



HAL
open science

Development of a large SNPs resource and a low-density SNP array for brown trout (*Salmo trutta*) population genetics

Keoni Saint-Pé, Maeva Leitwein, Laurence Tissot, Nicolas Poulet, Bruno Guinand, Patrick Berrebi, Geoffrey Marselli, Jean-Marc Lascaux, Pierre-Alexandre Gagnaire, Simon Blanchet

► To cite this version:

Keoni Saint-Pé, Maeva Leitwein, Laurence Tissot, Nicolas Poulet, Bruno Guinand, et al.. Development of a large SNPs resource and a low-density SNP array for brown trout (*Salmo trutta*) population genetics. *BMC Genomics*, 2019, 20 (1), 10.1186/s12864-019-5958-9 . hal-02395980

HAL Id: hal-02395980

<https://hal.science/hal-02395980v1>

Submitted on 5 Dec 2019

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

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

METHODOLOGY ARTICLE

Open Access



Development of a large SNPs resource and a low-density SNP array for brown trout (*Salmo trutta*) population genetics

Keoni Saint-Pé^{1*}, Maeva Leitwein^{2,3†}, Laurence Tissot⁵, Nicolas Poulet⁶, Bruno Guinand³, Patrick Berrebi^{3,4}, Geoffrey Marselli¹, Jean-Marc Lascaux⁷, Pierre-Alexandre Gagnaire^{3†} and Simon Blanchet^{1,8†}

Abstract

Background: The brown trout (*Salmo trutta*) is an economically and ecologically important species for which population genetic monitoring is frequently performed. The most commonly used genetic markers for this species are microsatellites and mitochondrial markers that lack replicability among laboratories, and a large genome coverage. An alternative that may be particularly efficient and universal is the development of small to large panels of Single Nucleotide Polymorphism markers (SNPs). Here, we used Restriction site Associated DNA sequences (RADs) markers to identify a set of 12,204 informative SNPs positioned on the brown trout linkage map and suitable for population genetics studies. Then, we used this novel resource to develop a cost-effective array of 192 SNPs (96 × 2) evenly spread on this map. This array was tested for genotyping success in five independent rivers occupied by two main brown trout evolutionary lineages (Atlantic -AT- and Mediterranean -ME-) on a total of 1862 individuals. Moreover, inference of admixture rate with domestic strains and population differentiation were assessed for a small river system (the Taurion River, 190 individuals) and results were compared to a panel of 13 microsatellites.

Results: A high genotyping success was observed for all rivers (< 1% of non-genotyped loci per individual), although some initially used SNP failed to be amplified, probably because of mutations in primers, and were replaced. These SNPs permitted to identify patterns of isolation-by-distance for some rivers. Finally, we found that microsatellite and SNP markers yielded very similar patterns for population differentiation and admixture assessments, with SNPs having better ability to detect introgression and differentiation.

Conclusions: The novel resources provided here opens new perspectives for universality and genome-wide studies in brown trout populations.

Keywords: Conservation, Management, Population genomics, Microarray, Admixture

Background

The brown trout (*Salmo trutta*) is one of the most widespread freshwater fish species in Eurasia, and it has been widely introduced in both the southern and northern hemispheres [1]. As part of the Salmonidae family, it is a scientifically interesting species because of its diversity in terms of ecology, life history strategies and habitat use

[2, 3]. Thanks to its wide ecological variability and excellent ability to spread and colonize new watersheds, the species is found both in fresh and salt waters over most of its range. The brown trout is also an economically major species in terms of farming, net fishing (for the sea-run form), and expenditure in recreational angling [4, 5], partly explaining its worldwide intentional introduction [6]. Because this species is strongly associated to human interests, wild brown trout populations are widely managed, either to sustain attractive leisure activities such as recreational angling or to conserve declining and/or emblematic populations. Moreover, the brown trout has been domesticated since the nineteenth

* Correspondence: keoni.saintpe@sete.cnrs.fr

†Keoni Saint-Pé and Maeva Leitwein are co-first authors

Pierre-Alexandre Gagnaire and Simon Blanchet are co-last authors

¹Centre National de la Recherche Scientifique (CNRS), Université Paul Sabatier (UPS), Station d'Ecologie Théorique et Expérimentale, SETE, UMR 5321, 2 route du CNRS, 09200 Moulis, France

Full list of author information is available at the end of the article



century [7], and hatchery strains have been largely used to sustain wild populations worldwide [8–10]. Instead of positive expected effects of these stocking activities, most have proven to have negative long term effects on wild brown trout populations in part because of the reduced fitness brought by hatchery fish in wild populations, and the loss of local genetic heritage caused by the replacement of local wild populations with genetically homogeneous hatchery strains [11–14].

The brown trout presents high levels of phenotypic and genetic polymorphism, with seven main mitochondrial (mtDNA) lineages with various geographical extents being generally recognized. These consist of (i) four sub-continental lineages: the Atlantic (AT), Mediterranean (ME), Danubian (DA) and Adriatic (AD) lineages, (ii) two regional lineages: Marmoratus (MA) in the north of the Adriatic Sea and North African (NA) in Morocco, Algeria and Sicily [15, 16] and “(iii) two local lineages limited to specific geographical areas: the Duero (DU) in Northwestern Iberian basins, and the Tigris (TI) in Turkey” [17–21]. Within these extent lineages that cover the whole range of *S. trutta*, high levels of genetic and phenotypic polymorphism are also observed at more local spatial scales within lineages (e.g. population scale; [22–24]). However, diversity patterns in brown trout have also been locally influenced by stocking practices that mostly relied on European hatchery strains of AT origin [25] to supplement local populations, with the exception of a few local strains stemming from local populations [26–28].

Genetic tools appeared as a key approach for scientists and local managers to optimize conservation efforts [29, 30] because they provide insight into both the ecological and the evolutionary dynamics of wild populations [10, 31]. For instance, assignment tests, fine-scale population structure, kinship analyses and genome-wide surveys [32–34] enable to monitor populations effectively, and have high potential applications for conservation and management in salmonids, including the brown trout [13, 28, 35].

Molecular studies on trout populations first used allozymes, mitochondrial markers, and then microsatellite loci [7, 36, 37]. These markers (notably microsatellites) are useful and adequate to answer many biological questions, but their genome coverage is generally weak, and replicability and universality are relatively low since each research group generally uses its own panel of markers. Single nucleotide polymorphisms (SNP) markers have been shown to potentially reduce these limitations [38–40]. They allow to uncover a relatively high number of annotated and mapped markers with low scoring-error rates [41, 42]. Also, SNPs markers can easily be chosen to represent both neutral genomic regions and regions under selection, at a genome-wide scale and across large samples [43]. Despite being biallelic markers, SNPs can be highly informative for most analyses used in population genetics,

as far as the number of loci is sufficiently high and evenly spread across the genome (> 50; [44–48]). Genome coverage is an important aspect in the choice of markers for population genetics: first, markers evenly spread across the genome are less likely to exhibit linkage disequilibrium [49], and second, it was shown that many population events, such as introgressive hybridization can only concern certain genomic blocks, a full coverage thus enables to capture these events [50, 51].

SNP arrays are commonly used for conservation purposes in salmonids [52–54], although rarely in brown trout, which is probably because only a handful of SNP markers were available for this species [55–57]. However, higher density resources were more recently developed. In particular, Linløkken [58] developed 3781 SNPs to analyse genetic differences between wild and hatchery brown trout in a tributary of Lake Savalen in central Norway. Moreover, a new lineage-specific high density linkage map for *S. trutta* comprising ancestry informative SNPs for both Atlantic (AT) and Mediterranean (ME) evolutionary lineages from Western Europe was also developed [14, 59]. This latter resource provides a novel baseline for the development of mapped SNPs that may be of prime interest for studies on brown trout population genetics across a wide spatial range.

The aim of this study was to develop a genome-wide, mapped and universal set of SNPs for the brown trout for both Atlantic and Mediterranean lineages (AT and ME lineages; sensu Bernatchez, 2001), which would be a useful and affordable tool for both scientists and environmental managers. By taking advantage of the new genomic resource available for brown trout [14, 51, 59], we developed a panel of 12,204 RADs containing 1 or 2 polymorphic SNPs evenly spread across the genome structured in 40 chromosomes, and ancestry informative for at least two of the main brown trout lineages, the Atlantic and the Mediterranean lineages. Among these, a sub-panel of 192 SNPs describing the whole genome was included in a low-density SNP array. The validity of this low-density SNP array was tested by quantifying genotyping success and number of polymorphic loci in five independent populations from the two lineages at a large spatial scale. Finally, the power of this array was compared to a panel of 13 microsatellites for answering classical population genetics questions: admixture with stocked domestic individuals and genetic differentiation among wild populations. All resources are made freely available for future users (Additional file 1: S1 and S2).

Results

Development of the large SNPs resource and characteristics of the low-density SNP array

After applying filters, we identified a RAD data set of 12,204 sequences each containing one or two SNPs that

met our specifications and that are made readily available to the scientific community (see Additional file 1: S1). The number of RADs per linkage groups (LGs) varies between 137 (LG 33) and 563 (LG 6). These RAD tags are spaced by 0.119 cM (+/- 0.039) in average. They were found in all LGs and they cover the linkage map relatively evenly, although there were large gaps on LGs 11 (top), 12 (top and bottom), 27 (bottom), 32 (top), 33 (bottom), 37 (bottom), and 39 (top) (Fig. 1a). One estimate of the local recombination rate for each SNP (i.e. one index of the relative power of markers for LD-based mapping approaches) is provided in Additional file 1: S1 and S2.

