HAL open science

Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding

Jonathan D'ambrosio, Florence Phocas, Pierrick Haffray, Anastasia Bestin, Sophie Brard-Fudulea, Charles Poncet, Edwige Quillet, Nicolas Dechamp, Clémence Fraslin, Mathieu Charles, et al.

To cite this version:

Jonathan D'ambrosio, Florence Phocas, Pierrick Haffray, Anastasia Bestin, Sophie Brard-Fudulea, et al.. Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding. Genetics Selection Evolution, 2019, 51 (1), pp.26. 10.1186/s12711-019-0468-4 . hal-02151750

HAL Id: hal-02151750

https://hal.science/hal-02151750v1

Submitted on 10 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

D'Ambrosio *et al. Genet Sel Evol* (2019) 51:26
https://doi.org/10.1186/s12711-019-0468-4

Genetics Selection Evolution

RESEARCH ARTICLE Open Access

Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding

Jonathan D'Ambrosio[1,2], Florence Phocas[1*], Pierrick Haffray[2], Anastasia Bestin[2], Sophie Brard-Fudulea[3], Charles Poncet[4], Edwige Quillet[1], Nicolas Dechamp[1], Clémence Fraslin[1,2], Mathieu Charles[1] and Mathilde Dupont-Nivet[1]

Abstract

Background: Selective breeding is a relatively recent practice in aquaculture species compared to terrestrial livestock. Nevertheless, the genetic variability of farmed salmonid lines, which have been selected for several generations, should be assessed. Indeed, a significant decrease in genetic variability due to high selection intensity could have occurred, potentially jeopardizing the long-term genetic progress as well as the adaptive capacities of populations facing change(s) in the environment. Thus, it is important to evaluate the impact of selection practices on genetic diversity to limit future inbreeding. The current study presents an analysis of genetic diversity within and between six French rainbow trout (*Oncorhynchus mykiss*) experimental or commercial lines based on a medium-density single nucleotide polymorphism (SNP) chip and various molecular genetic indicators: fixation index (F_{ST}), linkage disequilibrium (LD), effective population size (N_e) and inbreeding coefficient derived from runs of homozygosity (ROH).

Results: Our results showed a moderate level of genetic differentiation between selected lines (F_{ST} ranging from 0.08 to 0.15). LD declined rapidly over the first 100 kb, but then remained quite high at long distances, leading to low estimates of N_e in the last generation ranging from 24 to 68 depending on the line and methodology considered. These results were consistent with inbreeding estimates that varied from 10.0% in an unselected experimental line to 19.5% in a commercial line, and which are clearly higher than corresponding estimates in ruminants or pigs. In addition, strong variations in LD and inbreeding were observed along the genome that may be due to differences in local rates of recombination or due to key genes that tended to have fixed favorable alleles for domestication or production.

Conclusions: This is the first report on ROH for any aquaculture species. Inbreeding appeared to be moderate to high in the six French rainbow trout lines, due to founder effects at the start of the breeding programs, but also likely to sweepstakes reproductive success in addition to selection for the selected lines. Efficient management of inbreeding is a major goal in breeding programs to ensure that populations can adapt to future breeding objectives and SNP information can be used to manage the rate at which inbreeding builds up in the fish genome.

*Correspondence: florence.phocas@inra.fr
[1] GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France
Full list of author information is available at the end of the article

© The Author(s) 2019. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

Background

Rainbow trout (*Oncorhynchus mykiss*) is native to the Pacific drainages of North America and also to Kamchatka in Russia. This fish was introduced at the end of the nineteenth century to waters on all continents except Antarctica, for recreational angling and aquaculture purposes. Rainbow trout is one of the main species of fish reared in cold freshwater around the world, particularly in Europe, North America and Chile [1]. For several decades, the rainbow trout farming industry has been endeavoring to continually increase production efficiency and sales by increasing rearing densities, improving diets, water quality and recirculation technology, controlling sexual maturation and gender, or developing genetically superior lines of fish for improved growth, fillet quality and disease resistance. Recent access to the genome sequence of rainbow trout [2], genetic maps [3–5], and a medium-throughput genotyping chip [6] of 57,501 single nucleotide polymorphisms (SNPs) offer new perspectives for research and organization of trout breeding programs, which may be more effective than those that are historically based on phenotypic or genealogical selection of the broodstock.

Selective breeding can contribute to a significant decrease in the genetic variability of farmed populations, jeopardizing long-term genetic progress as well as reducing the adaptive capacities of populations in the event of a change in the environment [7]. Greater genetic variability within a population increases the likelihood that some of its individuals will have alleles that are better adapted to environmental fluctuations and are likely to survive and to transmit to their offspring alleles and favorable genetic characteristics.

Selection, mutation, migration between populations and genetic drift constitute the different evolutionary forces that can create linkage disequilibrium (LD), e.g. preferential association of alleles at different loci. The analysis of LD plays a central role in many areas of population genetics, including: the determination of genetic maps, ascertainment of levels of recombination at the population level, and estimation of effective population sizes (N_e). The N_e of a population is a concept that was developed by Wright [8] and defined as the size of an idealized population undergoing the same rate of genetic drift as the population under study. A number of methods to estimate N_e from demographic, pedigree, or molecular data have been proposed (e.g. Leroy et al. [9]). Most of the molecular estimates are derived from LD or temporal methods that give indirect estimators of N_e [10] through the use of a genetic index: the squared correlation r^2 of alleles at different gene loci using a single sample in the LD method or the standardized variance in allele frequency between two temporal samples in the temporal approach. The temporal method is based on the theory that N_e is the only parameter that is needed to determine rates of change in allele frequency at neutral loci in a population in Hardy–Weinberg equilibrium [11]. The LD method of estimating N_e, developed by Sved [12] and modified by Hill [13], is based on the principle that in closed finite populations in approximate drift–mutation–recombination equilibrium and constant census size, associations between alleles at different neutral loci are a function of the population's N_e. Therefore, if an estimate of the rate of recombination between loci is available, *Ne* can be derived from the expected level of allele association across loci $E\left(r^2\right)$.

The maintenance of genetic diversity within a population is achieved by maximizing N_e, or equivalently, by minimizing the increase in inbreeding across generations. A molecular estimate of the inbreeding coefficient can be based on measuring long stretches of consecutive homozygous genotypes in each individual, the so-called runs of homozygosity (ROH; McQuillan et al. [14]). Long homozygous regions throughout the genome result from mating between close relatives, reduction in population size, and selection. Thus, population structure and selection effects can be assessed based on the distribution and location of ROH. Several studies have shown that characterizing inbreeding based on ROH provides a better measure of individual autozygosity because parents transmit identical haplotypes to their offspring than estimating overall inbreeding based on pedigree information, because kinships between base animals are not accounted for in pedigree files [15, 16].

In the present study, we used medium-density SNP chips to analyze the genetic diversity within and between six lines of rainbow trout. The objective of this research was to evaluate the impact of selection practices on genetic diversity through basic population indices and individual molecular genetics statistics. Our results may help to evaluate breeding practices in the light of the genetic evolution of farmed fish populations under selection. In addition, it provides commercial breeders a better understanding of the genetic composition of their selected lines and allows them to design and implement effective genomic breeding programs.

Methods

Background of the selected lines

The INRA synthetic line was initially developed by intercrossing several domesticated lines of rainbow trout in order to create a population with a large genetic variability. The line was constituted from a mixture of French farmed populations with some new introductions from Denmark and USA in the early 1980s. The population was then closed to outside germplasm, and has since

been bred without any intentional selection in order to maintain genetic diversity using a full factorial mating design between 2-year-old breeding animals (about 60 dams and 80 sires, each year). Individuals genotyped for the current study were from the 2006 birth year class (SY_n), and from the 2016 birth year class (SY) (Fig. 1).

In early 2008, the first generation of the Suave (SU_1) line was spawned from a full factorial cross among SY_n founder parents (32 dams and 44 sires). The SU line in the current study is the 5th generation of selection for survival and growth on a totally plant-based diet provided from first feeding [17]. Each generation of the SU line is created from a full factorial design between at least 40 dams and 45 sires. The selection is a sequential phenotypic selection: among the surviving animals, fish with the longest body length are selected through three to four selection events during the first year. The proportion of selected fish at each generation is around 4 to 5%. After the first three generations of selection, survival of SU fish was increased by 15% compared to SY fish and their final weight (at 197 days post-fertilization) was increased by 48% due to a 19% increase in feed intake when fish were fed a plant-based diet [18].

Breeding schemes for the four commercial lines (SA, SB, SC and SD) have been based on closed populations that were selected for at least five generations at the time the cohorts were sampled for the study. The samples represented the 5th, 8th, 9th and 10th generation of selection in SB, SC, SA and SD breeding companies, respectively (Fig. 1). The history and the composition of these lines before the initiation of their structured selection program are unknown, however all four lines were created using sex-reversed males from INRA.

The four commercial lines are reproduced by an artificial fertilization protocol to balance the contribution of each parent by partial factorial mating of 6 to 10 dams and 10 sires according to the PROSPER procedure [19]. Each spawn is subdivided in 10 subgroups, with each subgroup being fertilized by a different sire before recombining the dam subgroups. Each of the 10 sires is used to fertilize subgroups from 6 to 10 different dams. At eyed stage, similar numbers of eggs from each dam are mixed together to balance the maternal contributions in the fry rearing tank (effectively also balancing the contribution of each sire). This procedure is expected to minimize inbreeding, since a very large number of families per generation are then created (> 600) limiting the risk of mating related fish from the same full-or half-sib families [20].

