BMC Genomics

Open Access

A high-density linkage map and sex-linked markers for the Amazon Tambaqui *Colossoma macropomum*

Eduardo Sousa Varela[1*], Michaël Bekaert[2], Luciana Nakaghi Ganeco-Kirschnik[1], Lucas Simon Torati[1], Luciana Shiotsuki[1], Fernanda Loureiro de Almeida[3], Luciana Cristine Vasques Villela[1], Fabrício Pereira Rezende[1], Aurisan da Silva Barroso[1], Luiz Eduardo Lima de Freitas[1], John Bernard Taggart[2] and Herve Migaud[2]

Abstract

Background: Tambaqui (*Colossoma macropomum*, Cuvier, 1818) is the most economically important native freshwater fish species in Brazil. It can reach a total length of over 1 m and a weight of over 40 kg. The species displays a clear sex dimorphism in growth performance, with females reaching larger sizes at harvest. In aquaculture, the production of monosex populations in selective breeding programmes has been therefore identified as a key priority.

Results: In the present study, a genetic linkage map was generated by double digest restriction-site associated DNA (ddRAD) sequencing from 248 individuals sampled from two F1 families. The map was constructed using 14,805 informative SNPs and spanned 27 linkage groups. From this, the tambaqui draft genome was improved, by ordering the scaffolds into chromosomes, and sex-linked markers were identified. A total of 235 markers on linkage group 26 showed a significant association with the phenotypic sex, supporting an XX/XY sex determination system in the species. The four most informative sex-linked markers were validated on another 206 sexed individuals, demonstrating an accuracy in predicting sex ranging from 90.0 to 96.7%.

Conclusions: The genetic mapping and novel sex-linked DNA markers identified and validated offer new tools for rapid progeny sexing, thus supporting the development of monosex female production in the industry while also supporting breeding programmes of the species.

Keywords: Linkage map, Sex-linked markers, Monosex, Sex determination, *Colossoma macropomum*, Aquaculture

Background

Over the past decade, the Brazilian aquaculture industry has undergone great changes, with expansion in the production of native species at a yearly growth rate of 22% mainly to supply regional markets [1, 2]. The high competitiveness of globally produced aquaculture species has driven technological innovations in the Brazilian

aquaculture sector to reduce production costs, increase productivity, and diversify fish products to meet market demands [2, 3]. The tambaqui *Colossoma macropomum* (Cuvier, 1818), also known as the "black pacu", has been identified as the most promising native species for the expansion of aquaculture in Brazil [4], with current production exceeding 139,000 t across approximately 4000 production units. The species accounts for more than 29% of fish aquaculture in Brazil [1], with an estimated increase in production of 68% by 2021 [5]. Alternatively, *C. macropomum* and its hybrids tambacu (female *C.*

* Correspondence: eduardo.varela@embrapa.br
[1]Embrapa Pesca e Aquicultura, Prolongamento da Av. NS 10, Cruzamento com AV. LO 18, Sentido Norte, loteamento Água Fria, CEP, Palmas, TO 77008-900, Brazil
Full list of author information is available at the end of the article

macropomum × male *Piaractus mesopotamicus*) and tambatinga (female *C. macropomum* × male *P. brachypomus*) are also produced in Brazilian aquaculture [1]. *C. macropomum* has been selected as the main candidate native species for genetic improvement programmes due to its many attributes for aquaculture, including a high market value, relatively easy and well controlled reproduction in captivity, excellent growth potential, good resilience to different rearing systems and an omnivorous diet [2–4].

C. macropomum growth and morphometric traits have great potential for selection since moderate to high heritability has been reported with genetic gains ranging from 8 to 31% [6–8]. However, the rate of genetic gain for body size is reduced in mixed sex populations due to a strong sex dimorphism in growth in favour of females. After sexual maturation, females can be 16% heavier that males of the same age [8, 9]. Importantly, differences in body weight between genders are not observed during the grow-out period (1.5–2.0 kg) when animals are usually phenotyped for selective breeding. For this reason, heritability and selection indices may be overestimated when gender is not accounted for [8].

Sex dimorphism, especially in growth rate, has been reported in many aquaculture finfish species with females usually reaching bigger sizes than males, including Atlantic halibut, *Hippoglossus hippoglossus* [10], Atlantic sea bass, *Dicentrarchus labrax* [11] and Japanese flounder, *Paralichthys olivaceus* [12] but also in crustaceans species like the giant freshwater prawn, *Macrobrachium rosenbergii* [13] and swimming crab, *Portunus trituberculatus* [14]. Therefore, the production of all-female stocks would significantly improve the productivity and profitability of these species [13, 15, 16]. Indirect hormonal manipulations are usually effective at producing all-female populations in species with a heterozygous sex determination system with females as the homogametic sex (i.e., XX/XY female/male [15];). Previous study has identified $2n = 54$ chromosomes in tambaqui and their hybrids by cytogenetic techniques, however any heteromorphism has been observed for sex chromosomes [17]. In *C. macropomum*, the sex determination system remains unknown, but a basic protocol for monosex female production through direct feminisation using 17 β-oestradiol has been published recently [18]. However, direct sex reversal is banned in many countries across the globe due to concerns for workers on farms handling hormones, discharge into the environment and food safety for consumers [19]. Indirect sex reversal protocols can be developed if tambaqui sex determination is proven to be heterozygous and the identification and validation of sex markers would fast track progeny testing and implementation in the industry.

Linkage mapping is critical for identifying the location of regions related to quantitative traits, such as those involved in disease resistance, growth, and sex determination. Previous study has obtained a large-scale SNP discovery to build a high-density linkage map in tambaqui (2811 cM), but they have not involved quantitative traits [20]. Marker-assisted selection and genomic selection using genetic markers linked to specific quantitative trait locus (QTL) affecting a trait of commercial interest have great potential to improve selection accuracy and accelerate genetic gain through selective breeding in many aquaculture species [21–23]. Marker-assisted selection and genomic selection have already been applied successfully to several aquaculture species for traits such as sexual maturity and disease resistance [23, 24]. Sex markers can be used as a tool to identify sex during the grow-out stage within selective breeding programmes and improve selection accuracy and genetic gains.

