.e. pool-sequencing, Schlötterer *et al.* 2014) has become increasingly viable. However, RAD-like techniques still hold a significant advantage for SNP discovery when (i) there is no reference genome available, and (ii) only a medium density SNP resource is required.

Linkage maps and reference genome assembly

Restriction-site associated DNA sequencing techniques have been widely used in aquaculture species for constructing genetic maps based on recombination events in defined crosses. Such medium density SNP linkage maps are useful tools for downstream applications such as QTL mapping, comparative genomic and gene mining, or population

genomic studies. For example, RAD-based linkage maps have been created for Atlantic salmon (Gonen *et al.* 2014), channel catfish (Li *et al.* 2014), Japanese flounder (Shao *et al.* 2015), turbot (Wang *et al.* 2015b) and Asian seabass (Wang *et al.* 2015a). Genetic maps based on RAD-Seq have also contributed to mapping and orientation of scaffolds for reference genome assemblies for key aquaculture species such as European sea bass (Tine *et al.* 2014), rainbow trout (Berthelot *et al.* 2014), Japanese eel (Kai *et al.* 2014), half-smooth tongue sole (Chen *et al.* 2014) and turbot (Figueras *et al.* 2016). While NGS technology has enabled rapid and cheap reference genome assemblies, they are typically fragmented and incomplete. Further, assembly errors are quite common, and linkage maps can also assist with resolving mis-assemblies (Fierst 2015; Tsai *et al.* 2016a). Aquaculture species typically have an amenable family structure for high-resolution linkage maps, due to the high fecundity resulting in large full and half sibling families. Linkage maps can also be used in conjunction with physical reference genome sequences to detect variation in recombination rates across the genome, with implications for downstream applications (e.g. LD between markers and QTL in association mapping studies).

Mapping QTL associated with traits of economic importance

The rate of application of genomic technology to aquaculture species tends to reflect the degree of scientific and commercial interest of those species. This is typically motivated by the interest of understanding the genetic basis of economically-important production traits, for example growth, disease resistance or sex determination. Researchers working in the high-value salmonid species were amongst the first to exploit RAD-Seq techniques, evaluating resistance to different pathogens causing high economic losses, including infectious pancreatic necrosis in Atlantic salmon (Houston *et al.* 2012), and infectious hematopoietic necrosis (Campbell *et al.* 2014) and bacterial cold water disease (Campbell *et al.* 2014; Liu *et al.* 2015a; Palti *et al.* 2015b) in rainbow trout. Based on early successes, and given the importance of disease resistance to modern aquaculture breeding programmes (Yáñez *et al.* 2014), large-scale projects have been established to apply RAD-like techniques to detect markers, and eventually the genes and causal mutations involved, for improving resistance. For example, the European Union funded FISHBOOST project (www.fishboost.eu) is using RAD sequencing techniques to genotype several thousand animals from large-scale disease challenge experiments in rainbow trout, common carp, European sea bass, gilthead sea bream (*Sparus aurata*) and turbot. These genotype and phenotype data will be used to estimate genetic parameters, map disease resistance QTL

and evaluate genomic prediction approaches for disease resistance breeding.

In addition to disease resistance, RAD-Seq association studies have been widely applied for mapping QTL affecting a range of other production-relevant traits, particularly in salmonid species. These include spleen size (Liu *et al.* 2015a) and cortisol response (Liu *et al.* 2015b) in rainbow trout, and thermal tolerance and growth in *Oncorhynchus nerka*, the sockeye salmon (Larson *et al.* 2016). Out with the salmonid genera, RAD-Seq has been performed to map loci affecting disease resistance in olive flounder (*Paralichthys olivaceous*, Shao *et al.* 2015), and growth in bighead carp (*Hypophthalmichthys nobilis*, Fu *et al.* 2016) and turbot (Wang *et al.* 2015b). In addition, RAD-like techniques have been very popular for marker discovery and QTL mapping in bivalve shellfish including Chinese scallop (*Argopecten irradians*; Jiao *et al.* 2014), Akoya pearl oyster (*Pinctata fucata*; Li & He 2014; Shi *et al.* 2014), variously coloured abalone (*Haliotis diversicolor*; Ren *et al.* 2016; Yesso scallop (*Patinopecten yessoensis*; Dou *et al.* 2016) and have also been applied in the shrimp kuruma prawn (*Mar-supenaes japonicas*; Lu *et al.* 2016) and one echinoderm, the sea cucumber (*Apostichopus japonicus*; Tian *et al.* 2015). Interestingly, 2b-RAD has been the most common technique in bivalves, while in finfish, traditional RAD has been more widely utilized.