Commercial lines are selected for growth traits (mainly weight and length at 18 months) by optimized within-group mass selection and a 3 to 10% selection pressure

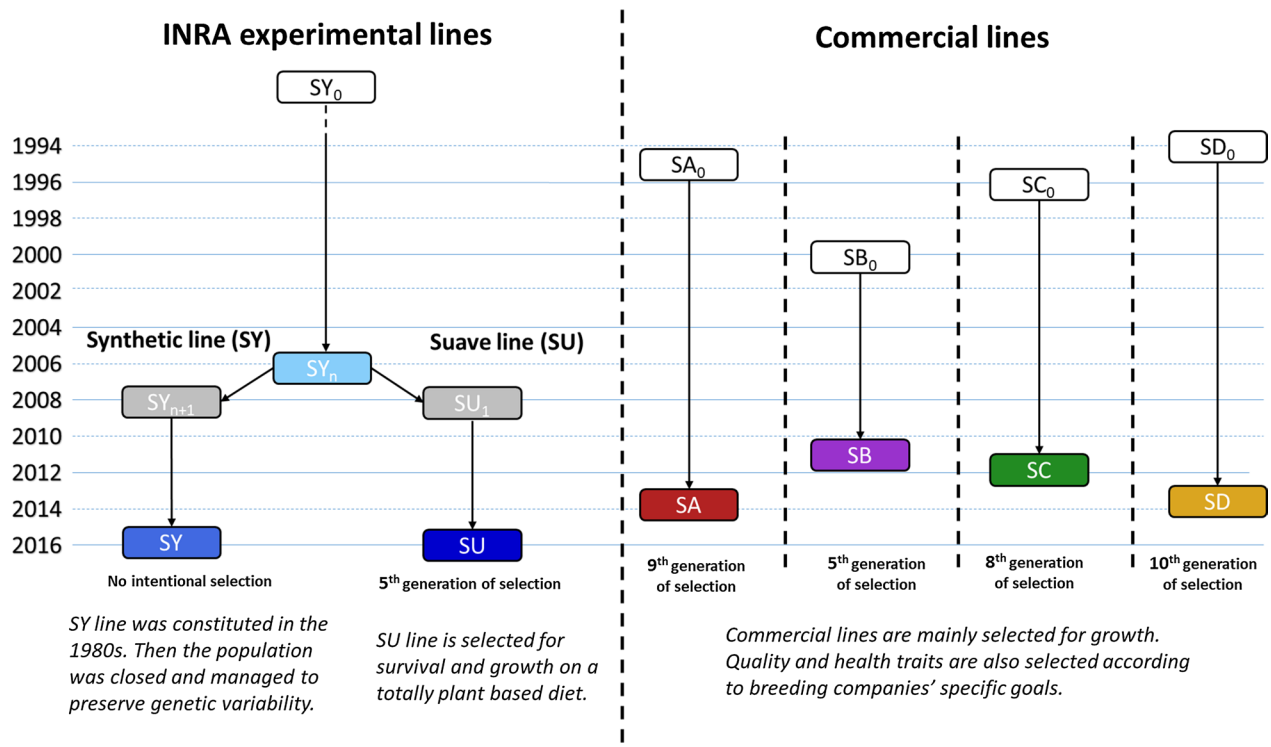

Fig. 1 Chronology of the breeding schemes that were set up for the six French rainbow trout lines

according to the PROSPER procedure [19]. To maintain the four commercial lines, a minimum number of broodstock (> 150) is selected at each generation. In addition, for the SA, SB and SC lines, a sib-based selection is performed based on carcass quality traits (carcass and fillet yields) assisted by ultrasound prediction [21, 22]. The management of inbreeding within these three lines has been based on DNA parentage assignment [23], and has since been improved over the last 10 years by using an optimal pedigree-based selection strategy [24].

Genotypes

The 57,501 SNPs (57K SNP) Axiom® Trout Genotyping array [6] was used to genotype 302 females from six French lines of rainbow trout at the INRA genotyping Platform Gentyane. Animals were sampled to represent the genetic diversity of their birth cohorts by avoiding full-sib relationships for the three commercial lines with pedigree information available (SA, SB, SC). Among the genotyped animals, eight individuals with more than 45% identity-by-state (IBS) with another individual were removed from the INRA and SD populations for which pedigrees are unknown. In addition, four animals with less than 95% of the SNPs genotyped were removed from the study. After editing, 290 genotyped animals were considered in the analysis, including 48, 48, 49, 48, 32, 32 and 33 fish from SA, SB, SC, SD, SU, SY_n and SY lines, respectively.

Subset of the 57K SNPs validated for the study

Quality control of SNPs was performed in several steps. First, all 57,501 SNP probes were positioned with a BLASTn® procedure on the second genome assembly at the chromosome level Omyk_1.0 [25, 26] and only 50,820 SNPs with a unique position were retained. Although *Oncorhynchus mykiss* is a diploid species, there is still residual tetraploidy in the 2.2 Gb of the rainbow trout genome due to the fourth round salmonid whole-genome duplication event that occurred approximatively 96 million years ago [2, 27]. In addition, intraspecific variation in diploid chromosome number (2n varying between 58 and 64) exists in rainbow trout [28]. The American reference genome is based on a set of 29 pairs of chromosomes. However, all six French lines are expected to carry a set of 30 pairs of chromosomes, the American Omy25 being split into two chromosomes (French Omy25a and Omy25b).

Second, we used the Axiom Analysis Suite 2.0 software [29] to control the quality of the remaining markers on a large dataset of 3418 genotyped individuals from the six French rainbow trout lines (including the 302 individuals of the current study). Edits consisted in discarding 7711 SNPs with probe polymorphism or a call rate lower than 97%; 2995 SNPs for which no homozygous individual was observed for the minor allele on the full genotyped set; and 1689 SNPs that were monomorphic in all lines.

Third, regarding the genotypes for the remaining 38,425 SNPs, we performed a final quality control using PLINK v1.9 software [30] on each line set considered in the current study. We discarded SNPs with a very significant deviation from Hardy–Weinberg equilibrium (p value < 0.0001) in one or several populations. Thus, only 38,350 SNPs remained for the analysis of genetic diversity between populations (PCA and F_{ST} analysis).

Finally, we only retained the SNPs that had a sufficiently high minor allele frequency (MAF) to obtain reasonable results for each of the analyses. For the ROH studies (MAF ≥ 1%), 34,077–37,340 SNPs were kept depending on the lines. For LD calculation (MAF ≥ 5%), 31,190–34,723 SNPs were considered depending on the lines (see Additional file 1: Table S1).

Genetic structure of the population

Observed heterozygosity (Ho) and expected heterozygosity (He) for a population under Hardy–Weinberg equilibrium were derived with the PLINK v1.9 software [30]. Levels of genetic variation in the different lines were compared using a Kruskal–Wallis non-parametric test on Ho values. Genetic differentiation between populations was measured with pairwise F_{ST} estimates [31], using the VCFtools v0.1.13 software [32]. In addition, a principal component analysis (PCA) was performed with the R package Adegenet [33] to visualize the genetic structure of the lines.

Linkage disequilibrium

To estimate LD, we used the squared correlation based on genotypic allele counts (number of non-reference alleles at each locus) using the PLINK v1.9 software [30]. This r^2 value does not necessitate phasing, it is very similar to but not identical to the r^2 estimate derived from haplotype frequencies [34].

Pairwise LD between adjacent SNPs and pairwise LD between all SNPs in a 60-Mb long window were derived for each chromosome and line. Mean r^2 values were calculated for each chromosome and line by considering the following average distances between SNPs: 10 kb with a 20 kb-window; 50 kb and 100 kb with a 50 kb-window; 1, 3, 5, 10 and 30 Mb within a 100-kb window for each distance.

Estimates of effective population size N_e

Two metrics were considered to estimate N_e: (N_{e_t}) was based on LD and could be derived for all populations, whereas (N_{e_F}) was based on temporal changes in allele

frequency and could therefore only be calculated for the SY and SU lines.

For all populations, N_e was derived using the formula proposed by Sved [12] and based on the expected LD: $E(r^2) \approx \frac{1}{4cN_e+1}$ in which c is the distance in Morgan between SNPs. Estimating r^2 by sampling from the populations produces another source of error [13]; consequently, for sample size S, the formula is $E(r^2) \approx \frac{1}{1+4cN_e} + \frac{1}{S}$, which can be rearranged as: $N_e \approx \frac{1+{}^{1}/_{S}-E(r^2)}{4c\left[E(r^2)-{}^{1}/_{S}\right]}$.

N_e was estimated by transforming physical distance between SNPs into genetic distance c, based on the genetic map recently established for a French rainbow trout population: 10 cM corresponded to six Mb on average across the 30 chromosomes [5].

We derived N_e at t past generations (N_{e_t}) based on the equation $t = 1/2c$ based on the coalescent theory and the assumptions of Wright–Fisher model [35], considering $E(r^2)$ as the mean r^2 across all the chromosomes and all SNP pairs across a distance of c.

To derive mean r^2, the window around c was determined for any t generation by considering the interval $]t-0.5; t+0.5]$; for instance at $t = 1$, c averaged 33.3 Mb between SNPs with possible values ranging from 20 Mb ($t = 1.5$) to 60 Mb ($t = 0.5$). For all populations, N_{e_t} was calculated up to the 10th generation back. Standard errors for these N_e estimates were derived according to equations (5) and (6) of Hill [13] considering the chromosome average $V(r^2)$ over all SNP pairs across a distance of c per chromosome.