The 245 SNPs selected from this set of RADs were spread over all linkage group so as to cover the linkage map as homogeneously as possible (Fig. 1b). The final

low-density 182 SNP array, to which we added 10 SNPs previously developed (see the Methods) is presented in Additional file 1: S2.

Genotyping success of the low-density SNPs array

Thirty of the 192 SNPs (among which a mitochondrial marker) that were initially genotyped in the five watersheds did not amplify, suggesting primers failure to properly bind their target DNA site. For the loci that successfully amplified, the overall genotyping success was high, with less than 1% of missing data per individual, irrespective of the river basin and the lineage (Table 1). Nonetheless, the number of polymorphic SNPs greatly varied from one river to another, ranging from 91 (for the Seuge River) to 160 (for the Aude River)

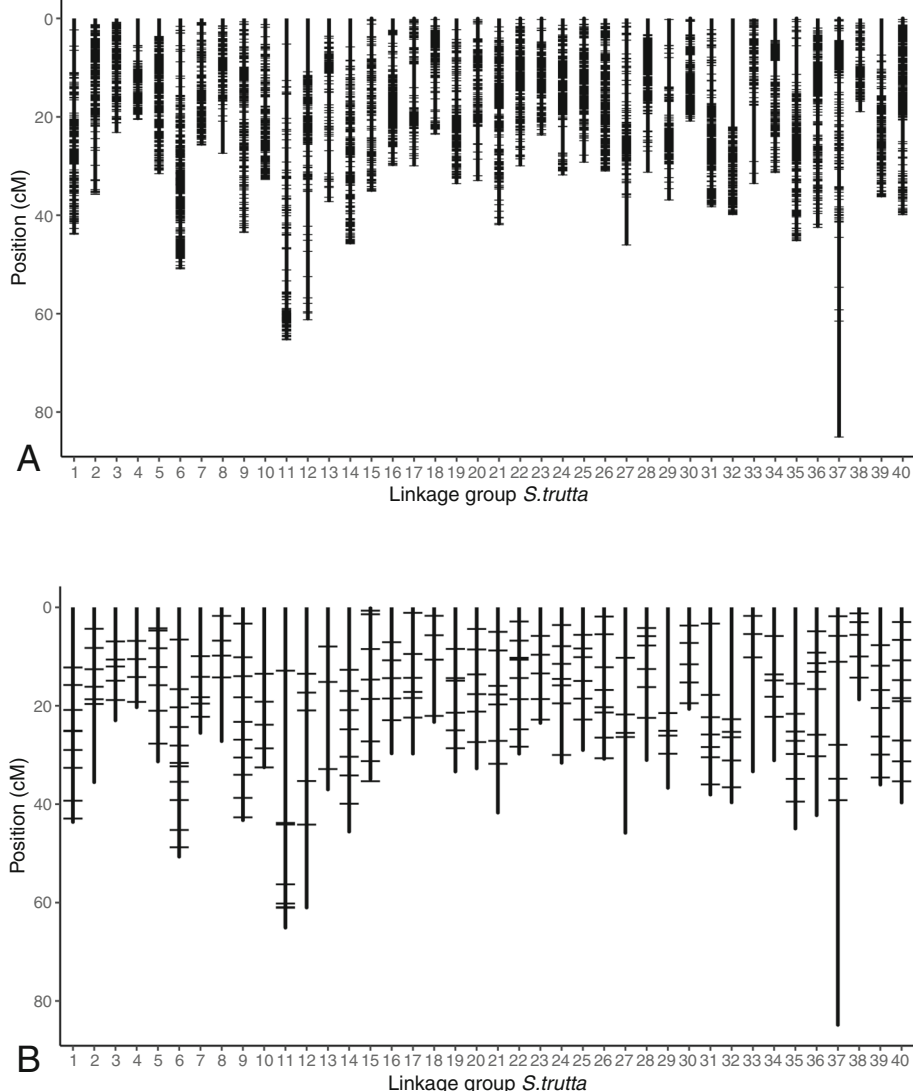


Fig. 1 Positions on the brown trout linkage map of the 12204 RADs (a; containing one or two SNPs with MAF ≥ 5% based on all individuals (AT + ME) and positions between the first or last 30 bp of the RAD with no undetermined nucleotides), and of the 245 SNPs (b; informative for at least the AT and ME lineages, positions between the first or last 50 bp of RADs (primer design zones), and spaced by at least 3.5 cM)

Table 1 Number of polymorphic loci, number of non-genotyped loci per individual considering polymorphic loci (average, min and max), expected heterozygosity (He) on polymorphic loci in each of the five river basins, and surface drained by the river basin (upstream from the lowest sampling site)

River basin (lineage)	Number of polymorphic loci	Average (±SD) number of non-genotyped loci per individual on polymorphic loci Min-max	He on polymorphic loci	Basin surface (Km ²)
Aude (Mediterranean)	160/162	1.04 (±1.68) 0–24	0.16	240
Doron de Bozel (Mediterranean)	153/162	0.71 (±0.99) 0–8	0.30	180
One (Atlantic)	119/162	0.87 (±1.23) 0–13	0.14	155
Roya (Mediterranean)	158/162	1.36 (±4.51) 0–63	0.30	360
Seuge (Atlantic)	91/162	0.66 (±1.06) 0–15	0.09	90
Average	136	0.95	0.20	205

out of the 162 SNPs that were successfully genotyped in that portion of our work (Table 1). The isolation-by-distance patterns were significant (pairwise *F_{st}* significantly correlated to pairwise riparian distance) for the Seuge River ($r_{Mantel} = 0.55$, p -value < 0.001) and the Roya River ($r_{Mantel} = 0.65$, p -value = 0.01). *F_{st}* and riparian distance were also positively although non-significantly correlated for the Aude River ($r_{Mantel} = 0.19$, p -value = 0.15), the One River ($r_{Mantel} = 0.28$, p -value = 0.07) and the Doron River ($r_{Mantel} = 0.02$, p -value = 0.36) (see Additional file 1: S3 for the plots).

Efficiency of the low-density SNPs array

Among the 30 replaced SNPs for the empirical test in the Taurion River, 10 failed at the amplification step (see Additional file 1: S2). This resulted in a set of 182 markers that successfully amplified, although only 92 of them were polymorphic in this river (among which only one mitochondrial marker of the 10 INRA and mitochondrial added markers). After removing individuals with more than a third of missing data, and keeping only individuals for which we had both SNPs and microsatellite data, our final dataset comprised 197 individuals (167

individuals from the Taurion basin and 30 from the Soueich trout farm). Expected heterozygosity ranged from 0.18 to 0.22 for SNPs and from 0.58 to 0.73 for microsatellites. *F_{st}* ranged from 0.022 to 0.070 for SNPs and from 0.019 to 0.040 for microsatellites (details for each site are shown in Table 2). We found one marker that was a significant *F_{st}* outlier, suggesting it may be influenced by selection (ID 295415; *F_{st}* > Expected heterozygosity), but we decided to keep it in our analyses (see the Discussion).

Individual inferences of hatchery ancestry measured with either SNPs or microsatellites are presented in Fig. 2 in the form of individual barplots. The distribution of captive-bred ancestry was bimodal, meaning that most individuals were either purely wild or captive-bred, with relatively low numbers of admixed genotypes (Fig. 2, Fig. 3a). Levels of hatchery ancestry were significantly and moderately correlated ($r_{Spearman} = 0.60$, d.f. = 193, $p < 0.001$; Fig. 3a), although for some individuals, there was a discrepancy between markers with one of the two marker types detecting introgressed fish while the other marker type assigned them as pure wild fish (Figs. 2, 3a). Pairwise *F_{st}* values are presented for both SNPs and microsatellites in Additional file 1: S4. They ranged from

Table 2 For each site, sample size (“N”), mean expected heterozygosity over all loci (“He”) and standard deviation between loci, mean observed heterozygosity over all loci (“Ho”) and standard deviation between loci, mean allelic richness computed using a rarefaction approach over all loci (“Ar”), mean *F_{is}* over all loci (“Fis”), mean *F_{st}* over all loci (i.e. uniqueness at the site level; $F_{st} = 1 - He_{Site}/He_{Total}$)