Using RAD to study sex determination

Sex determination (SD) is one of the most critical traits for many aquaculture species, as phenotypic sex is often not evident in juveniles and sexual dimorphism in growth rate is commonly observed. SD is complex in many fish species, often with polygenic control and an environmental component (reviewed in Martínez *et al.* 2014), and the application of large genotyping projects has been strongly recommended to screen for SD loci in fish (e.g. Pan *et al.* 2016). RAD-like techniques have clearly boosted our knowledge of SD in aquaculture, with studies in Nile tilapia (Palaiokostas *et al.* 2013a, 2015a), Atlantic halibut (*Hippoglossus hippoglossus*, Palaiokostas *et al.* 2013b), European sea bass (Palaiokostas *et al.* 2015b) and turbot (Wang *et al.* 2015b) finding putative sex determining loci. Controlling sex ratio is not only interesting to obtain higher growth rates, but also to avoid size dispersion or to delay sexual maturity. Further, there are some clear examples, like the sturgeon, where the commercial advantage of rearing fish of one sex over the other is obvious.

Genomic selection approaches

While QTL mapping and MAS approaches can be successful when the genetic architecture of a trait suggests a gene of major effect (e.g. IPNV resistance, Houston *et al.* 2008;

Moen *et al.* 2009), improvement of polygenic traits using genomic data is more effectively achieved using genomic prediction of breeding values (Meuwissen *et al.* 2001). Studies of genomic selection in aquaculture were first carried out in salmonid fish, with simulated (Sonesson & Meuwissen 2009; Lillehammer *et al.* 2013) and empirical (Ødegård *et al.* 2014; Tsai *et al.* 2015, 2016b; Vallejo *et al.* 2016) data, demonstrating the clear advantages over pedigree-based methods. Studies using varying marker densities for prediction in salmonids have highlighted that as few as a thousand SNPs may be adequate for achieving the gain in selection accuracy versus pedigree approaches (Ødegård *et al.* 2014; Tsai *et al.* 2015, 2016b). Therefore, it is reasonable to assume that RAD-like techniques may be useful for genomic selection in aquaculture breeding, as typical RAD SNP datasets comprise a few thousand SNPs. Indeed, the potential of this approach has already been highlighted for resistance to bacterial cold water disease in rainbow trout (Vallejo *et al.* 2016), for growth in Yesso scallop (Dou *et al.* 2016), and for resistance to pasteurellosis in gilthead sea bream (Palaiokostas *et al.* 2016).

Genetic traceability and aquaculture sustainability

One of the main concerns for aquaculture producers and consumers is to minimize the environmental impact of fish farming. In this sense, traceability tools are essential to assess the impact of aquaculture escapees in natural populations or distinguish between farmed and wild specimens. RAD-Seq has been utilized to obtain SNPs for sturgeon traceability and conservation (Ogden *et al.* 2013), which will contribute to enforce current legislation on aquaculture and fishing practices but also aid on the handling of wild stocks, critical for sustainable aquaculture. RAD-Seq is also the main tool of the European project AquaTrace (aquatrace.eu), the results of which have been recently presented in the European Aquaculture Society meeting in Edinburgh (Aquaculture Europe 2016). One of the AquaTrace objectives was to assess the impact of escapees on natural populations of European sea bass, gilthead sea bream and turbot, while also developing forensically validated tools for traceability purposes. The results highlighted the utility of RAD-Seq approaches to capture population or family specific variation making it a suitable tool for genetic traceability and conservation of natural populations. This is of the utmost importance for sustainable aquaculture growth, leading to lasting economic benefits, food safety and social acceptance.

RAD-Seq and SNP arrays, towards a peaceful co-existence

The development of NGS has greatly increased the amount of genomic resources available in the most important

aquaculture species, including genome assemblies for many of them. Alongside RNA-Seq and whole genome sequencing, RAD-Seq has contributed significantly to the availability of abundant genetic markers compared to a few years ago. While RAD-Seq and similar techniques are likely to remain the genotyping method of choice for species with few genomic resources, several medium and high-density SNP arrays are already available for aquaculture species (Atlantic salmon, Houston *et al.* 2014; Yáñez *et al.* 2016; channel catfish, Liu *et al.* 2014; common carp, Xu *et al.* 2014a; rainbow trout, Palti *et al.* 2015a; Pacific oyster and European flat oyster, Lapègue *et al.* 2014), and many more are unpublished or currently being produced and validated.