Second, N_e was derived by considering the temporal approach [11] and the formula proposed by Nei and Tajima [36]: $N_{e_F} = \frac{t}{2\hat{F}_k - 1/S_0 - 1/S_t}$, with $\hat{F}_k$ the standardized variance of allele frequency (corresponding to the Wright inbreeding coefficient), S_0 and S_t sample sizes at generation 0 and t, respectively.

$\hat{F}_k$ was derived as $\hat{F}_k = 2 * (x-y)^2 * \left[\frac{1}{x+y} + \frac{1}{2-x-y}\right]$ as proposed by Pollak [37] and considering only two alleles per SNP, where x is the major allele frequency at generation 0 and y is the allele frequency at generation t. Then, $\hat{F}_k$ was averaged for all SNPs to estimate N_{e_F}.

Runs of homozygosity

ROH were identified for each fish within all lines using PLINK v1.9 [30]. ROH were defined by sliding windows with a minimum length of one Mb containing at least 30 homozygous SNPs. The maximum gap between two consecutive homozygous SNPs in a run was set to the default value of one Mb. To ensure that the low SNP density did not falsify ROH length, a minimum density of one SNP every 100 kb was also set (the median and average distances between adjacent SNPs across all the genomes were 31 and 56 kb, respectively). No more than five SNPs with missing genotypes were allowed per window and up to one possible heterozygous genotype was permitted per ROH. These parameters are common practice when deriving ROHs in animal livestock populations with 2–3 Gb genome sizes and 50 K SNP chips. In most studies, the minimum number of SNPs to constitute a ROH is in-between 20 and 50 [38–40]. According to the formula proposed by Purfield et al. [15], the minimum number of SNPs that constituted a ROH should be about 35 SNPs in each line in order to minimize the number of ROH that may occur only by chance in the SNP panel (accepting a 5% false positive rate).

Mean number of ROH (N_{ROH}), number of SNP per ROH (S_{ROH}), length of ROH (L_{ROH}) and percentage of ROH segment longer or equal to 10 Mb were calculated per individual and line. To identify the genomic regions most commonly associated with ROH, the percentage of individuals with a SNP in a ROH segment was calculated by counting the number of times the SNP was detected in a ROH within the population. This count was plotted against the position of the SNP along the chromosome.

Estimates of total and recent inbreeding

Inbreeding coefficients (F_{ROH}) were calculated as the sum of ROH lengths of an individual divided by the total length of the autosomal genome covered by SNPs. The total size of the autosomal genome covered by SNPs was calculated within each line and chromosome. Genome regions with a gap between two adjacent SNPs larger than 1 Mb (authorized gap to derive a ROH) were deducted from the total size of the genome covered by SNPs. In addition, recent inbreeding ($F_{ROH>10Mb}$) was derived as the sum of the lengths of ROH segments longer than 10 Mb in order to estimate inbreeding occurring within the last three generations (10 Mb = 0.166c and $t = 1/2c$; therefore $t = 3$ generations for 10 Mb). Chromosome F_{ROH} was also derived as the sum of the lengths of ROH in a given chromosome of an individual divided by the total length of the chromosome genome covered by SNPs and then averaged for each line.

Results

Population genetic statistics and structure

When considering SNPs irrespective of their MAF, the observed heterozygosity (Ho) ranged from 33.5 to 35.3% (Table 1). On average, values for Ho went up to 36.5 to 38.0% when only SNPs with a MAF ≥ 5% within the population were considered (Table 1). Frequencies at almost all loci were in agreement with Hardy–Weinberg expectations in each line, although,

Table 1 Observed (Ho) and expected (He) heterozygosity for each rainbow trout line

Line	N	All SNPs		SNPs with a MAF ≥ 5%	
		Ho	He	Ho	He
SA	48	33.50*	32.47	36.54*	35.41
SB	48	34.24*	33.34	36.94	35.96
SC	49	33.92*	33.17	36.92	36.10
SD	48	35.01	34.18	37.47	36.58
SU	32	35.33	34.10	37.98	36.65
SY_n	32	35.30	34.90	37.49	37.06
SY	33	35.23	34.87	37.47	37.05

* Significant p value (< 1%) for Kruskal–Wallis test for Ho value in comparison to Ho value for SY line

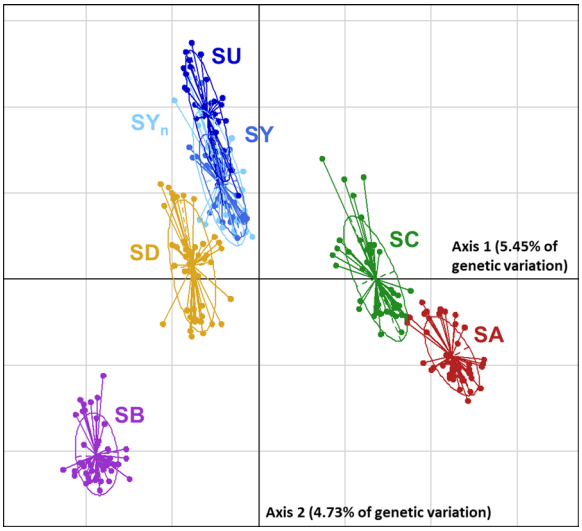

Fig. 2 Principal component analysis plot of the genetic diversity between the French rainbow trout lines. PCA was performed with 290 individuals and 38,350 SNPs

on average across all the SNPs the expected heterozygosity (He) tended to be one point below that of Ho in all selected populations. Kruskal–Wallis tests showed that there was a very significant difference in Ho (p value < 0.00001) between the lines when considering all SNPs and still some significant variation (p value < 1%) when considering SNPs with a MAF ≥ 5%. Comparing the Ho values two-by-two showed that the commercial SA, SB and SC lines had significantly lower heterozygosity values than the SD and INRA experimental lines.

Results showed moderate genetic differentiation between lines. Overall, 10% of the total genetic variation is explained by the first two PCA axes (Fig. 2). Except for the null differentiation between the two cohorts of the INRA unselected line (Table 2), F_{ST} values ranged from 0.02 (SY_n–SU) to 0.15 (SA–SB). All commercial lines were moderately distant from each other with F_{ST} ranging from 0.09 to 0.15. The SD line was the commercial line that was genetically the closest to the INRA SY line, whereas the SA and SB were the most distant.

Linkage disequilibrium analysis

As expected, average r^2 tended to decrease with increasing distance between pairs of SNPs in all the populations studied, the most rapid decline being over the first 100 kb (Fig. 3). On average, LD decreased from 0.34 (0.30–0.39) to 0.25 (0.22–0.30), 0.23 (0.19–0.28), 0.16 (0.13–0.20), 0.10 (0.09–0.12) and 0.07 (0.07–0.09) for distances between markers of 10 kb, 50 kb, 100 kb, 1 Mb, 5 Mb and 10 Mb, respectively (see Additional file 1: Table S2). The unselected lines SY_n and SY and the commercial line SD had the lowest LD with an average r^2 equal to 0.22 at 50 kb, whereas the selected lines SU, SC, SA and SB had corresponding r^2 values of 0.25, 0.27, 0.27 and 0.30, respectively (see Additional file 1: Table S2).

Average r^2 for the 50 kb-distant markers varied strongly between chromosomes: from 0.18 for Omy22 to 0.37 for Omy5 with an average value of 0.24 across chromosomes (see Additional file 2: Figure S1). Within chromosome, r^2 varied also between lines: for Omy5, average r^2 for 50 kb-distant markers varied from 0.34 for SY to 0.41 for the SU line; for Omy22, r^2 varied from 0.16 for SD to 0.22 for the SA and SB lines. For some other chromosomes, in

Table 2 Pairwise F_{ST} between lines of French rainbow trout

Line	SA	SB	SC	SD	SU	SY_n
SB	0.150					
SC	0.115	0.139				
SD	0.113	0.092	0.101			
SU	0.128	0.128	0.113	0.08		
SY_n	0.101	0.096	0.086	0.057	0.024	
SY	0.105	0.102	0.091	0.065	0.038	0.007

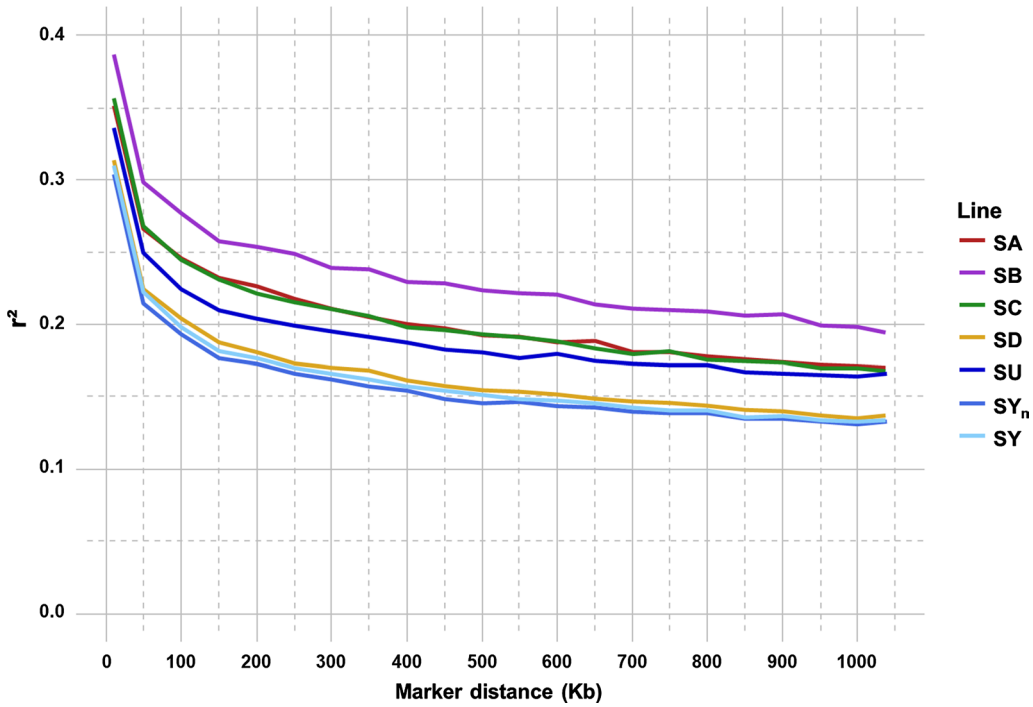

Fig. 3 Linkage disequilibrium (r^2) decay with physical distance between markers in each of the six French rainbow trout lines

particular for the sex chromosome Omy29, variation in average r^2 was more important across lines: from 0.18 for SD to 0.31 for SB.

