Significant Synteny and Colocalization of Ecologically Relevant Quantitative Trait Loci Within and Across Species of Salmonid Fishes

Arne Jacobs,* Robyn Womack,* Mel Chen,*,† Karim Gharbi,‡ and Kathryn R. Elmer*,1

*Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences and †School of Mathematics and Statistics, College of Science and Engineering, University of Glasgow, G12 8QQ, UK, and ‡Edinburgh Genomics, Ashworth Laboratories, University of Edinburgh, EH9 3FL, UK

ORCID IDs: 0000-0001-7635-5447 (A.J.); 0000-0003-4952-9848 (R.W.); 0000-0002-9219-7001 (K.R.E.)

ABSTRACT The organization of functional regions within genomes has important implications for evolutionary potential. Considerable research effort has gone toward identifying the genomic basis of phenotypic traits of interest through quantitative trait loci (QTL) analyses. Less research has assessed the arrangement of QTL in the genome within and across species. To investigate the distribution, extent of colocalization, and the synteny of QTL for ecologically relevant traits, we used a comparative genomic mapping approach within and across a range of salmonid species. We compiled 943 QTL from all available species [lake whitefish (*Coregonus clupeaformis*), coho salmon (*Oncorhynchus kisutch*), rainbow trout (*O. mykiss*), Chinook salmon (*O. tshawytscha*), Atlantic salmon (*Salmo salar*), and Arctic charr (*Salvelinus alpinus*)]. We developed a novel analytical framework for mapping and testing the distribution of these QTL. We found no correlation between QTL density and gene density at the chromosome level but did at the fine-scale. Two chromosomes were significantly enriched for QTL. We found multiple synteny blocks for morphological, life history, and physiological traits across species, but only morphology and physiology had significantly more than expected. Two or three pairs of traits were significantly colocalized in three species (lake whitefish, coho salmon, and rainbow trout). Colocalization and fine-scale synteny suggest genetic linkage between traits within species and a conserved genetic basis across species. However, this pattern was weak overall, with colocalization and synteny being relatively rare. These findings advance our understanding of the role of genomic organization in the renowned ecological and phenotypic variability of salmonid fishes.

KEYWORDS quantitative trait loci (QTL); comparative genetic mapping; genome organization; genome evolution; meta-analysis

The arrangement of functional regions within and across chromosomes has important implications for the evolution of genomes and phenotypes within and across species (Lynch and Walsh 2007; Yeaman 2013; Schwander *et al.* 2014). At the finest genomic resolution, functional regions are those genes and regulatory mechanisms that determine phenotypes; more broadly, those regions are considered to be QTL (Mackay *et al.* 2009). While considerable emphasis and research effort has sought to identify the genetic location of QTL (*e.g.*, McClelland and Naish 2010; Easton *et al.* 2011; Le Bras *et al.* 2011), less effort has been put into identifying the consistency of how those loci are functionally organized within genomes, and particularly the extent of synteny across closely related species.

Within species, the clustering or colocalization of QTL for ecologically relevant traits is thought to play an important role in facilitating phenotypic evolution (Yeaman 2013; Schwander *et al.* 2014; Charlesworth 2016), adaptation, and ultimately speciation (Hawthorne and Via 2001; Feder *et al.* 2012). The suppression of recombination is thought to be a prerequisite for the formation and maintenance of clusters of QTL in most cases (Roberts *et al.* 2009; Yeaman 2013; Schwander *et al.* 2014), although exceptions exist (Joron *et al.* 2006). Such regions of suppressed recombination can make up a substantial amount of the genome and spread over several megabase pairs (Mbp) (Schwander *et al.* 2014; Küpper *et al.* 2016;

doi: https://doi.org/10.1534/genetics.117.300093
Manuscript received January 31, 2017; accepted for publication July 21, 2017; published Early Online July 31, 2017.
Supplemental material is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.117.300093/-/DC1.
[1]Corresponding author: Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK. E-mail: kathryn.elmer@glasgow.ac.uk

Lamichhaney *et al.* 2016). The colocalization of QTL has been observed in several species affecting a range of traits. The most famous examples are "supergenes," clusters of loci affecting many complex (adaptive) phenotypes, such as floral heteromorphy in several plant species (Mather 1950; Gilmartin and Li 2010), plumage polymorphism in the white-throated sparrow (Thomas *et al.* 2008; Tuttle *et al.* 2016), alternative mating strategies in the ruff (Küpper *et al.* 2016; Lamichhaney *et al.* 2016), Batesian mimicry in butterflies (Joron *et al.* 2006, 2011), and the sex-chromosome system (Uno *et al.* 2008; Charlesworth and Mank 2010; Vicoso *et al.* 2013; Charlesworth 2016). However, even in less pronounced cases the colocalization of QTL can explain covariation in phenotypes observed in the wild, such as colocalizations of loci associated with growth rate and morphological traits (body weight and length) in coho salmon (*Oncorhynchus kisutch*) (McClelland and Naish 2010), eco-morphological traits (jaw shape and body shape) in cichlid fishes (Fruciano *et al.* 2016), or loci for flowering time and winter frost tolerance in pea (*Pisum sativum*) (Lejeune-Hénaut *et al.* 2008).

Although many well-described cases of colocalization within species exist in the literature (Mather 1950; Joron *et al.* 2006, 2011; Thomas *et al.* 2008; Gilmartin and Li 2010; Miller *et al.* 2014; Peichel and Marques 2017), the synteny of functional regions across closely related species has been less explored. There is some evidence for the reuse of QTL for a range of traits in closely related species, such as salinity tolerance and body weight in salmonid fishes, and components of drought resistance in grass species (Zhang *et al.* 2001; Reid *et al.* 2004; Norman *et al.* 2011; Swamy *et al.* 2011; Conte *et al.* 2012, 2015; Larson *et al.* 2016). Following this, the probability of gene or QTL reuse in examples of parallel and convergent phenotypic evolution in particular has been estimated to fall in the range of 0.32–0.47 (Conte *et al.* 2012, 2015), with the probability of gene reuse declining with increasing divergence time between species (Conte *et al.* 2012). For understanding the clustering and reuse of QTL across species, it is important to first identify their organization and second determine the factors influencing the location of such loci within genomes. For example, studies in sorghum wheat have shown that QTL-dense regions colocalize with gene-dense regions due to the suppression of recombination in heterochromatic regions (Mace and Jordon 2011). Such investigations are only possible with high-quality reference genomes and until now were limited to model organisms. However, the increasing availability of high-quality reference genomes for nonmodel organisms enables the investigation of factors influencing the distribution of QTL in a wide range of ecological and evolutionary species of interest, including salmonid fishes.

Fish of the family Salmonidae are an ideal biological system for investigating the origins and evolution of the genetic architecture of ecologically relevant traits. Highly variable traits include several morphological features, such as body shape, body size, and trophic morphology, as well as physiological and life history traits (Jonsson and Jonsson 2001; Klemetsen *et al.* 2003; Bernatchez *et al.* 2010; Hale *et al.* 2013; Hecht *et al.* 2013; Elmer 2016). Often this phenotypic diversity is replicated within and across species, such as trophic polymorphisms in Arctic charr (*Salvelinus alpinus*) (Hindar and Jonsson 1993; Adams *et al.* 1998, 2008; Alekseyev *et al.* 2002; Gordeeva *et al.* 2015) and whitefish (*Coregonus* species) (Bernatchez *et al.* 1996, 2010; Kahilainen and Østbye 2006; Vonlanthen *et al.* 2012; Siwertsson *et al.* 2013), or migratory behaviors in various anadromous species (Hale *et al.* 2013; Hecht *et al.* 2013). Colocalization of some traits has been postulated as a factor underpinning rapid replicated diversification (Bernatchez *et al.* 2010; Gagnaire *et al.* 2013; Elmer 2016; Larson *et al.* 2016). Compared to other teleosts, Salmonidae display an elevated speciation rate that falls into the top 10% of all extant ray-finned fishes (Rabosky *et al.* 2013), making them a powerful system for the analysis of the genomic basis of rapid and replicated diversification.

Salmonids are distinctive among teleosts in that their common ancestor underwent an additional whole-genome duplication (Ss4R) around 80–100 MYA (Berthelot *et al.* 2014; Lien *et al.* 2016). Mapping trait-specific QTL to homologous chromosomes or linkage groups suggests that the Ss4R had implications for the evolution of phenotypic traits in salmonids (Norman *et al.* 2012). In addition, genetic mapping studies have shown that most chromosome arms are syntenic between species, large collinear genomic blocks are retained, and that the chromosome arm number (NF) of ~100 is conserved within the Salmoninae subfamily (Ohno 1970; Danzmann *et al.* 2005; Phillips *et al.* 2009; Lien *et al.* 2011, 2016; Timusk *et al.* 2011; Guyomard *et al.* 2012), with the exception of Atlantic salmon, NF = 72–74 (Allendorf and Thorgaard 1984; Phillips and Ráb 2001). The conservation of gene arrangements, and the retention of large collinear blocks across species, allows for the comparison of QTL arrangements within those retained regions across species.

Because of their commercial and scientific importance, hundreds of quantitative loci for a range of phenotypic traits have been identified in salmonids over the last two decades, allowing the extensive comparison of QTL positions for a suite of ecologically relevant traits within and across species. Studies have found synteny of trait-specific QTL across three or four salmonid species at the chromosome level (Norman *et al.* 2012; Larson *et al.* 2016), but the extent of this phenomenon across traits and species more broadly has not been investigated with a genomic approach and little is known about the extent of colocalization within species.