In the present study, double digest restriction-site associated DNA (ddRAD) (ddRAD) sequencing was employed to construct a high-density genetic linkage map for association analysis of sex-linked QTL in tambaqui. Sex-linked markers were identified and validated by fluorescent based, allele specific PCR technology, to provide a tool for future development of monosex aquaculture and selective breeding programmes.

Results

Genome survey summary and markers assembly

High throughput sequencing of the 248 individuals from two families produced 933,251,367 raw paired-end reads in total. Reads were deposited at the EBI European Nucleotide Archive (ENA) project PRJEB33856. After removing low-quality reads and demultiplexing, 68.84% of the total reads were retained. The sequences were aligned with the *C. macropomum* genome scaffolds and genotypes for all samples were obtained using Stacks software, yielding 6,055,367 unique loci. Mean coverage per locus was 81.9x (Min: 4.4x, Max: 423.7x). Three samples (F01_143, F01_072 and F01_045) with a very low coverage and very high rate of missing sites were excluded from further analysis (Supplementary Table S1). Heterozygosity of the markers was low, as only 1.4% of the shared loci were polymorphic. A total of 19,293 polymorphic loci which were identified in the parents and at least 75% of the progeny were subsequently used to construct genetic linkage maps and perform a sex association analysis (Table 1).

Linkage map construction

In total, 14,805 informative SNPs were mapped to the expected 27 linkage groups, from the karyotyping [25], by using LepMap3 with a threshold logarithm of the odds (LOD) value of 11. Sex-averaged, female, and male

Table 1 Sequencing summary

Category	Number/length
Total number of reads	1,866,502,734
Total number of bases	279,975,410,100 nt
Read length	150 nt
Total number of filtered reads	1,285,083,893
Unique RAD tags	6,055,367
Polymorphic SNP markers	19,293
Informative SNP markers	14,805
Genetic map (*averaged sex*): loci	5459
Average-sex size	2751.81 cM

genetic linkage maps were constructed (Fig. 1, Supplementary Fig. S1). The sex-averaged map was spanning a total distance of 2752 cM with an average inter-locus distance of 0.51 cM (Table 2). The number of markers in a linkage group varied from 395 (LG 27) to 917 (LG 1) with an average of 548, and the genetic length per group ranged from 75.09 cM (LG26) to 130.20 cM (LG 1) with an average of 101.92 cM. The female map comprised 3147 loci and spanned 2733 cM, with an average loci interval of 0.90 cM; the male map consisted of 3022 loci and spanned 2925 cM with an average loci interval of 0.98 cM (Table 2 and Supplementary Table S2). The ordering and orientation of the *C. macropomum* genome scaffolds to reconstruct chromosomes (Supplementary Data S1) were performed using the linkage maps (sex average).

Association analysis and QTL mapping

R/SNPassoc software was used to conduct a quantitative trait locus (QTL) mapping analysis for sex determination association. QTL fine mapping based on high-density genetic linkage maps provided evidence for the existence of a major QTL in LG 26 for both families. The result for genome-wide significant QTL was identified on LG 26 (Fig. 2). From the 415 markers on LG 26, a total of 239 (57.6%) were strongly associated with sex with $P < 10^{-6}$. After Bonferroni correction for multiple tests, the significant LOD score threshold was 5.46. The highest LOD values (over 45) from 27 markers were observed in a region ranging from 15.9 cM (LOD = 48.6, scaffold NW_023494809.1:1051596) and 33.6 cM (LOD = 46.7, scaffold NW_023494809.1:15347236), representing an interval of 17.7 cM or 14,295,640 bp (based on alignment of the markers on the final genome).

Verification of SNP sex-association and validation

Primers for PACE assays were designed for the four potentially most associated sex-linked markers (Table 3 and Supplementary Table S3). All four markers were assessed in 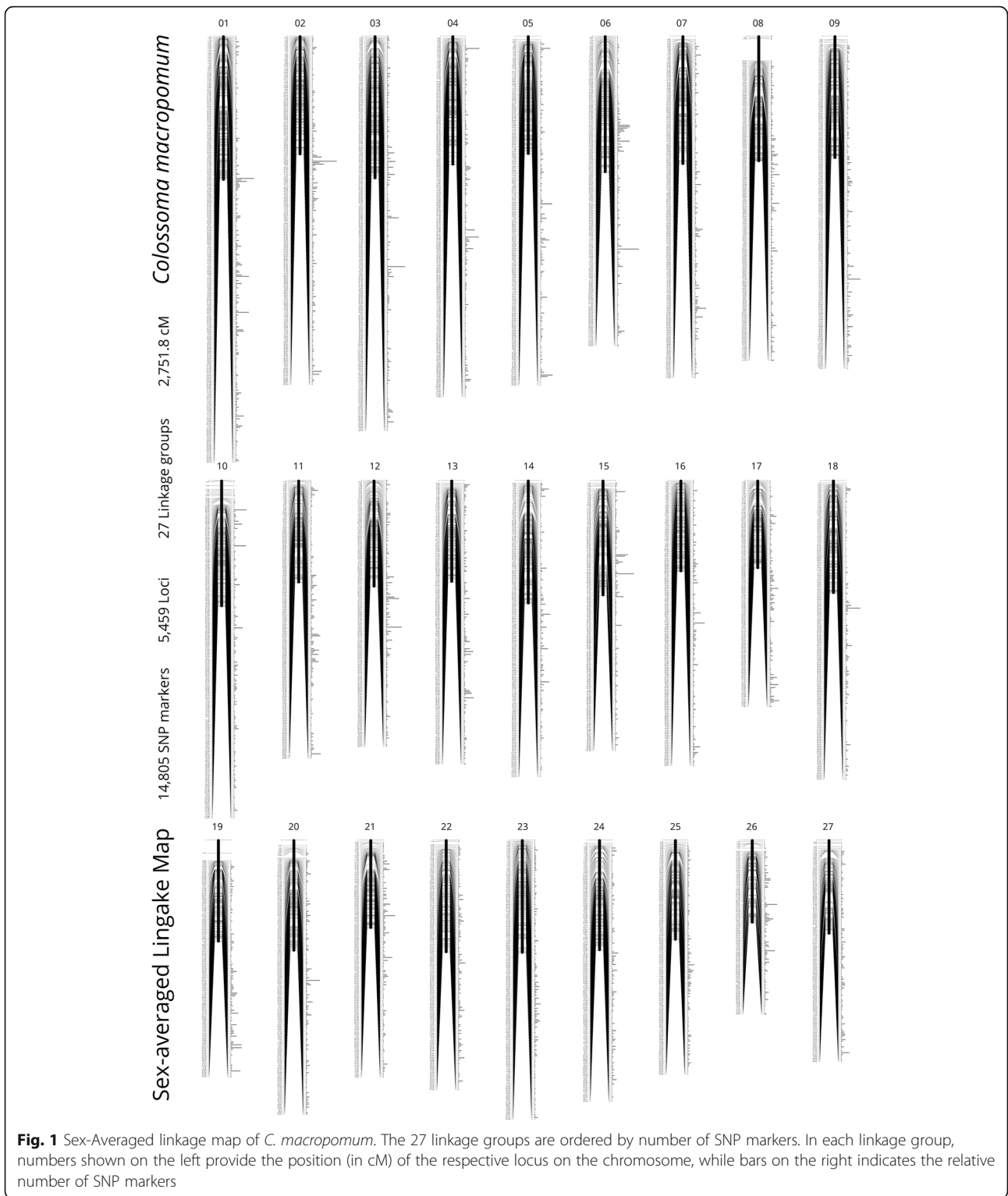extra samples collected for this validation step (22 broodstock and their 184 F1 progeny) using PACE SNP genotyping (Supplementary Table S4). The markers Cma511969, Cma5145911, Cma5119452 and Cma5117879 showed between 90.0 and 96.7% association to the phenotypic sex (Table 4).