Estimates of effective population size

The N_e of all lines showed a decreasing trend over the last 10 generations with a steeper slope for the INRA experimental lines and SD line (Fig. 4) than for the four other lines. N_e stabilized during the last three generations for the SA, SB and SC lines and even started to increase for the last two generations for the SA and SB lines.

In the last generation ($t = 1$), N_e ranged from only 24 for the SU line to 48 for the SA and SY lines, with intermediate values of 37, 38, 39 and 42 for the SB, SD, SC and SY_n lines, respectively. For the experimental lines SU and SY, estimates of N_e in the last generation ($t = 1$) due to variations in allele frequency were higher than estimates based on LD: 66 versus 24 and 216 versus 48 for the SU and SY lines, respectively (results not shown).

Analysis of runs of homozygosity

Statistics concerning the average number and size of ROH segments per individual according to the lines are reported in Table S3 (see Additional file 1: Table S3). The average number of ROH per individual varied from 46 for the SY_n to 68 for the SA line. Individuals with the smallest number (25) and the largest number (90) of ROH belonged to the SY_n and SA lines, respectively. The average size of ROH varied from 3.84 Mb for SY_n to 5.38 Mb for SB. The proportion of long ROH ($\geq$ 10 Mb) varied from 7% for SY_n to nearly 14% for SB.

Estimates of total and recent inbreeding

Average F_{ROH} varied from 10.0% for SY_n to 19.5% for SB [Fig. 5 and Table S3 (see Additional file 1: Table S3)]. Individual F_{ROH} varied from 4.6 to 31.7% (Fig. 5). Recent inbreeding ($F_{ROH>10Mb}$) ranged from 2.9 for SY_n to 7.9% for SB [Fig. 5 and Table S3 (see Additional file 1: Table S3)]. Individual $F_{ROH>10Mb}$ varied from 0 to 22.9% for an individual in SY (Fig. 5). An increase in inbreeding of 1 point (from 10 to 11% between SY_n and SY) was observed in the unselected SY line in five generations. All selected lines had higher average inbreeding levels than the unselected SY line, the highest levels for total (> 16%) and recent (> 5%) inbreeding being found for the SA, SB and SC lines that are selected based on pedigree information since seven, four and six generations, respectively.

Regarding F_{ROH} at the chromosome level (Fig. 6), the mean F_{ROH} across lines ranged from 11% for Omy13 to 20% for Omy5 with an average of 14% across chromosomes. The highest mean F_{ROH} (30%) was observed for SA on Omy23 and the lowest value (4%) for SY_n on

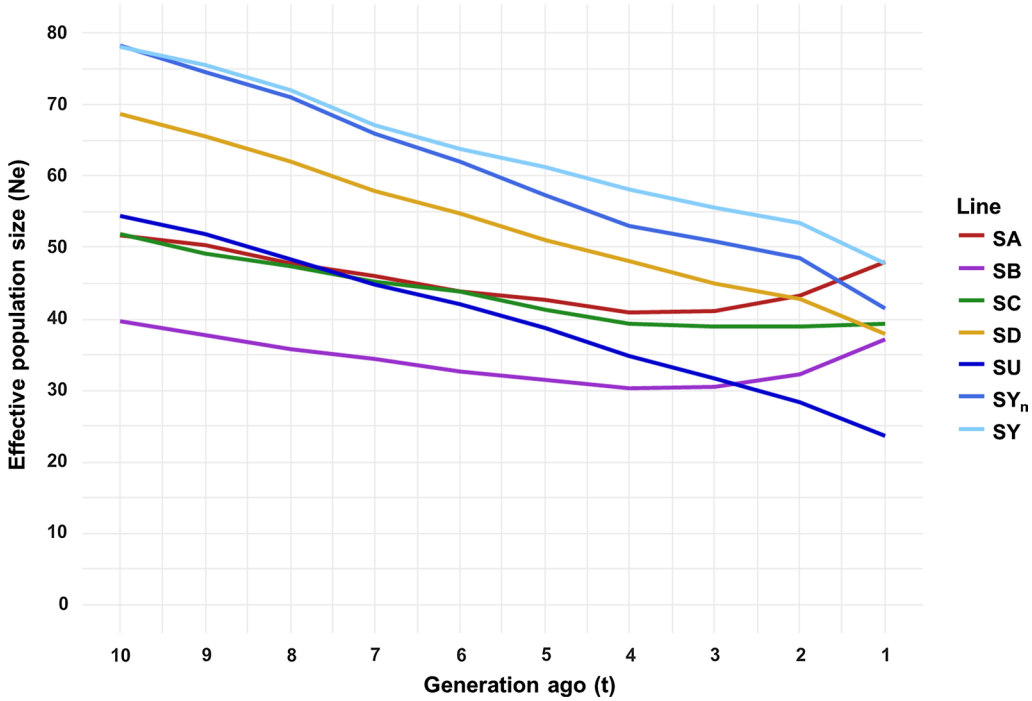

Fig. 4 LD-based estimates of effective population size (N_e) in the French rainbow trout lines over the last ten generations

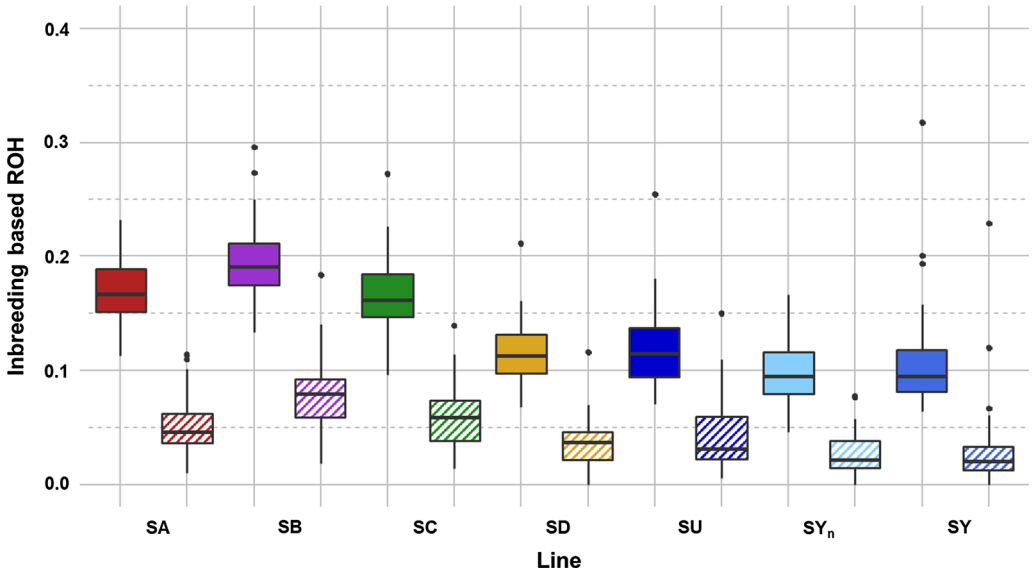

Fig. 5 Box plots of total inbreeding (F_{ROH}) and recent inbreeding ($F_{ROH>10}$) for each rainbow trout line. Plain box: total inbreeding (F_{ROH}); hatched box: recent inbreeding ($F_{ROH>10}$)

Omy3. Average F_{ROH} is sometimes relatively close from one line to another (for instance for Omy10, the lowest value is 11% for SD and the highest value is 17% for SB), whereas variation in F_{ROH} is much larger for Omy23, from 7% for SY to 30% for SA.

