Conservation Genetics Resources (2021) 13:369–373
https://doi.org/10.1007/s12686-021-01215-6

TECHNICAL NOTE

Development of SNP markers derived from RAD sequencing for Atlantic salmon (*Salmo salar* L.) inhabiting the rivers of southern England

R. Andrew King[1] · **Jamie R. Stevens**[1]

Received: 23 October 2020 / Accepted: 13 May 2021 / Published online: 20 May 2021

Abstract
The rivers of the Hampshire Basin, southern England contain a genetically unique group of Atlantic salmon that have suffered dramatic declines in numbers over the last 40 years. Knowledge of levels and patterns of genetic diversity is essential for effective management of these vulnerable populations. Using restriction site-associated DNA sequencing (RADseq) data, we describe the development and characterisation of a panel of 94 single nucleotide polymorphism (SNP) loci for salmon from this region and investigate their applicability and variability in both target (i.e. southern English) and non-target populations. The SNP loci will be useful for population genetic and assignment studies on Atlantic salmon within the UK and beyond.

Keywords Atlantic salmon · Conservation genetics · Management · *Salmo salar*

In recent years there has been an increase in the use of reduced representation sequencing data to develop cost-effective resources for monitoring genetic diversity (von Thaden et al 2020), with single nucleotide polymorphisms (SNPs) now used widely for non-model species such as salmonid fishes. Atlantic salmon (*Salmo salar* L.—hereafter referred to as salmon) is an iconic fish species that, due to its predominantly anadromous life history, is subject to multiple stressors in both its freshwater and marine habitats.

The chalk streams of the Hampshire Basin, southern England contain a genetically unique group of salmon (Finnegan et al. 2013; Ikediashi et al. 2018) which, like populations across the entire range of the species, have suffered severe declines in abundance over the last 40 years (ICES 2017). Within the Hampshire Basin chalk streams, salmon are found in only seven rivers and are threatened by agricultural pollution and over-abstraction of water (WWF 2014). Additionally, these populations have lower levels of genetic diversity (based on analysis of microsatellite loci) than neighbouring rivers in southwest England (Finnegan et al. 2013; Ikediashi et al. 2018). There is therefore a need for efficient genetic monitoring of these populations to aid in their management and conservation. Here, we develop a panel of SNP markers for use in population genetic and assignment studies in salmon populations from southern England and assess their variability in both southern English salmon and salmon from rivers outside of the target area.

DNA was extracted from adipose finclips using Qiagen Blood and Tissue kits for 84 individuals sampled from 11 rivers from southern England (Supplementary Table 1). RAD libraries were prepared in-house and 125 bp paired-end sequencing was undertaken on an Illumina HiSeq 2500 in rapid run mode (Supplementary Methods). Data was analysed using STACKS v2.41 (Rochette et al. 2019). Sequences were demultiplexed and trimmed to 122 bp using the process_radtags module. RAD loci were built using the denovo_map.pl pipeline using optimised parameters of $M = 1$ and $n = 1$ (Paris et al. 2017). The populations module was run to retain RAD loci found in at least 85% of the individuals in each of the 11 populations with a maximum observed heterozygosity of 0.7 and a minimum minor allele frequency of 0.05. A list of RAD loci containing only a single SNP were obtained and information on RAD locus length, sequence and variable SNP position were extracted from the population module outputs.

Loci were ranked based on G''_{ST}, calculated using the MMOD R library (Winter 2012). The top 700 loci were

✉ R. Andrew King
r.a.king@exeter.ac.uk

[1] Hatherly Laboratories, Department of Biosciences, College of Life and Environmental Sciences, University of Exeter, Hatherly Building, Prince of Wales Road, Exeter EX4 4PS, UK

Table 1 Summary statistics for 94 single nucleotide polymorphisms developed for Atlantic salmon (*Salmo salar*)