Discussion

In this study, we identified 14,805 informative SNPs and constructed linkage maps for *C. macropomum*, which represents the highest density genetic map so far generated for this species. The approach enabled to map sex-associated region on a single chromosome (LG 26) supporting an XX/XY sex determination system in tambaqui. A panel of four sex related markers was successfully validated in a wider population for use in future selection and monosexing within breeding programme in the species.

The construction of high-density, informative genetic linkage maps is an essential pre-requisite for resolving QTLs associated to traits of interest in aquaculture species. In recent years simultaneous discovery and genotyping of SNP loci by RAD-based technologies has been widely exploited for linkage and QTL mapping due to its relative simplicity (not requiring any prior genomic reference), multiplexing capacity, flexibility and cost efficiency [26]. One variant, ddRAD, has been deployed successfully across a range of aquaculture species to study population structure, fish sex determination and identify sex linked genetic markers for sex genotyping assays [10, 14, 27–30]. In the present study, we developed a high-density genetic linkage map by ddRAD sequencing in tambaqui, identified a strong sex associated QTL in LG 26, provided data to support an XX/XY female/male heterozygous sex determination system and validated a panel of SNPs to assign sex in future studies.

Three genetic linkage maps were constructed for tambaqui covering a total distance of 2752 cM, 2733 cM and 2925 cM for sex-averaged, female and male maps, respectively. No significant differences were observed between maps. The total number of markers in the consensus map was 14,805 SNPs, covering 5459 loci, and the average loci interval of the sex-averaged map was 0.51 cM. These results agree with a previous linkage map published for tambaqui based on a genotyping-by-sequencing approach made without accounting for sex (2811 cM of total distance coverage, 7192 mapped SNPs with a 0.39 cM interval) and without QTL mapping analyses [20]. Heterozygosity recorded was only of 1.4%. However, this is likely not representative of the species, but of breeding lines used in aquaculture. Before fish rearing in Embrapa, the founder populations originated from commercial captive broodstocks bred over several generations and therefore explaining the current population genetic polymorphism observed. This is an

Varela *et al. BMC Genomics* (2021) 22:709

Page 4 of 10

Fig. 1 Sex-Averaged linkage map of *C. macropomum*. The 27 linkage groups are ordered by number of SNP markers. In each linkage group, numbers shown on the left provide the position (in cM) of the respective locus on the chromosome, while bars on the right indicates the relative number of SNP markers

important consideration for future breeding programmes in the species as low heterozygosity may be a result of inbreeding.