Discussion

Population differentiation

Most of our rainbow trout populations were moderately differentiated according to the qualitative guidelines

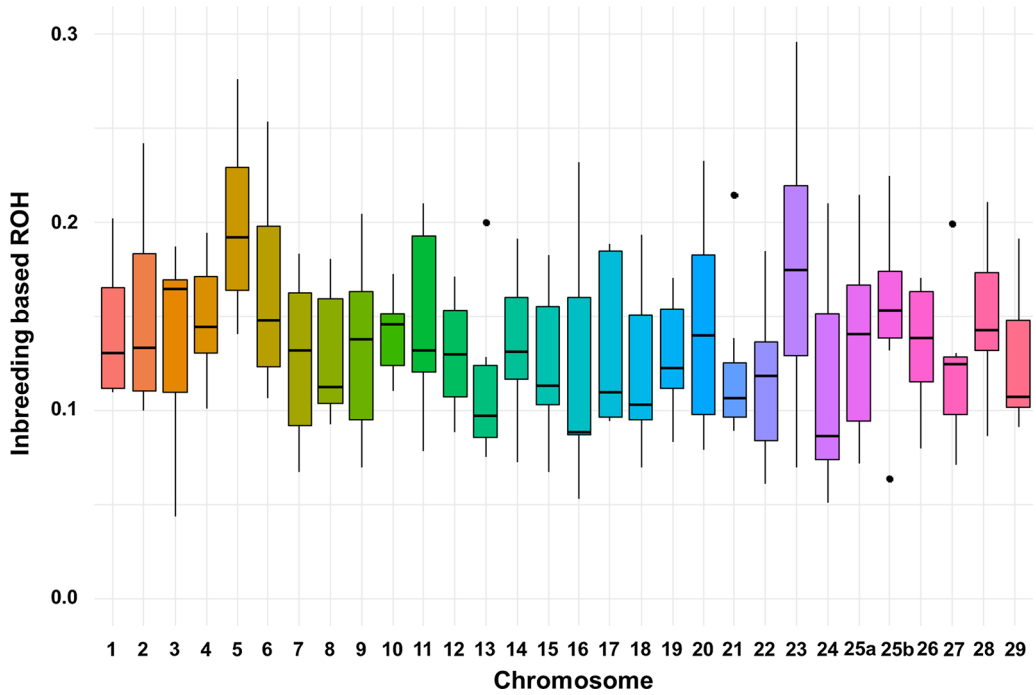

Fig. 6 Box plots of the line averages of inbreeding coefficients (F_{ROH}) derived from ROH per chromosome

proposed by Wright [41] for the interpretation of F_{ST}: 0–0.05 for little genetic differentiation, 0.05–0.15 for moderate genetic differentiation, 0.15–0.25 for large genetic differentiation, and above 0.25 for very large genetic differentiation, respectively.

Little genetic differentiation ($F_{ST} < 0.03$) was observed between the SU and SY experimental lines, which is reasonable given that the SU selected line was derived from the SY_n line. The SD commercial line was also genetically close to the SY INRA lines, partly due to the fact that about 50% of the male founders of the SD line came from the INRA sex-reversed males, 25 years ago. Although the other commercial lines are more distant from the INRA line than the SD line (Fig. 2), it is also important to note that all of these four lines were created using sex-reversed males from INRA. Among all the selected lines, F_{ST} values were quite similar to the observed differentiation levels between Chinese or Western pigs breeds [42], European cattle breeds [39, 43] or sheep breeds [44, 45]. As far as we know, only one F_{ST} study has been performed for rainbow trout based on a panel of 99 SNPs [46] and found a global F_{ST} of 0.13 across eight commercial lines and a pairwise F_{ST} between any two lines ranging from 0.056 to 0.195. Some older works based on microsatellites [47–49] or allozymes [50] also indicated a moderate degree of differentiation among wild population and/or farmed stock rainbow trout lines.

One way to assess the genetic diversity within populations is through observed (Ho) and expected (He) heterozygosity. We believe that these parameters should be derived without setting a threshold MAF value in order to keep all SNPs in the analysis and avoid any bias. In this way, we estimated an Ho of about 34% (± 1%); however, most of studies in the literature give results for a subset of SNPs with MAF above 5%. We also calculated Ho under this MAF threshold and found values close to 37% in all lines, which are very similar to the values estimated in pig lines [51] or cattle breeds [39]. Based on allozyme markers, Cárcamo et al. [50] indicated that commercial lines of rainbow trout showed a similar range of variation in heterozygosity to that of wild populations.

In our study, we observed a slight heterozygosity excess (Ho > He) with a large set of neutral loci for all lines. This observation may indicate a recent (< 100 generations) bottleneck [52] that is likely due to the recent domestication and selection process.

Linkage disequilibrium

The LD at short distances between markers (≤ 100 kb) is very similar to r^2 values reported for dairy cattle breeds [53] and slightly higher than values reported for beef cattle breeds or sheep breeds [45]; however, LD estimates for our trout lines at short distances are lower than in most pig breeds [51, 54]. At longer distances between markers (e.g. ~ 1 Mb), LD in French rainbow trout lines

(0.13–0. 20) is clearly higher than in ruminant breeds, but very similar to the LD observed in most pig breeds. This high LD, even at long distances, enables accurate genomic predictions for rainbow trout populations even with low density SNP chips [55].

Similar values of r^2 using long distance windows (5 Mb) were observed between the French rainbow trout lines and an American commercial line of rainbow trout [55], with a similar large variation in average r^2 among chromosomes (see Additional file 2: Figure S1), i.e. very high r^2 values on Omy5 and very low values on Omy21 and Omy22. The higher than average LD on Omy5 is likely caused by a large chromosomal double-inversion of 56 Mb [26], which prevents recombination in fish. This double-inversion contains key genes involved in photosensory processes, circadian rhythm, adiposity, and sexual differentiation [26]. This inversion is shown to mediate sex-specific migration through sex-dependent dominance. Quantitative trait loci (QTL) for spawning date and body weight have been detected on Omy5 [56] and may also explain the high LD because of selection of favorable haplotypes for these important traits for breeding. Furthermore, Omy5 is one major exception to the strong association observed between large metacentric chromosomes and high female:male recombination rate ratios [57]. In addition to having an equal female:male recombination rate ratio (average 9:1 ratio for metacentric chromosomes), there is evidence that the short arm of Omy5 is the homeologous linkage group to the Omy29 [58], sex chromosome for which we observed a large variation of r^2 values among lines.

Effective population size

The estimates of N_e for the selected lines were consistent with the reports in other aquaculture species such as the Pacific abalone [59], catfish [60], Atlantic salmon [61] for which N_e was less than 50 after a few generations of mass selection. The N_e across many livestock breeds is less than 100 [9, 51, 62]. Nevertheless, large and ongoing genetic gains for production traits are typically achieved in livestock, and there are no apparent signs of reaching selection plateaus [63]. Across both domestic and wild populations, the minimal N_e to avoid inbreeding depression in the short term has been estimated to be at least 50 [64]. For instance, in a long-term (120 generations) selection experiment in mice, several lines were kept at an average N_e of 60 with no apparent inbreeding problems [65].

In the ideal scenario, population genetics theory recommends keeping equal numbers of males and females and maintaining a constant population size over time. However in most livestock breeding programs, it is often impossible to maintain a 1:1 sex ratio, which greatly affects the N_e of a population. In the trout lines under study, the dams-sires ratios are relatively well balanced with values ranging from 6:10 to 10:10. In fish, an additional explanation for the small N_e may be asymmetric reproduction, e.g. high variance in individual reproductive success or survival rate of broodstock [66]. This last phenomenon has been described as "sweepstakes reproductive success" (SRS [67]), which maintains much less genetic diversity than expected on the basis of a large census size, and would increase the impact of inbreeding depression for aquaculture farms. Christie et al. [68] found in their review of salmon species that early-generation hatchery fish averaged only half the reproductive success of their wild-origin counterparts when spawning in the wild. For *Oncorhynchus mykiss*, the ratio of N_e to the estimated census population size (N) was estimated to range from 0.09 to 0.18 [47], 0.17 to 0.40 [69], and a 0.45 [48], with a large variance in reproductive success among individuals being the key factor to reduce the N_e/N ratio in salmonid species [68]. Fish breeding from multigenerational hatchery program without pedigree information resulted in a decrease in N_e, not only by decreasing the mean reproductive success but also by increasing the variance in reproductive success among breeding parents, whereas there was no reduction in N_e found in fish breeds for a single generation in a local hatchery [69].

Intense directional selection of relatively few animals often results in a skewed genetic contribution and may explain why N_e values less than 100 are observed for many livestock populations. In our study, the impact of intense selection can be quantified by comparing the evolution of N_e between the SU and SY lines, but also by taking into account that N_e is affected by the smaller number of broodstock used for SU than for SY breeding. Based on LD values, N_e at $t = 1$ generation was estimated to be equal to 24 and 48 for those two lines, respectively. The discrepancy was even greater when considering estimates of N_e based on allele frequency variation, 66 for the SU line and 216 for the SY line.

Due to the existence of two major inversions [26] on Omy5 and Omy20 that may influence our LD-based analysis (and to a lower extent the ROH), we removed these chromosomes from the derivation of N_e and F_{ROH}. Values for total F_{ROH} and recent inbreeding were almost unchanged, but N_e estimates were very sensitive to the absence of Omy5 and Omy20 although r^2 values were averaged per chromosome before deriving N_e. Removing these chromosomes increases our estimates of N_e across all lines, on average by 40% at generation $t = 1$ and by 14% at generation $t = 10$ (see Additional file 1: Table S4). However, LD-based estimates of N_e remained small for all the lines compared to census sizes that were

close to 200 for all commercial lines, 140 for the SY line and 80 for the SU line. Regardless of the line considered in the study (selected or not), we observed that in general N_e tended to decrease in the last 10 generations, which may be explained by drift and SRS phenomena. It should also be noted that in the three selected lines (SA, SB and SC) for which a mating optimization procedure to limit inbreeding was set up in the last 10 years by SYSAAF, N_e stabilized (SC) or even slightly increased (SA and SB) in the last three generations. This suggests that some mating optimization protocol (genomic-based pedigree) should be similarly introduced to limit the inbreeding increases in the SD, SU and SY lines.