Here, we conducted a comparative genomic analysis of all available ecologically relevant traits (specifically: morphology, physiology, swimming behavior, and life history) within and across salmonid species to assess the extent of colocalization and synteny of QTL. We focused on this suite of traits in particular to assess the potential impact of intra- and interspecific genome-wide QTL arrangements on the renowned ecologically relevant phenotypic diversity observed in salmonids (Jonsson and Jonsson 2001; Klemetsen *et al.* 2003; Bernatchez *et al.* 2010; Hale *et al.* 2013; Hecht *et al.* 2013;

Elmer 2016). To date, comparing loci across species has been complicated by methodological differences, such as the use of different marker types or genotyping protocols across studies. The recent publication of the rainbow trout and Atlantic salmon reference genomes (Berthelot *et al.* 2014; Lien *et al.* 2016) makes it only now possible to compare the relative positions of QTL to each other independent of marker type and genotyping protocols. First, we developed a comprehensive database of published QTL for salmonid fishes, which is made available with this study. Second, we assessed the genome-wide distribution of these ecologically relevant QTL, including their arrangement across chromosomes and relative to all available annotated genes. To date chromosome-level analysis has been the sole comparative level possible in the rich literature of nonmodel species such as salmonids. We advanced this by assessing fine-scale positional information and codistribution of QTL-linked markers and genes along the reference genome. Third, we used this QTL distribution to establish a quantitative framework for analyzing colocalization of loci within species and synteny of loci across species. Fourth, given this analytical framework, we investigated the extent of synteny across all available salmonid species and identified genomic regions that might play a conserved role in the evolution of adaptive phenotypic traits in salmonids, such as morphological, physiological, or life history traits. Finally, we quantified the extent of colocalization of QTL within species to detect phenotypic traits that might be controlled by tight linked genetic changes or by a single pleiotropic genetic change. We present evidence for low levels of QTL clustering across the genome relative to background gene density and explore factors behind this, identify significant synteny clustering for morphological and physiological traits across multiple salmonid species, and find low levels of significant QTL colocalization within species.

Materials and Methods

Data collection

A literature search was conducted for salmonid QTL information on the Web of Science and Google Scholar databases using the key terms "QTL," "quantitative trait loci," and "salmonid" individually and combined, and also scientific and common names for various salmonid species (studies published and accessible up to November 2015). QTL information was collected for the following ecologically relevant quantitative phenotypic traits grouped into four trait classes: (1) morphology: body weight, body length, body shape, Fulton's condition factor, head shape, and gill raker number; (2) life history: hatching time, spawning time, age at sexual maturation, growth rate, and embryonic development; (3) physiology: upper thermal tolerance, osmoregulatory ability, and salinity tolerance (Na^+/K^+-ATPase activity); and (4) behavior: directional change, depth selection, activity, and burst swimming. For each QTL, the following information was collected if available: a description of the phenotypic trait studied, linkage group on which the QTL was located, position of QTL on linkage group in centiMorgans, the names of the closest marker(s) located to the QTL, type of marker(s) used in the study (e.g. microsatellite markers, restriction site-associated DNA sequencing tags (RADtags), or SNPs), nucleotide sequences for the marker(s) closest to the QTL, logarithm of the odds (LOD)-score and/or percent of phenotypic variance explained (PVE) by the QTL, significance (*P*-value), and confidence interval (C.I.) [1.5-LOD interval, 2-LOD interval, or 95% C.I. in centiMorgans]. Studies using anonymous markers were excluded. Repeated uses of the same marker in different studies were retained in the database. These redundant markers were excluded from downstream analyses when they held redundant positional information within a species (*i.e.*, a marker in different studies on the same trait in the same species) but retained for synteny analysis when the same marker is used across species. QTL information was compiled in a spreadsheet database, where each QTL was assigned a unique code for reference, *e.g.*, Omy_BW_001 (full database available in Supplemental Material, Table S1).

Mapping of QTL sequences against the Atlantic salmon reference genome

For the six species combined, QTL-linked marker sequences (total $N = 1960$ sequences for 943 QTL) were aligned against the Atlantic salmon genome (Atlantic salmon reference genome ICSASG_v2) using three different approaches to overcome the caveats of different alignment strategies for the different marker types and sequence lengths. We used the recently published Atlantic salmon genome (Lien *et al.* 2016) as a consistent shared reference throughout as it is the highest quality reference available at present. First, Bowtie v1.1.1 (Langmead *et al.* 2009) was used for short microsatellite primer sequences while allowing for one mismatch, except in the case of lake whitefish, where three mismatches were permitted (-n), given its more distant phylogenetic relationship. Second, Bowtie2 v 2.2.6 (Langmead and Salzberg 2012) was used mainly for RAD loci and ESTs but also secondarily assessed microsatellite primer sequence mapping using the following settings: -D 5 -R 1 -N 1 -L 23 -i S,0,2.50. Furthermore, all QTL sequences were blasted against the reference genome using Blast+ v2.3 (Camacho *et al.* 2009) to increase alignment results for phylogenetically more distant species and long EST sequences. The best blast hits with a minimum e-value of e-13 were used for further analysis. When markers mapped using two or all methods, the bowtie1-alignment for primer sequences or bowtie2-alignment for RAD loci was used for further analyses. To avoid false placement of QTL on the genome, we removed QTL for which flanking markers mapped to different chromosomes and those for which markers mapped to different genome locations with equally high quality scores.

The C.I. for each QTL was determined in one of three different ways, depending on the information reported in the original study: (i) if only flanking markers were reported and mapped, we inferred the C.I. based on the physical distance

between flanking markers on the chromosome; (ii) if the peak marker and the C.I. were reported then we converted the C.I. into physical genome distance from recombination distance using an approximated conversion ratio of 1 cM = 1 Mbp; or (iii) if only the peak marker was reported without C.I.s, or only one flanking marker mapped to the genome, we estimated the C.I. based on the average C.I. length for a given species, since these differed between species. In a final filtering step, we removed redundant QTL-linked markers in cases where two or more markers for the same trait in the same species overlapped based on their C.I.s by only keeping the QTL-linked marker with the highest LOD-score or PVE. This filtered data set was used for all downstream analysis. We assessed if PVE differs between the different trait classes using a Kruskal–Wallis test with a *post hoc* Dunn test in R.

Analysis of QTL distribution across the genome

To evaluate the distribution of mapped QTL-linked markers across the reference genome, several statistical analyses in R software v3.2.2 (R Development Core Team 2016) were used. Peak marker position or the QTL-midpoint were used for all analyses of QTL distribution. First, we analyzed the correlation of chromosome length, number of genes per chromosome, and number of QTL to examine their relationship and influence on the distribution of QTL across chromosomes (Table S5). To correct for the impact of chromosome length on the number of QTL and genes, we also tested if QTL density (number of QTL per kbp) and gene density (number of genes per kbp) are correlated. Statistical significance of correlations was assessed using a permutation approach and Pearson correlations.

Second, to assess if certain chromosomes are enriched for QTL-linked markers or a specific trait class, we estimated the expected number of QTL per chromosome (null distribution) by distributing the number of QTL-linked markers relative to chromosome length for each chromosome (*i.e.*, 407 QTL for all trait classes and species combined across all 29 chromosomes). This model was found to fit best compared to models based on a uniform distribution of QTL across chromosomes or based on the number of genes on each chromosome. This leads to a (marginal) distribution for each chromosome given by a binomial distribution (N, chromosome length/total length). To test if certain chromosomes are significantly enriched for QTL, we calculated two-tailed *P*-values using the qbinom function in R, given the actual number of tested QTL-linked markers and adjusted for multiple testing using the Benjamini–Hochberg procedure using the *p.adjust* method in R. This test was performed for all trait classes combined and for each trait class separately, each time adjusting the number of distributed QTL.

Third, to analyze the codistribution of genes and QTL across the entire genome, the density of genes and QTL-linked markers was plotted along the genome in 5 Mbp windows, with a step size of 2.5 Mbp using the R-package *ggplot2* (Wickham 2009). The window size of 5 Mbp was empirically selected as it balances the number of QTL-linked markers and genes present in each window without returning too many windows without QTL-linked markers. A Pearson correlation was used to test if the number of QTL and number of genes per window are correlated.

Fourth, to identify clusters of QTL in the genome for all traits combined and for each specific trait class separately, we used a qbinom test to detect 5 Mbp windows with more QTL than expected based on the number of genes in each window. We used the number of genes per window as the null distribution because the number of QTL and genes are significantly correlated. We calculated a conservative probability of success in each trial for the qbinom test by using the following formula: 1 – (1 / number of windows tested). This gives a highly conservative cut-off probability of 0.9988 (Yeaman *et al.* 2016). Windows with more QTL-linked markers (for all traits combined or QTL by trait class) than expected based on the number of genes in the respective window were then defined as QTL clusters. The position of QTL and C.I.s for each trait class was plotted along the genome to visualize and analyze the distribution using *ggplot2*.

Optimal interval size for colocalization and synteny simulations

To identify the optimal interval size for defining colocalization and synteny in our simulations of null distributions, we examined the correlation between the observed synteny (based on overlapping C.I.s) with the inferred synteny (based on the distance of peak markers and/or midpoint positions for different interval sizes), testing across the first five chromosomes. We used peak markers or midpoints because simulations of QTL-linked marker distributions, colocalization, and synteny are unduly complex when calculated using empirical C.I.s. We defined QTL as colocalized if peak markers were closer to each other than the optimal interval distance. We derived the optimal interval distance by comparing different interval sizes spanning 25–100% the average empirical C.I. length (15 Mbp; Figure S1 in File S1): 3.75, 7.5, 10, 12, and 15 Mbp. Correlations were assessed using a Mantel test implemented in the *vegan* R-package (Oksanen *et al.* 2015). The 10 and 12 Mbp intervals were chosen because the Mantel *r* value peaked around those values in the initial analysis. The optimal interval size was determined to be 12 Mbp (see *Results* for details).