In the current study, one major QTL (LG 26) related to sex determination was supported by four markers (LOD > 45) that were found to peak in one single region (15.88 cM and 16.51 cM) of LG 26. Together, they contribute to 67.71% of the phenotypic variation, suggesting the existence of a sex-linked QTL in tambaqui. Further genotyping of the sex-linked QTLs revealed that male
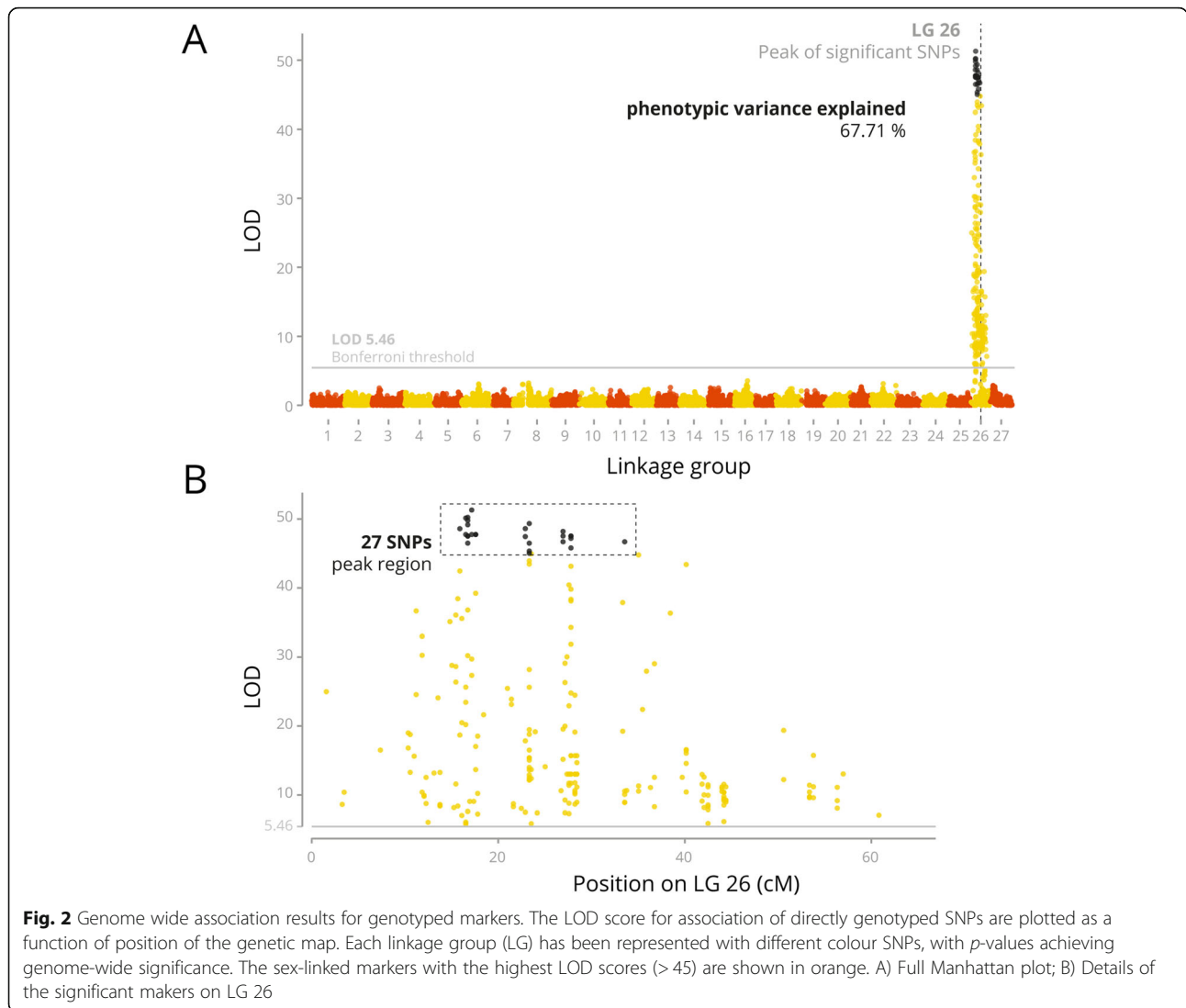
Table 2 Summary of the genetic linkage map of *C. macropomum*. * Estimated length (gap length between or within scaffold are unknown)

Chr./LG	Chr. Length* (bp)	Markers	Average-sex Loci	Average-sex Size (cM)	Female Loci	Female Size (cM)	Male Loci	Male Size (cM)
1	67,510,651	917	297	130.20	143	133.83	180	135.22
2	52,904,932	716	243	107.07	134	125.43	122	104.51
3	52,784,634	694	275	128.79	157	124.36	155	134.98
4	45,648,328	638	251	116.36	147	94.80	119	114.29
5	45,481,260	641	242	106.74	149	108.94	117	109.39
6	47,130,658	623	216	123.25	119	124.65	112	130.37
7	40,559,550	598	237	115.84	145	106.67	117	129.57
8	46,777,390	598	212	113.34	109	115.85	125	110.87
9	49,677,613	595	230	110.42	156	96.87	106	120.64
10	44,348,188	581	231	113.92	147	91.91	107	125.02
11	37,564,635	561	194	92.38	108	94.58	110	95.72
12	38,025,399	543	186	96.11	100	89.39	105	98.83
13	40,757,001	545	197	91.79	123	100.14	102	96.78
14	36,781,461	524	206	111.42	134	102.63	97	109.51
15	42,853,313	522	187	104.01	112	97.55	106	105.00
16	44,064,100	498	199	82.33	119	82.75	101	109.89
17	35,989,072	492	158	79.64	92	69.70	89	90.24
18	43,836,131	488	208	101.97	119	145.40	132	178.93
19	40,668,426	477	153	92.04	102	76.16	93	90.28
20	40,163,545	479	187	100.24	106	116.80	115	86.69
21	41,776,805	481	165	79.90	78	84.78	99	78.51
22	34,436,174	462	173	101.63	92	99.16	117	105.80
23	34,642,816	454	195	102.18	120	104.55	116	97.60
24	33,779,548	450	181	99.80	105	92.91	104	96.14
25	36,164,054	418	163	90.29	100	84.89	86	95.91
26	31,836,560	415	120	75.09	67	74.20	70	104.30
27	33,389,210	395	153	85.09	64	94.33	120	70.20
Unplaced	82,308,352	–	–	–	–	–	–	–
Total	1,221,859,806	14,805	5459	2751.81	3147	2733.20	3022	2925.19

genotypes were heterozygous, whereas females were homozygous. Validation of the four successfully amplified sex associated markers in broodstock ($n = 22$) and their progeny ($n = 184$) revealed an accuracy ranging from 90.0 to 96.7% at correctly predicting sex. Inconsistencies in sex assignment from SNP assays of sex-linked markers between different sires and dams have been reported in other species including Atlantic halibut [10], Nile tilapia, *Oreochromis niloticus* L. [27] and swimming crab [14]. This may be partly explained by the lack of recombination suppression in regions that include and flank the sex determining region in these species, so that the detected sex associated markers may frequently cross over during meiosis. Nevertheless, the sex-associated SNPs provide clear evidence that those markers are in strong linkage disequilibrium with the sex-determining genes and suggest that tambaqui has an XX/XY sex determination system.