Single nucleotide polymorphism arrays are a type of DNA microarray, where hybridization of allele-specific probes results in a fluorescent signal which can be measured to call a genotype in a given loci. They have both advantages and disadvantages over RAD-Seq approaches (Table 3). For instance, the experimental procedures and bioinformatic analyses are much simpler for the user of SNP arrays, requiring less technical knowledge and usually resulting in a faster turnaround. The genotype scoring method is more robust and amenable to automation, and therefore less prone to errors (Hong *et al.* 2012; Wall *et al.* 2014). The repeatability and reproducibility are higher for SNP arrays than RAD-Seq, and genotyped loci are known in advance. However, having a fixed set of loci on the chip is also a disadvantage, especially in species with strong population structure, because of ascertainment bias whereby the SNP set is biased to polymorphic markers in the discovery population(s). This presents a major issue where aquaculture strains for a specific species are highly variable, and the utility of a SNP array will vary hugely depending on the relationship to the discovery population. RAD-like approaches overcome this issue and also offer much greater flexibility to the researcher in terms of the targeted number of loci. Further, RAD-Seq captures variation that is specific to populations, families and individuals that is likely to be missed from SNP array, which are typically biased towards common variants. Another putative advantage of RAD-like

Table 3 General comparison of restriction-site associated DNA sequencing (RAD-Seq) and single nucleotide polymorphism (SNP) chips

	RAD-Seq	SNP arrays
Sample processing	Laborious	Straightforward
Bioinformatic analysis	Complex	Negligible
Turnaround time	Long	Medium
Accuracy	Medium-high	High
Repeatability	Medium	High
Design	Adjustable	Fixed
Cost	Low	Medium

techniques is that the direct cost of the experiment is cheaper, although the additional time required for library preparation and bioinformatics analyses should be considered into any comparison.

In the near future, genomic selection (GS) is likely to be a key technique for breeding programmes of many aquaculture species, due to the demonstrable increase in selection accuracy versus current pedigree-based methods. SNP arrays are now routinely used in livestock breeding programmes for GS and are increasingly utilized in technologically advanced aquaculture breeding. Several studies have shown that only moderate SNP marker density is required for effective GS in salmon (Ødegård *et al.* 2014; Tsai *et al.* 2015, 2016b). Vallejo *et al.* (2016) compared both RAD-Seq and SNP arrays for GS to BCWD resistance in rainbow trout, finding similar selection accuracies for both techniques despite higher marker density from the SNP chip (~40k SNP array versus ~10k RAD-Seq). This may reflect high levels of linkage disequilibrium in typical aquaculture family selection programmes, whereby trait recording is often performed on close relatives of the selection candidates. Therefore, the higher marker density associated with SNP chips may be advantageous when predicting breeding values in animals more distantly related to the training population (Tsai *et al.* 2016b), or in species with greater effective population sizes and/or lower levels of linkage disequilibrium. However, given the relatively short genomes of many nonsalmonid aquaculture species (i.e. European sea bass – ~763 Mb, or turbot – ~658 Mb; Atlantic salmon – ~2970 Mb), the typical marker density generated by RAD-like techniques may be perfectly adequate for effective GS. However, this needs to be tested, as the recombination frequency and patterns of linkage disequilibrium across the genome are pertinent to the question of adequate marker density. Further reductions in marker density requirements are likely to be observed when genotype imputation approaches are used, for example genotyping parents at high density, and offspring for a small subset of the markers. As already mentioned, RAD methods allow for substantial flexibility in terms of number of genotyped markers. In addition, lowering average sequence coverage in the offspring with parents sequenced at high coverage could be used to generate genotype data at a much lower cost.

Targeted GBS techniques

Both RAD-Seq and SNP arrays will also have to compete with recently developed genotyping methods based on targeted genotyping by sequencing. For example, genotyping-in-thousands by sequencing (GT-Seq, Campbell *et al.* 2015) is a method of targeted sequencing

which follows a multiplex PCR approach, where hundreds to thousands of loci (amplicons) are selected for genotyping. In this method, a multiplex PCR using loci-specific primers that also contain Illumina sequencing primers is used to amplify the targeted regions. Unique barcodes for each sample are added with a second PCR reaction, followed by pooling and sequencing of samples. Unlike RAD techniques, this method requires previous knowledge to design the assays, and the number of SNPs genotyped in a single run is limited to a few thousand. Similar technologies are now provided by major genotyping technology providers, and it appears likely to become one of the most cost-effective systems of genotyping targeted SNPs. Other GBS targeted-sequencing techniques have also been recently developed, for example RAD capture (Rapture), where preselected RAD tags are isolated using capture probes and then sequenced (Ali *et al.* 2016). These targeted GBS techniques have the potential to become major players in aquaculture breeding and genetics due to their simplicity and flexibility. However, in part, they suffer from the same limitation as SNP arrays that they require prior knowledge and selection of the SNPs that are useful in the population of interest.

