Genetic determinism of spontaneous masculinisation in XX female rainbow trout: new insights using medium throughput genotyping and whole-genome sequencing

Clémence Fraslin1,10*, Florence Phocas1,10*, Anastasia Bestin2, Mathieu Charles1,2, Maria Bernard1,2, Francine Krieg1, Nicolas Dechamp1, Céline Ciobotaru1, Chris Hozé1,4, Florent Petitprez2, Marine Milhes6, Jérôme Lluch6, Olivier Bouchez6, Charles Poncet7, Philippe Hocdé6, Pierrick Haffray2, Yann Guiguen9 & Edwige Quillet1

Rainbow trout has a male heterogametic (XY) sex determination system controlled by a major sex-determining gene, sdY. Unexpectedly, a few phenotypically masculinised fish are regularly observed in all-female farmed trout stocks. To better understand the genetic determinism underlying spontaneous maleness in XX-rainbow trout, we recorded the phenotypic sex of 20,210 XX-rainbow trout from a French farm population at 10 and 15 months post-hatching. The overall masculinisation rate was 1.45%. We performed two genome-wide association studies (GWAS) on a subsample of 1139 individuals classified as females, intersex or males using either medium-throughput genotyping (31,811 SNPs) or whole-genome sequencing (WGS, 8.7 million SNPs). The genomic heritability of maleness ranged between 0.48 and 0.62 depending on the method and the number of SNPs used for the estimation. At the 31K SNPs level, we detected four QTL on three chromosomes (Omy1, Omy12 and Omy20). Using WGS information, we narrowed down the positions of the two QTL detected on Omy1 to 96 kb and 347 kb respectively, with the second QTL explaining up to 14% of the total genetic variance of maleness. Within this QTL, we detected three putative candidate genes, fgfa8, cyp17a1 and an uncharacterised protein (LOC110527930), which might be involved in spontaneous maleness of XX-female rainbow trout.

The discovery of factors underlying sex determination in fish is a challenge for fundamental biological and evolutionary perspectives and for aquaculture purposes as in some species, managing the sex ratios of farmed stocks is essential for the production efficiency. Unlike birds and mammals which have highly conserved and simple heterogametic genetic sex determination systems, teleost fish species exhibit an amazing diversity of sex determination systems, with different master sex determining genes as well as many minor genetic determinants interacting with environmental effects (in particular temperature) and epigenetic mechanisms1–3. Among teleost fish, salmonids are known as strict gonochoric species with the individual sex determined at fertilization and remaining the same throughout their live1. In rainbow trout (Oncorhynchus mykiss), the genetic determinism of
sex has first been described as a male heterogametic system (males XY, females XX) based on the observations of sex-ratios in progenies of hormonally sex-reversed individuals and on the absence of males in gynogenetic offspring. The master gene controlling sex determination in salmonids is an unusual sex determining gene called sdY (sexually dimorphic Y chromosome) that evolved from the duplication of the irf9 (interferon regulatory factor 9) immune related gene. sdY is expressed during early gonad differentiation in the testis where it interacts with the conserved female differentiation factor Foxl2 to prevent cyp19a1a up-regulation and the subsequent oestrogen production needed for ovarian differentiation. However, despite a strict linkage between the phenotypic sex and the sdY locus in many salmonid species, spontaneous masculinisation of XX females have been reported in rainbow trout, first in experimental groups, but also in farmed populations (unpublished data). These spontaneous masculinisation and their transmission across generations have been characterized in gynogenetic families, where some of the minor genetic factors acting in addition to the major sdY sex determination system has been suspected. Later, Guyomard et al. detected QTL associated to masculinisation in two doubled haploid gynogenetic trout families. In addition to this genetic control, a few studies have also shown that high temperature treatments applied before the sexual differentiation and at the period of thermosensitivity can modulate sex differentiation and enhance the frequency of maleness in rainbow trout, as shown in some other species including for instance channel catfish (Ictalurus punctatus), Nile tilapia (Oreochromis niloticus) and seabass (Dicentrarchus labrax). Interestingly, in most of the studies in trout, intermediate phenotypes (intersex individuals with only partial masculinisation of gonads) were recorded.

In rainbow trout farming, the production of all-female populations is advantageous due to their late sexual maturation (most usually at 2 years-old) compared to males (about 1 year of age). Indeed, maturing males exhibit a reduced growth and flesh quality (reduction of muscle lipid content and discoloration) and are more sensitive to fungal saprolegniosis. All-female stocks are produced by mating standard XX-females with sex-reversed XX-males. Sex reversal is obtained by feeding young trout fry with a diet containing a masculinising hormone, the 17a-methyl testosterone before the sexual differentiation period. The administration of hormones to obtain sex-reversed males is carried out under strict veterinary control according to the European directive (99/22/CE, 29 April), and treated animals are euthanized after reproduction and discarded from the food chain.

Therefore, there is a dual interest in investigating factors that govern the spontaneous maleness in trout. Depending on the nature of these factors, they could be exploited either to reduce the occurrence of undesirable males observed in all-females populations or to produce sex-reversed males without using hormones, an asset for more sustainable aquaculture.

To further investigate the genetic architecture of the spontaneous maleness, we performed a Genome-Wide Association Study (GWAS) in a French commercial rainbow trout population, in which spontaneous maleness has been repeatedly reported. Genotype information from a medium-density trout genotyping array and whole genome sequencing (WGS) was combined to accurately map QTL- and identify candidate genes responsible for spontaneous maleness.

**Methods**

**Ethics statement.** Fish were reared at the farm « Les Fils de Charles Murgat » (Beaufort, France; UE approval number FR 38 032 001) under conditions complying with the Directive 98/58/CE on the protection of animals kept for farming purposes. The fish were euthanized following the approved method used in the farm for animals intended for marketing (electro-narcosis).

Article 3.1 of the EU Directive 2010/63/EU on the protection of animals used for scientific purposes "excludes the killing of animals solely for the use of their organs or tissues" from the procedures covered by the Directive. As such, the experiment was not subject to oversight by an institutional ethic committee.

Tissue sampling and necropsy to observe the gonads were carried out post-mortem, with the authorisation of an official veterinarian.

**Fish rearing.** In June 2017, gametes were collected from 2.5 years-old breeders that had been previously isolated in light-proofed tanks and exposed to an artificial photoperiod to induce maturation (constant light during two months followed by a 4-months period with partial (8/24 h) lighting, water temperature varied from 11 to 13.5 °C). In total, 50 dams (XX-trout, average body weight of 4.61 ± 0.84 kg) and 50 XX sex-reversed sires (average body weight of 3.28 ± 0.60 kg) were mated according to 5 full-factorial mating designs (10 dams × 10 sires per factorial) to produce all XX-eggs. Fertilized eggs were separated into two batches corresponding to two temperature treatments (25,000 and 45,000 eyed eggs in batch 1 and 2, respectively) and incubated in Mc Donald-type jars with a bottom-to-top water flow maintained at 10.5 °C until hatching. Hatching rate was 64% and 58% in the two batches, which is in the usual range when using progenitors whose spawning has been shifted by photo-period. At the end of yolk resorption (35 days post-fertilisation, dpf), larvae of batch 1 were kept at 12 °C (control group) while in batch 2, the water temperature was increased by 1 °C/day until it reached 18 °C. The water temperature was then maintained constant for 1,134 degree-days covering the expected window of gonad differentiation. It was then decreased by 1 °C/day during 6 days to the same temperature as in the control group (12 °C) and both groups were then reared between 12 and 14 °C. At 84 dpf, 13,000 fish from each temperature group were randomly sampled and retained for the experiment. In order to prevent too much heterogeneity in size within the groups and any risk of biased sampling at time of phenotyping, each group was temporarily split (at 214 dpf) into two subgroups according to the fish size and kept separated for 37 days, allowing the smaller fish to catch up in size with the bigger fish. Fish were then grouped again until sex was recorded. During the whole experiment, fish were fed a commercial diet according standard recommendations.
Phenotypic sex recording. In order to determine the phenotypic sex of the XX-offspring, 13,241 fish from control and 18 °C groups (6560 and 6681 fish, respectively) were sexed at 10 month-post fertilisation (mpf) and 6969 more fish were sexed 5 months later (3456 and 3513 fish, respectively). The overall mean body weight of fish was respectively 234 g at 10 mpf and around 860 g at 15 mpf. Fish were euthanized by electro-narcosis, according to standard procedures for commercial trout farming and both gonads were visually examined to determine sex. When necessary, visual observation was completed by observation of the gonads under a binocular magnifier (1079 fish) and, in some cases, by a histological control (23 fish).