Site	N	SNPs					Microsatellites				
		Ho	He	Fis	Fst	Ar	Ho	He	Fis	Fst	Ar
BEA-Rau	27	0.18 ± 0.19	0.18 ± 0.18	0.011	0.070	1.58	0.44 ± 0.20	0.58 ± 0.25	0.226	0.04	4.42
PON-Rau	12	0.22 ± 0.23	0.21 ± 0.20	−0.038	0.030	1.66	0.67 ± 0.19	0.70 ± 0.17	0.027	0.022	5.63
THA-Bar	26	0.22 ± 0.23	0.22 ± 0.20	−0.025	0.022	1.65	0.65 ± 0.17	0.67 ± 0.18	0.015	0.017	5.85
THA-Tcc	30	0.19 ± 0.20	0.20 ± 0.18	0.028	0.027	1.69	0.61 ± 0.18	0.69 ± 0.17	0.107	0.019	6.18
THA-Usi	29	0.22 ± 0.22	0.21 ± 0.19	−0.023	0.025	1.67	0.68 ± 0.12	0.73 ± 0.16	0.057	0.019	6.79
THA-Vig	21	0.21 ± 0.21	0.22 ± 0.20	0.056	0.028	1.68	0.64 ± 0.17	0.72 ± 0.12	0.125	0.024	6.40
VIG-Tex	22	0.21 ± 0.21	0.21 ± 0.19	−0.004	0.031	1.66	0.69 ± 0.14	0.70 ± 0.15	0.033	0.02	5.91

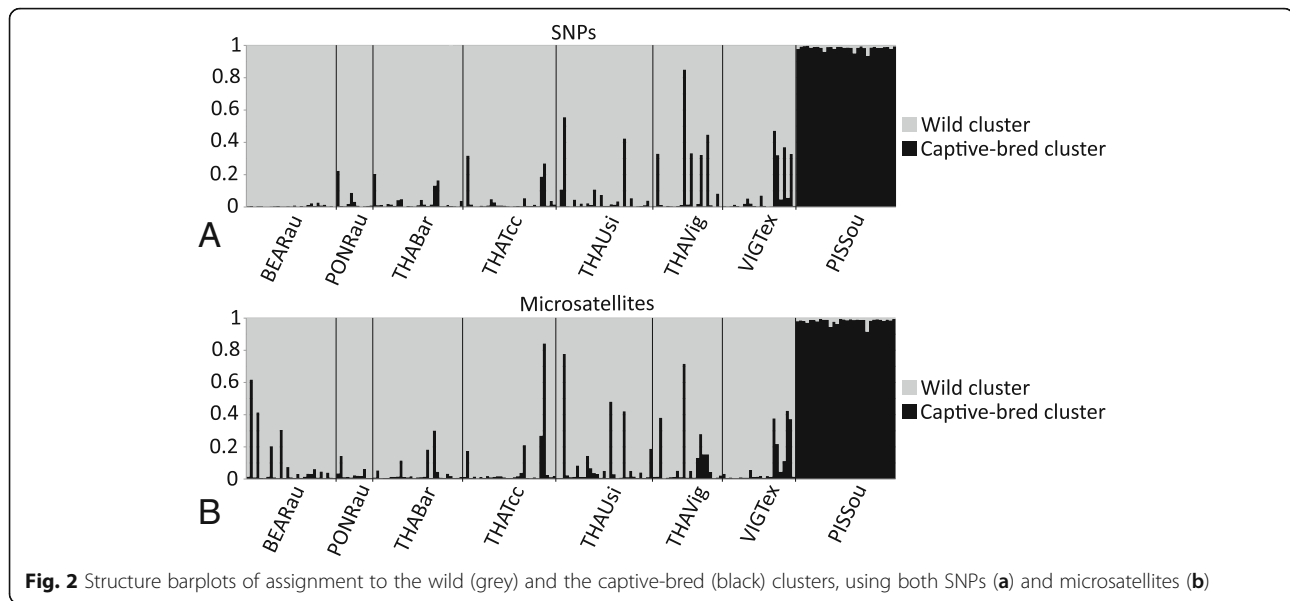


Fig. 2 Structure barplots of assignment to the wild (grey) and the captive-bred (black) clusters, using both SNPs (a) and microsatellites (b)

0.010 to 0.088 for SNPs (mean = 0.033 ± 0.025) and from 0.010 to 0.053 for microsatellites (mean = 0.025 ± 0.013). Pairwise F_{st} values between sites assessed with SNPs and microsatellites were strongly correlated ($r_{Mantel} = 0.92$, $p = 0.001$ based on 1000 permutations), and SNPs had higher values (Fig. 3b, the regression coefficient is significantly higher than the 1:1 line (dotted line on figure) since its 95% CI is 1.46–2.18).

Finally, within the Taurion River, the 92 SNPs displayed a slightly lower informativeness (I) than the 13 microsatellites (2.08 vs. 2.48; Fig. 4). We found that the informativeness of these 92 SNPs is actually equivalent to that of 10 of the microsatellites (Fig. 4). Based on the equation linking the number of our SNPs and informativeness ($I = 0.023 * N_{snps} + 0.0038$; $r^2 = 0.98$, p -value < 0.001), we extrapolated that 108 SNPs are required to be equivalent to the panel of 13 microsatellites in terms of informativeness for individual assignment.

Discussion

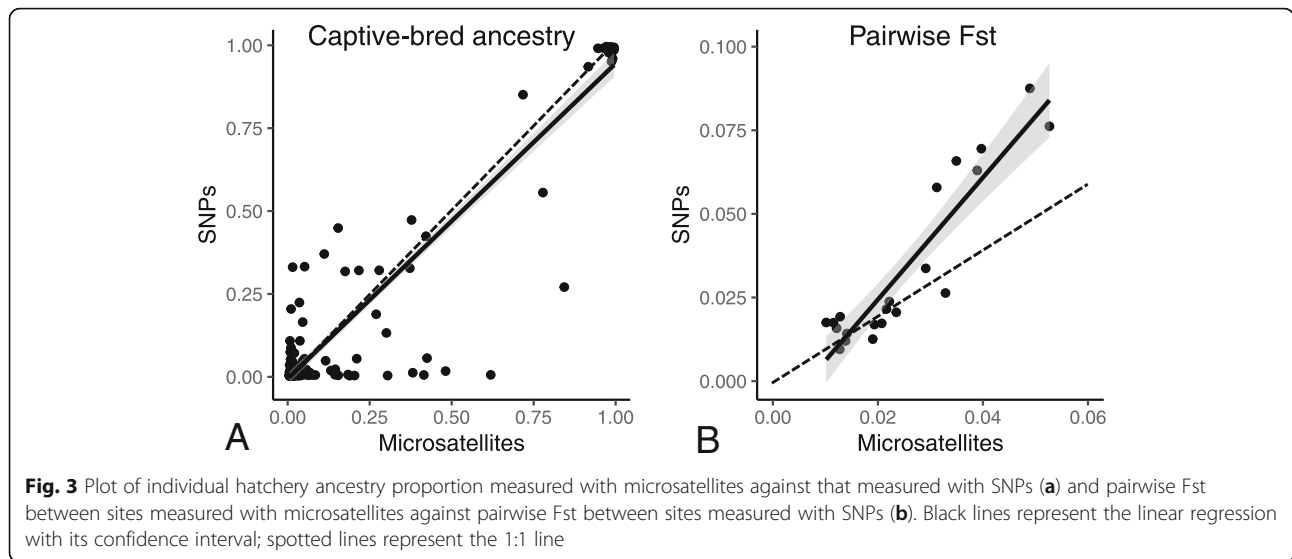
The SNP panel developed here was shown to be efficient to study the population genetics of Atlantic and Mediterranean brown trout lineages from Western Europe, which gathered a huge number of studies in the past decades [60–62]. We provided a panel of 12,204 RADs which are relatively evenly spread across the whole genome, and, from a sample of this panel, we proved its efficiency in terms of genotyping success, and measuring patterns of isolation-by-distance. We also proved that these SNPs successfully detect population structure, which opens new insights for many applications with great potential compared to commonly used markers. These advantages are multiple, going from lower error rates and a simple mutation model with low homoplasy [43] to

usability on poor quality samples. In terms of costs, SNPs and microsatellites are roughly similar, however in terms of time efficiency, SNPs are highly advantageous: samples are directly sent to the genotyping platform, and within a month, data is ready to be analyzed. Moreover, SNPs can be easily reused in other studies, and are more powerful in detecting hybridization [63].

Development of a large variation map for population genomic studies in brown trout

The 12,204 RAD-derived SNPs were relatively homogeneously distributed across the 40 linkage groups of *Salmo trutta* (with 137 to 563 RADs per LG), and showed an average spacing of 0.119 cM. There are some gaps in coverage, which does not necessarily mean that information is missing, they could be due to high recombination rates in these regions or to the positions of the centromeres. Although these gaps exist, genome coverage is satisfying, and with the marker density obtained, it should be sufficient for most genome-scale studies that need to tag a large fraction of genomic variation through linkage disequilibrium. However, this marker density may still be limited when a rapid decay of linkage disequilibrium leaves many genomic regions unattainable, as it might be the case for some applications like the search for loci underlying fitness in the wild [64]. Therefore, an estimate of the local recombination rate for each SNP as an index of the relative power of markers for LD-based mapping approaches is provided in Additional file 1: S1 and S2.