In addition to quantifying the important decreases in N_e due to selection, our results underscored the difficulty in acquiring accurate estimates for this parameter, since estimates based on the same molecular information can be doubled or tripled depending on the evolutionary model considered [70]. It is well-established that better estimates of inbreeding levels and N_e are obtained using molecular rather than genealogical information [16], at least when considering more than 10K SNPs [71]. In most cases, a full-depth pedigree is unavailable and most base populations include selected or partially inbred founders in the pedigree files. These unknown relationships between animals may lead to strong underestimation of the pedigree-based estimates of inbreeding [16]. In addition, by considering complete pedigree back to a base population in simulation studies, Liu et al. [72] and Forutan et al. [16] reported that pedigree-based estimates of inbreeding were lower than true inbreeding in selected populations. The pedigree-based inbreeding assumes neutral loci, i.e. that the two alleles at the same locus on two homologous chromosomes have an equal chance of being selected. In reality, for some loci, the two alleles may have different effects on a naturally or artificially selected trait, which leads to unequal selection probabilities between the two alleles.

Most methods applied to infer N_e from genomic population data rely on the Wright–Fisher model's assumptions of low fecundity non-skewed offspring distributions. Although proven to be robust to violations of most of these assumptions, these methods drastically failed to approximate the genealogies of species with high SRS, whereby few individuals contribute most of the offspring to the next generation [73]. As stated by Montano [70], the development of statistical tools based on models that consider SRS will substantially improve estimates of population demographic parameters.

Runs of homozygosity

The majority of metrics to estimate N_e assumes that the value remains constant across the genome. However, N_e varies across the genome, such that some regions have an increased loss of diversity compared with others [74, 75]. The N_e is expected to vary across the genome as a consequence of genetic hitchhiking due to selection [76] and negative selection acting on deleterious mutations (i.e., background selection [77]). The action of selection, particularly in regions of the genome with low rates of recombination, is expected to reduce N_e, leading to lower levels of genetic diversity and reduced effectiveness of selection. We could have derived N_e per chromosome based on chromosome-specific LD estimates; however, we preferred to study the heterogeneity of genetic diversity throughout the genome using ROH because N_e estimates have been proven unreliable for fish populations (see previous section), and quantifying the absolute level of inbreeding (in total and for each chromosome) in our rainbow trout lines was per se an objective of the study. ROH can be used to directly estimate inbreeding depression [78] and can identify the chromosome regions that are responsible for inbreeding depression along the genome [79].

While for selected rainbow trout lines, F_{ROH} varied between 12 and nearly 20%, estimates were a bit lower (~11%) for the INRA unselected line. These results are consistent with our N_e estimates for three to ten generations ago: the SY and SD lines had the largest N_e and the lowest F_{ROH} whereas SB had the smallest N_e and the highest F_{ROH} (Pearson correlation F_{ROH} and $N_{e_{t=10}}$: −0.90, p value < 0.005). It is difficult to compare our results to other broodstock programs because the few published estimations of levels of inbreeding for fish commercial populations are all based on pedigree information, and thus are likely to be underestimated due to unknown relationships and inbreeding in the base populations. In rainbow trout, Pante et al. [80] estimated inbreeding levels that varied between 3 and 10% at the 6th generation of selection for three commercial populations. A hierarchical mating system was used (one male mated to two to three females) with a very variable number of families from one generation to another; full-sib and half-sib matings were avoided and matings that would yield inbreeding coefficients of 12.5% or more were restricted. In Coho salmon, Myers et al. [81] reported inbreeding levels of about 15–16% in two lines after the 9th and 10 generations of the Domsea selection program based on a circular mating design with 60 families produced at each generation [82]. More recently, in two different lines of Coho salmon, Yáñez et al. [83] estimated inbreeding levels at 5 and 7% after 6 and 7 generations of selection, respectively, under a hierarchical design (one male mated to three to five females) with 100 families produced at each generation and inbreeding controlled by avoiding half- and full-sib matings.

While empirical information on genome-wide diversity in model species and livestock has been collected, we still lack a clear picture for farmed fish of the heterogeneity of genetic diversity across the genome. As far as we know, our study is the first to calculate ROH and estimate F_{ROH} in any fish species. F_{ROH} estimates from this study were higher than estimates for terrestrial livestock regardless of their breeding management (in breeds for ruminants or in lines for pigs). Estimates varied from 3 to 9% in dairy cattle breeds [15, 38, 39], from 2 to 11% in sheep breeds [84, 85] and from 3 to 11% in pig lines [51, 86, 87].

Although the SY line is not under selective breeding, inbreeding is relatively high (11%) and has increased by 1% in six generations (see Additional file 1: Table S3). This is probably partly due to a founder effect as well as to SRS, as previously discussed for N_e. In spite of the existence of mating plans for selected lines with known pedigree (SA, SB and SC), recent inbreeding is relatively high in these lines (5–8%). Nevertheless, we may be able to manage their genetic diversity along the entire genome. Because no ROH fragments are fixed within the populations (see Additional file 2: Figure S2), we can develop selection and mating strategies to specifically limit inbreeding in genome regions for which an important proportion of individuals (> 50% for instance) share the same ROH.

It is worthwhile to underscore that F_{ROH} estimates may vary depending on the parameters considered to define ROH segments. These parameters have to be tuned according to marker density and pan-genomic heterogeneity and the level of recombination along the genome. There is no clear consensus in the literature on how to choose these parameters and most studies used Plink default values. In a recent study [16], a gene-dropping simulation was performed and inbreeding estimates based on ROH and pedigree data were compared to true inbreeding. Inbreeding based on ROH was estimated using different software and different threshold parameters using 50 K chip data. While pedigree inbreeding underestimated true inbreeding, using ROH with a minimum window size of 20 to 50 SNPs provided the closest estimates to true inbreeding regardless of the software [16, 88]. In our study, we observed that inbreeding estimates were similar for window sizes of 30 and 50 SNPs, and estimates were almost identical when considering recent inbreeding (results not shown). Nevertheless the results were sensitive to the maximum gap between two successive SNPs (1000 kb vs. 250 kb) and, to a lesser extent, to the MAF threshold (see Additional file 1: Table S3). Due to the heterogeneity of the marker density along the genome, long ROH were fragmented into smaller ones when considering a maximum gap of 250 kb, leading to underestimated F_{ROH} in our lines. The largest differences were for recent inbreeding estimates, $F_{ROH \geq 10}$, that were less than 1% for the gap threshold of 250 kb, but varied between 3 and almost 8% for the 1000 kb threshold (see Additional file 1: Table S3).

As shown in previous studies on the sheep genome [45, 85, 89], we observed a strong variation in F_{ROH} values among chromosomes, these results being consistent with variation in LD among chromosomes. Chromosome ends are less represented in ROH segments (results not shown), probably due to the very low marker density in these areas. Hotspot areas for ROH were observed for each chromosome (see Additional file 2: Figure S2). In general, these areas and the proportions of individuals within a population that shared ROH hotspots in those areas were highly variable from one line to another. Nevertheless a few hotspot areas were commonly detected across lines, such as the position at ~ 48 Mb of Omy10 (see Additional file 2: Figure S3). At this specific position, is located a QTL for bacterial cold disease resistance that was detected in an American rainbow trout population [90] and in a French commercial line [91]. When hotspot areas are common to different lines, they may be due to local low recombination rates [85] or to key genes for domestication or production [39, 40, 85]. These hotspot areas could then help identify important genes for domestication or production.

The accumulation of inbreeding is heterogeneous across the genome, such that certain regions are being inbred at a faster rate than other regions of the genome. Currently, SNP information is predominantly used to predict breeding values for genomic selection in livestock populations, now including farmed fish (catfish, tilapia and salmonids). However, this information could also be used to manage the rate at which inbreeding builds up in the genome of livestock populations. Characterizing and efficiently managing inbreeding levels is a major goal to ensure that populations can adapt to future breeding goals while maintaining genetic diversity and avoiding the accumulation of detrimental effects associated with inbreeding. Identifying the main regions of the genome that contribute to inbreeding depression may help to manage inbreeding not only at the individual level but directly at the genome level.

Conclusions

Our findings provide the breeding companies a better understanding of the genetic diversity in their rainbow trout lines in order to implement efficient breeding programs. The availability of a genome-wide SNP chip allowed us to characterize genetic diversity between and within lines. Lines are moderately differentiated across commercially developed trout lines, with similar F_{ST} values as those reported across ruminant breeds or pig lines.

Within each line, effective population size seems rather small and inbreeding levels are higher than in terrestrial livestock selected populations. This may be explained by founder effects, sweepstakes reproductive success, and intense selection in some lines. The impact of these significant levels of inbreeding on rainbow trout performance should be quantified in order to assess potential inbreeding depression phenomena and risks for future genetic gains. The levels of molecular inbreeding derived from the identification of homozygous genomic segments could also be used to directly identify the regions that are responsible for inbreeding depression along the genome. We could then expect a more efficient purging allowing for higher relatedness between selected individuals without inbreeding drawbacks in breeding programs.

Additional files

Additional file 1. **Table S1**: Number of SNPs per minimum allele frequency (MAF) category in each line. **Table S2**: Average $r^2 \pm$ SD between SNPs according to different distances. **Table S3**: Sensitivity analysis of ROH estimates to MAF and maximal distance gap between two SNPs and derived inbreeding coefficient. **Table S4**: Estimates of effective population size (standard errors in brackets) for each line with or without Omy5 and Omy20.