The morphology of tambaqui chromosomes is metacentric and submetacentric; nevertheless, no heteromorphism in sex chromosomes has been observed [17]. In fish, sex determining QTLs are often found in sub-telomeric regions, suggesting that chromosome ends are areas of accelerated evolution developing non-recombining via rearrangements [31]. As a result of recombination suppression, these chromosomes are derived from autosomal regions of the genomes with higher plasticity and lower density of core genes, and are rich in transposable elements and other repetitive sequences [31, 32]. Our results are consistent with the

Fig. 2 Genome wide association results for genotyped markers. The LOD score for association of directly genotyped SNPs are plotted as a function of position of the genetic map. Each linkage group (LG) has been represented with different colour SNPs, with *p*-values achieving genome-wide significance. The sex-linked markers with the highest LOD scores (> 45) are shown in orange. A) Full Manhattan plot; B) Details of the significant makers on LG 26

above with blocks of suppression found at around 11 cM of distance on the LG 26, without enzymes restriction sites used by ddRAD, however, this region is not large enough to indicate sexual chromosome heteromorphism.

Conclusions

The PACE-validated sex marker panel used in this study is a powerful tool for gender monitoring and increasing genetic gains in future selective breeding programmes of tambaqui. As a result, the sex-linked markers identified in this study provide a valuable molecular tool for discriminating genders early in the production cycle. Furthermore, sex markers will expedite the identification of neomales following indirect hormonal sex reversal, which appears to be achievable given the species' apparent XX/XY sex determination mechanism.

Table 3 Selected SNP markers. The four markers significantly associated with phenotypic sex based on LOD score and position on LG 26

Marker	Chr./LG	Chr. position	Position Average	Position Female	Position Male	LOD	Male	Female
Cma5119693	26	1,051,596	15.88	58.77	49.96	48.59	C/G	G/G
Cma5119452	26	1,002,040	16.73	58.77	51.24	49.79	A/G	A/A
Cma5117879	26	4,967,169	16.51	60.90	59.32	50.13	A/G	G/G
Cma5145911	26	6,424,206	16.51	58.77	49.96	47.75	C/T	C/C

Table 4 Validation data of sex markers in *C. macropomum* broodstock and their offspring archived by PACE and ddRAD showing accuracy

Genotyping Approach	Sample type	Marker	Accuracy (%)
ddRAD assay	Offspring	Cma5145911	96.06
ddRAD assay	Offspring	Cma5119693	97.30
ddRAD assay	Offspring	Cma5119452	95.74
ddRAD assay	Offspring	Cma5117879	94.22
PACE assay	Broodstock	Cma5145911	90.00
PACE assay	Broodstock	Cma5119693	90.00
PACE assay	Broodstock	Cma5119452	90.00
PACE assay	Broodstock	Cma5117879	90.00
PACE assay	Offspring	Cma5145911	96.15
PACE assay	Offspring	Cma5119693	96.17
PACE assay	Offspring	Cma5119452	96.72
PACE assay	Offspring	Cma5117879	96.71

Methods

Family construction and rearing conditions

Five full-sib *C. macropomum* families were reared at the facilities of Embrapa Fisheries and Aquaculture, Palmas-TO, Brazil. The families, FAM01, AM01, EMB01, EMB02 and EMB04, were reared in separate 1 m^3 tanks for 175 and 215 days, respectively. Fish were fed twice a day ad libitum with a commercial dry pellet diet containing 32% crude protein during the first 60 days, and 28% crude protein thereafter (Guabi S. A, Brazil).

Sample collection for QTL mapping and sex-linked SNP markers validation

At sampling, a total of 244 F1 juveniles from two families (142 from FAM01 and 102 from EMB01) were randomly culled by lethal anaesthesia (10% Benzocaine solution; Merck & Co., USA), and a fin clip was collected, fixed in 100% ethanol, and kept at − 20 °C until DNA extraction. Fish were then dissected and gonad samples for histological identification of sex were fixed in Bouin's solution for 24 h, then washed in distilled water and dehydrated in a series of ethanol solutions (70, 80, 90 and 100%) according to Almeida et al. [9]. Gonad samples were embedded in paraffin wax, sectioned (5 μm), mounted on slides, stained with haematoxylin-eosin, and analysed under a light microscope (Leica DM500, Heerbrugg, Switzerland).

For sex-QTL validation, fin clips from another 206 individuals were further collected and fixed in 100% ethanol. It included, 22 adult broodstocks (11 males and 11 females) from commercial broodstocks from Mato Grosso State (Brazil) and maintained in Embrapa's Germplasm Active Bank-BAG (Palmas-TO, Brazil), which were sampled after anaesthesia (10% Benzocaine solution; Merck & Co., USA). As well as an extra 184 F1

individuals, belonging to the five full-sib families (AM01 (*n* = 69), EMB01 (*n* = 20), EMB02 (*n* = 39), EMB04 (*n* = 36), FAM01 (*n* = 20)) reared in a communal tank after being tagged intramuscularly with PIT tags (Animal Tag, Brazil). Fish were then sacrificed (378 days post hatch) to dissect their gonad and identify sex histologically, as described above.

Genomic DNA extraction

All fin clips samples were stored at 4 °C in 100% ethanol prior to use. Total DNA was extracted by a modified salt-based extraction protocol according to A Blanquer [33], as described by Taslima et al. [34]. Each sample was initially quantified by spectrophotometry (Nanodrop 1000, Thermo Fisher Scientific, USA). Genomic DNAs were and stored at 4 °C. Agarose gel electrophoresis (0.8%) was used to check the integrity of the genomic DNA. For those samples used for library construction, the DNAs were re-quantified by fluorimetry, (Qubit 2.0, Thermo Fisher Scientific, USA) and diluted further, to 7 ng/μL, in 5 mM Tris, pH 8.0.