Future outlook

Restriction-site associated DNA sequencing techniques have driven a major increase in the application of genomics to aquaculture species. While the catalogue of SNP arrays for aquaculture species will increase in the coming years, it is likely that RAD techniques will continue to be widely applied. We anticipate that both techniques will co-exist for several years, and the choice of RAD-Seq or SNP chip will depend on the species and project-specific factors. For example, it may be that high-value aquaculture species with larger genomes (e.g. salmonids) are more suitable for SNP arrays, while lower-value species with smaller genomes (and/or higher levels of LD) are more suitable for RAD techniques, although it will also depend on the resources available for each particular project. Targeted GBS techniques like GT-Seq are likely to find a niche in genotyping hundreds to several thousands of previously identified SNPs across many samples. Further, RAD techniques are likely to remain the gold standard for new aquaculture species and/or those produced on a smaller scale, where SNP arrays are not available, and genomic resources are scarce. Eventually the cost of generating and analysing sequence data may drop to a level where genome complexity reduction is no longer required, but it seems unlikely in the short term. Therefore, RAD sequencing will continue to flourish in aquaculture research in the following years and is likely to be routinely applied to deliver the benefits of genomic

selection to selective breeding of many different aquaculture species.

Concluding remarks

The appearance of genotyping by sequencing technologies has provided the aquaculture research community with a hugely valuable method for identifying and concurrently genotyping large numbers of genetic markers in species with limited genomic resources. Further, these techniques have become multi-purpose tools for addressing several topics of research and commercial interest like genetic diversity, population and family structure, association analyses with traits of economic interest, and genomic selection. Despite the increasing availability of genomic resources and the increasing number of SNP arrays, RAD techniques will continue being important for aquaculture research and application to selective breeding in the next few years. RAD sequencing and other genotyping by sequencing currently offer unequalled versatility and cost-effectiveness for meeting the needs of many diverse research projects.

Acknowledgements

The authors are supported by funding from the European Union's Seventh Framework Programme (FP7 2007-2013) under grant agreement no. 613611 (FISHBOOST) and no.: 311920 (AquaTrace), Biotechnology and Biological Science Research Council (BBSRC) grants (BB/N024044/1, BB/M028321/1), and by BBSRC Institute Strategic Funding Grants to The Roslin Institute (BB/J004235/1, BB/J004324/1).