Fish were distributed into four sex classes (no matter the degree of maturation) as followed:

- Female, for fish with two ovaries and no visible sign of testis area (illustration in Supplementary Figure S1, panel a).
- Intersex, for fish with either one entirely female gonad and one entirely male gonad, or with both male and female areas in at least one of the gonads (the other gonad may be male, female, intersex or undetermined) (n = 132). The presence of ovarian lamellae was the criterion used to declare a female area in a gonad, whether oocytes were present or not (illustration in Supplementary Figure S1, panel b).
- Male, for fish with two testis, or for five fish, one testis only (the other gonad being undifferentiated) and no visible sign of any ovary area (n = 162) (illustration in Supplementary Figure S1, panel c).

Fish with undifferentiated gonads that could not be sexed after binocular or histological observation were classified as undetermined and removed from the analysis (n = 84).

All intersex individuals (n = 132), all males (n = 162) as well as 858 females with well-developed ovaries were kept for QTL detection (563 from the control group, at both 10 and 15 mpf, and 295 from the 18 °C group at 15 mpf). Sex was recorded as a categorical variable with three levels: sex = 1 for females, sex = 2 for intersex, sex = 3 for males.

Genotyping and sequencing. Pieces of caudal fin sampled from those 1152 fish were sent to Gentane genotyping platform (INRAE, Clermont-Ferrand, France) for DNA extraction using the DNAAdvance kit from Beckman Coulter following manufacturer instructions. Genotyping was performed with the Axiom Trout Genotyping Array from Thermofisher\textsuperscript{19} that contains 57,501 SNPs.

Quality controls of genotyped SNPs were performed as described in D’Ambrosio et al.\textsuperscript{20}, in particular to remove SNPs with probe polymorphism and multiple locations on the genome. In addition, only the 31,811 SNPs with a call rate higher than 0.97, a test of deviation from Hardy–Weinberg equilibrium with a p-value > 0.0001 and a minor allele frequency (MAF) higher than 0.05 were kept for further analyses. Individuals with a call rate lower than 0.90 were removed from the genotype dataset (n = 5). All missing genotypes were imputed using the FImpute software\textsuperscript{21} (version 2.2) in order to get the full 31,811 SNPs genotypes for all the animals considered in the analyses.

Samples of caudal fin were also collected from a set of 60 females (the 50 dams + 10 relatives) for DNA extraction and whole genome sequencing. DNA was extracted with a Promega Relia Prep dDNA tissue miniprep system (A2051) following manufacturer instructions. DNA sequencing was performed at the GeT-PlaGe core facility and whole genome sequencing. DNA was extracted with a Promega Relia Prep dDNA tissue miniprep system.

Imputation to whole-genome-sequence. Imputation of the 31K SNPs genotypes of the progeny to whole-genome sequence (WGS) was performed chromosome by chromosome using the latest version of FImpute software\textsuperscript{21} (version 3.0) based on the reference population constituted by the 60 sequenced females. After imputation, 8,765,613 SNPs with a MAF higher than 1% were kept for the WGS analysis. In order to reduce the variant dataset to produce a genomic relationship matrix, SNPs in linkage disequilibrium were filtered out with the indel-pairwise option of the PLINK software\textsuperscript{22} (version 1.09). The filtering was performed first on 50 bp sliding windows. In every window, the $r^2$ between each pair of SNPs was calculated and one SNP of the pair was removed if $r^2$ was higher than 0.7. A second round of filtering was performed using 100 bp sliding windows and $r^2$ higher than 0.3. The reduced dataset obtained was composed of 275,283 SNP.
The estimated genetic variance

\[ G \]

and a probability \( 1 - \pi \) that \( \delta_{jk} \) is equal to 0.

At the 31K SNPs level, we performed the same GWAS using the WGS dataset. This GWAS at the sequence level referred to as GCTA-seq (see Table 1). In order to refine the position and size of the QTL detected by the analysis on the 31K SNPs level with a GRM constructed using all 31K SNPs included in the model, this analysis will be estimated genetic variance and

as for the marker-by-marker approach, a first GWAS using the BayesC\( \pi \) model was performed on the 31K SNPs level with a GRM constructed using all 31K SNPs included in the model, this analysis will be referred to as GCTA-chip (see Table 1). In order to refine the position and size of the QTL detected by the analysis at the 31K SNPs level, we performed the same GWAS using the WGS dataset. This GWAS at the sequence level was performed on the whole 8.7 million SNPs with a GRM constructed with 275K SNPs according to Yang et al.\(^27\). This GWAS will be referred to as GCTA-seq (Table 1).

The second GWAS approach was performed using the Bayesian Stochastic Search Variable Selection approach BayesC\( \pi \)\(^28\) implemented in the BESSiE software\(^29\) (version 1.0). In this BayesC\( \pi \) approach, at each iteration, a proportion \( \pi \) of SNP is assumed to have a non-zero-effect on the trait, thus at each iteration the number of SNP effects to be estimated is lower than the number of phenotypic records. The statistical model used is defined according to Eq. (2):

\[ y_i = \mu + T_i + a_j x_{ij} + u_i + \epsilon_i \]

\[ y_i = \mu + T_i + u_i + \sum_{j=1}^{n} \delta_{jk} a_j g_{ij} + \epsilon_{ik} \]

with \( y_i \) the observed phenotype for the \( i \)th individual (sex in 3 levels), \( \mu \) the overall mean in the population, \( T_i \) the fixed effect of the temperature treatment (2 levels), \( a_j \) the additive effect of the reference allele of the all the SNP (j) to be tested as fixed effect for sex association and \( x_{ij} \) the reference allele count (0, 1, or 2) for the SNP j for individual i, \( u_i \) is the random polygenic additive value of individual i and \( \epsilon_i \) the residual effect for individual i. The vector of residual effects is normally and independently distributed \( \epsilon \sim N(0, \Sigma\epsilon) \) with \( \sigma\epsilon^2 \) the residual variance. The random vector of polygenic effects follows a normal distribution \( u \sim N(0, G\Sigma_u) \) with \( \sigma_u^2 \) the estimated genetic variance and G a GRM constructed using only SNPs information. We first performed a GWAS on the 31K SNPs level with a GRM constructed using all 31K SNPs included in the model, this analysis will be referred to as GCTA-chip (see Table 1). In order to refine the position and size of the QTL detected by the analysis at the 31K SNPs level, we performed the same GWAS using the WGS dataset. This GWAS at the sequence level was performed on the whole 8.7 million SNPs with a GRM constructed with 275K SNPs according to Yang et al.\(^27\). This GWAS will be referred to as GCTA-seq (Table 1).

The second GWAS approach was performed using the Bayesian Stochastic Search Variable Selection approach BayesC\( \pi \)\(^28\) implemented in the BESSiE software\(^29\) (version 1.0). In this BayesC\( \pi \) approach, at each iteration, a proportion \( \pi \) of SNP is assumed to have a non-zero-effect on the trait, thus at each iteration the number of SNP effects to be estimated is lower than the number of phenotypic records. The statistical model used is defined according to Eq. (2):

\[ y_i = \mu + T_i + u_i + \sum_{j=1}^{n} \delta_{jk} a_j g_{ij} + \epsilon_{ik} \]

with \( y_i \) the observed phenotype for the \( i \)th individual (sex in 3 levels), \( \mu \) the overall mean in the population, \( T_i \) the fixed effect of the temperature treatment (2 levels), \( u_i \) the additive polygenic effect for individual i, \( a_j \) is the additive effect of the reference allele for the jth SNP with its genotype for individual i, \( g_{ij} \) coded as 0, 1 or 2 and \( n \) the number of SNPs in the analysis; \( \delta_{jk} \) is the indicator variable for the non-zero effect of the jth SNP at the kth iteration; \( \epsilon_{ik} \) the residual effect for the ith individual at the kth iteration. As in the previous model, the random vector of polygenic effects follows a normal distribution \( u \sim N(0, G\Sigma_u) \) with \( \sigma_u^2 \) the estimated genetic variance and G the GRM constructed using information from the 275K SNPs. The vector of residual effects is normally and independently distributed \( \epsilon \sim N(0, \Sigma\epsilon) \) with \( \sigma\epsilon^2 \) the residual variance.