Locus	SNP	Linkage Group	H_o	H_e	MAF	F_{IS}	P_{HWE}
Ssa_76064	C/A	ssa01	0.219	0.211	0.186	−0.036	0.409
Ssa_21324	T/C	ssa01	0.447	0.388	0.322	−0.151	0.021
Ssa_10857	T/A	ssa01	0.319	0.302	0.194	−0.054	0.220
Ssa_30724	C/G	ssa01	0.297	0.329	0.253	0.097	0.165
Ssa_68872	A/G	ssa01	0.427	0.427	0.322	0.001	0.465
Ssa_75585	A/G	ssa01	0.180	0.202	0.123	0.109	0.128
Ssa_22383	A/C	ssa01	0.214	0.199	0.130	−0.073	0.305
Ssa_70410	T/G	ssa01	0.293	0.287	0.196	−0.021	0.409
Ssa_87021	G/T	ssa01	0.397	0.424	0.412	0.064	0.176
Ssa_16534	A/G	ssa01	0.375	0.404	0.428	0.071	0.208
Ssa_83524	A/G	ssa01	0.224	0.220	0.168	−0.020	0.488
Ssa_73614	T/C	ssa01	0.401	0.409	0.319	0.021	0.371
Ssa_70574	T/A	ssa02	0.281	0.217	0.144	−0.293	0.000
Ssa_26579	T/C	ssa03	0.381	0.413	0.318	0.078	0.138
Ssa_82536	C/T	ssa04	0.267	0.286	0.199	0.066	0.219
Ssa_11045	G/A	ssa04	0.385	0.362	0.293	−0.066	0.213
Ssa_12205	C/T	ssa04	0.536	0.464	0.476	−0.155	0.021
Ssa_16361	G/C	ssa05	0.516	0.446	0.497	−0.157	0.018
Ssa_57180	G/A	ssa05	0.411	0.379	0.495	−0.086	0.157
Ssa_24657	G/A	ssa05	0.448	0.425	0.423	−0.053	0.278
Ssa_677051	C/A	ssa05	0.143	0.131	0.072	−0.091	0.283
Ssa_13920	T/G	ssa05	0.229	0.195	0.122	−0.176	0.029
Ssa_59923	A/G	ssa06	0.388	0.419	0.439	0.074	0.181
Ssa_69544	G/T	ssa06	0.344	0.359	0.299	0.042	0.295
Ssa_six6	A/G	ssa09	0.343	0.315	0.251	−0.089	0.117
Ssa_31387	A/G	ssa09	0.417	0.391	0.348	−0.066	0.173
Ssa_81011	G/C	ssa09	0.278	0.292	0.263	0.049	0.320
Ssa_45675	C/G	ssa09	0.375	0.331	0.394	−0.134	0.078
Ssa_92668	G/C	ssa10	0.385	0.377	0.356	−0.021	0.434
Ssa_63331	T/G	ssa10	0.235	0.230	0.162	−0.023	0.425
Ssa_38395	G/T	ssa10	0.427	0.446	0.346	0.043	0.313
Ssa_16383	T/C	ssa10	0.054	0.085	0.048	0.358	0.010
Ssa_23042	A/T	ssa10	0.290	0.273	0.186	−0.061	0.276
Ssa_87214	T/A	ssa10	0.142	0.163	0.098	0.127	0.116
Ssa_73787	G/T	ssa10	0.198	0.185	0.122	−0.072	0.240
Ssa_72732	G/A	ssa10	0.157	0.176	0.112	0.110	0.074
Ssa_11025	A/C	ssa10	0.508	0.434	0.436	−0.169	0.012
Ssa_62568	T/C	ssa11	0.228	0.202	0.310	−0.132	0.083
Ssa_889	T/A	ssa12	0.297	0.252	0.178	−0.180	0.016
Ssa_73710	C/G	ssa12	0.130	0.130	0.077	−0.001	0.669
Ssa_69553	T/G	ssa12	0.394	0.368	0.439	−0.070	0.221
Ssa_11069	C/A	ssa13	0.219	0.203	0.133	−0.079	0.280
Ssa_69001	G/T	ssa13	0.061	0.064	0.035	0.042	0.547
Ssa_24091	G/A	ssa13	0.176	0.182	0.165	0.035	0.348
Ssa_69865	G/C	ssa13	0.290	0.294	0.449	0.014	0.433
Ssa_84255	A/T	ssa13	0.361	0.349	0.311	−0.033	0.322
Ssa_24528	C/T	ssa13	0.250	0.211	0.149	−0.181	0.045
Ssa_67740	G/C	ssa14	0.366	0.333	0.255	−0.101	0.143
Ssa_52153	C/A	ssa14	0.328	0.272	0.162	−0.204	0.002
Ssa_62807	C/T	ssa15	0.474	0.449	0.449	−0.056	0.242
Ssa_58789	C/T	ssa15	0.179	0.161	0.128	−0.112	0.141