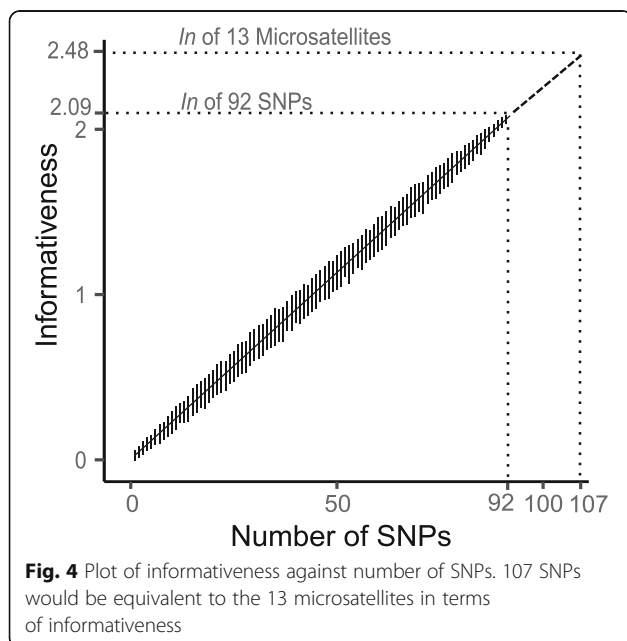
The proposed SNP resource only includes SNPs with a MAF higher than 0.05 (i.e. removing of rare variants), a criterion that has been set for two main reasons. First, SNPs with very low MAF can in many cases be



genotyping errors [65, 66]. Second, we chose a relatively high MAF because this panel of SNPs is primarily designed for studying populations for most of the species range. Therefore, SNPs that are discovered in the populations used to develop the panel are more likely to be polymorphic in other populations from which they were not developed if their MAF is high [67]. Although filtering SNPs on their MAF could lead to ascertainment bias [40, 68, 69], we suggest this is not an issue in our case study because we applied the MAF criteria regarding distinct glacial lineages, which limits marker choice bias. Moreover, for individually-centered investigations such as population structure, kinship and individual assignment,

SNPs with higher MAF were shown to generally be the most powerful [42, 70, 71].

In the test panel, and for the Taurion river (test basin), one locus might be potentially affected by selection. Outlier markers are usually removed before analyses for inferring neutral evolutionary processes, such as genetic drift and gene flow [72]. In our analyses, we decided not to remove it because in the case of detection of admixture between strains, it was shown that these ancestry informative markers can increase accuracy for detecting differentiation and assignment of individuals to populations [68, 73, 74]. Removing loci displaying selection is thus up to the users of the resource, depending on the aim of the study. For instance, if the aim is to determine if a population is at HWE, or to quantify gene flow/inbreeding, or calculate effective population sizes, loci affected by selection should be removed, whereas more individual-based questions do not necessarily require removing these loci.



Genotyping success

Genotyping success was very high in all basins, and except for two basins, the number of polymorphic SNPs among the 162 amplified (from the 192 set) was satisfactory (73 to 99%), confirming the potential versatility of this tool. This set of 162 SNPs already benefits to the field of population genetics for the species as to our knowledge, most studies used less than 40 SNPs [46–48]. Therefore, the 12,204 RAD panel is a promising tool for genome wide studies on brown trout. It would be of interest however to further test this panel on other evolutionary lineages, or on populations which have been shown to have diverged from the continental populations such as western Mediterranean populations found in Iberian Peninsula [25], Italian populations [75], and

Corsican populations [76], as well as on the other main lineages (e.g. Adriatic and Danube lineages; [15]) or remote populations inhabiting at the edges of the species' range (Iran [77, 78], Morocco [16]).

As a first approach we found this panel to be efficient for detecting patterns of isolation-by-distance, although we had no other markers to compare with. Patterns of isolation-by-distance were found to be significant only in the Seuge and the Roya Rivers, although there was a tendency in the other three basins. Interestingly, it seems that even when the number of polymorphic SNPs was low (Seuge River: 92 polymorphic loci), detecting a pattern of isolation-by-distance was still possible. In other basins, we suggest that the strength of the relation between genetic and riparian distance may also be affected by stocking events, or by characteristics of the watershed and demographic histories. However, we did not investigate these issues further in the present manuscript.

Tests of the low-density SNP array

Studies in which low-density SNP arrays equal or outperform a handful of microsatellites for population structure and differentiation are common, particularly when sample sizes are large and populations are strongly structured [47, 79, 80]. Although we found no literature on this particular aspect in brown trout, it has been shown in Atlantic Salmon: genetic divergence, structuring and isolation-by-distance were assessed as successfully using only 7 SNPs and 14 microsatellites, although genetic diversity estimates were less concordant [81]. Twenty-six SNPs were also shown to be nearly as efficient as 16 microsatellites for parentage assignment in this species [82], which makes our set of SNPs appear very promising. The panel of 192 SNPs tested here performed well in terms of detection of admixture and population differentiation, although only 92 SNPs were polymorphic in the Taurion basin. The low level of polymorphism is probably due to the fact that the study scale is extremely low (less than 15 km between the two most extreme sampling sites; Fig. 5) and/or that the biogeographic area in which this river basin is situated has historically low level of diversity (see below).

We found similar introgression level when measured with SNPs compared to microsatellites (Fig. 3a), and higher pairwise F_{st} values between sites with SNPs, suggesting that SNPs have – at least – a similar discriminatory and assignment power. However, the low number of polymorphic SNPs in this panel (92 polymorphic SNPs) compared to other SNP-microsatellite comparison case studies [45, 79, 83] lowered its' informativeness: it was outperformed by the 13 microsatellites, and we found that it would actually require 107 SNPs to be equivalent to microsatellites in terms of informativeness for assignment. However, the advantage of SNPs may here not be accounted for when calculating informativeness. Indeed,

they enable to better detect introgression and admixture compared to microsatellites, and show that individuals considered as « pure » with microsatellites may in fact be introgressed [84]. This difference in individual admixture proportions calculated with microsatellites and SNPs might results from the fact that our panel is characterized by an even repartition of SNPs along the genome of the brown trout, which is expected to improve the global assessment of genome-wide admixture proportions [51]. Hence, the strong advantage of a SNP panel of this type is that it ensures a better representativeness of the entire genome of the brown trout. Moreover, although informativeness is lower, the 92 SNPs still give sufficient information on admixture and differentiation, highly correlated with that given by the microsatellites: the trade-off between cost and power must also be taken into account in regard of the questions asked and the means of the user.

For other empirical case studies, such as the five river basins on which genotyping success was measured, and in which individuals are variable on more loci (except for the Seuge river, other basins showed 119 to 160 polymorphic loci, Table 1), we can expect higher informativeness. We even expect SNPs to equal or outperform microsatellites in terms of informativeness for these basins (indeed, around 107 SNPs should be equivalent to the panel of 13 microsatellites, as pairwise F_{st} values were higher for SNPs than microsatellites), and with the advantages of SNPs mentioned previously. This is particularly true in Mediterranean rivers, probably because they contain both domestic Atlantic and natural Mediterranean ancestries. Indeed, these SNPs were discovered using a mix of Mediterranean and Atlantic individuals, and are therefore more likely to be polymorphic if both lineages are present. Additionally, the mitochondrial and INRA markers were also developed to have a fixed allele in each lineage, explaining why in the Taurion for instance, in which only the Atlantic lineage is present, they were not polymorphic (except for one).

We make this SNP panel freely available as a resource. As it contains many more untested SNPs, future users will be able to choose the number, density, and position of markers in the linkage map, and considering the local recombination rate around each SNP in order to adjust their own panel to their objectives. This should also be considered for the Massif Central Rivers (in our case Taurion and Seuge Rivers), in which the number of polymorphic loci is lower than in the other basins studied here, probably because of the past demographic histories (colonization, connectivity, stocking, bottlenecks, population sizes and habitat). However, as this subpanel was tested on only 7 populations (190 individuals), we first suggest that increasing the sample size should increase statistical power better than increasing the number of SNPs, especially when F_{st} is low (< 0.01 ; [67, 85]).