At each cycle k, the decision to include the SNP j in the model depends on the indicator variable \( \delta_{jk} \) with the effect \( a_j \) of the SNP j estimated if \( \delta_{jk} \) is equal to 1, and not estimated if \( \delta_{jk} \) is equal to 0. This indicator variable\(^28\) is sampled from a binomial distribution with a probability \( \pi \) that \( \delta_{jk} \) is equal to 1 (the SNP as a non-zero-effect) and a probability \( 1 - \pi \) that \( \delta_{jk} \) is equal to 0.

As for the marker-by-marker approach, a first GWAS using the BayesC\( \pi \) model was performed on the 31K SNPs level with parameters determined in order to have 1% of the markers included in the model at each cycle. This analysis will be referred to as BC\( \pi \)-chip (Table 1). To refine the localisation of the QTL, the second GWAS was performed at the WGS level using only SNPs located on a portion of one chromosome (selected using the results of the GCTA-seq analysis) with the effect of all other SNPs (from this chromosome and the others) included in the polygenic effect calculated with the GRM based on 275K SNPs. For this analysis, which will be referred to as BC\( \pi \)-seq (Table 1), the proportion \( \pi \) of SNPs to be included at each cycle was constrained to 0.02% (i.e. 3 SNPs included at each iteration).

Both BayesC\( \pi \) analyses were performed with a MCMC algorithm. In total, 500,000 iterations were used for the BC\( \pi \)-chip and 1 million iterations were used for the BC\( \pi \)-seq, with, for both analyses, a burn-in period of 50,000 cycles and results saved every 20 cycles. In order to control the reproducibility of our analyses, both were run twice, with two different seeds for the random number generator to initialize MCMC algorithm. Convergence

| Name          | Software | Number of SNPs in the analysis | Number of SNPs in GRM | \( \pi \) |
|---------------|----------|---------------------------------|------------------------|--------|
| GCTA_chip     | GCTA     | 30,811                          | 30,811                 | -      |
| GCTA_seq      | GCTA     | 8,765,613                       | 275,283                | -      |
| BC\( \pi \)_chip | BESSiE   | 30,811                          | 275,283                | 1%     |
| BC\( \pi \)_seq | BESSiE   | 21,149                          | 275,283                | 0.02%  |

Table 1. Summary of GWAS analyses used to detect QTLs associated with spontaneous maleness in XX-rainbow trout. \( \pi \) is the proportion of SNPs with a non-zero-effect included at each cycle k of the MCMC algorithm. GRM genomic relationship matrix.
was assessed by visual inspection of trace plots of the SNP effects and variance components estimates for all different seeds.

**Estimation of genetic parameters.** Variance components and heritability were estimated using the average information of restricted maximum likelihood analysis (AI-REML) implemented in GCTA software\(^2^6\) with the model described in Eq. (1). The two GRMs built with 31K and 257K SNPs for GWAS were used to estimate the genetic parameters at the genome and the sequence level, respectively.

The values obtained with the 31K SNPs GRM were used as prior for both BayesC\(\pi\) models and the genetic variance \((\sigma^2_g)\) was calculated as the sum of the polygenic variance \((\sigma^2_a\), estimated by BESSiE) and the genomic variance \((\sigma^2_q\) estimated by \(n\) SNPs calculated according to Eq. (3):

\[
\sigma^2_g = \sum_{i=1}^{n} 2p_i q_i a_i^2
\]

with \(p_i\) and \(q_i\) the allele frequencies for the \(i\)th SNP and \(a_i\) the estimated additive effect of the \(i\)th SNP.

Genomic heritability (\(h^2_g\)) was estimated as \(h^2_g = \frac{\sigma^2_g}{\sigma^2_g + \sigma^2_a}\) with \(\sigma^2_a\) the estimated genetic variance and \(\sigma^2_g\) the residual variance.

**QTL definition.** For the GCTA-chip and GCTA-seq analysis, we determined chromosome-wide suggestive and genome-wide significance thresholds using a Bonferroni correction with \(\pi = 1\%\), i.e. genome-wide threshold = \(\log_{10}(\alpha/[n/30])\), with \(n\) the number of SNPs in the analysis (30,811 or 8.7 millions). Only SNPs with \(\log_{10}(P\text{-value})\) over the chromosome-wide threshold for GCTA-chip and over the genome-wide threshold for GCTA-seq were considered. For each QTL with a peak SNP value over the significance threshold, approximate confidence intervals (CIs) were estimated using a drop-off limit\(^3^6\) of 1.5 unit of \(\log_{10}(P\text{-value})\) and a maximum distance of 1 Mb for the GCTA-chip and 200 kb for the GCTA-seq between two successive SNPs, with \(\log_{10}(P\text{-value})\) over the drop-off limit starting from the position of the peak SNP.

In the GCTA-seq, when two successive QTLs were detected at a distance lower than 50 bp and with less than 250 kb between their peak SNPs, the CIs were cumulated in a credibility interval (see Supplementary Table S1 for detailed CIs).

For the Bayesian approaches, the degree of association between a SNP and the phenotype was assessed using the Bayes Factor\(^3^4\) (BF):

\[
BF = \frac{P_j/(1 - P_j)}{\pi/(1 - \pi)}
\]

with \(P_j\) the probability of the SNP \(j\) having a zero effect and \(\pi\) the proportion of SNPs having a non-zero effect on the trait (in our case \(\pi = 1\%\) for BC\(\pi\)-chip and \(\pi = 0.02\%\) for BC\(\pi\)-seq). BF was transformed into logBF computed as twice the natural logarithm in order to obtain values of the same usual range as the P-value, thus facilitating the visual appraisal of QTL and the comparison between methods\(^3^2\). In order to define QTL with the BayesC\(\pi\) approaches, we considered two categories to classify the strength of the evidence in favour of a QTL\(^3^3\): strong evidence for \(8 \leq \log_{10}(\text{BF}) < 10\) and very strong evidence for \(\log_{10}(\text{BF}) \geq 10\). Because the BF is not a statistic test, true confidence intervals cannot be derived, but credibility intervals can be built as defined in Michenet et al.\(^3^4\). Credibility intervals were determined using the threshold \(\log_{10}(\text{BF}) \geq 8\) for defining a peak SNP showing evidence for a QTL in either a BC\(\pi\)-chip or a BC\(\pi\)-seq analysis. For the BC\(\pi\)-chip approach, the credibility interval included all SNPs with a logBF \(> 3\) within a 1 Mb sliding window from the peak SNP. For the BC\(\pi\)-seq approach, it included all SNPs with a logBF \(> 5\) within a 200 kb sliding window from the peak SNP.

The proportion of the total genetic variance explained by each SNP was derived from the Bayesian analyses and calculated according to the ratio \(\frac{\text{logBF}}{\text{total variance}}\). The proportions of genetic variance explained by all the SNPs within the credibility interval of a QTL were cumulated to estimate the proportion of variance explained by the QTL.

**Candidate genes and SNP annotation.** Candidate genes located within a reduced interval determined as the intersection of the confidence and the credibility intervals of the main QTLs were listed from the NCBI *Oncorhynchus mykiss* Annotation Release 100 (GCF_002163495.1).

Annotation of SNPs within those intervals was performed using the SNPEff software\(^3^5\) (version 4.3) with the NCBI *Oncorhynchus mykiss* Annotation Release 100 (GCA_002163495.1)\(^3^2\) as a reference. SNPs were then filtered according to their estimated putative impact, all SNPs with a modifier putative impact were filtered out and only SNP with a low, moderate, or high putative impact were conserved.