Table 1 (continued)

Locus	SNP	Linkage Group	H_o	H_e	MAF	F_{IS}	P_{HWE}
Ssa_28584	G/A	ssa16	0.083	0.077	0.043	−0.079	0.537
Ssa_71340	A/C	ssa16	0.438	0.454	0.476	0.036	0.297
Ssa_35217	G/T	ssa16	0.227	0.260	0.173	0.126	0.098
Ssa_81295	A/G	ssa16	0.327	0.334	0.285	0.023	0.368
Ssa_87179	T/C	ssa16	0.375	0.408	0.471	0.080	0.157
Ssa_49994	A/G	ssa16	0.444	0.409	0.463	−0.086	0.139
Ssa_27186	T/C	ssa16	0.485	0.419	0.374	−0.157	0.022
Ssa_25589	C/A	ssa16	0.334	0.330	0.255	−0.013	0.453
Ssa_15660	C/T	ssa17	0.439	0.395	0.327	−0.111	0.077
Ssa_42784	A/G	ssa18	0.441	0.387	0.346	−0.141	0.032
Ssa_73197	C/G	ssa18	0.282	0.270	0.213	−0.043	0.347
Ssa_21426	A/C	ssa18	0.120	0.110	0.061	−0.092	0.352
Ssa_585746	C/T	ssa18	0.305	0.320	0.234	0.047	0.322
Ssa_47781	C/A	ssa19	0.246	0.270	0.178	0.087	0.124
Ssa_1354	A/G	ssa19	0.228	0.250	0.165	0.088	0.204
Ssa_88229	C/T	ssa19	0.177	0.177	0.117	−0.001	0.477
Ssa_69384	G/A	ssa19	0.443	0.467	0.405	0.051	0.245
Ssa_41749	T/A	ssa19	0.408	0.397	0.293	−0.028	0.347
Ssa_25077	G/A	ssa19	0.143	0.117	0.069	−0.219	0.061
Ssa_88509	C/T	ssa20	0.241	0.214	0.128	−0.125	0.088
Ssa_13237	G/C	ssa21	0.330	0.327	0.246	−0.007	0.515
Ssa_55742	G/A	ssa21	0.422	0.380	0.463	−0.110	0.091
Ssa_74950	A/C	ssa21	0.252	0.271	0.181	0.070	0.161
Ssa_61219	G/A	ssa22	0.245	0.267	0.208	0.084	0.138
Ssa_94352	A/G	ssa22	0.473	0.491	0.449	0.036	0.327
Ssa_62594	G/A	ssa22	0.305	0.295	0.184	−0.031	0.385
Ssa_45327	G/A	ssa23	0.314	0.366	0.404	0.143	0.033
Ssa_55100	G/A	ssa24	0.230	0.223	0.209	−0.031	0.440
Ssa_75012	G/C	ssa24	0.439	0.467	0.418	0.060	0.215
Ssa_18895	T/G	ssa24	0.123	0.140	0.177	0.120	0.239
Ssa_vgll3	T/C	ssa25	0.419	0.413	0.399	−0.013	0.395
Ssa_94577	G/A	ssa25	0.346	0.323	0.250	−0.071	0.190
Ssa_32312	T/C	ssa25	0.327	0.312	0.229	−0.050	0.306
Ssa_24312	T/G	ssa25	0.277	0.265	0.173	−0.044	0.354
Ssa_8495	T/C	ssa28	0.371	0.421	0.439	0.118	0.048
Ssa_5024	A/G	ssa28	0.464	0.476	0.481	0.026	0.399
Ssa_72208	A/C	ssa28	0.361	0.356	0.285	−0.014	0.454
Ssa_30875	T/G	ssa28	0.356	0.385	0.396	0.075	0.194
Ssa_36501	T/C	ssa29	0.136	0.182	0.120	0.250	0.011
Ssa_64369	G/T	ssa29	0.281	0.261	0.191	−0.076	0.253
Ssa_34412	T/C	ssa29	0.302	0.296	0.335	−0.019	0.397
Ssa_70399	A/G	ssa29	0.226	0.289	0.278	0.218	0.008
Ssa_36430	A/G	ssa29	0.357	0.359	0.271	0.006	0.445