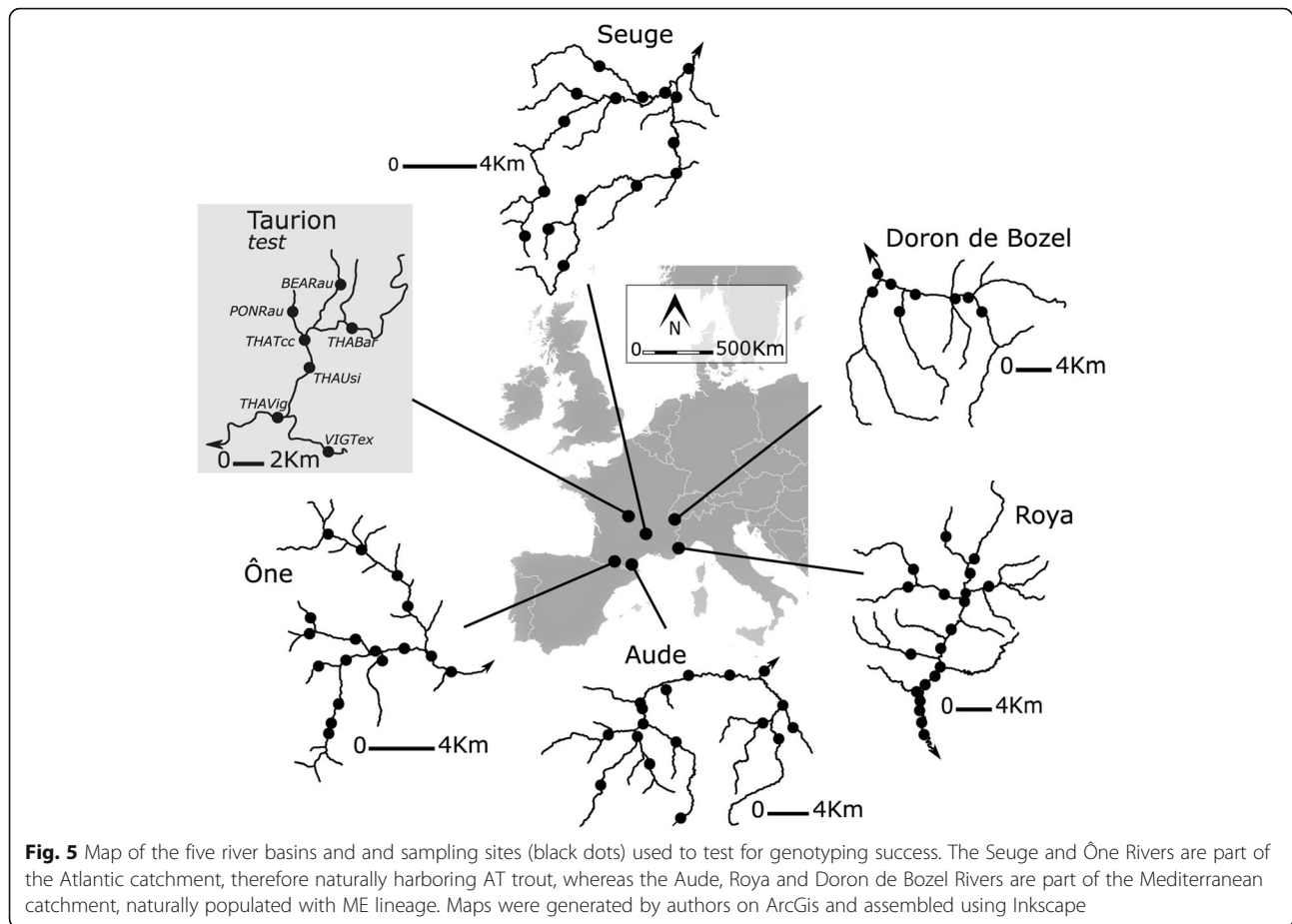


Fig. 5 Map of the five river basins and and sampling sites (black dots) used to test for genotyping success. The Seuge and Ône Rivers are part of the Atlantic catchment, therefore naturally harboring AT trout, whereas the Aude, Roya and Doron de Bozel Rivers are part of the Mediterranean catchment, naturally populated with ME lineage. Maps were generated by authors on ArcGis and assembled using Inkscape

Conclusion

The SNP panel presented here appears as a novel tool to study diverse aspects of population genetics in the brown trout. The possibility to genotype many loci in a fast and affordable way will open many perspectives. It opens new insights into the species life history, with many potential applications both for fundamental population genetics, conservation and management questions, but also for more biological questions such as mapping of quantitative trait loci, or investigating links between genetic and environmental divergence. This resource has the potential to offer high flexibility for many possible applications, outperforming previously used markers in many ways: genome coverage and ancestry detection for instance, but also in terms of cost and efficiency to obtain individual genotypes. We hope that it will be useful to the population geneticists' community working on brown trout and call for future studies across the species' range.

Methods

Development of a genome-wide reference SNP panel

The panel of SNPs markers identified in this study was filtered from the variants described in Leitwein [51] using

double-digest RAD sequencing. Restriction enzymes *EcoRI*-HF and *MspI* were used to digest individual genomic DNA and create the dd-RAD library, which was submitted to size selection in order to retain fragments of 200 to 700 bp using CleanPCR beads. The library was then amplified by PCR and sequenced with Illumina HiSeq2500, producing 125-bp paired-end reads. The initial set contained 75,000 SNPs discovered from 82 wild Mediterranean *S. trutta* from tributaries of the Orb River catchment in southern France, and 102 captive-bred individuals from farms formerly used for stocking in this region (41 and 61 hatchery fish from the ME and AT lineages respectively; see Leitwein [51] for details). These SNPs were anchored to the high density *S. trutta* linkage map using an intermediate step of physical mapping to the Atlantic salmon reference genome: using their relative positions on the Atlantic salmon reference genome, it was possible to determine the relative mapping positions of a large number of additional RAD loci that were not present on the brown trout linkage map [59].

We applied a series of filters which allowed for selecting a panel of SNP markers that were (i) likely to be

highly polymorphic, (ii) mapped on the linkage map, and (iii) present in the two brown trout lineages. To do so, we removed from the initial database all SNPs with a minimum allele frequency (MAF) of 5% or less using *vcftools* [86] based on all individuals (AT + ME). Then, in order to have clean sequences, straightforward to be genotyped, RAD sequences with more than two SNPs, and SNP positions falling in the first or last 30 bp of the RAD sequences were also removed, as well as sequences with undetermined (N) nucleotides. Finally, we kept only RAD sequences for which mapping positions on the *S. trutta* linkage map were determined (see above).

Development of the low-density SNP array

We used this large SNPs resource to develop a low-density SNP array containing 192 SNPs. The goal was to propose a cost-effective tool which holds on two 96-wells genotyping plates for analyzing a large number of individuals. Moreover, SNPs were selected to be informative for population genetics analyses of brown trout populations from at least the AT and ME lineages (sensu Bernatchez [15]) from Western Europe. The array was genotyped using the KASPAR technology® [87] that allows a rapid and cost-effective genotyping service for such a number of markers. We hence applied further filters to fulfill KASPAR genotyping constraints. In particular, RADs with more than one SNP between the primer designing zones (50 bp at the two extremities) were excluded. To ensure a good representativeness of the genome, we selected SNPs evenly spread and spaced by at least 3.5 cM across the 1453 cM estimated length of the *S. trutta* linkage map [59]. This resulted in 245 SNPs (average $1453/245 = 5.9$ cM) among which, in order to retain 182 SNPs, we randomly removed 63 SNPs using the *sample()* R function. Then, we added five ancestry informative SNPs developed by the Institut National de Recherche en Agronomie and the University of Savoie (UMR-0042 CARRTEL, France) that were used in previous studies to distinguish individuals from the AT and ME lineages (OMM1164, OMM1105, OMM1154, Str541INRA, Str591INRA; [29, 87–89]). Finally, we added five mitochondrial SNPs previously used to differentiate among the five main brown trout lineages (mitoDA10Proline, mitoDA10ProlineB, mitoCytob, mitoATPaseIVA, mitoATPaseIVB; [89, 90]; Additional file 1: S1). These numbers of five ancestry informative nuclear or mtDNA markers were chosen to represent approx. 5% of the markers present on the SNP array. This resulted in a low-density array of 192 SNPs markers. Information and sequences are available on Figshare, DOI: <https://doi.org/10.6084/m9.figshare.8174708>

Genotyping success of the low-density SNP array

This low-density SNP array was first evaluated for genotyping success using individuals from five independent

French river basins: two from the Pyrénées mountains (the Ône and the Aude Rivers), two from the Alps mountains (the Roya and the Doron de Bozel Rivers) and one from the Massif Central mountains (the Seuge River) (Fig. 5). Three of these rivers belong to the Mediterranean lineage (ME; the Aude, the Doron and the Roya Rivers), whereas the two others belong to the Atlantic lineage (AT; the One and the Seuge Rivers) (see Fig. 5).

The sampling sessions were performed in 2016, using a single-pass electrofishing approach from a total of 79 sites (between 8 and 21 sites per river basin, Additional file 1: S5), with an aim of sampling 30 individuals of brown trout per site. In total, we captured 1862 individuals (26 individuals per site in average; see Additional file 1: S5) from which a fin clip was taken (after Eugenol anesthesia), and kept in 70% TE Ethanol for genotyping. All individuals were released alive to their site of capture. Fin samples were sent to the LGC Genomics company for DNA extraction and multilocus genotyping of the 192 SNPs markers using KASPAR® [87]. Genotyping success was measured at the individual level, by the proportion of SNPs which were not genotyped (either because not amplified or because the allele could not be read). Finally, these basins were tested for patterns of isolation-by-distance, by using a mantel test with 1000 permutations on pairwise *Fst* matrices (calculated with the pairwise.fst *adegenet* function) and riparian distances (measured in meters with STARS ArcGis package).

Efficiency of the low-density SNPs array

The low-density SNP array was further used for classic population genetic questions in order to compare its efficiency with thirteen microsatellite markers previously used in brown trout population genetic studies (e.g. [91–93]). A total of 190 brown trout individuals were sampled in a small river basin (the Taurion River in the Massif Central Mountains; Fig. 5, “test”) in 2017 using electrofishing. Seven sites were sampled with 21 to 30 individuals per site (Additional file 1: S6). For each individual, a pelvic fin clip was taken for genetic analyses. All individuals were released to their original sampling site. We additionally sampled 30 individuals of domestic Atlantic brown trout from a local hatchery used for stocking purposes (the Soueich trout hatchery), to quantify genetic admixture with wild populations from the Taurion basin. Fin samples were sent to the LGC Genomics Company for DNA extraction and for multilocus genotyping to 192 SNPs markers using the KASPAR® [87]. Note that 30 SNP markers from the initial 192 SNP panel did not amplify (see the Results), and were hence replaced by 30 other SNPs from the 245 SNPs filtered (see methods) to improve the SNP array (see Additional file 1: S2 for details). Additionally, all individuals were genotyped at thirteen microsatellites assembled in PCR multiplexes (see Additional file 1: S7, and Saint-Pé et

al., 2018 for details). We tested whether the selected SNPs were likely influenced by selection using the *F_{st}* outlier detection method implemented in the *fsthet* R package [95], in which outlier values of *F_{ST}* can be identified in a plot of *F_{ST}* vs. heterozygosity [72].