Additional file 2. **Figure S1**: Line mean linkage disequilibrium at 50 kb for each chromosome. **Figure S2**: Proportion of individuals per line with a SNP in a ROH along the genome. **Figure S3**: Proportion of individuals per line with a SNP in a ROH along chromosome 10.

Acknowledgements
We thank the staff of the INRA experimental facilities (PEIMA, Sizun) and the French breeding companies who are members of SYSAAF (Aqualande, Bretagne Truite, Les Fils de Charles Murgat and Viviers de Sarrance) for their participation in the fish rearing and for providing biological samples for genotyping individuals. We are very grateful to D. Laloë and G. Restoux for valuable advice on principal component analysis and N_e estimation, respectively.

Authors' contributions
JDA managed the validation of genotypes, wrote some scripts, estimated and interpreted the population genetic parameters and wrote the manuscript. FP wrote some scripts, supervised the analysis and the interpretation of the results and wrote the manuscript. PH and AB managed the project, supervised data collection and contributed to interpreting the results. SB proposed some scripts. CP performed the genotyping of all animals. EQ contributed to the design of the study. ND wrote some scripts. CF and MC blasted the markers on the rainbow trout genome and edited duplicate markers. MDN initiated the project, managed the data collection and preparation for genotyping, and contributed to the design of the study and interpretation of the results. All authors read and approved the final manuscript.

Funding
This study was supported by the European Maritime and Fisheries Fund and FranceAgrimer as part of "57K-Truite" Project (No. 2015-0638) and "SG-Truite" Project (RFEA47 0016 FA 1000016, No. 2017-0239).

Availability of data and materials
The datasets generated and/or analyzed during the current study are not publicly available because they belong to commercial breeding companies but are available from the corresponding author on reasonable request and with permission of the relevant companies.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France. [2] SYSAAF Section Aquacole, Campus de Beaulieu, 35000 Rennes, France. [3] SYSAAF Section Avicole, Centre INRA Val de Loire, 37380 Nouzilly, France. [4] GDEC, INRA, Université Clermont-Auvergne, 63039 Clermont-Ferrand, France.

Received: 1 October 2018 Accepted: 22 May 2019
Published online: 06 June 2019

References

1. Cowx IG. Cultured aquatic species information programme—*Oncorhynchus mykiss*. Rome: FAO Fishries and Aquaculture Department. 2005. http://www.fao.org/fishery/culturedspecies/Oncorhynchus_mykiss/en. Accessed 2 May 2019.
2. Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B, et al. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nat Commun. 2014;5:3657.
3. Guyomard R, Boussaha M, Krieg F, Hervet C, Quillet E. A synthetic rainbow trout linkage map provides new insights into the salmonid whole genome duplication and the conservation of synteny among teleosts. BMC Genet. 2012;13:15.
4. Gonzalez-Pena D, Gao G, Baranski M, Moen T, Cleveland BM, Kenney PB, et al. Genome-wide association study for identifying loci that affect fllet yield, carcass, and body weight traits in Rainbow trout (*Oncorhynchus mykiss*). Front Genet. 2016;7:203.
5. Fraslin C, Dechamp N, Bernard M, Krieg F, Hervet C, Guyomard R, et al. Quantitative trait loci for resistance to *Flavobacterium psychrophilum* in rainbow trout: effect of the mode of infection and evidence of epistatic interactions. Genet Sel Evol. 2018;50:60.
6. Palti Y, Gao G, Liu S, Kent MP, Lien S, Miller MR, et al. The development and characterization of a 57K SNP array for rainbow trout. Mol Ecol Resour. 2015;15:662–72.
7. Felsenstein J. The effect of linkage on directional selection. Genetics. 1965;52:349–63.
8. Wright S. Evolution in mendelian populations. Genetics. 1931;16:97–159.
9. Leroy G, Mary-Huard T, Verrier E, Danvy S, Charvolin E, Danchin-Burge C. Methods to estimate effective population size using pedigree data: examples in dog, sheep, cattle and horse. Genet Sel Evol. 2013;45:1.
10. Waples RS. Tiny estimates of the N_e/N ratio in marine fishes: Are they real? J Fish Biol. 2016;89:2479–504.
11. Waples RS. A generalized approach for estimating effective population size from temporal changes in allele frequency. Genetics. 1989;121:379–91.
12. Sved JA. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theor Popul Biol. 1971;2:125–41.
13. Hill WG. Estimation of effective population size from data on linkage disequilibrium. Genet Res (Camb). 1981;38:209–16.
14. McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of homozygosity in European populations. Am J Hum Genet. 2008;83:359–72.
15. Purfield DC, Berry DP, McParland S, Bradley DG. Runs of homozygosity and population history in cattle. BMC Genet. 2012;13:70.
16. Forutan M, Ansari Mahyari S, Baes C, Melzer N, Schenkel FS, Sargolzaei M. Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. BMC Genomics. 2018;19:98.
17. Le Boucher R, Dupont-Nivet M, Vandeputte M, Kerneïs T, Goardon L, Labbé L, et al. Selection for adaptation to dietary shifts: towards sustainable breeding of carnivorous fish. PLoS One. 2012;7:e44989.

18. Callet T, Médale F, Larroquet L, Surget A, Aguirre P, Kerneis T, et al. Successful selection of rainbow trout (*Oncorhynchus mykiss*) on their ability to grow with a diet completely devoid of fishmeal and fish oil, and correlated changes in nutritional traits. PLoS One. 2017;12:e0186705.
19. Chevassus B, Quillet E, Krieg F, Hollebecq M-G, Mambrini M, Fauré A, et al. Enhanced individual selection for selecting fast growing fish: the 'PROSPER' method, with application on brown trout (*Salmo trutta fario*). Genet Sel Evol. 2004;36:643–61.
20. Dupont-Nivet M, Vandeputte M, Haffray P, Chevassus B. Effect of different mating designs on inbreeding, genetic variance and response to selection when applying individual selection in fish breeding programs. Aquaculture. 2006;252:161–70.
21. Haffray P, Bugeon J, Rivard Q, Quittet B, Puyo S, Allamelou JM, et al. Genetic parameters of in vivo prediction of carcass, head and fillet yields by internal ultrasound and 2D external imagery in large rainbow trout (*Oncorhynchus mykiss*). Aquaculture. 2013;410–411:236–44.
22. Haffray P, Enez F, Bugeon J, Chapuis H, Dupont-Nivet M, Chatain B, et al. Accuracy of BLUP breeding values in a factorial mating design with mixed families and marker-based parentage assignment in rainbow trout *Oncorhynchus mykiss*. Aquaculture. 2018;490:350–4.
23. Haffray P, Pincent C, Rault P, Coudurier B. Domestication et amélioration génétique des cheptels piscicoles français dans le cadre du SYSAAF. Prod Anim. 2004;17:243–52.
24. Chapuis H, Pincent C, Colleau JJ. Optimizing selection with several constraints in poultry breeding. J Anim Breed Genet. 2016;133:3–12.
25. Omyk_1.0. https://www.ncbi.nlm.nih.gov/assembly/GCF_002163495.1/.
26. Pearse DE, Barson NJ, Nome T, Gao G, Campbell MA, Abadía-Cardoso A, et al. Sex-dependent dominance maintains migration supergene in rainbow trout. bioRxiv. 2018. https://doi.org/10.1101/504621.
27. Macqueen DJ, Johnston IA. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proc Biol Sci. 2014;281:20132881.
28. Thorgaard GH. Chromosomal differences among rainbow trout populations. Copeia. 1983;1983:650–62.
29. Affymetrix. Axiom™ Analysis Suite 2.0: UserGuide. 2016. https://media.affymetrix.com/support/downloads/manuals/axiom_analysis_suite_user_guide.pdf. Accessed 2 May 2019.
30. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.
31. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution (NY). 1984;38:1358–70.
32. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27:2156–8.
33. Jombart T, Ahmed I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics. 2011;27:3070–1.
34. Hill WG, Robertson A. Linkage disequilibrium in finite populations. Theor Appl Genet. 1968;38:226–31.
35. Hayes BJ, Visscher PM, Mcpartlan HC, Goddard ME. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Res. 2003;13:635–43.
36. Nei M, Tajima F. Genetic drift and estimation of effective population size. Genetics. 1981;98:625–40.
37. Pollak E. A new method for estimating the effective population size from allele frequency changes. Genetics. 1983;104:531–48.
38. Rodríguez-Ramilo ST, Fernández J, Toro MA, Hernández D, Villanueva B. Genome-wide estimates of coancestry, inbreeding and effective population size in the spanish holstein population. PLoS One. 2015;10:e0124157.
39. Signer-Hasler H, Burren A, Neuditschko M, Frischknecht M, Garrick D, Stricker C, et al. Population structure and genomic inbreeding in nine Swiss dairy cattle populations. Genet Sel Evol. 2017;49:83.
40. Mastrangelo S, Ciani E, Sardina MT, Sottile G, Pilla F, Portolano B. Runs of homozygosity reveal genome-wide autozygosity in Italian sheep breeds. Anim Genet. 2018;49:71–81.
41. Wright S. Evolution and the genetics of populations, volume 4. Variability within and among natural populations. Chicago: The University of Chicago Press; 1984.
42. Ai H, Huang L, Ren J. Genetic diversity, linkage disequilibrium and selection signatures in Chinese and Western pigs revealed by genome-wide SNP markers. PLoS One. 2013;8:e56001.
43. Gautier M, Laloë D, Moazami-Goudarzi K. Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. PLoS One. 2010;5:e13038.
44. Mastrangelo S, Di Gerlando R, Tolone M, Tortorici L, Sardina MT, Portolano B. Genome wide linkage disequilibrium and genetic structure in Sicilian dairy sheep breeds. BMC Genet. 2014;15:108.
45. Al-Mamun HA, Clark SA, Kwan P, Gondro C. Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. Genet Sel Evol. 2015;47:90.
46. Liu S, Palti Y, Martin KE, Parsons JE, Rexroad CE. Assessment of genetic differentiation and genetic assignment of commercial rainbow trout strains using a SNP panel. Aquaculture. 2017;468:120–5.
47. Heath DD, Busch C, Kelly J, Atagi DY. Temporal change in genetic structure and effective population size in steelhead trout (*Oncorhynchus mykiss*). Mol Ecol. 2002;11:197–214.
48. Rexroad CE III, Vallejo RL. Estimates of linkage disequilibrium and effective population size in rainbow trout. BMC Genet. 2009;10:83.
49. Gross R, Lulla P, Paaver T. Genetic variability and differentiation of rainbow trout (*Oncorhynchus mykiss*) strains in northern and Eastern Europe. Aquaculture. 2007;272:S139–46.
50. Carcamo CB, Diaz NF, Winkler FM. Genetic diversity in Chilean populations of rainbow trout, *Oncorhynchus mykiss*. Lat Am J Aquat Res. 2015;43:59–70.
51. Grossi DA, Jafarikia M, Brito LF, Buzanskas ME, Sargolzaei M, Schenkel FS. Genetic diversity, extent of linkage disequilibrium and persistence of gametic phase in Canadian pigs. BMC Genet. 2017;18:6.
52. Cornuet JM, Luikart G. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics. 1996;144:2001–14.
53. Hozé C, Fouilloux MN, Venot E, Guillaume F, Dassonneville R, Fritz S, et al. High-density marker imputation accuracy in sixteen French cattle breeds. Genet Sel Evol. 2013;45:33.
54. Wang L, Sørensen P, Janss L, Ostersen T, Edwards D. Genome-wide and local pattern of linkage disequilibrium and persistence of phase for 3 Danish pig breeds. BMC Genet. 2013;14:115.
55. Vallejo RL, Silva RMO, Evenhuis JP, Gao G, Sixin L, Parsons JE, et al. Accurate genomic predictions for BCWD resistance in rainbow trout are achieved using low-density SNP panels: evidence that long-range LD is a major contributing factor. J Anim Breed Genet. 2018;135:263–74.
56. O'Malley KG, Sakamoto T, Danzmann RG, Ferguson MM. Quantitative trait loci for spawning date and body weight in rainbow trout: testing for conserved effects across ancestrally duplicated chromosomes. J Hered. 2003;94:273–84.
57. Danzmann RG, Cairney M, Davidson WS, Ferguson MM, Gharbi K, Guyomard R, et al. 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. 2005;48:1037–51.
58. Phillips RB, Nichols KM, DeKoning JJ, Morasch MR, Keatley KA, Rexroad C, et al. Assignment of rainbow trout linkage groups to specific chromosomes. Genetics. 2006;174:1661–70.
59. Chen N, Luo X, Lu C, Ke C, You W. Effects of artificial selection practices on loss of genetic diversity in the Pacific abalone, *Haliotis discus* hannai. Aquac Res. 2017;48:4923–33.
60. Garcia ALS, Bosworth B, Waldbieser G, Misztal I, Tsuruta S, Lourenco DAL. Development of genomic predictions for harvest and carcass weight in channel catfish. Genet Sel Evol. 2018;50:66.
61. Barria A, López ME, Yoshida G, Carvalheiro R, Lhorente JP, Yáñez JM. Population genomic structure and genome-wide linkage disequilibrium in farmed Atlantic salmon (*Salmo salar* L.) using dense SNP genotypes. Front Genet. 2018;9:649.
62. Mastrangelo S, Tolone M, Di Gerlando R, Fontanesi L, Sardina MT, Portolano B. Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds. Animal. 2016;10:746–54.
63. Hill WG, Kirkpatrick M. What animal breeding has taught us about evolution. Annu Rev Ecol Evol Syst. 2010;41:1–19.
64. FAO. Secondary guidelines for development of national farm animal genetic resources management plans: management of small populations at risk. 1998. http://www.fao.org/3/a-w9361e.pdf. Accessed 2 May 2019.