H_o—observed heterozygosity; H_e—expected heterozygosity; MAF—minor allele frequency; F_{IS}—inbreeding coefficient; P_{HWE}—*p* value for tests of Hardy–Weinberg equilibrium

aligned against the Atlantic salmon reference genome (ICSASG_v2—Lien et al. 2016) using BLAST. To determine the genic nature of each RAD locus, i.e. non-coding, intron, exon etc., we used the Genome Browser on SalmoBase (https://salmobase.org, Samy et al. 2017). A whole genome duplication event in the ancestor of salmonid species approximately 80 MYA has resulted in high levels of gene duplication and synteny between different chromosomes in Atlantic

Table 2 Basic diversity metrics for eight populations of British Atlantic salmon

Pop	N	Year	Tissue	Ho	He	F_{IS}	%P
Eden	24	2009	finclip	0.289	0.275	−0.051	88.30%
Wye	24	2012	finclip	0.309	0.314	0.016	92.55%
Fowey	24	2004	scale	0.332	0.305	−0.089	90.43%
Tamar	24	2018	finclip	0.316	0.319	−0.009	91.49%
Exe	24	2013	finclip	0.323	0.317	−0.019	92.55%
Frome	22	2018	scale	0.301	0.293	−0.027	86.17%
Test	24	2017	finclip	0.295	0.293	−0.007	89.36%
Bresle	22	2019	scale	0.322	0.327	0.015	94.68%

N number of individuals, *Ho* Observed Heterozygosity, *He* expected Heterozygosity, F_{IS} inbreeding coefficient; *%P* percentage polymorphic loci

salmon (Lien et al. 2016). We therefore retained only those loci aligning to a single genomic location.

SNP genotyping was undertaken on the Fluidigm EP1 Genotyping System using 96.96 Dynamic Genotyping Arrays and scored using the Fluidigm SNP Genotyping analysis software. We tested the loci on DNA extracted from both contemporary and archive tissues. Of the 106 loci tested, 12 did not give the expected genotype clusters (Supplementary Fig. 1) resulting in a final panel of 94 loci (Table 1, Supplementary Table 2). To demonstrate the utility of our SNP loci to detect variation in salmon populations, we screened loci for variation in salmon from eight rivers – five 'target' rivers from southern England (Tamar, Fowey, Exe, Frome and Test) and three 'non-target' rivers: Bresle (northern France), Wye (south Wales) and Eden (northwest England). The number of loci per linkage group ranged from 0 to 12 (Table 1). All loci were polymorphic in at least three populations. Observed and expected heterozygosity ranged from 0.054 to 0.536 and 0.064 to 0.491, respectively (Table 1). Minor allele frequency and F_{IS} ranged from 0.035 to 0.497 and -0.293 to 0.358, respectively (Table 1). At the population level, observed and expected heterozygosity ranged from 0.289 to 0.332 and 0.275 to 0.327, respectively (Table 2). Percentage of polymorphic loci ranged from 86.17% to 94.68% (Table 2). Three significant cases of linkage between pairs of markers were found: two pairs of loci (Ssa_41748-Ssa_25077 and Ssa_25077-Ssa1354) on linkage group ssa19 in the Bresle sample and one pair in the Frome sample on linkage group ssa01 (Ssa_76064-Ssa_68872). Population pairwise G''st (Supplementary Table 3) values ranged from 0.016 (Frome v Test) to 0.440 (Eden v Frome).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12686-021-01215-6.