From both SNP and microsatellite datasets, we removed individuals with more than a third of missing data, and kept only individuals for which we had both SNPs and microsatellites genotypes. We first compared genetic admixture between wild and captive-bred strains using STRUCTURE 2.3.1 [96] with the admixture model and the correlated allele frequency model, without prior population information. Twenty runs assuming two clusters ($K = 2$, in order to discriminate between wild and captive-bred individuals, see [94]) were performed with a burn-in period of 200,000 and 200,000 subsequent MCMC repetitions. The ten best runs (highest $\text{LnP}(D)$ values) were compiled using CLUMPP [97] to obtain final averaged individual Q-values. Individuals were assigned to one of the two clusters with the greatest Q-value, provided that value exceeded 0.7 (as in [94]). Individuals with intermediate Q-values were considered genetically admixed individuals between hatchery and wild strains. Individual assignment Q-values to the cluster containing all Soueich hatchery individuals (i.e. degree of assignment to the “captive-bred” cluster = individual level of “hatchery ancestry”) were compared between SNPs and microsatellites using a Spearman correlation test, as admixed ancestry was not normally distributed.

We then compared population differentiation assessment between markers by calculating pairwise *F_{st}* between sites using the *adegenet* R package and *mantel* R function. Finally, we compared the informativeness of both sets of markers for population structure by calculating the informativeness for assignment (I_n , [98]). A higher index indicates a higher informativeness of the set of markers. It was calculated with R as follows: for $i = 1, 2, \dots, K$ populations and $m = 1, 2, \dots, L$ loci, with $K \geq 2$ and $L \geq 1$. Locus m has alleles $j = 1, 2, \dots, N_{(m)}$. The average frequency of allele j at locus m across the K populations is defined as $P_j = \sum_{i=1}^K \frac{P_{ij(m)}}{K}$, where $P_{ij(m)}$ is the relative frequency for allele j of locus m in population i . The informativeness is defined as:

$$I_n = \sum_{i=1}^N \left(-P_j \log P_j + \sum_{i=1}^K \frac{P_{ij}}{K} \log P_{ij} \right)$$

Additional file

Additional file 1: S1. RAD data set of 12204 sequences each containing one or two SNPs, with minimum allele frequency of 5% or higher, no SNPs in the first or last 30 bp of the RAD, no unsequenced nucleotides, and which can be positioned on the *S. trutta* linkage map. S2. SNPs used for low density arrays (5 basins and Taurion test). S3. Isolation by distance patterns for each river basin, represented by plots of pairwise *F_{st}* values against pairwise riparian distance between sites. S4. Pairwise *F_{st}*

values (calculated with *adegenet* R package) between sites measured with SNPs and with microsatellites. S5. Sample sizes and mean (\pm standard error) body length in mm of fish sampled in each site and each basin for testing the 192 SNPs panel's genotyping success. S6. Map of the Taurion River showing sampling points (black dots). Sample sizes by site, and mean (\pm SE) body length in mm from the Taurion River in table. Map was generated by authors on ArcGIS and assembled using Inkscape. S7. Genotyping microsatellites (DOCX 171 kb)

Abbreviations

INRA: Institut national de la recherche agronomique; KASPAR: Kompetitive allele-specific PCR amplification of target sequences and endpoint fluorescence genotyping; LD: Linkage disequilibrium; MAF: Minor allele frequency; PCR: polymerase chain reaction; RAD: Restriction site associated DNA; SNP: Single nucleotide polymorphism

Acknowledgments

We warmly thank the AAPPMA of Luchon (J. Lériné, D. Estrade and colleagues) and of Hérault (E. Ravel and colleagues), ECOGEA (P. Baran and colleagues) and members of EDF (A. Barillier, V. Gouraud and colleagues) for field sampling support. This work was undertaken at SETE, which forms part of the “Laboratoire d'Excellence” (LABEX) entitled TULIP (ANR-10-LABX-41), and M. Leitwein was supported by a grant from LABEX CeMEB (Centre Méditerranéen de l'Environnement et de la Biodiversité, Montpellier), using genotyping and sequencing facilities of ISEM (Institut des Sciences de l'Evolution-Montpellier)

Authors' contributions

KSP, SB, NP and LT designed the experiment and coordinated the study; KSP, ML, BG, LT, NP, GM, JML, PB, and SB conducted sampling; ML, PAG, BG and PB generated the variation map used for developing the SNP panel; KSP, ML, GM and SB carried out the experimental lab work; KSP and ML ran the statistical analyses; KSP, ML, PAG, GB, LT, NP, and SB interpreted the data; KSP, ML, PAG, and SB wrote the first draft of the manuscript; BG, LT, NP, GM, JML and PB read, commented and corrected the initial draft, and all authors gave final approval for publication.

Funding

This work was funded by the French Agency for Biodiversity (n° 136406), Electricité De France (n°8619–5920011137). Two of the authors are attached to the funding body (Tissot Laurence, EDF; Poulet Nicolas, AFB), and have contributed to the design of the study, the collection, analysis, and interpretation of data, and in writing the manuscript.

Availability of data and materials

SNPs sequences and information are available on Figshare. DOI: <https://doi.org/10.6084/m9.figshare.8174708>

Ethics approval and consent to participate

The Direction Départementale des Territoires et de la Mer of departments gave their ethical approval for the field sessions and trout capture.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Centre National de la Recherche Scientifique (CNRS), Université Paul Sabatier (UPS), Station d'Ecologie Théorique et Expérimentale, SETE, UMR 5321, 2 route du CNRS, 09200 Moulis, France. ²Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, QC, Canada. ³ISEM, CNRS, Univ. Montpellier, IRD, EPHE, Montpellier, France. ⁴Present address: Genome - Research and Diagnostic, 697 avenue de Lunel, 34400 Saint-Just, France. ⁵EDF R and D LNHE - Laboratoire National d'Hydraulique et Environnement, 6 quai Watier, 78401 Chatou Cedex, France. ⁶Pôle écohydraulique AFB-IMT, allée du Pr Camille Soula, 31400 Toulouse, France. ⁷ECOGEA, 352 Avenue Roger Tissandier, 31600 Muret, France.

⁸Université de Toulouse, UPS, UMR 5174 (EDB), 118 route de Narbonne, F-31062 Toulouse cedex 4, France.