ddRAD library preparation and sequencing

Two ddRAD libraries were constructed, one for each family pedigree. DNAs from 142 F1 progeny and both parents were used for the FAM01 library and another 102 F1 progeny and both parents were used for EMB01 library. The methodology followed the original protocol reported by Peterson et al. [35], with some modifications/refinements as described in full by Manousaki et al. [36]. Each progeny sample was processed in duplicate while five samples from each parent were processed. Briefly, using 96 well microplates, each DNA sample (15 ng DNA per sample) was simultaneously digested by two high fidelity restriction enzymes: *Sbf*I (CCTG CA|GG recognition site), and *NIa*III (CATG| recognition site), both sourced from New England Biolabs (NEB, UK). Digestions were incubated at 37 °C for 60 min using 0.3 U of each enzyme per sample (20 U/μg DNA equivalent) in 1× CutSmart Buffer (NEB, UK), in a final reaction volume of 6 μL. A heat inactivation step was then performed at 65 °C for 25 min. After cooling the reactions to room temperature, 3 μL of a premade, unique barcode / adapter mix was added to each digested DNA sample, and incubated at 22 °C for 10 min. This adapter mix comprised individual-specific barcoded combinations of P1 (*Sbf*I-compatible) and P2 (*Nla*III-compatible) adapters at 6 nM and 1.54 μM concentrations respectively, in 1x reaction buffer 2 (NEB, UK). Adapters were compatible with Illumina sequencing chemistry [35]. The adapters included an inline five- or seven-base barcode for sample identification (Supplementary Table S1). Following a 10-min incubation, 3 μL of ligation mix was then added, comprising 4

Varela *et al. BMC Genomics*   (2021) 22:709

Page 8 of 10

mM rATP (Promega, UK), and 3000 cohesive-end units of T4 ligase (NEB, UK) in 1x CutSmart buffer. An optimised three steps ligation incubation was performed, i.e., 1.5 h at 14 °C, 1.5 h at 22 °C, and then overnight (16 h) at 5 °C. To stop ligation, 35 μL P1 buffer (Qiagen, UK) was added to each reaction well, the samples then being combined into a single pool. This pooled sample was column purified (MinElute PCR Purification Kit, Qiagen, UK), being eluted in 55 μL EB buffer (Qiagen, UK). Size selection of fragments was performed by agarose gel separation and excision of the 400 to 700 bp range. Following gel purification, the eluted size-selected template DNA was amplified by PCR (13 cycles). The PCR reactions were column-purified in MinElute PCR Purification Kit (Qiagen, UK). The 55 μL eluted in EB buffer was then subjected to an additional clean-up / concentration using AMPure XP magnetic beads (Beckman Coulter), being eluted in 17 μL EB buffer, quantified by fluorimetry and size checked by agarose gel electrophoresis. The libraries were commercially sequenced (Novagene, UK) on an Illumina HiSeq 2500 platform using 150 base paired-end reads (v3 chemistry).

Genotyping ddRAD alleles

The sequences were pre-processed to discard low quality (i.e., with a quality score of less than 20), missing tag structure or ambiguous bases. Remaining reads were aligned to *C. macropomum* genome scaffolds (Assembly GCA_904425465.1) using bwa v0.7.17 [37], then sorted into loci and genotypes using Stacks v2.41 [38]. Informative markers were kept only when presenting at least two alleles with a minor allele frequency over 0.01 and which were present in both parents and at least 75% of the offspring to minimise the amount of missing or erroneous data. Only one SNP (selected at random) was reported for each RAD locus.

Construction of linkage maps

Based on the SNP genotypes obtained, sex-specific and sex-averaged linkage maps were constructed with LepMap3 [39]. SNPs deviating from expected Mendelian segregation ($P < 0.001$) were excluded. Based on available karyotyping data from Nakayama et al. [25], the number of linkage groups (LG) was set to 27. The total length of maps in centi-Morgans (cM) was estimated using the Kosambi mapping function [40]. Maps generated with the Order-Marker2 module were checked for contiguous sequence (contig) continuity. Genetic maps were drawn using Genetic-Mapper v0.13 [41].

Genome scaffold ordering

The ordering and orientation of the *C. macropomum* genome scaffolds to reconstruct chromosomes were performed with ALLMAPS [42] using the linkage maps (sex average).

Association analysis and QTL mapping

Using phenotypic sex data, an association analysis was performed within the package R/SNPassoc v1.9–2 according to González et al. [43] to test for associations between SNP genotypes and phenotypic sex under a binary codominant genetic model, using the function *WGassociation*. Bonferroni correction was used to counteract the problem of multiple comparisons when determining the significance of observed results. The phenotypic variance explained (PVE) was defined as:

$$PVE = 100 \times \left(1 - 10^{\frac{-2 \times LOD}{N}} \right)$$

With N the number of sexed individual ($N = 206$) and LOD the peak LOD value.

Verification of SNP sex association by PACE assay

Marker sex association was assessed for four SNPs that were commonly found in the mapping families to span the region of the highest association with sex from ddRAD-seq, using fluorescent, allele specific, endpoint-genotyping assays: PCR Allele Competitive Extension (PACE™ genotyping assays, 3CR Bioscience, UK). SNP-specific primer sets were designed by 3CR Bioscience. DNA was extracted from fin clips from the 206 individuals (22 broodstock and 184 offspring) using a HOTSHOT protocol adapted from Truett et al. [44]. Each genotyping assay was run in 10 μL volume containing approximately 15 ng of target genomic DNA incorporated with a PACE master mix reaction. All assays were run with the same touchdown thermal cycling programme using a Biometra TGradient thermal cycler (Biometra GmbH, Goettingen, Germany) as follows: 94 °C for 15 min followed by 10 cycles of 94 °C for 20 s melt, 61–57 °C for 1 min anneal and extension (decreasing of 0.6 °C per cycle) followed by 25 cycles of 94 °C for 20 s melt, 57 °C for 1 min anneal and extension. Thereafter, assays results were read at 25 °C using an endpoint genotyping programme on a Techne Quantica qPCR thermal cycler (Bibby Scientific Ltd., Stone, UK) in which unknown genotypes were assigned based on fluorescent output in comparison to non-template control wells containing DNA/RNA free H_2O.

Abbreviations
LG: Linkage groups; cM: Centi-Morgan; LOD: Logarithm of the odds; PACE: PCR Allele Competitive Extension; ddRAD: Double digest restriction-site associated DNA; QTL: Quantitative trait locus

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-021-08037-8.

Additional file 1: Supplementary Figure S1 Linkage map of *C. macropomum*. The 27 linkage groups are ordered by number of SNP markers. In each linkage group, numbers shown on the left provide the position (in cM) of the respective locus on the chromosome, while bars

Varela *et al. BMC Genomics*   (2021) 22:709

Page 9 of 10

on the right indicates the relative number of SNP markers. A) Female-only linkage map; B) Male-only linkage map.