Synteny analysis of QTL clusters across species

To identify conserved syntenic regions for trait classes within the genome, we looked for genomic regions in which QTL-linked markers for the same trait or from the same trait class in two or more species were syntenic more often than expected compared to a null distribution of markers. We calculated a null distribution by simulating a chromosome with the length of the total anchored reference genome (1322.7 Mb, Atlantic salmon reference genome ICSASG_v2: GenBank: GCA_000233375.4 https://www.ncbi.nlm.nih.gov/assembly/GCF_000233375.1/), randomly distributing QTL-linked marker points along the simulated chromosome,

and determining how often those markers for two, three, four, five, or six species are found within a 12 Mbp interval (optimal interval size, see previous section). The simulations were performed using peak marker positions or the midpoints of C.I.s (see above) as the interval is the effective C.I. Multiple hits on the same comparisons were only counted once and at the highest level of sharing (*i.e.*, QTL synteny across two species also found across a three-species set was only counted in the three-species synteny). The simulation was repeated 10,000 times for each trait class and the mean and standard deviation (SD) were calculated. Synteny of trait class-specific QTL-linked markers was defined as significant if the number of colocalized QTL-linked markers between two and six species was higher than two SDs from the calculated null distribution. The R-script for the simulation of synteny is attached as a supplemental file (File S2).

Since the suppression of recombination in most parts of the genome in male salmonids might cause false-positive synteny (Ohno 1970; Danzmann *et al.* 2005), we checked *post hoc* if QTL markers in synteny blocks originated from both sexes or only males. However, markers within all synteny blocks originated from both sexes, and therefore we did not take sex-specific recombination differences further into account.

Analysis of colocalization within species

To test if QTL for two traits are colocalized more often than expected under a random distribution, we conducted pairwise analyses of colocalization between two traits within each species. We did this by counting how often two QTL-linked markers for two traits were colocalized, defined as having overlapping C.I.s. We calculated a null distribution of QTL across the genome for each pairwise comparison within a species. The simulation of null distributions for colocalization analyses followed the same rationale as those for the synteny analysis. In brief, we simulated a chromosome with the length of the total anchored reference genome (1322.7 Mb), randomly placed each set of QTL-linked markers for each pairwise comparison, and counted how often QTL-linked markers are within 12 Mbp of each other (File S3). Simulations were performed based on peak marker positions or the midpoints of C.I.s (as above). Each pairwise comparison was simulated 10,000 times for each trait comparison and the average was used as the expected frequency of colocalization under a random distribution (Table S4). A Pearson's χ^2 test was used to test if markers for traits of interest are colocalized more often than expected by chance. We corrected for multiple testing using a Benjamini–Hochberg correction implemented in the R function *p.adjust*. Colocalization matrices for each species were visualized as heatmaps using the *heatmap.2* function implemented in the *gplots* R-package (Warnes *et al.* 2014).

Data availability

All data analyzed and custom R-scripts used during this study are included in this published article and as supplemental information files. All supplemental figures are included in File S1. Information about all supplemental files is included in File S1.

Results

Creation of an extensive QTL database and mapping to reference genome

From a literature search of publication databases, information on 943 ecologically relevant QTL was collected from 28 salmonid studies (Table S1). The selected traits were for morphology (body weight, body length, body shape, Fulton's condition factor, head shape, and gill raker number), life history (hatching time, spawning time, age at sexual maturation, growth rate, and embryonic development), physiology (upper thermal tolerance, osmoregulatory ability, and salinity tolerance), and swimming behavior (directional change, depth selection, activity, and burst swimming). Information was available for six species spanning two subfamilies (Coregoninae and Salmoninae) and four genera: lake whitefish (*C. clupeaformis*) (204 QTL-linked markers), coho salmon (*O. kisutch*) (66 QTL-linked markers), rainbow trout (*O. mykiss*) (168 QTL-linked markers), Chinook salmon (*O. tshawytscha*) (24 QTL-linked markers), Atlantic salmon (*Salmo salar*) (94 QTL-linked markers), and Arctic charr (*Salvelinus alpinus*) (387 QTL-linked markers). This selection of species captures all major salmonid phylogenetic lineages except the graylings (Thymallinae) (Figure 1) (Crête-Lafrenière *et al.* 2012). The QTL were inferred by molecular methods for polymorphic markers: SNPs (either RADseq or EST derived), microsatellite loci, or sequence characterized amplified regions (Table S1).

A total of 407 nonredundant QTL-linked markers could be successfully mapped to the Atlantic salmon genome (ICSASG_v2) (Table 1, positional information in Table S2). This mapping success differed between alignment methods depending on species and marker type. Between 63.2 and 87.2% of markers could be mapped to the Atlantic salmon reference genome depending on the species (Table 1). Across species, 25–68% of mapped QTL-linked markers were redundant and removed, meaning those markers overlapped in their C.I.s with other QTL-linked markers for the same trait in the same species and had a smaller effect size or a lower LOD-score (Table 1 and Table S1). Interestingly, the average effect size of QTL, measured as PVE, was significantly higher for physiological traits compared to morphological and life history traits (Figure S2 in File S1).

Chromosome level: QTL distribution and enrichment of chromosomes for QTL

To examine how QTL drawn from six salmonid species are distributed across the genome, we analyzed the chromosomal location of all QTL-linked markers relative to the distribution of annotated genes in the Atlantic salmon reference genome. We found a significant positive correlation between chromosome length and the number of QTL from all species mapped on that chromosome (Figure S3B in File S1: $R^2 = 0.835$, $P < 0.001$)

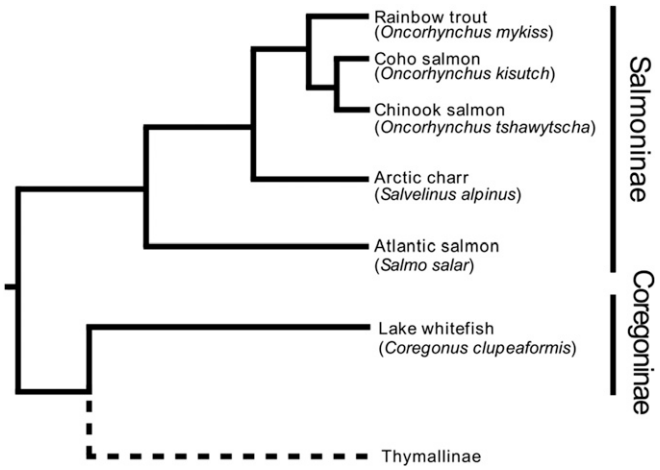

Figure 1 Phylogenetic relationship between study species [(following relationships inferred in Crête-Lafrenière *et al.* (2012) and Macqueen and Johnston (2014)].

and the number of genes and number of QTL-linked markers per chromosome (Figure S3C in File S1: $R^2 = 0.733$, $P < 0.001$). The number of genes per chromosome was also highly correlated to chromosome length (Figure S3A in File S1: $R^2 = 0.898$, $P < 0.001$). However, we found no correlation between the number of QTL-linked markers per kbp (hereafter, QTL density) and chromosome length (Figure 2A; $R^2 = 0.224$, $P = 0.249$). QTL density was not significantly correlated with the number of genes per kbp (gene density), suggesting that the density of genes does not predict the density of QTL on any given chromosome (Figure 2B: $R^2 = 0.078$, $P = 0.687$). However, we observed a heterogeneity in the distribution of QTL and gene density across chromosomes, with pronounced peaks on chromosomes 8, 17, 26, 27, 28, and 29 (Figure 3A and Figure S4 in File S1). Such local peaks in QTL density can be observed on many chromosomes and in some cases coincide with decreased gene density (*e.g.*, chromosome 4) (Figure S4 in File S1).

After controlling for chromosome length, we found that chromosome 1 and chromosome 6 are significantly enriched for QTL of all traits combined (Figure 4A; $P = 0.04$) and that chromosome 22 is significantly enriched for swimming behavior QTL ($P = 0.04$). However, no chromosomes were enriched for morphology, physiology, or life history QTL.

Genomic level: salmonid QTL distribution within chromosomes

To investigate the observation that local peaks of QTL density do not seem to coincide with regions of increased gene density, we examined the fine-scale relationship between QTL-linked marker and functional gene codistribution across the genome. We found that the number of genes and QTL-linked markers per 5 Mbp window are highly correlated (Figure 3B; $R^2 = 0.118$, $P < 0.001$). We used this correlation to identify clusters of QTL-linked markers ("QTL clusters") compared to the gene background distribution by comparing the number of QTL-linked markers to the number of genes in each window, and used the qbinom function to detect windows that contain more QTL than expected. We detected 12 QTL clusters in the genome for all traits combined (Figure 5). In most but not all cases, QTL from multiple species fall in the QTL cluster. Those clusters generally contained QTL-linked markers for different traits and trait classes. However, compared to the overall number of tested windows ($N_{windows} = 853$), these results suggest a generally even distribution of QTL-linked markers in salmonids, with a few elevated regions.

We then compared the positions of QTL clusters for all traits combined to the positions of peaks in QTL density. All detected QTL clusters overlap with local peaks in QTL density but not necessarily the most pronounced peaks, *i.e.*, we did not detect a QTL cluster within the peak on chromosome 8 (Figure 5 and Figure S4 in File S1). Also, not every peak in QTL density is associated with a significant clustering of QTL-linked markers. Peaks in QTL density that are not associated with QTL clusters do not contain more QTL-linked markers than expected by chance or are associated with increased background gene density. These were not detected in our clustering analysis, since our criterion for detecting QTL clusters was based on elevated QTL density compared to gene density.