Acknowledgements This research was funded by the European Union Interreg France England Channel programme project 'Salmonid Management Around the Channel' (SAMARCH), with additional funding from Southern Water. We thank Karen Moore and Audrey Farbos of the Exeter Sequencing Service for construction and sequencing of the RAD libraries and Tim Sykes, Georgina Samoluk, Dominic Longley and Rob Hillman (Environment Agency) and Rasmus Lauridsen and Celine Artero (Game and Wildlife Conservation Trust) for coordinating sample collection.

Author contributions JRS obtained funding; RAK undertook all wet lab work, bioinformatics and data analyses; RAK & JRS wrote the manuscript.

Funding European Union Interreg France England Channel programme project 'Salmonid Management Around the Channel' (SAMARCH), with additional funding from Southern Water.

Data availability SNP genotypes are available upon request from the lead author.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

Finnegan AK, Griffiths AM, King RA, Machado-Schaffino G, Porcher JP, Garcia-Vazquez E, Bright D, Stevens JR (2013) Use of multiple markers demonstrates a cryptic western refugium and postglacial colonisation routes of Atlantic salmon (*Salmo salar* L.) in western Europe. Heredity 111:34–43. https://doi.org/10.1038/hdy.2013.17

ICES (2017) Report of the Working Group on North Atlantic Salmon (WGNAS), 29 March–7 April 2017, Copenhagen, Denmark. ICES CM 2017/ACOM:20. 296 pp.

Ikediashi C, Paris JR, King RA, Beaumont WRC, Ibbotson A, Stevens JR (2018) Atlantic salmon *Salmo salar* in the chalk streams of England are genetically unique. J Fish Biol 92:621–641. https://doi.org/10.1111/jfb.13538

Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T et al (2016) The Atlantic salmon genome provides insights into rediploidization. Nature 533:200–205. https://doi.org/10.1038/nature17164

Paris JR, Stevens JR, Catchen JM (2017) Lost in parameter space: a road map for stacks. Methods Ecol Evol 8:1360–1373. https://doi.org/10.1111/2041-210X.12775

Rochette NC, Rivera-Colón AG, Catchen JM (2019) Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol Ecol 28:4737–4754. https://doi.org/10.1111/mec.15253

Samy JKA, Mulugeta TD, Nome T, Sandve SR, Grammes F, Kent MP et al (2017) SalmoBase: an integrated molecular data resource for Salmonid species. BMC Genomics 18:482. https://doi.org/10.1186/s12864-017-3877-1

von Thaden A, Nowak C, Tiesmeyer A, Reiners TE, Alves PC, Lyons LA et al (2020) Applying genomic data in wildlife monitoring: development guidelines for genotyping degraded samples with reduced single nucleotide polymorphism panels. Mol Ecol Res 20:662–680. https://doi.org/10.1111/1755-0998.13136

Winter DJ (2012) MMOD: an R library for the calculation of population differentiation statistics. Mol Ecol Res 12:1158–1160. https://doi.org/10.1111/j.1755-0998.2012.03174.x

WWF—World Wide Fund for Nature (2014) The State of England's Chalk Streams. Available from: https://www.wwf.org.uk/updates/state-englands-chalk-streams