Additional file 2: Supplementary Table S1 Origin of samples and barcode. Details of each sample used: barcode, file name ID, family, gender, status number of extracted and retained reads.

Additional file 3: Supplementary Table S2 Genetic maps. Ordered markers: marker ID, linkage group and Position Average (cM), Scaffold, Scaffold position, and Marker sequence.

Additional file 4: Supplementary Table S3 Primer used in PACE validation.

Additional file 5: Supplementary Table S4 Details of the PACE assay results. Genotypes of 206 samples.

Additional file 6: Supplementary Data S1 Structure of chromosome-level genome assemblies ordered from the linkage map. AGP (A Golden Path) file formatted, reporting the scaffold structure of each linkage group into 27 chromosomes.

Authors' contributions
Conceptualisation: ESV, JBT and HM; Data curation: ASB, LS, LNGK, LST, MB and ESV; Formal analysis: MB, JBT and ESV; Funding acquisition: ESV, LS FPR and LELF; Investigation: HM, JBT, MB, ESV, LNGK, LST, LS, FLA, LCVV, FPR and LELF; Methodology: ESV, LNGK, FLA, LST, FPR and LELF; Project administration: HM and ESV; Resources: LNGK, LST, LS, FLA, LCVV, FPR; Software: MB and ESV; Supervision: ESV and HM; Validation: ESV, HM, JBT, MB; Visualisation: ESV, HM, JBT, MB; Writing original draft: ESV; Writing: Review & editing: HM, JBT, MB, ESV, LNGK, LST, LS, FLA, LCVV, FPR and LELF. The author(s) read and approved the final manuscript.

Availability of data and materials
The datasets generated and analysed during the current study are available in the supplementary information files of the manuscript.

Declarations

Ethics approval and consent to participate
All animal handling and procedures were approved by the Embrapa Animal Ethics Committee, in accordance with the national guidelines under the Animal Welfare Act 6899 (CONCEA) of Brazil. All methods are reported in accordance with ARRIVE guidelines [45].

Consent for publication
Not Applicable.

Competing interests
The authors declare that they have no competing interests as defined by BMC, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Author details
[1]Embrapa Pesca e Aquicultura, Prolongamento da Av. NS 10, Cruzamento com AV. LO 18, Sentido Norte, loteamento Água Fria, CEP, Palmas, TO 77008-900, Brazil. [2]Institute of Aquaculture, Faculty of Natural Sciences, University of Stirling, Stirling, Scotland FK9 4LA, UK. [3]Embrapa Amazônia Ocidental, Manaus, AM CEP 69010-970, Brazil.

References
1. PeixeBR. In: Corporativa TC, editor. Anuário Brasileiro da Piscicultura Peixe BR 2020. São Paulo-SP: PeixeBR; 2020. p. 135.
2. Pincinato RBM, Asche F. The development of Brazilian aquaculture: introduced and native species. Aquac Econ Manag. 2016;20:312–23.
3. Saint-Paul U. Native fish species boosting Brazilian's aquaculture development. Acta Fisheries Aquatic Resour. 2017;5(1):1–9.
4. Valladão GMR, Gallani SU, Pilarski F. South American fish for continental aquaculture. Rev Aquac. 2016;0:1–19.
5. OECD/FAO. OECD/Food and Agriculture Organization of the United Nations (2015). In: OECD-FAO agricultural outlook 2015. Paris: OECD Publishing; 2017.
6. De Mello F, Felippe D, Godoy LC, Lothhammer N, Guerreiro LRJ, Streit DP Jr. Morphological and morphometric analysis of skeletal muscle between male and female young adult *Colossoma macropomum* (Characiformes: Serrasalmidae). Neotrop Ichthyol. 2016;14. https://doi.org/10.1590/1982-0224-20150149.
7. Perazza CA, Ferraz JBS, Almeida-Val VMF, Hilsdorf AWS. Genetic parameters for loin eye area and other body traits of an important Neotropical aquaculture species, *Colossoma macropomum* (Cuvier, 1818). Aquac Res. 2019;50:2907–16.
8. Campos EC, Oliveira CAL, Araújo FCT, Todesco H, Souza FN, Rossi RM, et al. Genetic parameters and response to selection for growth in tambaqui. Animal. 2020;14(9):1–9.
9. Almeida FL, Lopes JS, Crescencio R, Izel ACU, Chagas EC, Boijink C. Early puberty of farmed tambaqui (*Colossoma macropomum*): possible influence of male sexual maturation on harvest weight. Aquaculture. 2016;452:224–32.
10. Palaiokostas C, Bekaert M, Davie A, Cowan ME, Oral M, Taggart JB, et al. Mapping the sex determination locus in the Atlantic halibut (*Hippoglossus hippoglossus*) using RAD sequencing. BMC Genomics. 2013;14(566):1–12.
11. Saillant E, Fostier A, Menu B, Haffray P, Chatain B. Sexual growth dimorphism in sea bass *Dicentrarchus labrax*. Aquaculture. 2001;202:371–87.
12. Yoneda M, Kurita Y, Kitagawa D, Ito M, Tomiyama T, Goto T, et al. Age validation and growth variability of Japanese flounder *Paralichthys olivaceus* off the Pacific coast of northern Japan. Fish Sci. 2007;73:585–92.
13. Levy T, Rosen O, Eilam B, Azulay D, Zohar I, Aflalo ED, et al. All-female monosex culture in the freshwater prawn *Macrobrachium rosenbergii* - a comparative large-scale field study. Aquaculture. 2017;479:857–62.
14. Li R, Bekaert M, Lu J, Lu S, Zhang Z, Zhang W, et al. Mapping and validation of sex-linked SNP markers in the swimming crab *Portunus trituberculatus*. Aquaculture. 2020;524:735–228.
15. Piferrer F. Endocrine sex control strategies for the feminization of teleost fish. Aquaculture. 2001;197:229–81.
16. Singh AK. Introduction of modern endocrine techniques for the production of monosex population of fishes. Gen Comp Endocrinol. 2013;181:146–55.
17. Hashimoto DT, Senhorini JA, Foresti F, Martínez P, Porto-Foresti F. Genetic identification of F1 and post-F1 serrasalmid juvenile hybrids in Brazilian aquaculture. PLoS One. 2014;9:e89902.
18. Reis VR, FLA A. Effect of 17β-oestradiol on the sex ratio of tambaqui, *Colossoma macropomum*. Aquac Res. 2019;50:154–61.
19. Hoga CA, Almeida FL, Reyes FGR. A review on the use of hormones in fish farming: analytical methods to determine their residues. CyTA J Food. 2018;16:679–91.
20. Nunes JRS, Liu S, Pértille F, Perazza CA, Villela PMS, Almeida-Val VMF, et al. Large-scale SNP discovery and construction of a high-density genetic map of *Colossoma macropomum* through genotyping-by-sequencing. Sci Rep. 2017;7:46112.
21. Zhang S, Zhang X, Chen X, Xu T, Wang M, Qin Q, et al. Construction of a high-density linkage map and QTL fine mapping for growth- and sex-related traits in channel catfish (*Ictalurus punctatus*). Front Genet. 2019;10:1–14.
22. Zhou Y, Liu H, Wang X, Fu B, Yu X, Tong J. QTL fine mapping for sex determination region in bighead carp (*Hypophthalmichthys nobilis*) and comparison with silver carp (*Hypophthalmichthys molitrix*). Mar Biotechnol. 2020;22:41–53.
23. Houston RD, Bean TP, Macqueen DJ, Gundappa MK, Jin YH, Jenkins TL, et al. Harnessing genomics to fast- track genetic improvement in aquaculture. Nat Rev Genet. 2020;21:389–409.