Received: 29 January 2019 Accepted: 4 July 2019
Published online: 15 July 2019

References

- Lobón-Cervía J, Sanz N, editors. Brown trout: biology, ecology and management. Hoboken: Wiley; 2018.
- Klemetsen A, Amundsen P-A, Dempson JB, Jonsson B, Jonsson N, O'Connell MF, et al. Atlantic salmon *Salmo salar* L., brown trout *Salmo trutta* L. and Arctic charr *Salvelinus alpinus* L.: a review of aspects of their life histories. *Ecol Freshw Fish*. 2003;12:1–59.
- Rahel FJ, Nibbelink NP. Spatial patterns in relations among brown trout (*Salmo trutta*) distribution, summer air temperature, and stream size in Rocky Mountain streams. *Can J Fish Aquat Sci*. 1999;56:43–51.
- Butler JRA, Radford A, Riddington G, Loughton R. Evaluating an ecosystem service provided by Atlantic salmon, sea trout and other fish species in the river Spey, Scotland: the economic impact of recreational rod fisheries. *Fish Res*. 2009;9:6259–66.
- Mills D. Conservation and management of brown trout, *Salmo trutta*, in Scotland: a historical review and the future. *Freshw Biol*. 1989;21:87–98.
- Rahel FJ. Biogeographic barriers, connectivity and homogenization of freshwater faunas: it's a small world after all. *Freshw Biol*. 2007;52:696–710.
- Laike L, Antunes A. Concerted action on identification M and E of Genetic R in the BT (*Salmo trutta*) (projekt). Conservation genetic management of brown trout (*Salmo trutta*) in Europe. Danmarks Fiskeriundersøgelser, Afd. for Ferskvandsfiskeri: Silkeborg; 1999.
- Berrebi P, Poteaux C, Fissier M, Cattaneo-Berrebi G. Stocking impact and allozyme diversity in brown trout from Mediterranean southern France. *J Fish Biol*. 2000;56:949–60.
- Bohling JH. Strategies to address the conservation threats posed by hybridization and genetic introgression. *Biol Conserv*. 2016;203:321–7.
- Elliott JM. Quantitative ecology and the brown trout. Oxford. New York: Oxford University Press; 1994.
- Araki H, Schmid C. Is hatchery stocking a help or harm? *Aquaculture*. 2010;308:52–11.
- Christie MR, Marine ML, French RA, Waples RS, Blouin MS. Effective size of a wild salmonid population is greatly reduced by hatchery supplementation. *Heredity*. 2012;109:254–60.
- Hansen MM. Estimating the long-term effects of stocking domesticated trout into wild brown trout (*Salmo trutta*) populations: an approach using microsatellite DNA analysis of historical and contemporary samples. *Mol Ecol*. 2002;11:1003–15.
- Leitwein M, Gagnaire P-A, Desmarais E, Guendouz S, Rohmer M, Berrebi P, et al. Genome-wide nucleotide diversity of hatchery-reared Atlantic and Mediterranean strains of brown trout *Salmo trutta* compared to wild Mediterranean populations: genome-wide nucleotide diversity of *s. trutta*. *J Fish Biol*. 2016;89:2717–34.
- Bernatchez L. The evolutionary history of brown trout (*Salmo trutta* L.) inferred from phylogeographic, nested clade, and mismatch analyses of mitochondrial DNA variation. *Evolution*. 2001;55:351.
- Tougaard C, Justy F, Guinand B, Douzery EJP, Berrebi P. *Salmo macrostigma* (Teleostei, Salmonidae): nothing more than a brown trout (*S. trutta*) lineage? *J Fish Biol*. 2018;93:302–10.
- Suárez J, Bautista JM, Almodóvar A, Machordom A. Evolution of the mitochondrial control region in Palaearctic brown trout (*Salmo trutta*) populations: the biogeographical role of the Iberian Peninsula. *Heredity*. 2001;87:198–206.
- Susnik S, Schoffmann J, Weiss S. Genetic verification of native brown trout from the Persian Gulf (Catak Cay River, Tigris basin). *J Fish Biol*. 2005;67:879–84.
- Bardakci F, Degerli N, Ozdemir O, Basibuyuk HH. Phylogeography of the Turkish brown trout *Salmo trutta* L.: mitochondrial DNA PCR-RFLP variation. *J Fish Biol*. 2006;68(A):36–55.
- Vera M, Cortey M, Sanz N, García-Marín J-L. Maintenance of an endemic lineage of brown trout (*Salmo trutta*) within the Duero river basin. *J Zool Syst Evol Res*. 2010;48:181–7.
- Vera M, García-Marín JL, Martínez P, Bouza C. Phylogenetic diversity within the endemic brown trout Duero lineage: implications for conservation and management. *Mar Freshw Res*. 2015;66:1066.
- Swatdipong A, Vasemägi A, Niva T, Koljonen M-L, Primmer CR. High level of population genetic structuring in lake-run brown trout, *Salmo trutta*, of the Inari Basin, northern Finland. *J Fish Biol*. 2010;77:2048–71.
- Vøllestad LA, Serbezov D, Bass A, Bernatchez L, Olsen EM, Taugbøl A. Small-scale dispersal and population structure in stream-living brown trout (*Salmo trutta*) inferred by mark-recapture, pedigree reconstruction, and population genetics. *Can J Fish Aquat Sci*. 2012;69:1513–24.
- Berrebi P. Three brown trout *Salmo trutta* lineages in Corsica described through allozyme variation: *salmo trutta* genetics in Corsica. *J Fish Biol*. 2015;86:60–73.
- Cortey M, Pla C, García-Marín J-L. Historical biogeography of Mediterranean brown trout. *Mol Phylogenet Evol*. 2004;33:831–44.
- Cortey M, García-Marín J-L. Evidence for phylogeographically informative sequence variation in the mitochondrial control region of Atlantic brown trout. *J Fish Biol*. 2002;60:1058–63.
- Bohling J, Haffray P, Berrebi P. Genetic diversity and population structure of domestic brown trout (*Salmo trutta*) in France. *Aquaculture*. 2016;462:1–9.
- Caudron A, Champigneulle A, Guyomard R. Assessment of restocking as a strategy for rehabilitating a native population of brown trout *Salmo trutta* L. in a fast-flowing mountain stream in the northern French Alps. *J Fish Biol*. 2006;69:sa:127–39.
- Altukhov IP, Salmenkova EA, Omel'chenko VT. Salmonid fishes: population biology, genetics, and management. Oxford [England], Malden: Blackwell Science; 2000.
- Sato T, Harada Y. Loss of genetic variation and effective population size of Kirikuchi charr: implications for the management of small, isolated salmonid populations. *Anim Conserv*. 2008;11:153–9.
- Pelletier F, Garant D, Hendry AP. Eco-evolutionary dynamics. *Philos Trans R Soc B Biol Sci*. 2009;364:1483–9.
- Araki H, Waples RS, Ardren WR, Cooper B, Blouin MS. Effective population size of steelhead trout: influence of variance in reproductive success, hatchery programs, and genetic compensation between life-history forms: effective population size of steelhead trout. *Mol Ecol*. 2007;16:953–66.
- Bourret V, Kent MP, Primmer CR, Vasemägi A, Karlsson S, Hindar K, et al. SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Mol Ecol*. 2013;22:532–51.
- Stinchcombe JR, Hoekstra HE. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*. 2008;100:158–70.
- Waples RS, Hendry AP. Editorial: special issue: evolutionary perspectives on salmonid conservation and management: evolutionary perspectives on salmonids. *Evol Appl*. 2008;1:183–8.
- Apostolidis AP, Triantaphyllidis C, Kouvatzi A, Economidis PS. Mitochondrial DNA sequence variation and phylogeography among *Salmo trutta* L. (Greek brown trout) populations. *Mol Ecol*. 1997;6:531–42.
- Carlsson J, Olsen KH, Nilsson J, Overli O, Stabell OB. Microsatellites reveal fine-scale genetic structure in stream-living brown trout. *J Fish Biol*. 1999;55:1290–303.
- Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet*. 2011;12:499–510.
- Etter PD, Preston JL, Bassham S, Cresko WA, Johnson EA. Local De novo assembly of RAD paired-end Contigs using short sequencing reads. *PLoS One*. 2011;6:e18561.
- Helyar SJ, Hemmer-Hansen J, Bekkevold D, Taylor MI, Ogdén R, Limborg MT, et al. Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges: Analytical approaches. *Mol Ecol Resour*. 2011;11:123–36.
- Brumfield RT, Beerli P, Nickerson DA, Edwards SV. The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol Evol*. 2003;18:249–56.
- Morin PA, Luikart G, Wayne RK. The SNP workshop group. SNPs in ecology, evolution and conservation. *Trends Ecol Evol*. 2004;19:208–16.
- Morin PA, McCarthy M. Highly accurate SNP genotyping from historical and low-quality samples. *Mol Ecol Notes*. 2007;7:937–46.
- Bradbury IR, Hamilton LC, Dempson B, Robertson MJ, Bourret V, Bernatchez L, et al. Transatlantic secondary contact in Atlantic Salmon, comparing microsatellites, a single nucleotide polymorphism array and restriction-site associated DNA sequencing for the resolution of complex spatial structure. *Mol Ecol*. 2015;24:5130–44.