**Pedigree and dams’ genotypes.** Genomic pedigree information was recovered using identity by descent (IBD) estimates from PLINK software\(^3^1\) (version 1.9) based on information from the 31K SNPs. The percentage of masculinised fish in the genotyped offspring of each dam was calculated (see Supplementary Table S2). The 50 dams used in the mating scheme were labelled according to the masculinisation rate in their genotyped offspring, with the AA-dam and the BX-dam having the higher and the lower proportions of masculinised offspring, respectively (Fig. 1). Among the 40 dams with at least 10 progeny recorded, 22 dams with extreme proportions of the different phenotypes in their genotyped progeny were selected: the 11 dams (labelled from BH to BR) that had less than 8% of masculinised offspring and the 11 dams (labelled from AA to AK) that had more than 35% of
masculinised offspring. Those 22 dams with contrasted masculinisation rates in their offspring were selected for in-depth analysis of their genotypes at the QTL regions.

**Results**

**Phenotyping, genotyping and pedigree.** In this experiment, we phenotyped 20,210 XX-rainbow trout from the French trout farm « Les Fils de Charles Murgat ». Among them 98.1% were females (Table 2) and only 1.4% of fish were partially (intersex) or fully masculinised. From the 294 masculinised individuals, 45.0% were intersex, with both female and male gonadal tissues. Among these intersex individuals, 60.6% had a right gonad completely masculinised or more masculinised than the left gonad. A significantly (p = 3.201e−12) higher male-ness (males + intersex) was observed in the group that was reared at 12° (2.0%) than in the group reared at 18° (0.9%) (Table 2).

From the 1152 genotyped XX-rainbow trout, 1139 fish genotyped for 30,811 SNPs were retained after quality controls. Pedigree was recovered for all of them but four. In average, the 50 dams had 22.7 genotyped offspring (min = 2, max = 63). The average percentage of masculinised fish in the genotyped offspring sample was 20.7%, ranging from 0% for eight dams to 71.4% for the AA-dam (see Fig. 1 and Supplementary Table S2). The 11 dams (labelled from dam AA to dam AK, Fig. 1) that had the highest proportion of masculinised offspring bred 370 fish and accounted for 68% of masculinised genotyped offspring. Conversely, the 11 dams with a low masculinised rate among their genotyped offspring sample (dam BH to dam BR, Fig. 1) accounted for only 5% of the masculinised genotyped offspring. Among those 22 dams, we detected two couples of related (IBD coefficients of 0.36) dams (AA-BK and AI-BR). In those two pairs, one "sister" produced a highly

---

**Figure 1.** Total number of genotyped offspring (females, males and intersex) for each of the 50 dams (AA to BX) used in the mating design. The 22 dams with more than 10 genotyped offspring and extreme masculinisation rate selected for in-depth analysis of their genotypes are identified by a black underlying line.

**Table 2.** Phenotyping and genotyping dataset from the XX-rainbow trout produced at “Les fils de Charles Murgat” farm. mpf months post-fertilisation. Total number of fish are highlighted in bold.
Based on the BCπ-chip model, we estimated that the genetic variance ($\sigma_g^2$) explained by the 31K SNPs was 4.3%, corresponding to 16% of the total genetic variance explained ($\sigma_g^2 + \sigma_e^2$) and 1.9% of the total genetic variance explained by the polygenic variance ($\sigma_p^2$) ($\sigma_g^2$/$\sigma_g^2 + \sigma_p^2$). The residual variance ($\sigma_e^2$) accounted for 0.4% of the total genetic variance ($\sigma_g^2 + \sigma_p^2 + \sigma_e^2$). Using the BCπ-seq model, we focused the analysis on the most relevant genomic regions and showed that the genetic variance explained ($\sigma_g^2$) increased with the number of SNPs (from 31K to WGS) included in the model and with the addition of a polygenic variance component (Table 3). The estimate values of genetic variance explained ($\sigma_g^2$) were 0.48 ± 0.04, 0.56 ± 0.05, 0.62 ± 0.06, and 0.59 ± 0.05 for GCTA-chip, BCπ-chip, GCTA-seq, and BCπ-seq respectively. The highest value (0.26) was obtained with the GCTA-seq analysis. The estimate values of genetic variance explained ($\sigma_g^2$) with either the 30,811 SNPs of the chip or the 8,784,147 SNPs from the WGS. The estimated heritability ($h^2$) was consistent across all models, with the lower value obtained for GCTA-chip ($h^2 = 0.48$ ± 0.03) and the highest value ($h^2 = 0.59$ ± 0.05) obtained with the BayesCπ-seq analysis. The estimate values of genetic variance explained ($\sigma_g^2$) increased with the number of SNPs (from 31K to WGS) included in the model and with the addition of a polygenic effect under a Bayesian approach, showing that the 31K SNPs genotyped were insufficient to capture all the genetic variation. Therefore, the actual heritability of maleness in rainbow trout is expected to be about 0.6.

### Table 3. Estimates of genetic and genomic parameters for spontaneous maleness under the different statistical models.

| Analysis          | $h^2$ ± SE | $\sigma_g^2$ ± SE | $\sigma_p^2$ ± SE | $\sigma_e^2$ ± SE | $\sigma_g^2$ ± SE | $\sigma_p^2$ ± SE |
|-------------------|-----------|------------------|------------------|------------------|------------------|------------------|
| GCTA-chip         | 0.48 ± 0.04 | 0.43 ± 0.02       | 0.22 ± 0.01       | 0.21 ± 0.03       | -                | -                |
| BCπ-chip          | 0.56 ± 0.05 | 0.43 ± 0.02       | 0.19 ± 0.02       | 0.24 ± 0.03       | 0.10 ± 0.03      | 0.14 ± 0.03      |
| GCTA-seq          | 0.62 ± 0.06 | 0.42 ± 0.02       | 0.16 ± 0.02       | 0.26 ± 0.03       | -                | -                |
| BCπ-seq*          | 0.59 ± 0.05 | 0.44 ± 0.02       | 0.17 ± 0.02       | 0.25 ± 0.03       | 0.21 ± 0.03      | 0.04 ± 0.01      |

**Estimation of genetic parameters for spontaneous maleness in XX-rainbow trout.** Table 3 presents the variance components and the heritability of the sexual phenotype estimated with the different genomic models using either the 30,811 SNPs of the chip or the 8,784,147 SNPs from the WGS. The estimated heritability of the sexual phenotype was high, ranging from 0.48 with the GCTA-chip analysis to 0.62 with the GCTA-seq analysis. The estimated genetic variance ($\sigma_g^2$) was consistent across all models, with the lower value obtained for GCTA-chip (0.21), intermediate values obtained with BayesCπ models (0.24 and 0.25 for BCπ-chip and BCπ-seq, respectively) and the highest value (0.26) obtained with the GCTA-seq analysis. The estimate values of genetic variance explained with the number of SNPs (from 31K to WGS) included in the model and with the addition of a polygenic effect under a Bayesian approach, showing that the 31K SNPs genotyped were insufficient to capture all the genetic variation. Therefore, the actual heritability of maleness in rainbow trout is expected to be about 0.6.

Based on the BCπ-chip model, we estimated that the genetic variance ($\sigma_g^2$) explained by the 31K SNPs was 4.3%, corresponding to 16% of the total genetic variance explained ($\sigma_g^2 + \sigma_e^2$) and 1.9% of the total genetic variance explained by the polygenic variance ($\sigma_p^2$) ($\sigma_g^2$/$\sigma_g^2 + \sigma_p^2$). The residual variance ($\sigma_e^2$) accounted for 0.4% of the total genetic variance ($\sigma_g^2 + \sigma_p^2 + \sigma_e^2$). Using the BCπ-seq model, we focused the analysis on the most relevant genomic regions and showed that the genetic variance explained ($\sigma_g^2$) increased with the number of SNPs (from 31K to WGS) included in the model and with the addition of a polygenic effect under a Bayesian approach, showing that the 31K SNPs genotyped were insufficient to capture all the genetic variation. Therefore, the actual heritability of maleness in rainbow trout is expected to be about 0.6.