References

- Ali OA, O'Rourke SM, Amish SJ, Meek MH, Luikart G, Jeffres C *et al.* (2016) RAD capture (rapture): flexible and efficient sequence-based genotyping. *Genetics* **202**: 389–400.
- Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* **17**: 81–92.
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**: e3376.
- Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B *et al.* (2014) The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature Communications* **5**: 3657.
- Bradic M, Teotónio H, Borowsky RL (2013) The population genomics of repeated evolution in the blind cavefish *Astyanax mexicanus*. *Molecular Biology and Evolution* **30**: 2383–2400.
- Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S *et al.* (2015) The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**: 375–381.
- Brown JK, Taggart JB, Bekaert M, Wehner S, Palaiokostas C, Setiawan AN *et al.* (2016) Mapping the sex determination locus in the hapuku (*Polyprion oxygeneios*) using ddRAD sequencing. *BMC Genomics* **17**: 448.
- Campbell NR, LaPatra SE, Overturf K, Towner R, Narum SR (2014) Association mapping of disease resistance traits in rainbow trout using restriction site associated DNA sequencing. *G3 (Bethesda)* **4**: 2473–2481.
- Campbell NR, Harmon SA, Narum SR (2015) Genotyping-in-thousands by sequencing (GT-seq): a cost-effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources* **15**: 855–867.
- Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) *Stacks*: building and genotyping Loci de novo from short-read sequences. *G3 (Bethesda)* **1**: 171–182.
- Chavanne H, Janssen K, Hofherr J, Contini F, Haffray P, Aquatrace Consortium *et al.* (2016) A comprehensive survey on selective breeding programs and seed market in the European aquaculture fish industry. *Aquaculture International* **24**: 1287–1307.
- Chen S, Zhang G, Shao C, Huang Q, Liu G, Zhang P *et al.* (2014) Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nature Genetics* **46**: 253–260.
- Chen X, Zhong L, Bian C, Xu P, Qiu Y, You X *et al.* (2016) High-quality genome assembly of channel catfish, *Ictalurus punctatus*. *Gigascience* **5**: 39.
- DaCosta JM, Sorenson MD (2014) Amplification biases and consistent recovery of loci in a double-digest RAD-Seq protocol. *PLoS One* **9**: e106713.
- Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**: 499–510.
- Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML (2013) Special features of RAD sequencing data: implications for genotyping. *Molecular Ecology* **22**: 3151–3164.
- Dou J, Li X, Fu Q, Jiao W, Li Y, Li T *et al.* (2016) Evaluation of the 2b-RAD method for genomic selection in scallop breeding. *Scientific Reports* **6**: 19244.
- Eaton DAR (2014) PyRA: assembly of denovo RADseq loci for phylogenetic analysis. *Bioinformatics* **30**: 1844–1849.
- Everett MV, Seeb JE (2014) Detection and mapping of QTL for temperature tolerance for body size in Chinook salmon (*Oncorhynchus tshawytscha*) using genotyping by sequencing. *Evolutionary Applications* **7**: 480–492.
- Etter PD, Preston JL, Bassham S, Cresko WA, Johnson EA (2011) Local de novo assembly of RAD paired-end contigs using short sequencing reads. *PLoS One* **6**: e18561.
- Fierst JL (2015) Using linkage maps to correct and scaffold de novo genome assemblies: methods, challenges, and computational tools. *Frontiers in Genetics* **6**: 220.