45. Kaiser SA, Taylor SA, Chen N, Sillett TS, Bondra ER, Webster MS. A comparative assessment of SNP and microsatellite markers for assigning parentage in a socially monogamous bird. *Mol Ecol Resour.* 2017;17:183–93.
46. Paschou P, Ziv E, Burchard EG, Choudhry S, Rodriguez-Cintron W, Mahoney MW, et al. PCA-correlated SNPs for structure identification in worldwide human populations. *PLoS Genet.* 2007;3:e160.
47. Freamo H, O'Reilly P, Berg PR, Lien S, Boulding EG. Outlier SNPs show more genetic structure between two bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs: SNP genotyping and applications. *Mol Ecol Resour.* 2011;11:254–67.
48. Karlsson S, Moen T, Lien S, Glover KA, Hindar K. Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip: SNP genotyping and applications. *Mol Ecol Resour.* 2011;11:247–53.
49. Scheet P, Stephens M. Linkage disequilibrium-based quality control for large-scale genetic studies. *PLoS Genet.* 2008;4:e1000147.
50. Lamaze FC, Sauvage C, Marie A, Garant D, Bernatchez L. Dynamics of introgressive hybridization assessed by SNP population genomics of coding genes in stocked brook charr (*Salvelinus fontinalis*): SNP population genomics in brook charr. *Mol Ecol.* 2012;21:2877–95.
51. Leitwein M, Gagnaire P-A, Desmarais E, Berrebi P, Guinand B. Genomic consequences of a recent three-way admixture in supplemented wild brown trout populations revealed by ancestry tracts. *bioRxiv.* 2018. <https://doi.org/10.1101/302380>.
52. Dominik S, Henshall JM, Kube PD, King H, Lien S, Kent MP, et al. Evaluation of an Atlantic salmon SNP chip as a genomic tool for the application in a Tasmanian Atlantic salmon (*Salmo salar*) breeding population. *Aquaculture.* 2010;308:S56–61.
53. Kochzius M, Nölte M, Weber H, Silkenbeumer N, Hjørleifsdottir S, Hreggvidsson GO, et al. DNA microarrays for identifying fishes. *Mar Biotechnol.* 2008;10:207–17.
54. Lien S, Gidskehaug L, Moen T, Hayes BJ, Berg PR, Davidson WS, et al. A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homologies and striking differences in sex-specific recombination patterns. *BMC Genomics.* 2011;12. <https://doi.org/10.1186/1471-2164-12-615>.
55. Drywa A, Poćwierz-Kotus A, Waś A, Dobosz S, Kent MP, Lien S, et al. Genotyping of two populations of southern Baltic Sea trout *Salmo trutta m. trutta* using an Atlantic salmon derived SNP-array. *Mar Genomics.* 2013;9:25–32.
56. Pustovrh G, Sušnik Bajec S, Snoj A. A set of SNPs for *Salmo trutta* and its application in supplementary breeding programs. *Aquaculture.* 2012;370–371:102–8.
57. Sušnik Bajec S, Pustovrh G, Jesenšek D, Snoj A. Population genetic SNP analysis of marble and brown trout in a hybridization zone of the Adriatic watershed in Slovenia. *Biol Conserv.* 2015;184:239–50.
58. Linløkken AN, Haugen TO, Kent MP, Lien S. Genetic differences between wild and hatchery-bred brown trout (*Salmo trutta* L.) in single nucleotide polymorphisms linked to selective traits. *Ecol Evol.* 2017;7:4963–72.
59. Leitwein M, Guinand B, Pouzadoux J, Desmarais E, Berrebi P, Gagnaire P-A. A Dense Brown Trout (*Salmo trutta*) Linkage map reveals recent chromosomal rearrangements in the *Salmo* genus and the impact of selection on linked neutral diversity. *G3 (Bethesda).* 2017;7:1365–76.
60. Almodóvar A, Nicola GG, Ayllón D, Elvira B. Global warming threatens the persistence of Mediterranean brown trout. *Glob Change Biol.* 2012;18:1549–60.
61. Aurelle D, Berrebi P. Genetic structure of brown trout (*Salmo trutta*, L.) populations from South-Western France: data from mitochondrial control region variability. *Mol Ecol.* 2001;10:1551–61.
62. Poteaux C. Microsatellite polymorphism and genetic impact of restocking in Mediterranean brown trout (*Salmo trutta* L.). *Heredity.* 1999;82:645.
63. McFarlane SE, Pemberton JM. Detecting the true extent of introgression during anthropogenic hybridization. *Trends Ecol Evol.* 2019;34:315–26.
64. Kaeuffer R, Réale D, Coltman DW, Pontier D. Detecting population STRUCTURE using STRUCTURE software: effect of background linkage disequilibrium. *Heredity.* 2007;99:374–80.
65. Ahn K, Gordon D, Finch SJ. Increase of rejection rate in case-control studies with the differential genotyping error rates. *Stat Appl Genet Mol Biol.* 2009;8:1–9.
66. Wang S, Sha Z, Sonstegard TS, Liu H, Xu P, Somridhivej B, et al. Quality assessment parameters for EST-derived SNPs from catfish. *BMC Genomics.* 2008;9:450.
67. Morin PA, Martien KK, Taylor BL. Assessing statistical power of SNPs for population structure and conservation studies. *Mol Ecol Resour.* 2009;9:66–73.
68. Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D. Population genomics of marine fishes: identifying adaptive variation in space and time. *Mol Ecol.* 2009;18:3128–50.
69. Rosenblum EB, Novembre J. Ascertainment bias in spatially structured populations: a case study in the eastern fence lizard. *J Hered.* 2007;98:331–6.
70. Krawczak M. Informativity assessment for biallelic single nucleotide polymorphisms. *Electrophoresis.* 1999;20:1676–81.
71. Ryman N, Palm S, André C, Carvalho GR, Dahlgren TG, Jorde PE, et al. Power for detecting genetic divergence: differences between statistical methods and marker loci: power for detecting divergence. *Mol Ecol.* 2006;15:2031–45.
72. Beaumont MA, Nichols RA. Evaluating loci for use in the genetic analysis of population structure. *Proc R Soc Lond B Biol Sci.* 1996;263:1619–26.
73. Funk WC, McKay JK, Hohenlohe PA, Allendorf FW. Harnessing genomics for delineating conservation units. *Trends Ecol Evol.* 2012;27:489–96.
74. Gagnaire P-A, Broquet T, Aurelle D, Viard F, Souissi A, Bonhomme F, et al. Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evol Appl.* 2015;8:769–86.
75. Splendiani A, Ruggeri P, Giovannotti M, Pesaresi S, Occhipinti G, Fioravanti T, et al. Alien brown trout invasion of the Italian peninsula: the role of geological, climate and anthropogenic factors. *Biol Invasions.* 2016;18:2029–44.
76. Berrebi P, Caputo Barucchi V, Splendiani A, Muracciole S, Sabatini A, Palmas F, et al. Brown trout (*Salmo trutta* L.) high genetic diversity around the Tyrrhenian Sea as revealed by nuclear and mitochondrial markers. *Hydrobiologia.* 2019;826:209–31.
77. Osinov AG. Iranian samples of brown trout *Salmo trutta* or rainbow trout *Oncorhynchus mykiss*: comparison of evidence from allozyme and mitochondrial DNA control region sequence analysis. *J Ichthyol.* 2009;49:825–8.
78. Hashemzadeh Segherloo I, Farahmand H, Abdoli A, Bernatchez L, Primmer CR, Swatdipong A, et al. Phylogenetic status of brown trout *Salmo trutta* populations in five rivers from the southern Caspian Sea and two inland lake basins, Iran: a morphogenetic approach. *J Fish Biol.* 2012;81:1479–500.
79. Gärke C, Ytornel F, Bed'hom B, Gut I, Lathrop M, Weigend S, et al. Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations: population differentiation: SNPs vs SSRs. *Anim Genet.* 2012;43:419–28.
80. Liu N, Chen L, Wang S, Oh C, Zhao H. Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genet.* 2005;6(Suppl 1):S26.
81. Ryyanen HJ, Tonteri A, Vasemagi A, Primmer CR. A comparison of Biallelic markers and microsatellites for the estimation of population and conservation genetic parameters in Atlantic Salmon (*Salmo salar*). *J Hered.* 2007;98:692–704.
82. Rengmark AH, Slettan A, Skaala Ø, Lie Ø, Lingaas F. Genetic variability in wild and farmed Atlantic salmon (*Salmo salar*) strains estimated by SNP and microsatellites. *Aquaculture.* 2006;253:229–37.
83. Hauser L, Baird M, Hilborn R, Seeb LW, Seeb JE. An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population: ANALYTICAL APPROACHES. *Mol Ecol Resour.* 2011;11:150–61.
84. Haasl RJ, Payseur BA. Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. *Heredity.* 2011;106:158–71.
85. Kalinowski ST. Do polymorphic loci require large sample sizes to estimate genetic distances? *Heredity.* 2005;94:33–6.
86. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27:2156–8.
87. Smith SM, Maughan PJ. SNP genotyping using KASPar assays. In: Batley J, editor. *Plant genotyping.* New York: Springer New York; 2015. p. 243–56. https://doi.org/10.1007/978-1-4939-1966-6_18.
88. Harrang E, Guyomard R, Caudron A. Ready-to-use diagnostic molecular markers for assessment and monitoring of hybridization between native and non-native domesticated brown trout (*Salmo trutta*): case studies in French alpine streams. *Girona: Advances in the Population Ecology of Stream Salmonids – IV;* 2015.
89. Bernatchez L, Guyomard R, Bonhomme F. DNA sequence variation of the mitochondrial control region among geographically and

- morphologically remote European brown trout *Salmo trutta* populations. *Mol Ecol.* 1992;1:161–73.
90. Giuffra E, Bernatchez L, Guyomard R. Mitochondrial control region and protein coding genes sequence variation among phenotypic forms of brown trout *Salmo trutta* from northern Italy. *Mol Ecol.* 1994;3:161–71.
 91. Estoup A, Rousset F, Michalakis Y, Cornuet J-M, Adriamanga M, Guyomard R. Comparative analysis of microsatellite and allozyme markers: a case study investigating microgeographic differentiation in brown trout (*Salmo trutta*). *Mol Ecol.* 1998;7:339–53.
 92. Grimholt U, Drabløs F, Jørgensen S, Høyheim B, Stet R. The major histocompatibility class I locus in Atlantic salmon (*Salmo salar* L.): polymorphism, linkage analysis and protein modelling. *Immunogenetics.* 2002;54:570–81.
 93. Slettan A, Olsaker I, Lie O. Atlantic salmon, *Salmo salar*, microsatellites at the *SSOSL25*, *SSOSL85*, *SSOSL311*, *SSOSL417* loci. *Anim Genet.* 1995;26:281–2.
 94. Saint-Pé K, Blanchet S, Tissot L, Poulet N, Plasseraud O, Loot G, et al. Genetic admixture between captive-bred and wild individuals affects patterns of dispersal in a brown trout (*Salmo trutta*) population. *Conserv Genet.* 2018;19:1269–79.
 95. Flanagan SP, Jones AG. Constraints on the FST-heterozygosity outlier approach. *J Hered.* 2017;108:561–73.
 96. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155:945–59.
 97. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics.* 2007;23:1801–6.
 98. Rosenberg NA, Li LM, Ward R, Pritchard JK. Informativeness of genetic markers for inference of ancestry. *Am J Hum Genet.* 2003;73:1402–22.

Publisher's Note

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

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

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

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