### Table 4. Detection of QTLs associated with spontaneous maleness in XX-rainbow trout with GCTA-chip and BCπ-chip methods. The QTL start and end positions correspond to the confidence and credibility intervals for GCTA-chip and BCπ-chip respectively. For BCπ-chip, the logBF is calculated as twice the natural logarithm of the Bayes Factor. % $\sigma_g^2$: proportion of total genetic variance explained by the QTL calculated as the sum of the variance of the QTL estimated by the BCπ-chip analysis of all SNPs within the QTL credibility interval.

| Chromosome | QTL start (Mb) | QTL end (Mb) | Peak SNP name | Peak SNP position (Mb) | Peak − $\log_{10}(P)$ or logBF | % $\sigma_g^2$ explained by QTL | Method |
|------------|----------------|--------------|---------------|------------------------|---------------------------------|-------------------------------|--------|
| 1          | 62.93206       | 63.73809     | Affx-88953453 | 63.53900               | 7.9                             | –                             | GCTA-chip |
| 1          | 64.42011       | 64.63252     | Affx-88916383 | 64.42011               | 11.1                            | –                             | GCTA-chip |
| 1          | 63.45244       | 65.16399     | Affx-88950822 | 64.63252               | 11.3                            | 3.86%                         | BCπ-chip |
| 12         | 6.10355        | 6.79990      | Affx-88940013 | 6.79990                | 8.1                             | 0.62%                         | BCπ-chip |
| 20         | 31.35259       | 31.35289     | Affx-88916019 | 31.35259               | 8.2                             | 0.40%                         | BCπ-chip |

**Genome wide association studies for maleness in XX-female rainbow trout.** **GCTA-chip and BCπ-chip approaches at the 31K genotyping level.** Based on the 31K SNPs, we detected four QTLs associated with maleness on three different chromosomes (Omy1, 12 and 20) with the GCTA and the BayesCπ analyses (Table 4). With the GCTA-chip analysis, we detected two QTLs on Omy1, the first one being suggestive only and explaining less than 0.2% of the total genetic variance. The second QTL was significant at the genome-wide level, under GCTA-chip analysis ($-\log_{10}(P-value) = 11.1$) and had a very strong evidence under BCπ-chip analysis (logBF = 11.3) (Table 4). Even if the two peak SNPs of this QTL differed across the two GWAS approaches, the GCTA peak SNP being located 212 kb before the BCπ peak SNP, they were in close vicinity with only three markers in-between the two peak SNPs. This second QTL explained 3.9% of the total genetic variance. We did not distinguish the two QTL with the BCπ-chip approach, as the credibility interval estimated with BCπ-chip was 1.7 Mb and contained the two peak SNPs detected with GCTA-chip.

The two other QTLs located on Omy12 and Omy20 were detected using the BCπ-chip approach only (Table 4). The QTL located on Omy12 explained 0.6% of the total genetic variance. The QTL on Omy20, defined by a single SNP located at 31.352 Mb, explained 0.4% of the total genetic variance. This QTL was suggestive at 5% at the chromosome-wide level under GCTA-chip analysis ($-\log_{10}(P-value) = 4.2$).
Table 5. Summary of the characteristics of the two QTLs associated with maleness in XX-rainbow trout detected using GCTA-seq and BCπ-seq methods. For GCTA-seq, the start and stop positions of the first QTL correspond to a CI estimated with a drop-off method and the start and stop positions of the second QTL correspond to the credibility interval estimated as the union of confidence intervals of four QTLs. The BCπ-seq1 and BCπ-seq2 correspond to the same analysis run with two different seeds for MCMC initialization.

| Chromosome | QTL start (Mb) | QTL end (Mb) | QTL size (kb) | Peak SNP name | Peak SNP position (Mb) | Peak SNP – log10(P) or logBF | % σ^2_a explained by QTL | Method |
|------------|----------------|--------------|---------------|---------------|------------------------|-----------------------------|-------------------------|--------|
| 1          | 63.229379      | 63.74478     | 545.1         | Omy1-6349395  | 63.49395               | 10.61                       | –                       | GCTA-seq |
| 1          | 64.360291      | 64.707100    | 341.6         | Omy1-64597800 | 64.597800              | 13.74                       | –                       | GCTA-seq |
| 1          | 63.459553      | 63.55809     | 96.27         | Omy1-63520900 | 63.542090              | 12.97                       | 0.56%                   | BCπ-seq1 |
| 1          | 64.266321      | 64.649580    | 383.26        | Omy1-64607018 | 64.607018              | 14.67                       | 13.55%                  | BCπ-seq1 |
| 1          | 63.459545      | 63.727071    | 267.53        | Omy1-63493395 | 63.493395              | 11.55                       | 0.51%                   | BCπ-seq2 |
| 1          | 64.087947      | 64.632497    | 544.55        | Omy1-64610233 | 64.610233              | 17.32                       | 14.56%                  | BCπ-seq2 |

GCTA-seq and BCπ-seq approaches at the whole genome sequence level. Using the WGS data and GCTA-seq approach, significant QTLs were detected only on Omy1, in a restricted region (in-between 62 and 66 Mb). In total six peak SNPs corresponding to six putative QTLs (see Supplementary Table S1 and Supplementary Figure S2) were detected in this region. Therefore, the BCπ-seq approach was performed considering only the 21 K SNP spanning this 4 Mb-window on Omy1 in addition to a genome-wide polygenic component. Using this approach, we identified only two QTLs (Table 5 and Supplementary Figure S3). Out of the six putative QTLs detected by the GCTA-seq analysis, the first one was not confirmed with the BCπ-seq analyses. Therefore, it will not be further discussed, as it is likely that the association is only due to linkage disequilibrium with the SNPs included in the second QTL region. The second putative QTL detected between 63.229 and 63.774 Mb with the GCTA-seq analysis was also detected with the BCπ-seq analyses in a reduced credibility interval (between 63.459 and 63.556 Mb, Table 5); it was hereafter considered as the first true QTL on Omy1.

The four remaining putative QTLs from the GCTA-seq analysis were very close to each other, with less than 250 kb between two successive peak SNPs, and the limits of their confidence intervals (CIs) approximated by a drop-off approach were all distant from less than 3 kb. Therefore these four CIs were fused into a single credibility interval of 342 kb. This second QTL was also detected with the BCπ-seq analyses (Table 5). While the peak SNP of the first QTL was the same under the two BCπ-seq analyses, this was not the case for this second QTL. However, the two peak SNPs were separated by 3.2 kb only and their credibility intervals were strongly overlapping (Table 5).

Based on the average of the two BCπ-seq analyses, the first QTL explained about 0.5% and the second QTL about 14% of the total genetic variance.

At the end of the second QTL, we identified a haplotype block of 15 consecutive SNPs (spanning 745 kb from 64.632011 to 64.632756 Mb) that have a significant effect on masculinisation (all the 15 SNPs have − log10(P-value) between 12.2 and 13.6 and among them five SNPs with logBF > 9). Interestingly, for this 15 SNPs-haplotype, eight of the 11 dams with almost no masculinised offspring carried two copies of the reference genome haplotype (the three remaining dams being heterozygous at the 15 SNPs). Among the 11 dams with a high rate of masculinised offspring, only three were homozygous for the reference haplotype, six were heterozygous and the last two (labelled AA and A) in Fig. 1) were homozygous for the alternative haplotype. In the offspring, the homozygous genotype for the alternative haplotype was overrepresented in masculinised fish (51.5% of masculinised fish) and underrepresented in females (8.7% of females). Conversely, the reference haplotype was observed either at the heterozygous or homozygous state for respectively 50.6% and 40.2% of the female progeny.

Positional candidate genes and SNP annotation. Five genes were located within the first QTL region spanning from 63.459 to 63.556 Mb (Supplementary Table S3). The peak SNP from GCTA-seq (at 63.493 Mb) was located within the pygb gene (glycogen phosphorylase, brain form, from 63.487 to 63.508 Mb), and the peak SNP from both BCπ-seq runs (6.542 Mb) was located within the ninl gene (ninein-like protein, from 63.542 to 63.594 Mb). Among the 669 annotated SNPs spanning the QTL region, only 23 SNPs were indicated with either a low (16 SNPs) or a moderate (7 SNPs) potential effect on gene expressions when annotated with the SNPEff software (See Fig. 2 and Supplementary Table S4). The only SNP that was both significant (− log10(P-value) = 9.6) and annotated with a potential moderate effect was located at 63 543 061 bp within the ninl gene and annotated as a missense variant (Fig. 2). This SNP was very close to the peak SNP detected by the two BCπ-seq analyses (Table 5).