24. Regan T, Bean TP, Ellis T, Davie A, Carboni S, Migaud H, et al. Genetic improvement technologies to support the sustainable growth of UK aquaculture. Rev Aquac. 2021;13(4):1958–85.
25. Nakayama CM, Feldberg E, Bertollo LAC. Karyotype differentiation and cytotaxonomic considerations in species of Serrasalmidae (Characiformes) from the Amazon basin. Neotrop Ichthyol. 2012;10:53–8.
26. Robledo D, Palaiokostas C, Bargelloni L, Martínez P, Houston R. Applications of genotyping by sequencing in aquaculture breeding and genetics. Rev Aquac. 2017;10(3):670–82.
27. Palaiokostas C, Bekaert M, Khan MG, Taggart JB, Gharbi K, BJ MA, et al. Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) Using RAD sequencing. PLoS One. 2013;8(7): e68389.
28. Palaiokostas C, Bekaert M, Khan MG, Taggart JB, Gharb K, McAndrew BJ, et al. A novel sex-determining QTL in Nile tilapia (*Oreochromis niloticus*). BMC Genomics. 2015;16(171):1–10.
29. Brown JK, Taggart JB, Bekaert M, Wehner S, Palaiokostas C, Setiawan AN, et al. Mapping the sex determination locus in the hāpuku (*Polyprion oxygeneios*) using ddRAD sequencing. BMC Genomics. 2016;17(448):1–12.
30. Torati LS, Taggart JB, Varela ES, Araripe J, Wehner S, Migaud H. Genetic diversity and structure in *Arapaima gigas* populations from Amazon and Araguaia-Tocantins river basins. BMC Genet. 2019;20(13):1–13.
31. Volff J-N, Nanda I, Schmid M, Schartl M. Governing sex determination in fish: regulatory putsches and ephemeral dictators. Sex Dev. 2007;1:85–99.
32. Graves JA. Sex chromosome specialization and degeneration in mammals. Cell. 2006;124:901–14.
33. Blanquer A. Phylogeographie intraspecifique d'un poisson marin, le flet *Platichthys flesus* L. (Heterosomata). Polymorphisme des marqueurs nucleaires et mitochondriaux. Montpellier: University of Montpellier; 1990.
34. Taslima K, Wehner S, Taggart JB, Verdal H, Benzie JAH, Bekaert M, et al. Sex determination in the GIFT strain of tilapia is controlled by a locus in linkage group 23. BMC Genet. 2020;21(49):1–15.
35. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: an inexpensive method for SNP discovery and genotyping in model and non-model species. PLoS One. 2012;7(5):1–11.
36. Manousaki T, Tsakogiannis A, Taggart JB, Tsaparis D, Lagnel J, Chatziplis D, et al. Exploring a nonmodel teleost genome through RAD sequencing—linkage mapping in common Pandora, *Pagellus erythrinus* and comparative genomic analysis. G3 (Bethesda). 2016;6:509–19.
37. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics. 2009;25(14):1754–60.
38. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: building and genotyping loci de novo from short-read sequences. Genes, Genomes and Genetics. 2011;1:171–82.
39. Rastas P. Lep-MAP 3: robust linkage mapping even for low-coverage whole genome sequencing data. Bioinformatics. 2017;33:3726–32.
40. Kosambi DD. The estimation of map distances from recombination values. Ann Eugenics. 1943;12:172–5.
41. Bekaert M. Genetic-Mapper: vectorial genetic map drawer. F1000Res. 2016;5: 1301.
42. Tang H, Zhang X, Miao C, Zhang J, Ming R, Schnable JC, et al. ALLMAPS: robust scaffold ordering based on multiple maps. Genome Biol. 2015;16(3): 1–15.
43. González JR, Armengol L, Solé X, Guinó E, Mercader JM, Estivill X, et al. SNPassoc: an R package to perform whole genome association studies. Bioinformatics. 2007;23:644–5.
44. Truett GE, Heeger P, Mynatt RL, Truett AA, Walker JA, Warman ML. Preparation of PCR-quality mouse genomic dna with hot sodium hydroxide and tris (HotSHOT). Biotechniques. 2000;29:52–4.
45. Percie du Sert N, Ahluwalia A, Alam S, Avey MT, Baker M, Browne WJ, et al. Reporting animal research: Explanation and elaboration for the ARRIVE guidelines 2.0. PLoS Biol. 2020;18(7):e3000411.

Publisher's Note

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