- Figueras A, Robledo D, Corvelo A, Hermida M, Pereiro P, Rubiolo JA *et al.* (2016) Whole genome sequencing of turbot (*Scophthalmus maximus*; Pleuronectiformes): a fish adapted to demersal life. *DNA Research* **23**: 181–192.
- Fu B, Liu H, Yu X, Tong J (2016) A high-density genetic map and growth related QTL mapping in bighead carp (*Hypophthalmichthys nobilis*). *Scientific Reports* **6**: 28679.
- Gautier M, Gharbi K, Cezard T, Foucaud J, Kerdelhué C, Pudlo P (2013) The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology* **22**: 3165–3178.
- Gjedrem T, Robinson N, Rye M (2012) The importance of selective breeding in aquaculture to meet future demands for animal protein: a review. *Aquaculture* **350–353**: 117–129.
- Gonen S, Lowe NR, Cezard T, Gharbi K, Bishop SC, Houston RD (2014) Linkage maps of the Atlantic salmon (*Salmo salar*) genome derived from RAD sequencing. *BMC Genomics* **15**: 166.
- Gonen S, Bishop SC, Houston RD (2015) Exploring the utility of cross-laboratory RAD-sequencing datasets for phylogenetic analysis. *BMC Research Notes* **8**: 299.
- Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science* **92**: 433–443.
- Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6**: e1000862.
- Hong H, Xu L, Liu J, Jones WD, Su Z, Ning B *et al.* (2012) Technical reproducibility of genotyping SNP arrays used in genome-wide association studies. *PLoS One* **7**: e44483.
- Houston RD, Haley CS, Hamilton A, Guy DR, Tinch AE, Taggart JB *et al.* (2008) Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). *Genetics* **178**: 1109–1115.
- Houston RD, Davey JW, Bishop SC, Lowe NR, Mota-Velasco JC, Hamilton A *et al.* (2012) Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC Genomics* **13**: 244.
- Houston RD, Taggart JB, Cézard T, Bekaert M, Lowe NR, Downing A *et al.* (2014) Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). *BMC Genomics* **15**: 90.
- Janssen K, Chavanne H, Berentsen P, Komen H (2016) Impact of selective breeding on European aquaculture. *Aquaculture*. doi: 10.1016/j.aquaculture.2016.03.012.
- Jiao W, Fu X, Dou J, Li H, Su H, Mao J *et al.* (2014) High-resolution linkage and quantitative trait locus mapping aided by genome survey sequencing: building up an integrative genomic framework for a bivalve mollusc. *DNA Research* **21**: 85–101.
- Kai W, Nomura K, Fujiwara A, Nakamura Y, Yasuike M, Ojima N *et al.* (2014) A ddRAD-based genetic map and its integration with the genome assembly of Japanese eel (*Anguilla japonica*) provides insights into genome evolution after the teleost-specific genome duplication. *BMC Genomics* **15**: 233.
- Kakioka R, Kokita T, Kumada H, Watanabe K, Okuda N (2013) A RAD-based linkage map and comparative genomics in the gudgeons (genus *Gnathopogon*, Cyprinidae). *BMC Genomics* **14**: 32.
- Lapègue S, Harrang E, Heurtebise S, Flahauw E, Donnadiou C, Gayral P *et al.* (2014) Development of SNP-genotyping arrays in two shellfish species. *Molecular Ecology Resources* **14**: 820–830.
- Larson WA, McKinney GJ, Limborg MT, Everett MV, Seeb LW, Seeb JE (2016) Identification of multiple QTL hotspots in sockeye salmon (*Oncorhynchus nerka*) using genotyping-by-sequencing and a dense linkage map. *Journal of Heredity* **107**: 122–133.
- Lepais O, Weir JT (2014) SimRAD: an R package for simulation-based prediction of the number of loci expected in RAD-seq and similar genotyping by sequencing approaches. *Molecular Ecology Resources* **14**: 1314–1321.
- Li Y, He M (2014) Genetic mapping and QTL analysis of growth-related traits in *Pinctada fucata* using restriction-site associated DNA sequencing. *PLoS One* **9**: e111707.
- Li Y, Liu S, Qin Z, Waldbieser G, Wang R, Sun L *et al.* (2014) Construction of a high-density, high-resolution genetic map and its integration with BAC-based physical map in channel catfish. *DNA Research* **22**: 39–52.
- Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T *et al.* (2016) The Atlantic salmon genome provides insights into rediploidization. *Nature* **533**: 200–205.
- Lillehammer M, Meuwissen TH, Sonesson AK (2013) A low-marker density implementation of genomic selection in aquaculture using within-family genomic breeding values. *Genetics Selection Evolution* **45**: 39.
- Liu S, Sun L, Li Y, Sun F, Jiang Y, Zhang Y *et al.* (2014) Development of the catfish 250K SNP array for genome-wide association studies. *BMC Research Notes* **7**: 135.
- Liu S, Vallejo RL, Gao G, Palti Y, Weber GM, Hernandez A *et al.* (2015a) Identification of single-nucleotide polymorphism markers associated with cortisol response in rainbow trout. *Marine Biotechnology* **17**: 328–337.
- Liu S, Vallejo RL, Palti Y, Gao G, Marancik DP, Hernandez AG *et al.* (2015b) Identification of single nucleotide polymorphism markers associated with bacterial cold water disease resistance and spleen size in rainbow trout. *Frontiers in Genetics* **6**: 298.
- Lu X, Luan S, Hu LY, Mao Y, Tao Y, Zhong SP *et al.* (2016) High-resolution genetic linkage mapping, high-temperature tolerance and growth-related quantitative trait locus (QTL) identification in *Marsupenaeus japonicus*. *Molecular Genetics and Genomics* **291**: 1391–1405.
- Manousaki T, Tsakogiannis A, Taggart JB, Palaiokostas C, Tsaparis D, Lagnel J *et al.* (2015) Exploring a nonmodel teleost genome through RAD sequencing-linkage mapping in common pandora, *Pagellus erythrinus* and comparative genomic analysis. *G3 (Bethesda)* **6**: 509–519.