Within the second QTL region spanning from 64.360 to 64.633 Mb, 13 genes were detected (Supplementary Table S3). The three peak SNPs (Table 5) identified by the different WGS analyses were all located in the intergenic region between the hells gene (helicase, lymphoid specific) and an uncharacterized protein (LOC110527930, see Supplementary Table S3). Among the 1498 SNPs annotated in the QTL region, only 44 had a potential effect (moderate for 15 SNPs and low for 29 SNPs) on gene expression (Fig. 3 and Supplementary table S5). Among
Figure 2. Annotated SNPs located within the first QTL region from 63.549 to 63.556 Mb on Omy1. The significance of SNP effect (dots) is given by the value of $-\log_{10}(p$-value) estimated with the GCTA-seq method. The dark blue line corresponds to the 1% threshold at the genome wide level. Only SNPs with at least low putative effects on genes (estimated by SNPEff) are represented. The positions of the five genes located within the QTL region are figured by rectangles: banf1 (barrier-to-autointegration factor-like), pygb (glycogen phosphorylase, brain form), abhd12 (alpha/beta-Hydrolase domain containing 12), gins1 (DNA replication complex GINS protein PSF1), ninl (ninein-like protein).

Figure 3. Annotated SNPs located within the second QTL region, from 64.360 to 64.633 Mb on Omy1. The significance of SNP effect (dots) is given by the $-\log_{10}(p$-value) estimated with the GCTA-seq method. The dark blue line corresponds to the 1% threshold at the genome wide level. Only SNPs with at least low putative effects on genes (estimated by SNPEff) are represented. Genes located within this QTL are represented by rectangle: slc2a15a (solute carrier family 2, facilitated glucose transporter member 9-like), fgf8 (fibroblast growth factor 8), fbxw4 (F-box/WD repeat-containing protein 4), wbp1l (WW domain binding protein 1-like), cyp17a1 (Cytochrome P450 Family 17 Subfamily A Member 1), borc7 (BLOC-1-relat ed complex subunit 7), m5c2 (cytosolic purine 5'-nucleotidase), ina (alpha-internexin), pcf6 (polycomb group RING finger protein 6), zgc:175214 (RING finger protein 122-like), calhm2 (calcium homeostasis modulator protein 2), hel2 (lymphocyte-specific helicase-like), LOC110527930 (uncharacterized protein).
completely overlap the critical thermosensitive period of sex differentiation\(^36,37\), inducing a repression instead of
development (at 32 dpf, instead of 35 dpf), so the temperature treatment applied in our study might have not
ment (18 °C vs 12 °C). However, in their experiment, the temperature treatment was applied a little bit earlier dur-
ing development (at 32 dpf, instead of 35 dpf), so the temperature treatment applied in our study might have not
completely overlap the critical thermosensitive period of sex differentiation\(^36,37\), inducing a repression instead of
an induction of masculinisation. This hypothesis deserves further investigation to be confirmed. Different genetic
backgrounds could also contribute to the observed differences in response to thermal treatment, as the families
tested in Valdivia et al.\(^15\) (2014) belonged to the INRA XX-\(\text{mal}\) lineage, an experimental line different from
the commercial population used in this study. The hypothesis is further supported by the fact that distinct maleness
QTLs were detected in INRA XX-\(\text{mal}\) families\(^19\) and in this study (see below), suggesting that masculinising
factors may differ according to populations, resulting in differences in susceptibility to temperature. Indeed, a
range of sex ratios in response to temperature treatments have been reported in rainbow trout populations with
various genetic backgrounds\(^14,38\). In other fish species such as Atlantic silverside\(^39\) or Nile tilapia\(^40\), sex differentia-
tion is also known to depend both on the water temperature and on the genetic background of the population.

Among masculinised individuals, the extent of the gonad masculinisation was variable, with about 40 and
61% of intersex fish in the 12 °C and 18 °C groups, respectively. Within those intersex individuals, whatever
the rearing temperature at early stages, the right gonad was very often more masculinised than the left gonad.
This particularity was first described by Quillet et al. (2004)\(^41\) and confirmed by Valdivia et al. (2013, 2014)\(^12,15\),
in families originating from the INRA XX-\(\text{mal}\) carrying line. Under standard rearing conditions, the left–right
asymmetry (LR) is almost undetectable, as intersex fish remain rare and poorly described. However, the observa-

In this study, we reported high genomic heritability estimates for spontaneous maleness (from 0.48 up to
0.62). The QTL we detected explained only a small proportion (up to 14%) of the estimated genetic variance.
Using the BC\(\text{\text{\text{\text{-}}}}\)chip method, we estimated that the pan-genomic 31 K SNPs explained 58% of the genetic vari-
ance of maleness. The remaining 42% of the genetic variance were probably not captured by the SNPs due to
their heterogeneous density distribution across the genome as well as to lack of accuracy their effects (regressed
towards 0) with only 1139 phenotyped and genotyped individuals.

Concerning the identification of the genetic determinants of maleness, we detected four QTLs on three dif-
cerent chromosomes, two QTLs on Omy1, one on Omy12 and the last one on Omy20. Both QTLs on Omy12
and Omy20 were only detected in the 31 K SNPs analyses and were not confirmed with WGS analyses. The QTL
on Omy20 might be the result of a spurious association as it was defined by a single SNP in both BC\(\text{\text{\text{\text{-}}}}\)chip and
GCTA-chip analyses. It is worth noting that in their study in gynogenetic families from INRA XX-\(\text{mal}\) lineage, Guy-
omard et al.\(^17\) (2014) detected four QTLs associated with spontaneous maleness on four different chromosomes
and that one of them was located on Omy20 (linkage group RT17) too. However, the QTL had a wide confidence
interval, covering the entire chromosome, so it is not possible to say if the QTL we detected in the present study
on Omy20 is the same. As we did not detect QTL on the three other chromosomes reported in Guyomard et al.\(^17\)
(2014), maleness in rainbow trout with different genetic backgrounds is likely to be controlled by a diversity of
underlying genetic mechanisms.

The two main QTLs associated with masculinisation in our population were both detected on Omy1, which has
never been reported as associated with neither sex determinism nor maleness in rainbow trout. Using WGS infor-
mation, we were able to reduce the confidence interval of the first QTL from 806 kb (GCTA-chip, Table 4) down
to 545 kb (GCTA-seq analysis) and even 96 kb for the credibility interval estimated with one of the BC\(\text{\text{\text{\text{-}}}}\)seq runs
(Table 5). While highly significant, this QTL explained only 0.5% of the total genetic variance of the trait. This low
proportion of variance explained may be the result of a strong linkage disequilibrium of those SNPs with some SNPs
of the second QTL, which was mapped at a close distance (< 1.2 Mb between the peak SNPs of the two QTLs), making
the first QTL an artefact. Within this first QTL region, no gene seems to have a functional role in gonad develop-
ment; however, this QTL may contain regulating factors involved in the control of genes contained in the second QTL.
For the second QTL, the credibility interval ranged from 268 to 545 kb depending on the WGS analyses (Table 5 and Supplementary Table S1). Based on the BCπ-seq analysis, we estimated that this second QTL explained about 14% of the total genetic variance while its contribution was reduced to less than 4% with the BCπ-chip analysis. It is difficult to know whether the proportion of variance explained was overestimated under the BCπ-seq model or underestimated under the BCπ-chip model because we ignored the covariance between close SNPs estimates when deriving this proportion. Nevertheless, this QTL explains a sufficient part of the genetic variance to deserve specific attention in breeding programs.