Genomic level: synteny of QTL distribution within chromosomes

Using the newly available salmon reference genome and a novel analytical framework, we identified genomic regions enriched in QTL-linked markers for the same trait class and shared across salmonid species. The aim of this analysis was to identify genomic regions with conserved synteny for specific trait classes across salmonids. To accomplish this, we looked for QTL-linked markers from the same trait class (*i.e.*, traits

Table 1 QTL markers identified, mapped and retained for each of six salmonid species

Species	Number of Markers Compiled	Number and Percentage of Markers Mapped	Number and Percentage of Redundant Markers	Total Nonredundant Markers Retained
Rainbow trout	168	132 (78.6%)	38 (28.8%)	94 (55.9%)
Chinook salmon	24	20 (83.0%)	5 (25.0%)	15 (62.5%)
Coho salmon	66	44 (66.7%)	30 (68.2%)	14 (21.2%)
Arctic charr	387	268 (69.3%)	117 (43.7%)	151 (39.0%)
Atlantic salmon	94	82 (87.2%)	31 (37.8%)	51 (54.3%)
Lake whitefish	204	129 (63.2%)	47 (36.4%)	82 (40.2%)

For each species, the number of QTL markers that were compiled from the literature, the number and percentage of markers that were uniquely mapped to the Atlantic salmon (*Sa. salar*) genome, and the total number of retained markers after removing redundant markers.

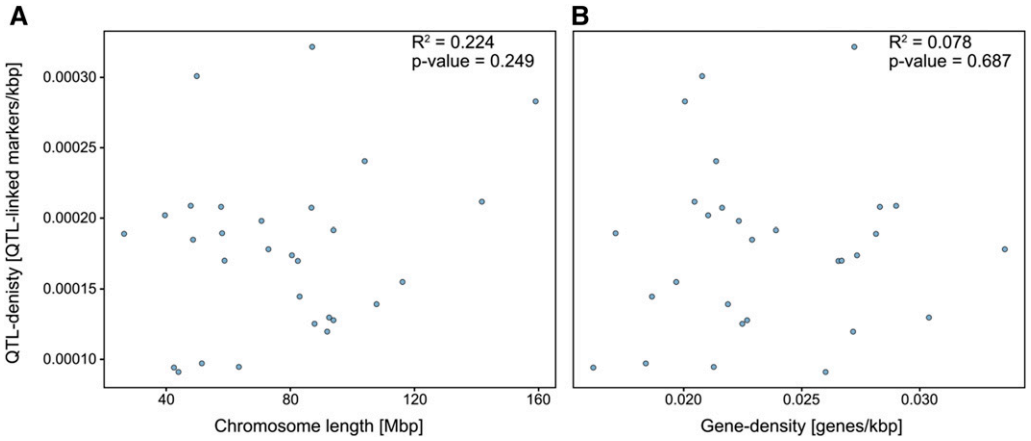

Figure 2 QTL density not correlated with chromosome length and gene density. QTL density per chromosome, defined as the number of QTL-linked markers per kbp, is not correlated (A) with chromosome length nor (B) with gene density.

within the categories of morphology, life history, or physiology), that were syntenic with QTL-linked markers from the same trait class in at least one other species ("syntenic QTL-linked clusters"). We identified a total of 179 genomic regions that contain syntenic QTL-linked markers for a given trait class (life history, morphological, or physiological traits) across multiple (two or more) salmonid species, with the highest number of clusters for morphology ($N = 138$) and the lowest for physiology ($N = 9$) (Table S3). If we only focus on regions highly conserved across species, such with QTL-linked markers from at least three species, we find 49 different syntenic clusters (Table S3).

To assess if some trait classes have more syntenic QTL clusters than expected by chance, we compared the number of observed syntenic clusters to a simulated null distribution. As the null distribution was simulated using peak markers/QTL-midpoints, we empirically determined the optimal interval length for detecting colocalization of syntenic markers (see *Materials and Methods*: Optimal interval size). We found the highest correlation between colocalization determined by peak markers and overlapping C.I.s for an interval length of 12 Mbp (Mantel's $r = 0.7599$). The correlation for the other tested interval lengths ranged from 0.5499 for 3.75 Mbp to 0.7589 for 15 Mbp.

Using this interval, we found that the number of syntenic clusters for morphology (specifically two-, four-, and five-species synteny) and for physiology (three-species synteny) is significantly higher than expected by chance (Figure 4B). However, the syntenic clustering of life history QTL-linked markers was not significant in our analyses.

We checked *post hoc* if significant syntenic QTL-linked clusters were associated with a specific trait or several different traits. However, syntenic clusters did not seem to be enriched for any particular traits but rather contain a mix of different traits. Six out of the 49 syntenic clusters highly conserved across species (≥ three species) overlapped with identified QTL clusters (*i.e.*, QTL-enriched relative to background gene density): three of the morphology synteny clusters (chromosome 1, 20, and 23), two life history synteny clusters (chromosome 1 and 17), and one physiology synteny cluster (chromosome 9) (Table S3). These regions are candidates for containing QTL that affect several different traits across salmonid species.

Colocalization of QTL within species

To identify potential genetic linkage between phenotypic traits within species, *i.e.*, colocalization, we analyzed the genome-wide distribution of QTL-linked markers in each of the six species separately. QTL were considered colocalized if the C.I.s of associated QTL-linked markers overlapped (Figure 5). Overall, we found that lake whitefish shows the lowest frequency of colocalization, with only 25% of traits being colocalized (7 of 28 trait comparisons). Arctic charr had the second highest frequency of colocalization among traits (25 of 36 trait comparisons, 69.4%) and Chinook salmon the highest, with all three traits being colocalized (Figure 6). Given the high possible number of pairwise comparisons, we calculated a null model for the chance that two traits are colocalized within a species (see *Materials and Methods*). This is therefore an empirically derived definition of colocalization and its significance is assessed statistically compared to an empirically derived null model.

Only 8 out of 134 trait pairs were significantly colocalized and those significant colocalizations were found in three species (Figure 6). In lake whitefish, depth selection and gill raker number as well as directional change and burst swimming were significantly colocalized (Figure 6B). In coho salmon, 3 out of 10 trait pairs were significantly colocalized: growth rate with body weight and body length, and body length with body weight (Figure 6D). In rainbow trout, body length was significantly colocalized with embryonic developmental rate, timing of sexual maturation, and upper thermal tolerance (3 of 36 pairwise comparisons; Figure 6E). No significant colocalizations were observed for Arctic charr, Atlantic salmon, or Chinook salmon.

We also found that in lake whitefish, body shape and directional change were colocalized significantly less often than expected by chance [false discovery rate (FDR) < 0.1; Table S4].

Discussion

In this study, we assessed for the first time the extent of colocalization and synteny of QTL within and between six salmonid species by mapping QTL-linked markers associated with ecologically

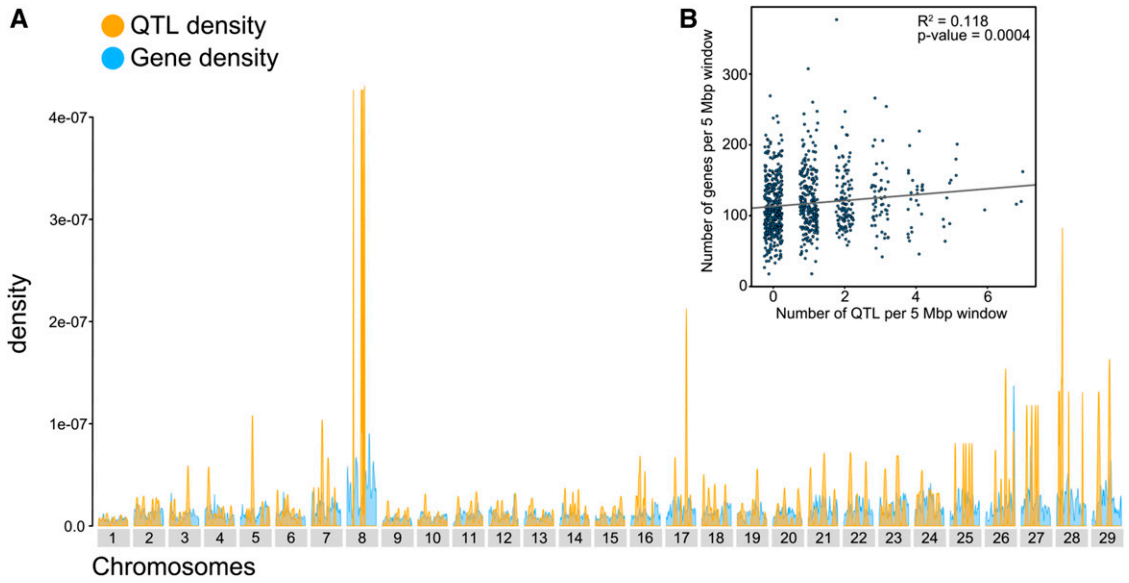

Figure 3 Characterization of QTL and gene codistribution. (A) The distribution of QTL and gene densities along the Atlantic salmon genome. (B) The number of QTL per 5 Mbp window is significantly correlated to the number of genes per 5 Mbp window.

relevant traits to the Atlantic salmon reference genome. These species represent the breadth of salmonid phylogenetic diversity, including four genera (*Oncorhynchus*, *Salmo*, *Salvelinus*, and *Coregonus*). Of these, Arctic charr and lake whitefish are particularly renowned for their morphological and ecological variability (Bernatchez *et al.* 2010; Vonlanthen *et al.* 2012; Klemetsen 2013). We detected two chromosomes that are enriched for QTL but only found a low amount of local clustering of QTL-linked markers across the genome. Furthermore, we found significant synteny for morphological and physiological traits, indicating the clustering of homologous QTL-linked markers for the same trait class across species. Focusing within species, our analyses revealed a statistically significant colocalization of QTL-linked markers for several phenotypic traits within lake whitefish, coho salmon, and rainbow trout. This study provides an important advance by synthesizing across multiple species to infer general patterns of genomic organization and QTL reuse, and how this relates to the famous morphological, physiological, and ecological phenotypic variability in salmonids.