- Martínez P, Viñas A, Sánchez L, Díaz N, Ribas L, Piferrer F (2014) Genetic architecture of sex determination in fish: applications to sex ratio control in aquaculture. *Frontiers in Genetics* **5**: 340.
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome wide dense marker maps. *Genetics* **157**: 1819–1829.
- Miller MR, Atwood TS, Eames BF, Eberhart JK, Yan YL, Postlethwait JH *et al.* (2007) RAD marker microarrays enable rapid mapping of zebrafish mutations. *Genome Biology* **8**: R105.
- Moen T, Baranski M, Sonesson AK, Kjøglum S (2009) Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): population-level associations between markers and trait. *BMC Genomics* **10**: 368.
- Murgarella M, Puiu D, Novoa B, Figueras A, Posada D, Canchaya C (2016) A first insight into the genome of the filter-feeding mussel *Mytilus galloprovincialis*. *PLoS One* **11**: e0151561.
- Ødegård J, Moen T, Santi N, Korsvoll SA, Kjøglum S, Meuwissen THE (2014) Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). *Frontiers in Genetics* **5**: 402.
- Ogden R, Gharbi R, Mugue N, Martinsohn J, Senn H, Davey JW *et al.* (2013) Sturgeon conservation genomics: SNP discovery and validation using RAD sequencing. *Molecular Ecology* **22**: 3112–3123.
- Palaiokostas C, Bekaert M, Davie A, Cowan ME, Oral M, Taggart JB *et al.* (2013a) Mapping the sex determination locus in Atlantic halibut (*Hippoglossus hippoglossus*) using RAD sequencing. *BMC Genomics* **14**: 566.
- Palaiokostas C, Bekaert M, Khan MG, Taggart JB, Gharbi K, McAndrew BJ *et al.* (2013b) Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) using RAD sequencing. *PLoS One* **8**: e68389.
- Palaiokostas C, Bekaert M, Khan MG, Taggart JB, Gharbi K, McAndrew BJ *et al.* (2015a) A novel sex-determining QTL in Nile tilapia (*Oreochromis niloticus*). *BMC Genomics* **16**: 171.
- Palaiokostas C, Bekaert M, Taggart JB, Gharbi K, McAndrew BJ, Chatain B *et al.* (2015b) A new SNP-based vision of the genetics of sex determination in European sea bass (*Dicentrarchus labrax*). *Genetics Selection Evolution* **47**: 68.
- Palaiokostas C, Ferrareso S, Franch R, Houston RD, Bargelloni L (2016) Genomic prediction of resistance to pasteurellosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. *G3 (Bethesda)* **6**: 3693–3700.
- Palti Y, Gao G, Miller MR, Vallejo RL, Wheeler PA, Quillet E *et al.* (2014) A resource of single-nucleotide polymorphisms for rainbow trout generated by restriction-site associated DNA sequencing of doubled haploids. *Molecular Ecology Resources* **14**: 588–596.
- Palti Y, Gao G, Liu S, Kent MP, Lien S, Miller MR *et al.* (2015a) The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. *Molecular Ecology Resources* **15**: 662–672.
- Palti Y, Vallejo RL, Gao G, Liu S, Hernandez AG, Rexroad CE 3rd *et al.* (2015b) Detection and validation of QTL affecting bacterial cold water disease resistance in rainbow trout using restriction-site associated DNA sequencing. *PLoS One* **10**: e0138435.
- Pan Q, Anderson J, Bertho S, Herpin A, Wilson C, Postlethwait JH *et al.* (2016) Vertebrate sex-determining genes play musical chairs. *Comptes Rendus Biologies* **339**: 258–262.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7**: e37135.
- Puritz JB, Matz MV, Toonen RJ, Weber JN, Bolnick DI, Bird CE (2014a) Demystifying the RAD fad. *Molecular Ecology* **23**: 5937–5942.
- Puritz JB, Hollenbeck CM, Gold JR (2014b) dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* **2**: e431.
- Recknagel H, Elmer KR, Meyer A (2013) A hybrid genetic linkage map of two ecologically and morphologically divergent Midas cichlid fishes (*Amphilophus* spp.) obtained by massively parallel DNA sequencing (ddRADSeq). *G3 (Bethesda)* **3**: 65–74.
- Ren P, Peng W, You W, Huang Z, Guo Q, Chen N *et al.* (2016) Genetic mapping and quantitative trait loci analysis of growth-related traits in the small abalone *Haliotis diversicolor* using restriction-site-associated DNA sequencing. *Aquaculture* **454**: 163–170.
- Rondeau EB, Minkley DR, Leong JS, Messmer AM, Jantzen JR, von Schalburg KR *et al.* (2014) The genome and linkage map of the Northern pike (*Esox Lucius*): conserved synteny revealed between the salmonid sister group and the Neoteleostei. *PLoS One* **9**: e102089.
- Schlötterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals – mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics* **15**: 749–763.
- Schweyen H, Rozenberg A, Leese F (2014) Detection and removal of PCR duplicates in population genomic ddRAD studies by addition of a degenerate base region (DBR) in sequencing adapters. *Biological Bulletin* **227**: 146–160.
- Shao C, Niu Y, Rastas P, Liu Y, Xie Z, Li H *et al.* (2015) Genome-wide SNP identification for the construction of a high-resolution genetic map of Japanese flounder (*Paralichthys olivaceus*): applications to QTL mapping of *Vibrio anguillarum* disease resistance and comparative genomic analysis. *DNA Research* **22**: 161–170.
- Shi Y, Wang S, Gu Z, Lv J, Zhan X, Yu C *et al.* (2014) High-density nucleotide polymorphisms linkage and quantitative trait locus mapping of the pearl oyster, *Pinctada fucata martensii* Dunker. *Aquaculture* **434**: 376–384.
- Sonesson A, Meuwissen T (2009) Testing strategies for genomic selection in aquaculture breeding programs. *Genetics Selection Evolution* **41**: 37.

- Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrøm M, Gregers TF *et al.* (2011) The genome sequence of Atlantic cod reveals a unique immune system. *Nature* **477**: 207–210.
- Sun X, Liu D, Zhang X, Li W, Liu H, Hong W *et al.* (2013) SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS One* **8**: e58700.
- Tian M, Li Y, Jing J, Mu C, Du H, Doi J *et al.* (2015) Construction of a high-density genetic map and quantitative trait locus mapping in the sea cucumber *Apostichopus japonicus*. *Scientific Reports* **5**: 14852.
- Tine M, Kuhl H, Gagnaire PA, Louro B, Desmarais E, Martins RS *et al.* (2014) European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications* **5**: 5770.
- Toonen RJ, Puritz JB, Forsman ZH, Whitney JL, Fernandez-Silva I, Andrews KR *et al.* (2013) ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ* **1**: e203.
- Tsai HY, Hamilton A, Tinch AE, Guy DR, Gharbi K, Stear MJ *et al.* (2015) Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics* **16**: 969.
- Tsai HY, Robledo D, Lowe NR, Bekaert M, Taggart JB, Bron JE *et al.* (2016a) Construction and annotation of a high density SNP linkage map of the Atlantic salmon (*Salmo salar*) genome. *G3 (Bethesda)* **6**: 2173–2179.
- Tsai HY, Hamilton A, Tinch AE, Guy DR, Bron JE, Taggart JB *et al.* (2016b) Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. *Genetics Selection Evolution* **48**: 47.
- Vallejo RL, Leeds TD, Fragonmeni BO, Gao G, Hernandez AG, Misztal I *et al.* (2016) 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. *Frontiers in Genetics* **7**: 96.
- Vandeputte M, Haffray P (2014) Parentage assignment with genomic markers: a major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. *Frontiers in Genetics* **5**: 432.
- Vij S, Kuhl H, Kuznetsova IS, Komissarov A, Yurchenko AA, Van Heusden P *et al.* (2016) Chromosomal-level assembly of the Asian seabass genome using long sequence reads and multi-layered scaffolding. *PLoS Genetics* **12**: e1005954.
- Wall JD, Tang LF, Zerbe B, Kvale MN, Kwok PY, Shaefer C *et al.* (2014) Estimating genotype error rates from high-coverage next-generation sequence data. *Genome Research* **24**: 1734–1739.
- Wang S, Meyer E, McKay JK, Matz MV (2012) 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods* **9**: 808–810.
- Wang L, Wang ZY, Bai B, Huang SQ, Chua E, Lee M *et al.* (2015a) Construction of a high-density linkage map and fine mapping of QTL for growth in Asian seabass. *Scientific Reports* **5**: 16358.
- Wang W, Hu Y, Ma Y, Xu L, Guan J, Kong J (2015b) High-density genetic linkage mapping in turbot (*Scophthalmus maximus* L.) based on SNP markers and major sex- and growth-related regions detection. *PLoS One* **13**: e0120410.
- Xu J, Zhao Z, Zhang X, Zheng X, Li J, Jiang Y *et al.* (2014a) Development and evaluation of the first high-throughput SNP array for common carp (*Cyprinus carpio*). *BMC Genomics* **15**: 307.
- Xu P, Zhang X, Wang X, Li J, Liu G, Kuang Y *et al.* (2014b) Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nature Genetics* **46**: 1212–1219.
- Yáñez JM, Houston RD, Newman S (2014) Genetics and genomics of disease resistance in salmonid species. *Frontiers in Genetics* **5**: 415.
- Yáñez JM, Naswa S, López ME, Bassini L, Correa K, Gilbey J *et al.* (2016) Genomewide single nucleotide polymorphism discovery in Atlantic salmon (*Salmo salar*): validation in wild and farmed American and European populations. *Molecular Ecology Resources* **16**: 1002–1011.
- Zhang G, Fang X, Guo X, Li L, Luo R, Xu F *et al.* (2012) The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**: 49–54.