Within this main QTL region, three candidate genes possibly involved in spontaneous masculinisation were identified. First, the fgf8a gene might be involved in the left–right (LR) asymmetrical gonad development observed in the intersex fish. This gene is part of the fibroblast growth factor family which plays a major role in vertebrate development and has been reported to be fundamental for proper LR asymmetric development of multiples organs (brain, heart and gut) in zebrafish. The second and biologically more convincing candidate gene is cyp17a1 as it is a major enzyme of the steroidogenic pathway and it is involved in the synthesis of biologically active gonadal androgens and oestrogens. Indeed, cyp17a1 has been identified as potentially involved in spontaneous maleness in common carp (Cyprinus carpio) and was recently characterized as being involved in the gonad differentiation in zebrafish with complete masculinisation phenotype in cyp17a1 deficient zebrafish. However, in our analysis, the only SNP with a putative missense effect for this gene was significant in the GCTA-seq analysis only (−log10(P-value) = 10.2) and had very low logBF values in both BCπ-seq approach (0 or 2.4).

However, even if the cyp17a1 gene was not within the most significant region of the QTL (located in-between the helfs gene and LOC110527930) in terms of statistical significance/evidence tests, we cannot exclude its implication, for instance through long distance regulation of its expression. Based on the GWAS results as well as the SNP annotation, the uncharacterized LOC110527930 protein was the most relevant candidate found in our analysis. Within this gene, we identified a haplotype block of 15 consecutive SNPs (from Omy1-64632011 to Omy1-64632756) that were all significant with the GCTA-seq analysis (−log10(P-value) > 12) and presented alternative genotypes for four dams with opposite proportions of masculinised progeny (2 dams with highly masculinised offspring and 2 dams with low masculinised offspring). In addition, seven of those 15 SNPs were annotated with a putative missense effect on the LOC110527930 protein. However, there is no published evidence that this uncharacterized protein could play a role in gonadal sex differentiation and more work is still needed to characterise the protein and its potential role in gonadal differentiation.

The high heritability (up to 0.62) of spontaneous maleness estimated in this study opens up opportunities to manage maleness in all-female trout populations. If the two highly significant QTLs detected on Omy1 in the present population were confirmed to play a role in spontaneous maleness in other rainbow trout populations with diverse genetic origins, the identified SNPs could then be used in a cost-efficient genotyping test to identify female broodstock with a higher propensity to transmit male or intersex phenotype in their progeny and help breeders willing to limit the occurrence of undesirable masculinised individuals in their commercial all-female stocks to discard those breeders from reproduction. In order to optimize the efficiency of such a test, i.e. maximising the detection of masculinised fish while keeping low the number of true females, we tested various combinations of SNPs located within the two QTLs identified on Omy1. Within the main QTL (from 64.360 to 64.707 Mb), all significant SNPs would give similar yields, any SNP homozygous for the alternative allele would allow the detection of 44.7 to 51.5% of masculinised fish in our sample, whereas only 6.8 to 13.7% of the homozygous fish for the alternative allele would be females. In particular, using the alternative allele of any SNP from the haplotype block of 15 SNPs described previously (between 64,632,011 and 64,632,756 bp) would allow detecting 51.3% of masculinised fish in our sample. Among those 15 SNPs, the SNP Affx-88950822 is present on the commercial 57K chip and could be of immediate use, provided its effect is confirmed in other trout populations. Every combination of two or more SNPs within this main QTL would yield in a smaller proportion of females discarded but also in a slightly smaller proportion of males identified. Always considering our sample, pairing the SNP Affx-88950822 (or any other SNP from the associated 15-SNPs haploblock) with SNPs from the first QTL on Omy1 (from 63,459 to 63,556 Mb) would slightly increase the test sensitivity (identifying more masculinised fish) without eliminating too many true females. Before applying the same test in other populations, the effect of those SNPs should be confirmed in a large set of populations with diverse genetic backgrounds, or at least, in the target population. Our preliminary results regarding maleness QTLs in either the XX-mal families from the INRA experimental population and the population used in this study suggest that this might not be the case.

From a breeder perspective, additional research is needed to improve knowledge about the genetic basis and the environmental factors determining spontaneous maleness in all-female stocks before any industrial application. In addition to the confirmation that QTLs are present in other rainbow trout populations, the expected genetic and phenotypic responses under different thermal regimes and either pedigree-based or marker-assisted selection should be quantified to assess the efficiency of a strategy aiming at limiting the rate of spontaneous males in all-female stocks. Further investigation is also needed with regard to the production of XX sex-reversed male breeders based on the use of spontaneous sexual inversion to prevent the use of hormones. Indeed, the high heritability of maleness suggests that the use of spontaneously masculinised individuals as progenitors of all-female populations would increase the frequency of undesirable masculinised progeny. Combining masculinising genetic factors together with an environmental (temperature) control of gonad masculinisation according to the destination of the fish (broodstock vs all-female production stock) and/or the rearing environment might offer a solution to manage the trade-off. However, because of the overall low masculinisation rates recorded in this study, we were not able to detect any potential interaction between rearing temperature and genotype. We have no suspicion of the existence of such an interaction, as a GWAS with GCTA-chip, carried out with only the individuals from the 12 °C group, detected the same QTL as for the overall population (results not shown). In the current state of knowledge on the effect of temperature on sexual differentiation, it is too early to propose an efficient management of the rearing environment to either enhance or limit spontaneous maleness in rainbow trout.
Although the prospect of using hormone-free XX-neomales is very attractive, a potential important drawback may be to fall back into the fish quality and animal welfare issues (flesh loss of lipids, color and firmness due to precocious maturation; mortality due to Sparocephalus fungus) encountered in commercial stocks with XY-males.

In conclusion, in this study we detected minor genetic factors involved in maleness in rainbow trout in the absence of the master gene sdY. Two QTLs detected on Osmyl explained up to 15% of the total genetic variance of maleness in the population used in this study and we identified three candidate genes that might be involved in the masculinisation of XX-rainbow trout. Among those three genes, one uncharacterised protein (LOC110527930) was the most relevant candidate and more work would be needed to characterise the potential role of this protein on gonadal differentiation.

**Data availability**

The datasets for this manuscript are not publicly available because data belongs partly to a private company. The data can be made available for reproduction of the results from Florence Phocas and Charles Murgat Pisciculture on request via a material transfer agreement.

Received: 30 April 2020; Accepted: 6 October 2020

**References**

1. Devlin, R. H. & Nagahama, Y. Sex determination and sex differentiation in fish: An overview of genetic, physiological, and environmental influences. *Aquaculture* **208**, 191–364 (2002).

2. Martínez, P. et al. Genetic architecture of sex determination in fish: Applications to sex ratio control in aquaculture. *Front. Genet.* **5** (2014).

3. Manl, J. E. & Avice, J. C. Evolutionary diversity and turn-over of sex determination in teleost fishes. *Sex. Dev. Genet. Mol. Biol. Evol. Endocrinol. Embryol. Pathol. Sex Determin. Differ.* **3**, 60–67 (2009).

4. Johnstone, R., Simpson, T. H., Youngson, A. F. & Whitehead, C. Sex reversal in salmonid culture: Part II. The progeny of sex-reversed rainbow trout. *Aquaculture* **18**, 13–19 (1979).

5. Chevassus, B., Devaux, A., Chourront, D. & Jalabert, B. Production of YY rainbow trout males by self-fertilization of induced hermaphrodites. *J. Hered.* **79**, 89–92 (1988).

6. Chourront, D. & Quillet, E. Induced gynogenesis in the rainbow trout: Sex and survival of progenies production of all-triploid populations. *Theor. Appl. Genet.* **63**, 201–205 (1982).

7. Yano, A. et al. An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss*. *Curr. Biol.* **22**, 1423–1428 (2012).

8. Yano, A. et al. The sexually dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. *Evol. Appl.* **6**, 486–496 (2012).

9. Bertho, S. et al. Fosf2 and its relatives are evolutionary conserved partners in gonadal sex determination. *Sex. Dev. Genet. Mol. Biol. Evol. Endocrinol. Embryol. Pathol. Sex Determin. Differ.* **10**, 111–129 (2016).

10. Bertho, S. et al. The unusual rainbow trout sex determination gene hijacked the canonical vertebrate gonadal differentiation pathway. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 12781–12786 (2018).

11. Quillet, E., Aubard, G. & Queau, I. Mutation in a sex-determining gene in rainbow trout: detection and genetic analysis. *J. Hered.* **93**, 91–99 (2002).

12. Valdivia, K. et al. Sex Differentiation in an all-female (XX) rainbow trout population with a genetically governed masculinization phenotype. *Sex. Dev.* **7**, 196–206 (2013).

13. Guyomard, R. et al. RAD-seq mapping of spontaneous masculinization in XX doubled haploid rainbow trout lines (2014).

14. Mayerhans, A., Müller-Belecke, A. & Hörstgen-Schwark, G. Effect of rearing temperatures post hatching on sex ratios of rainbow trout (*Oncorhynchus mykiss*) populations. *Aquaculture* **294**, 25–29 (2009).

15. Valdivia, K. et al. High temperature increases the masculinization rate of the all-female (XX) rainbow trout “mal” population. *PLoS ONE* **9**, e113355 (2014).

16. Ospina-Álvarez, N. & Pifferer, F. Temperature-dependent sex determination in fish revisited: Prevalence, a single sex ratio response pattern, and possible effects of climate change. *PLoS ONE* **3**, e2837 (2008).

17. Baroller, J. F., D’Cotta, H. & Saillant, E. Environmental influences on fish sex determination and differentiation. *Sex. Dev. Genet. Mol. Biol. Evol. Endocrinol. Embryol. Pathol. Sex Determin. Differ.* **10**, 111–129 (2016).

18. Jalabert, B., Billard, R., Chevassus, B., Escaille, A. & Carpentier, M. Preliminary experiments on sex control in trout: production of sterile fishes ans dimultidaneous self-fertilizable hermaphrodites. *Annu. Biol. Anim. Biochim. Biophys.* **15**, 19–28 (1975).

19. Palti, Y. et al. The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. *Mol. Ecol. Resour.* **15**, 662–672 (2015).

20. D’Ambrosio, J. et al. Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding. *Genet. Sel. Evol.* **51**, 26 (2019).

21. Sargolzaei, M., Chesnais, J. P. & Schenkel, F. S. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* **15** (2014).

22. Pearse, D. E. et al. Sex-dependent dominance maintains migration supergene in rainbow trout. *Nat. Ecol. Evol.* **3**, 1731–1742 (2019).

23. Gao, G. et al. A new single nucleotide polymorphism database for rainbow trout generated through whole genome resequencing. *Front. Genet.* **9** (2018).

24. Daneczek, P. et al. The variant call format and VCFTools. *Bioinformatics* **27**, 2156–2158 (2011).

25. Chang, C. C. et al. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).

26. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

27. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).

28. Habier, D., Fernando, R. L., Kizilkaya, K. & Garrick, D. J. Extension of the bayesian alphabet for genomic selection. *BMC Bioinform.* **12**, 186 (2011).

29. Boerner, V. & Tier, B. BESSiE: A software for linear model BLUP and Bayesian MCMC analysis of large-scale genomic data. *Genet. Sel. Evol.* **48**, 63 (2016).

30. Li, H. A quick method to calculate QTL confidence interval. *J. Genet.* **90**, 355–360 (2011).

31. Schurink, A., Jans, L. L. & Heuven, H. C. Bayesian variable selection to identify QTL affecting a simulated quantitative trait. *BMC Proc.* **6**, S8 (2012).

32. Kass, R. E. & Raftery, A. E. Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
33. Vidal, O. et al. Identification of carcass and meat quality quantitative trait loci in a Landrace pig population selected for growth and leanness. *J. Anim. Sci.* **83**, 293–300 (2005).
34. Michenet, A., Barbat, M., Saintilan, R., Venot, E. & Phocas, F. Detection of quantitative trait loci for maternal traits using high-density genotypes of Blonde d'Aquitaine beef cattle. *BMC Genet.* **17**, 88 (2016).
35. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnPEff: SNPs in the genome of Drosophila melanogaster strain w 118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
36. Azuma, T. et al. The influence of temperature on sex determination in sockeye salmon *Oncorhynchus nerka*. *Aquaculture* **234**, 461–473 (2004).
37. Craig, I. K., Foote, C. J. & Wood, C. C. Evidence for temperature-dependent sex determination in sockeye salmon (*Oncorhynchus nerka*). *Can. J. Fish. Aquat. Sci.* **53**, 141–147 (1996).
38. Magerhans, A. & Hörsten-Schwark, G. Selection experiments to alter the sex ratio in rainbow trout (*Oncorhynchus mykiss*) by means of temperature treatment. *Aquaculture* **306**, 63–67 (2010).
39. Conover, D. O. & Kynard, B. E. Environmental sex determination: Interaction of temperature and genotype in a fish. *Science* **213**, 577–579 (1981).
40. Baroller, J. F., D’Cotta, H., Bezault, E., Wessels, S. & Hörsten-Schwark, G. Tilapia sex determination: Where temperature and genetics meet. *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* **153**, 30–38 (2009).
41. Quillet, E., Labbe, L. & Queau, I. Asymmetry in sexual development of gonads in intersex rainbow trout. *J. Fish Biol.* **64**, 1147–1151 (2004).
42. Yu, Z. H. Asymmetrical testicular weight in mammals, birds, reptiles and amphibia. *Int. J. Androl.* **21**, 53–55 (1998).
43. Takahashi, E. et al. Migration behavior of PGCs and asymmetrical gonad formation in pond smelt *Hypomesus nipponensis*. *Int. J. Dev. Biol.* **61**, 397–405 (2017).
44. Hayakawa, Y. & Kohbayashi, M. Histological observations of early gonadal development to form asymmetrically in the dwarf gourami *Colisa lalia*. *Zool. Sci.* **29**, 807–814 (2012).
45. Strussmann, C. A. & Ito, L. S. Where does gonadal sex differentiation begin? Gradient of histological sex differentiation in the gonads of pejerrey, *Odontesthes bonariensis* (Pisces, Atherinidae). *J. Morphol.* **265**, 190–196 (2005).
46. Albertson, R. C. & Yelick, P. C. Roles for fgf8 signaling in left–right patterning of the visceral organs and craniofacial skeleton. *Dev. Biol.* **283**, 310–321 (2005).
47. Ruane, N. M., Lambert, J. G. D., Goos, H. J. T. & Komen, J. Hypocorticism and interrenal hyperplasia are not directly related to masculinization in XX mas−1/mas−1 carp. *Cyprinus carpio. Gen. Comp. Endocrinol.* **143**, 66–74 (2005).
48. Zhai, G. et al. Characterization of sexual trait development in cyp17a1-deficient zebrafish. *Endocrinology* **159**, 3549–3562 (2018).

**Acknowledgements**

The authors are grateful to R. Murgat and the staff of “Les Fils de Charles Murgat” farm for their help and care of the fish. The authors thank C. Blay, J. D’Ambrosio and F. Enez (SYSAAF), J. Bobe, E. Cardonna, A. Fostier, L. Les Fils de Charles Murgat and the staff of “farm for their help and care of the fish. The authors thank C. Blay, J. D’Ambrosio and F. Enez (SYSAAF), J. Bobe, E. Cardonna, A. Fostier, L.

**Author contributions**

C.F. performed the data curation, interpretation and formal analysis, prepared all figures and wrote the original draft. E.P. was involved in the conceptualisation and supervision of the project, the development of statistical methodology and data interpretation and wrote the original draft. A.B., C.P., P.H., N.D., C.C., Y.G. and E.Q. were involved in animal sampling and phenotyping. F.K., N.D., C.C., M.M, J.L. and O.B. were involved in laboratory works. M.C., M.B, C.H. and F.P. were involved in the data curation, and helped with programming and software utilisation. A.B., C.P., P.H., P.H. and Y.G. were involved in the conceptualisation of the project, in funding acquisition and in the design of the protocol. E.Q. was project coordinator involved in the conceptualisation and coordination of the whole study, funding acquisition, design of the protocol, data interpretation and revision of the original draft. All authors reviewed and approved the manuscript.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-74757-8.

Correspondence and requests for materials should be addressed to C.F. or E.P.

Reprints and permissions information is available at www.nature.com/reprints.

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

© The Author(s) 2020