A new quantitative definition of QTL colocalization

As part of this study, we developed a general statistical framework for detecting QTL colocalization and fine-scale synteny using a heterospecific reference genome. This may be useful for the research community in general because it is independent of the way in which QTL were identified and reported in their original study; therefore, our method can be applied to any QTL data set and related or conspecific reference genome. Our framework is based on comparing observed colocalization of QTL-linked markers with different interval sizes, using the average distribution of C.I. lengths as a starting point, to inform simulations for testing the statistical significance of QTL and gene distributions. Overall, this criterion allowed us to investigate several aspects of QTL distributions in this study: the extent of colocalization within species, the identification of highly conserved QTL within species combined across several studies, and the identification of conserved synteny regions for trait classes across species.

How are QTL and genes distributed across the genome?

The codistribution of genes and QTL across the genome is important for understanding the factors shaping the genomic architecture underlying phenotypic traits and potential reuse or conservation of QTL for traits across species. To date, there has been little research on the genome-wide relationship between QTL regions and local gene density in any species. This dearth has existed because such analyses require genetic markers that can be mapped, a biological system with a relatively large number of quantitative trait studies linking genotype and phenotype in related species, and a genetically closely related, high-quality, and well-annotated reference genome. Agricultural crop plants are a biological system that tends to meet these requirements, and a study in sorghum wheat showed that gene and QTL enrichment in localized genomic regions is due to suppressed recombination in heterochromatic regions (Mace and Jordon 2011), leading to the association of QTL- and gene-dense regions. A major finding of our study is that although QTL density is not correlated with gene density on a chromosome level, we find an overall significant correlation between the number of QTL-linked markers and genes on a finer scale (Figure 2B and Figure 3, A and B). However, as no information is currently available on the distribution of euchromatic and heterochromatic regions along the Atlantic salmon genome, we cannot test the association of gene and QTL densities with chromatin distribution. While there are many local peaks, only 12 clusters were significant (Figure 5). Local peaks in QTL density across the genome, which in

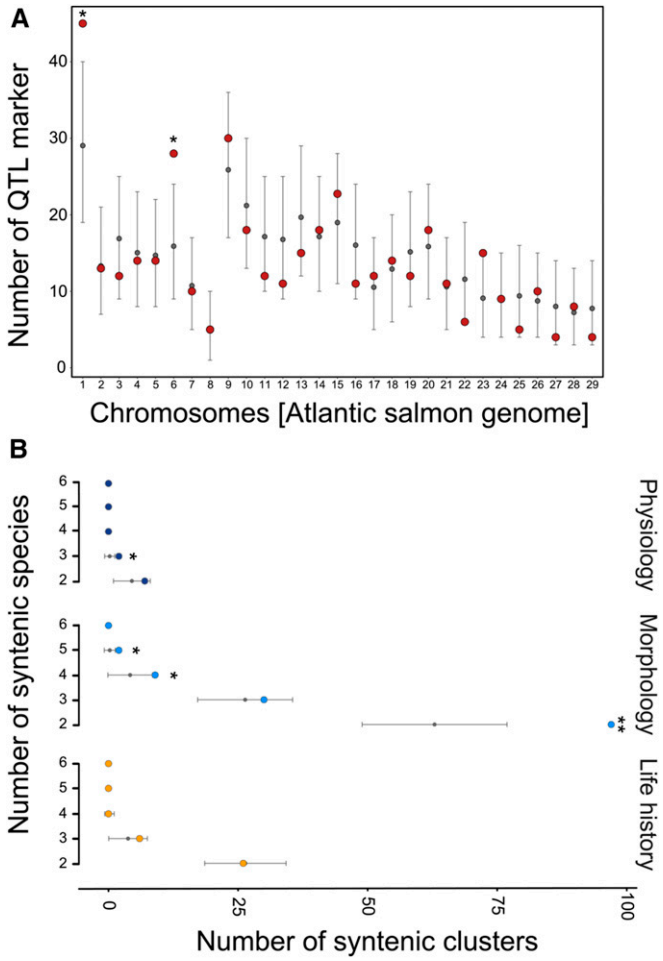

Figure 4 QTL enrichment and synteny analysis. (A) Number of QTL-linked markers on each chromosome. Red dots indicate the observed number of QTL-linked markers per chromosome. (B) Number of clusters found to be syntenic across 2–6 species for physiology, morphology, and life history traits separately. The observed number of syntenic QTL clusters per trait class are shown by dots (dark blue = physiology; light blue = morphology; and orange = life history). In (A and B) gray dots show the mean and bars the 95% C.I.s of the inferred null distribution; statistical significance is indicated by asterisks: * $P < 0.05$ and ** $P < 0.01$.

most cases do not coincide with regions of increased gene density, suggest that linkage of causal variants and/or pleiotropy as well as structural genomic rearrangements (*i.e.*, inversions) might play a role in the colocalization of QTL generally (Yeaman 2013; Schwander *et al.* 2014; Peichel and Marques 2017).

Our finding that some chromosomes are significantly enriched for QTL (Figure 4A) supports the hypothesis that QTL-linked markers are not randomly distributed and that certain genetic factors, such as linkage, pleiotropy, and structural rearrangements, may drive their clustering in salmonids. Similar amounts of QTL enrichment were found in stickleback (Miller *et al.* 2014; Peichel and Marques 2017) and suggest that the enrichment of QTL on some chromosomes is not unique to salmonids. However, similar to salmonids, sticklebacks diversify rapidly (McKinnon and Rundle 2002), which raises the question of whether the significant clustering of QTL in genomic regions or on certain chromosomes is a shared feature of rapidly diversifying lineages. Detailed comparisons to closely related nondiversifying or more slowly diversifying lineages are needed to address this question.

Understanding the genomic arrangement and linkage of QTL within the genome is important for understanding the correlated evolution of traits during adaptation, especially in the face of gene flow (Feder *et al.* 2012; Yeaman 2013). The colocalization of QTL for developmentally correlated genes is a common biological phenomenon, as many previous comparative mapping studies have shown (Badani *et al.* 2006; Lejeune-Hénaut *et al.* 2008; McClelland and Naish 2010; Sauge *et al.* 2011; Larson *et al.* 2016). By combining QTL data from several independent studies with a common reference genome and a simulated null model of expected colocalization events under the assumption of random distribution, we detected significant colocalizations of QTL-linked markers for different traits within species (Figure 6 and Table S4).

The significant colocalization of trait pairs within species suggests genetic linkage of these traits. However, the generally very low level of significant colocalizations detected in this study suggests that genetic linkage of ecologically relevant traits within salmonid species is considerably less common than expected.

Our data set does not allow us to determine whether linkage, pleiotropy, or genomic arrangements (or combinations of those) are the driving force behind the clustering of QTL-linked markers, as the causal variants underlying those traits are not known. Further, we lack the genomic resources to identify the role of small- and large-scale structural variations in salmonids. Given the new breadth of tools and resources becoming available, future research linking ecologically relevant phenotypes with the identification of causal variants and associated genomic architectures may elucidate the interplay of genome composition and quantitative traits.

QTL synteny: conserved functional regions across salmonids?

Our analyses across several species revealed significant synteny for morphology and physiology QTL-linked markers across species. Although we also detected several syntenic QTL-linked markers for life history, they were not significant compared to the null expectation (Figure 4B). However, not finding significantly syntenic QTL-linked markers for some traits does not necessarily mean that they are not syntenic across species, as genomic rearrangements might have affected the position of those markers in Atlantic salmon. The significant synteny of QTL-linked markers for morphology and physiology suggests that the genetic basis of these trait classes is conserved across salmonid species. Interestingly, we did not find that these clusters were enriched for a specific trait but were rather a mix of different traits across species. The presence of a few regions with large effect on phenotype, *e.g.*, pleiotropic QTL clusters that affect several phenotypes in one (significant colocalization) or several species (syntenic QTL clusters), but numerous small effect, unclustered QTL, follows the predictions of the geometric model of adaptation (Fisher 1930; Albert *et al.* 2008).

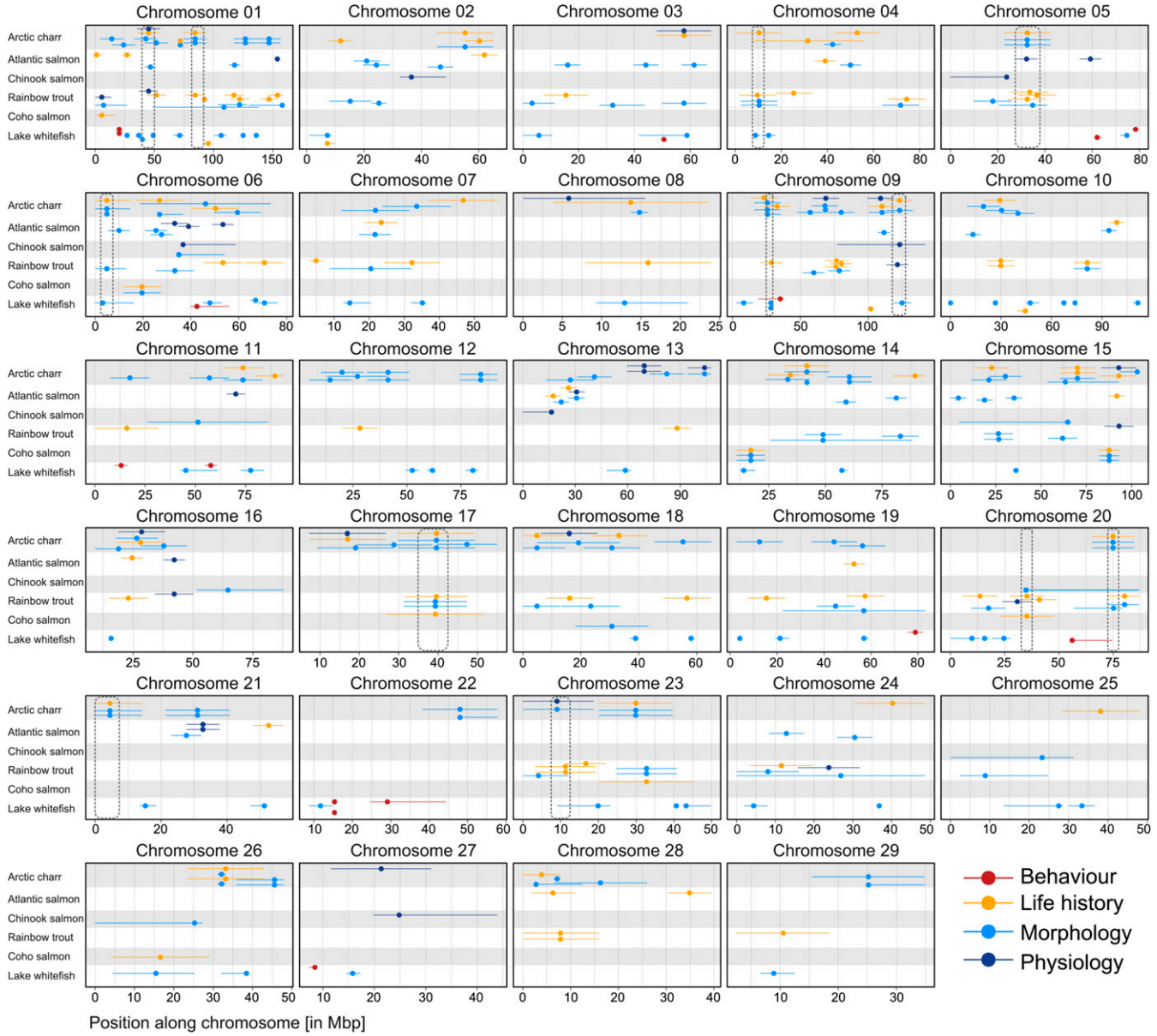

Figure 5 QTL and synteny blocks for each functional group on the Atlantic salmon genome. Distribution of all QTL-linked markers (± C.I.) for each trait class (swimming behavior in red, life history in yellow, morphology in light blue, and physiology in dark blue) for each of six species (Arctic charr, Atlantic salmon, Chinook salmon, rainbow trout, coho salmon, and lake whitefish) across 29 chromosomes. We identified 12 significant QTL clusters for all traits combined across the genome (dashed rectangle outline).

However, even without a significant excess in syntenic QTL-linked markers, conserved homologous QTL for osmoregulation, trophic morphology, or life history are of special interest in salmonid fishes generally (Carlson and Seamons 2008; Elmer 2016) due to the variability in anadromous behavior (Norman *et al.* 2011; Hale *et al.* 2013; Hecht *et al.* 2013) and the great diversity in ecology, morphology, and life history (Jonsson and Jonsson 2001; Klemetsen *et al.* 2003; Adams *et al.* 2008; Kahilainen *et al.* 2011; Küttner *et al.* 2014). Since these QTL-linked markers are conserved across several species, it is likely that the associated genomic organization plays a major functional role in the variability of ecologically important traits more generally.

Based on the close but variable phylogenetic distances between the species in this study (Figure 1), we expected to find more instances of synteny between more closely related species. However, this relationship is not clear in our study, probably due to the strong sampling bias and uneven research effort into different species and traits. Overall, our power to infer synteny is influenced by the combined number of mapped QTL across species.

Limitations due to unequal research effort across species

In this study, we developed valuable resources and analytical frameworks with a great potential for application in future

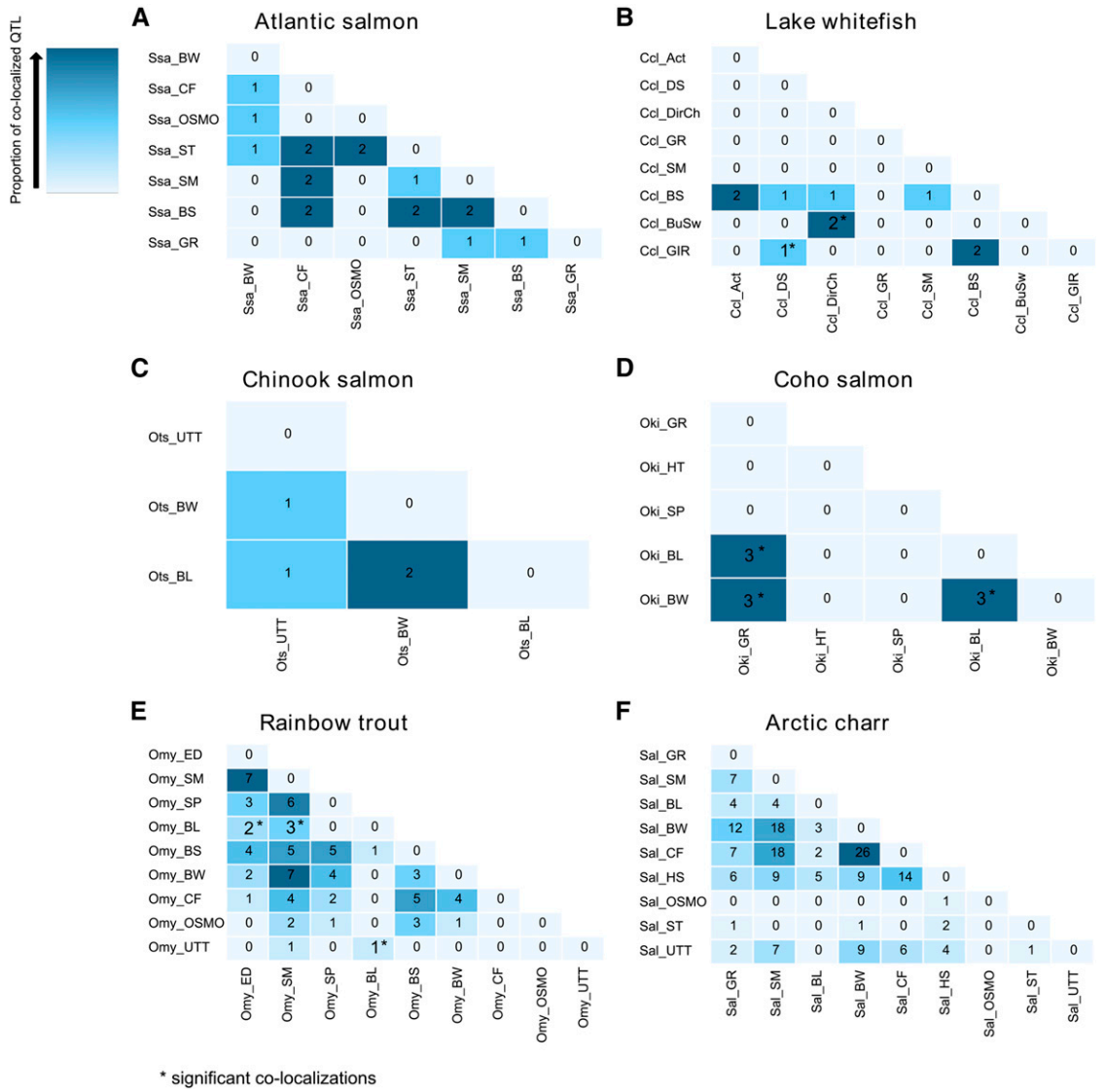

Figure 6 Colocalizations of ecologically relevant traits within species. For each species (A–F) the pairwise comparison of traits is shown, with a heatmap of the relative proportion of genomic colocalizations that occurred between two traits. The numbers within the cells refer to the number of colocalized QTL-linked markers, as phenotypic traits can be represented by multiple markers. Significance was tested against a simulated random distribution [False discovery rate (FDR) < 0.1] and significant colocalizations between traits are shown with asterisks (Table S4).

studies on the genetic architecture of phenotypic traits and their genomic basis. However, there are several limitations that should not be ignored. First, there is a species-specific information bias since there is a huge difference in the amount, type, and accuracy of QTL information that is available for different salmonid species (Table 1 and Table S1). Studies are usually biased toward economically important traits such as body weight, condition factor, growth rate, and thermal tolerance, and aquaculture species such as rainbow trout, Atlantic salmon, or Arctic charr. This bias is clear in the number of QTL markers and traits observed: with 649 (68.8%) of the 943 QTL used in this study originating from studies on these three species and 374 (39.6%) QTL for those four economically important traits.

One important limitation in comparing across QTL studies has been the heterogeneity in the amount and quality of reported information. We suggest that it would be helpful if information such as C.I.s, LOD-scores, PVE, and marker sequences were reported and accessible. This would increase the reproducibility of QTL-mapping studies and facilitate empirical comparisons and comparative genomic analyses.

We may see these limitations be overcome in the future with the increasing amount of QTL information available and the increasing quality and availability of reference genomes for nonmodel species in combination with high reporting standards (Murray *et al.* 2008).

Applications across evolutionary scales

This study has important applications to other species and evolutionary scales regarding a quantitative approach to assess colocalization. Specifically, as the availability of high-throughput

genomic data for wild populations increases, the results of this study provide a useful framework for investigating the colocalization and annotation of population genomic markers or linkage maps. They could be used *a priori* to identify genomic regions associated with the trait of interest or *a posteriori* to annotate, *e.g.*, genome scans to find out if outlier loci overlap with certain QTL or QTL-rich genomic regions (Marques *et al.* 2016). The use of reference genomes for closely related species to facilitate the comparison of genomic architectures between species will become more common with the increasing availability of such reference genomes (Macqueen *et al.* 2016). Furthermore, the QTL database we developed here for salmonids is made available for the revision and addition of new species, additional traits, and new QTL studies. This will facilitate future extensions of this study to other salmonid species or different suites of traits depending on research interests.

Conclusions

We mapped 407 QTL-linked markers for functional traits associated with morphology, life history, and physiology from six salmonid fish species to the Atlantic salmon reference genome. We were able to identify significant QTL synteny clusters across species and colocalized traits within species. We showed that it is possible to use existing reference genomes of closely related species to analyze the genomic arrangement of QTL-linked markers within and across species and developed a transferable analytical framework for such studies across different species and reference genomes.

The identification of significant QTL synteny clusters for physiological and morphological traits across the genome highlights the potentially conserved genetic basis of diverse traits across salmonid species. Contrary to expectation, we only detected low levels of colocalization between ecologically relevant traits. Overall, the low levels of clustering, synteny, and colocalization suggest that genetic linkage and pleiotropy may play a role in the overall phenotypic diversification of salmonid fishes but that this is not a prevailing phenomenon across the genome and across species. However, the presence of a few regions of large effect on phenotype (*e.g.*, syntenic QTL clusters) and numerous small effect loci (*e.g.*, unclustered QTL scattered across the genome) follows the predictions of the geometric model of adaptation. In the future, comparing the degree of fine-scale synteny and colocalization between salmonids and other less diverse lineages of fishes, in combination with high-quality reference genomes, will make it possible to better assess the role of genomic architecture of QTL on phenotypic variability and diversification. This study enables detailed analyses of genomic mechanisms and QTL reuse that may facilitate the rapid adaptive diversification found in salmonid species.

Acknowledgments

We thank the Institute of Biodiversity, Animal Health & Comparative Medicine Evolutionary Analysis Group for constructive feedback and K. Peichel, S. Yeaman, and an anonymous reviewer for comments and suggestions that improved the manuscript. This research was funded by FP7 Marie Curie Career Integration grant PCIG11-GA-2012-321999 to K.R.E. R.W. was supported by the University of Glasgow Medical, Veterinary and Life Sciences Doctoral Training Programme. M.C. was supported by a University of Glasgow Lord Kelvin-Adam Smith Studentship with L. Watson and K.R.E.

Literature Cited

Adams, C. E., D. Fraser, F. A. Huntingford, R. B. Greer, C. M. Askew *et al.*, 1998 Trophic polymorphism amongst Arctic charr from Loch Rannoch, Scotland. J. Fish Biol. 52: 1259–1271.

Adams, C. E., A. J. Wilson, and M. M. Ferguson, 2008 Parallel divergence of sympatric genetic and body size forms of Arctic charr, *Salvelinus alpinus*, from two Scottish lakes. Biol. J. Linn. Soc. Lond. 95: 748–757.

Albert, A. Y., S. Sawaya, T. H. Vines, A. K. Knecht, C. T. Miller *et al.*, 2008 The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. Evolution 62: 76–85.

Alekseyev, S. S., V. P. Samusenok, A. N. Matveev, and M. Y. Pichugin, 2002 Diversification, sympatric speciation, and trophic polymorphism of Arctic charr, *Salvelinus alpinus* complex, in Transbaikalia. Environ. Biol. Fishes 64: 97–114.

Allendorf, F. W., and G. H. Thorgaard, 1984 Tetraploidy and the evolution of salmonid fishes, pp. 1–53 in *Evolutionary Genetics of Fishes*, edited by B. J. Turner. Plenum Press, New York.

Badani, A. G., R. J. Snowdon, B. Wittkop, F. D. Lipsa, R. Baetzel *et al.*, 2006 Colocalization of a partially dominant gene for yellow seed colour with a major QTL influencing acid detergent fibre (ADF) content in different crosses of oilseed rape (*Brassica napus*). Genome 49: 1499–1509.

Bernatchez, L., J. A. Vuorinen, R. A. Bodaly, and J. J. Dodson, 1996 Genetic evidence for reproductive isolation and multiple origins of sympatric trophic ecotypes of whitefish (*Coregonus*). Evolution 50: 624–635.

Bernatchez, L., S. Renaut, A. R. Whiteley, N. Derome, J. Jeukens *et al.*, 2010 On the origin of species: insights from the ecological genomics of lake whitefish. Phil. Trans. R. Soc. 365: 1783–1800.

Berthelot, C., F. Brunet, D. Chalopin, A. Juanchich, M. Bernard *et al.*, 2014 The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nat. Commun. 5: 3657.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos *et al.*, 2009 BLAST+: architecture and applications. BMC Bioinformatics 10: 1–9.

Carlson, S. M., and T. R. Seamons, 2008 A review of quantitative genetic components of fitness in salmonids: implications for adaptation to future change. Evol. Appl. 1: 222–238.

Charlesworth, D., 2016 The status of supergenes in the 21st century: recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. Evol. Appl. 9: 74–90.

Charlesworth, D., and J. E. Mank, 2010 The birds and the bees and the flowers and the trees: lessons from genetic mapping of sex determination in plants and animals. Genetics 186: 9–31.

Conte, G. L., M. E. Arnegard, C. L. Peichel, and D. Schluter, 2012 The probability of genetic parallelism and convergence in natural populations. Proc. Biol. Sci. 279: 5039–5047.

Conte, G. L., M. E. Arnegard, J. Best, Y. F. Chan, F. C. Jones *et al.*, 2015 Extent of QTL reuse during repeated phenotypic divergence of sympatric threespine stickleback. Genetics 201: 1189–1200.

Crête-Lafrenière, A., L. K. Weir, and L. Bernatchez, 2012 Framing the Salmonidae family phylogenetic portrait: a more complete picture from increased taxon sampling. PLoS One 7: e46662.

Danzmann, R. G., M. Cairney, W. S. Davidson, M. M. Ferguson, K. Gharbi *et al.*, 2005 A comparative analysis of the rainbow trout genome with 2 other species of fish (Arctic charr and Atlantic salmon) within the tetraploid derivative Salmonidae family (subfamily: Salmoninae). Genome 48: 1037–1051.

Easton, A. A., H. K. Moghadam, R. G. Danzmann, and M. M. Ferguson, 2011 The genetic architecture of embryonic developmental rate and genetic covariation with age at maturation in rainbow trout *Oncorhynchus mykiss*. J. Fish Biol. 78: 602–623.

Elmer, K. R., 2016 Genomic tools for new insights to variation, adaptation, and evolution in the salmonid fishes: a perspective for charr. Hydrobiologia 783: 191–208.

Feder, J. L., S. P. Egan, and P. Nosil, 2012 The genomics of speciation-with-gene-flow. Trends Genet. 28: 342–350.

Fisher, R. A., 1930 *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

Fruciano, C., P. Franchini, V. Kovacova, K. R. Elmer, F. Henning *et al.*, 2016 Genetic linkage of distinct adaptive traits in sympatrically speciating crater lake cichlid fish. Nat. Commun. 7: 12736.

Gagnaire, P.-A., E. Normandeau, S. A. Pavey, and L. Bernatchez, 2013 Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the Lake Whitefish (*Coregonus clupeaformis*). Mol. Ecol. 22: 3036–3048.

Gilmartin, P. M., and J. Li, 2010 Homing in on heterostyly. Heredity (Edinb) 105: 161–162.

Gordeeva, N. V., S. S. Alekseyev, A. N. Matveev, and V. P. Samusenok, 2015 Parallel evolutionary divergence in Arctic charr *Salvelinus alpinus* (L.) complex from Transbaikalia: variation in differentiation degree and segregation of genetic diversity between sympatric forms. Can. J. Fish. Aquat. Sci. 72: 96–115.

Guyomard, R., M. Boussaha, F. Krieg, C. Hervet, and E. Quillet, 2012 A synthetic rainbow trout linkage map provides new insights into the salmonid whole genome duplication and the conservation of synteny among teleosts. BMC Genet. 13: 1–12.

Hale, M. C., F. P. Thrower, E. A. Berntson, M. R. Miller, and K. M. Nichols, 2013 Evaluating adaptive divergence between migratory and nonmigratory ecotypes of a salmonid fish, *Oncorhynchus mykiss*. G3 (Bethesda) 3: 1273–1285.

Hawthorne, D. J., and S. Via, 2001 Genetic linkage of ecological specialization and reproductive isolation in pea aphids. Nature 412: 904–907.

Hecht, B. C., N. R. Campbell, D. E. Holecek, and S. R. Narum, 2013 Genome-wide association reveals genetic basis for the propensity to migrate in wild populations of rainbow and steelhead trout. Mol. Ecol. 22: 3061–3076.

Hindar, K., and B. Jonsson, 1993 Ecological polymorphism in Arctic charr. Biol. J. Linn. Soc. Lond. 48: 63–74.

Jonsson, B., and N. Jonsson, 2001 Polymorphism and speciation in Arctic charr. J. Fish Biol. 58: 605–638.

Joron, M., R. Papa, M. Beltrán, N. Chamberlain, J. Mavárez *et al.*, 2006 A conserved supergene locus controls colour pattern diversity in *heliconius* butterflies. PLoS Biol. 4: e303.

Joron, M., L. Frezal, R. T. Jones, N. L. Chamberlain, S. F. Lee *et al.*, 2011 Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature 477: 203–206.

Kahilainen, K., and K. Østbye, 2006 Morphological differentiation and resource polymorphism in three sympatric whitefish *Coregonus lavaretus*(L.) forms in a subarctic lake. J. Fish Biol. 68: 63–79.

Kahilainen, K. K., A. Siwertsson, K. O. Gjelland, R. Knudsen, T. Bøhn *et al.*, 2011 The role of gill raker variability in adaptive radiation of Coregonid fish. Evol. Ecol. 25: 573–588.

Klemetsen, A., 2013 The most variable vertebrate on Earth. J. Ichthyol. 53: 781–791.

Klemetsen, A., P.-A. Amundsen, J. B. Dempson, B. Jonsson, N. Jonsson *et al.*, 2003 Atlantic salmon *Salmo salar* L., brown trout *Salmo trutta* L. and Arctic charr *Salvelinus alpinus* (L.): a review of aspects of their life histories. Ecol. Freshwat. Fish 12: 1–59.

Küpper, C., M. Stocks, J. E. Risse, N. dos Remedios, L. L. Farrell *et al.*, 2016 A supergene determines highly divergent male reproductive morphs in the ruff. Nat. Genet. 48: 79–83.

Küttner, E., K. J. Parsons, A. A. Easton, S. Skúlason, R. G. Danzmann *et al.*, 2014 Hidden genetic variation evolves with ecological specialization: the genetic basis of phenotypic plasticity in Arctic charr ecomorphs. Evol. Dev. 16: 247–257.

Lamichhaney, S., G. Fan, F. Widemo, U. Gunnarsson, D. S. Thalmann *et al.*, 2016 Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). Nat. Genet. 48: 84–88.

Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nat. Methods 9: 357–359.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg, 2009 Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10: 1–10.

Larson, W. A., G. J. McKinney, M. T. Limborg, M. V. Everett, L. W. Seeb *et al.*, 2016 Identification of multiple QTL hotspots in Sockeye Salmon (*Oncorhynchus nerka*) using genotyping-by-sequencing and a dense linkage map. J. Hered. 107: 122–133.

Le Bras, Y., N. Dechamp, F. Krieg, O. Filangi, R. Guyomard *et al.*, 2011 Detection of QTL with effects on osmoregulation capacities in the rainbow trout (*Oncorhynchus mykiss*). BMC Genet. 12: 46.

Lejeune-Hénaut, I., E. Hanocq, L. Béthencourt, V. Fontaine, B. Delbreil *et al.*, 2008 The flowering locus Hr colocalizes with a major QTL affecting winter frost tolerance in *Pisum sativum* L. Theor. Appl. Genet. 116: 1105–1116.

Lien, S., L. Gidskehaug, T. Moen, B. J. Hayes, P. R. Berg *et al.*, 2011 A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. BMC Genomics 12: 615.

Lien, S., B. F. Koop, S. R. Sandve, J. R. Miller, M. P. Kent *et al.*, 2016 The Atlantic salmon genome provides insights into rediploidization. Nature 533: 200–205.

Lynch, M., and B. Walsh, 2007 *The Origins of Genome Architecture*. Sinauer Associates, Sunderland, MA.

Mace, E. S., and D. R. Jordan, 2011 Integrating sorghum whole genome sequence information with a compendium of sorghum QTL studies reveals uneven distribution of QTL and of gene-rich regions with significant implications for crop improvement. Theor. Appl. Genet. 123: 169–191.

Mackay, T. F., E. A. Stone, and J. F. Ayroles, 2009 The genetics of quantitative traits: challenges and prospects. Nat. Rev. Genet. 10: 565–577.

Macqueen, D. J., and I. A. Johnston, 2014 A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proc. Biol. Sci. 281: 20132881.

Macqueen, D. J., C. R. Primmer, R. D. Houston, B. F. Nowak, L. Bernatchez *et al.*, 2016 Functional analysis of all Salmonid genomes (FAASG): an international initiative supporting future salmonid research, conservation and aquaculture. bioRxiv DOI: https://doi.org/10.1101/095737.

Marques, D. A., K. Lucek, J. I. Meier, S. Mwaiko, C. E. Wagner *et al.*, 2016 Genomics of rapid incipient speciation in sympatric threespine stickleback. PLoS Genet. 12: e1005887.

Mather, K., 1950 The genetical architecture of heterostyly in *Primula sinensis*. Evolution 4: 340–352.

McClelland, E. K., and K. A. Naish, 2010 Quantitative trait locus analysis of hatch timing, weight, length and growth rate in coho salmon, *Oncorhynchus kisutch*. Heredity (Edinb) 105: 562–573.

McKinnon, J., and H. Rundle, 2002 Speciation in nature: the threespine stickleback model systems. Trends Ecol Evol 17: 480–488.

Miller, C. T., A. M. Glazer, B. R. Summers, B. K. Blackman, A. R. Norman *et al.*, 2014 Modular skeletal evolution in sticklebacks is controlled by additive and clustered quantitative trait loci. Genetics 197: 405–420.

Murray, S. C., W. L. Rooney, S. E. Mitchell, A. Sharma, and P. E. Klein, 2008 Genetic improvement of sorghum as a biofuel feedstock: II. QTL for stem and leaf structural carbohydrates. Crop Sci. 48: 2180–2193.

Norman, J. D., R. G. Danzmann, B. Glebe, and M. M. Ferguson, 2011 The genetic basis of salinity tolerance traits in Arctic charr (*Salvelinus alpinus*). BMC Genet. 12: 1–12.

Norman, J. D., M. Robinson, B. Glebe, M. M. Ferguson, and R. G. Danzmann, 2012 Genomic arrangement of salinity tolerance QTLs in salmonids: a comparative analysis of Atlantic salmon (*Salmo salar*) with Arctic charr (*Salvelinus alpinus*) and rainbow trout (*Oncorhynchus mykiss*). BMC Genomics 13: 420.

Ohno, S., 1970 The enormous diversity in genome sizes of fish as a reflection of nature's extensive experiments with gene duplication. Trans. Am. Fish. Soc. 99: 120–130.

Oksanen J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, *et al.* 2015 vegan: Community Ecology Package. R package version 2. Available at: https://CRAN.R-project.org/package=vegan.

Peichel, C. L., and D. A. Marques, 2017 The genetic and molecular architecture of phenotypic diversity in sticklebacks. Phil. Trans. R. Soc. B 372: 20150486.

Phillips, R., and P. Ráb, 2001 Chromosome evolution in the Salmonidae (Pisces): an update. Biol. Rev. Camb. Philos. Soc. 76: 1–25.

Phillips, R. B., K. A. Keatley, M. R. Morasch, A. B. Ventura, K. P. Lubieniecki *et al.*, 2009 Assignment of Atlantic salmon (*Salmo salar*) linkage groups to specific chromosomes: conservation of large syntenic blocks corresponding to whole chromosome arms in rainbow trout (*Oncorhynchus mykiss*). BMC Genet. 10: 1–11.

Rabosky, D. L., F. Santini, J. Eastman, S. A. Smith, B. Sidlauskas *et al.*, 2013 Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nat. Commun. 4: 1958.

R Development Core Team, 2016 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. Available at: http://www.R-project.org.

Reid, D. P., A. Szanto, B. Glebe, R. G. Danzmann, and M. M. Ferguson, 2004 QTL for body weight and condition factor in Atlantic salmon (*Salmo salar*): comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Arctic charr (*Salvelinus alpinus*). Heredity (Edinb) 94: 166–172.

Roberts, R. B., J. R. Ser, and T. D. Kocher, 2009 Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. Science 326: 998–1001.

Sauge, M.-H., P. Lambert, and T. Pascal, 2011 Co-localisation of host plant resistance QTLs affecting the performance and feeding behaviour of the aphid Myzus persicae in the peach tree. Heredity 108: 292–301.

Schwander, T., R. Libbrecht, and L. Keller, 2014 Supergenes and complex phenotypes. Curr. Biol. 24: R288–R294.

Siwertsson, A., R. Knudsen, C. E. Adams, K. Præbel, and P.-A. Amundsen, 2013 Parallel and non-parallel morphological divergence among foraging specialists in European whitefish (*Coregonus lavaretus*). Ecol. Evol. 3: 1590–1602.

Swamy, B. P., P. Vikram, S. Dixit, H. U. Ahmed, and A. Kumar, 2011 Meta-analysis of grain yield QTL identified during agricultural drought in grasses showed consensus. BMC Genomics 12: 1–18.

Thomas, J. W., M. Cáceres, J. J. Lowman, C. B. Morehouse, M. E. Short *et al.*, 2008 The chromosomal polymorphism linked to variation in social behavior in the white-throated sparrow (*Zonotrichia albicollis*) is a complex rearrangement and suppressor of recombination. Genetics 179: 1455–1468.

Timusk, E. R., M. M. Ferguson, H. K. Moghadam, J. D. Norman, C. C. Wilson *et al.*, 2011 Genome evolution in the fish family Salmonidae: generation of a brook charr genetic map and comparisons among charrs (Arctic charr and brook charr) with rainbow trout. BMC Genet. 12: 68.

Tuttle, E. M., A. O. Bergland, M. L. Korody, M. S. Brewer, D. J. Newhouse *et al.*, 2016 Divergence and functional degradation of a sex chromosome-like supergene. Curr. Biol. 26: 344–350.

Uno, Y., C. Nishida, S. Yoshimoto, M. Ito, Y. Oshima *et al.*, 2008 Diversity in the origins of sex chromosomes in anurans inferred from comparative mapping of sexual differentiation genes for three species of the Raninae and Xenopodinae. Chromosome Res. 16: 999–1011.

Vicoso, B., J. J. Emerson, Y. Zektser, S. Mahajan, and D. Bachtrog, 2013 Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. PLoS Biol. 11: e1001643.

Vonlanthen, P., D. Bittner, A. G. Hudson, K. A. Young, R. Müller *et al.*, 2012 Eutrophication causes speciation reversal in whitefish adaptive radiations. Nature 482: 357–362.

Warnes G. R., B. Bolker, L. Bonebakker, R. Gentleman, W. H. A. Liaw, *et al.*, 2016 gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1. Available at: https://CRAN.R-project.org/package=gplots.

Wickham, H., 2009 *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.

Yeaman, S., 2013 Genomic rearrangements and the evolution of clusters of locally adaptive loci. Proc. Natl. Acad. Sci. USA 110: E1743–E1751.

Yeaman, S., K. A. Hodgins, K. E. Lotterhos, H. Suren, S. Nadeau *et al.*, 2016 Convergent local adaptation to climate in distantly related conifers. Science 353: 1431–1433.

Zhang, J., H. G. Zheng, A. Aarti, G. Pantuwan, T. T. Nguyen *et al.*, 2001 Locating genomic regions associated with components of drought resistance in rice: comparative mapping within and across species. Theor. Appl. Genet. 103: 19–29.

Communicating editor: K. Peichel