65. Holt M, Meuwissen T, Vangen O. Long-term responses, changes in genetic variances and inbreeding depression from 122 generations of selection on increased litter size in mice. J Anim Breed Genet. 2005;122:199–209.
66. Lallias D, Boudry P, Lapègue S, King JW, Beaumont AR. Strategies for the retention of high genetic variability in European flat oyster (*Ostrea edulis*) restoration programmes. Conserv Genet. 2010;11:1899–910.
67. Hedgecock D, Pudovkin AI. Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. Bull Mar Sci. 2011;87:971–1002.
68. Christie MR, Ford MJ, Blouin MS. On the reproductive success of early-generation hatchery fish in the wild. Evol Appl. 2014;7:883–96.
69. Araki H, Waples RS, Ardren WR, Cooper B, Blouin MS. Effective population size of steelhead trout: influence of variance in reproductive success, hatchery programs, and genetic compensation between life-history forms. Mol Ecol. 2007;16:953–66.
70. Montano V. Coalescent inferences in conservation genetics: Should the exception become the rule? Biol Lett. 2016. https://doi.org/10.1098/rsbl.2016.0211.
71. Wang J. Pedigrees or markers: Which are better in estimating relatedness and inbreeding coefficient? Theor Popul Biol. 2016;107:4–13.
72. Liu H, Sørensen AC, Meuwissen THE, Berg P. Allele frequency changes due to hitch-hiking in genomic selection programs. Genet Sel Evol. 2014;46:8.
73. Eldon B, Wakeley J. Coalescent processes when the distribution of offspring number among individuals is highly skewed. Genetics. 2006;172:2621–33.
74. Gossmann TI, Woolfit M, Eyre-Walker A. Quantifying the variation in the effective population size within a genome. Genetics. 2011;189:1389–402.
75. Jiménez-Mena B, Hospital F, Bataillon T. Heterogeneity in effective population size and its implications in conservation genetics and animal breeding. Conserv Genet Resour. 2016;8:35–41.
76. Smith JM, Haigh J. Hitch-hiking effect of a favorable gene. Genet Res (Camb). 1974;23:23–35.
77. Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. Genetics. 1993;134:1289–303.
78. Silió L, Rodríguez MC, Fernández A, Barragán C, Benítez R, Óvilo C, et al. Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. J Anim Breed Genet. 2013;130:349–60.
79. Pryce JE, Haile-Mariam M, Goddard ME, Hayes BJ. Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. Genet Sel Evol. 2014;46:71.
80. Pante MJR, Gjerde B, McMillan I. Inbreeding levels in selected populations of rainbow trout, *Oncorhynchus mykiss*. Aquaculture. 2001;192:213–24.
81. Myers JM, Heggelund PO, Hudson G, Iwamoto RN. Genetics and broodstock management of coho salmon. Aquaculture. 2001;197:43–62.
82. Iwamoto RN, Towner R, Myers J, Hudson G, Munsell P. Coho salmon broodstock development: a case study of the Domsea Coho salmon (1977 to 2015). Bull Jpn Fish Res Educ Agency. 2017;45:107–12.
83. Yáñez JM, Bassini LN, Filp M, Lhorente JP, Neira R, Ponzoni RW, Neira R. Inbreeding and effective population size in a coho salmon (*Oncorhynchus kisutch*) breeding nucleus in Chile. Aquaculture. 2013;420–421:S15–9.
84. Mastrangelo S, Portolano B, Di Gerlando R, Ciampolini R, Tolone M, Sardina MT. Genome-wide analysis in endangered populations: a case study in Barbaresca sheep. Animal. 2017;11:1107–16.
85. Purfield DC, McParland S, Wall E, Berry DP. The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. PLoS One. 2017;12:e0176780.
86. Zanella R, Peixoto JO, Cardoso FF, Cardoso LL, Biegelmeyer P, Cantão ME, et al. Genetic diversity analysis of two commercial breeds of pigs using genomic and pedigree data. Genet Sel Evol. 2016;48:24.
87. Yang B, Cui L, Perez-Enciso M, Traspov A, Crooijmans RPMA, Zinovieva N, et al. Genome-wide SNP data unveils the globalization of domesticated pigs. Genet Sel Evol. 2017;49:71.
88. Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan TV, Kane JM, et al. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. Proc Natl Acad Sci USA. 2007;104:19942–7.
89. Chitneedi PK, Arranz JJ, Suarez-Vega A, García-Gámez E, Gutiérrez-Gil B. Estimations of linkage disequilibrium, effective population size and ROH-based inbreeding coefficients in Spanish Churra sheep using imputed high-density SNP genotypes. Anim Genet. 2017;48:436–46.
90. Vallejo RL, Liu S, Gao G, Fragomeni BO, Hernandez AG, Leeds TD, et al. Similar genetic architecture with shared and unique quantitative trait loci for bacterial cold water disease resistance in two rainbow trout breeding populations. Front Genet. 2017;8:156.
91. Fraslin C, Brard-Fudulea S, D'Ambrosio J, Bestin A, Charles M, Haffray P, et al. Rainbow trout resistance to bacterial cold water disease: two new quantitative trait loci identified after a natural disease outbreak on a French farm. Anim Genet. 2019. https://doi.org/10.1111/age.12777.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC