

# Rapid and effective isolation of candidate sequences for development of microsatellite markers in 30 fish species by using kit-based target capture and multiplexed parallel sequencing

Hirohiko Takeshima<sup>1,2</sup> · Nozomu Muto<sup>2,3</sup> · Yasuyuki Sakai<sup>4</sup> · Naoya Ishiguro<sup>4</sup> · Keiichiro Iguchi<sup>5,6</sup> · Satoshi Ishikawa<sup>2</sup> · Mutsumi Nishida<sup>1,7</sup>

Received: 16 May 2016 / Accepted: 3 February 2017 / Published online: 27 February 2017  
© The Author(s) 2017. This article is an open access publication

**Abstract** Recent advances in next-generation sequencing (NGS) technology have accelerated the development of microsatellite markers for wildlife conservation genetics. Although the discovery of microsatellite-containing sequences based on NGS is more efficient with sequencing of a microsatellite-enriched library than with whole-genome shotgun sequencing, the process of constructing a microsatellite-enriched library is somewhat complicated. Therefore, many researchers prefer to use external services for the microsatellite-enrichment, which requires more time. To facilitate the rapid and effective development of

novel microsatellite markers, we attempted to simplify the process of constructing a microsatellite-enriched library for multiplexed parallel sequencing. To capture microsatellite-containing sequences, we applied an easy-to-use commercially available kit for the hybridization and wash steps. After preparing shotgun libraries of 30 fish species for NGS, we captured microsatellite-containing DNA fragments directly from the shotgun libraries by using the commercially available kit. Next, three runs of multiplexed parallel sequencing were conducted on the 454 GS Junior platform. The resulting sequences for each species included high proportions of microsatellite-containing sequences (from 46 to 79%). Thus, sufficient numbers of primer sets, ranging from 1029 to 6606, were effectively designed for each species. Microsatellite capture and sequencing were completed in about a week, so the time required was substantially reduced. To validate the effectiveness of our strategy, we screened 44 potential primer sets designed for ayu (*Plecoglossus altivelis*). The results of polymorphisms revealed that allelic variability at 23 markers will be useful for studying population structure. These results prove the effectiveness of our improved approach for microsatellite marker development.

The original version of this article was revised due to a retrospective Open Access order.

✉ Hirohiko Takeshima  
takeshim@chikyu.ac.jp

- <sup>1</sup> Atmosphere and Ocean Research Institute, University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa, Chiba 277-8564, Japan
- <sup>2</sup> Research Institute for Humanity and Nature, 457-4 Motoyama, Kamigamo, Kita-ku, Kyoto 603-8047, Japan
- <sup>3</sup> School of Biological Science, Tokai University, 1-1 1-Chome 5-Jo Minami-sawa, Minami-ku, Sapporo-shi, Hokkaido 005-8601, Japan
- <sup>4</sup> Department of Chemistry, Faculty of Science, Josai University, 1-1 Keyakidai, Sakado-shi, Saitama 350-0295, Japan
- <sup>5</sup> National Research Institute of Fisheries Science, Fisheries Research Agency, Komaki 1088, Ueda, Nagano 386-0031, Japan
- <sup>6</sup> Graduate School of Fisheries Science and Environmental Studies, Nagasaki University, 1-14 Bunkyo-machi, Nagasaki 852-8521, Japan
- <sup>7</sup> University of the Ryukyus, 1 Senbaru, Nakagami-gun, Nishihara-cho, Okinawa 903-0213, Japan

**Keywords** Microsatellite DNA · Target capture · Next generation sequencing · Multiplexed sequencing · Ayu

## Introduction

Microsatellite DNA is one of the most powerful genetic markers for wildlife conservation genetics and population genetics (Frankham et al. 2010; Guichoux et al. 2011). Recent advances in next-generation sequencing (NGS) technology have accelerated the development of novel

microsatellite markers for target species (Gardner et al. 2011; Schoebel et al. 2013; Wei et al. 2014 and references therein). There are two main approaches in NGS-based microsatellite isolation: whole-genome shotgun sequencing (e.g., Abdelkrim et al. 2009) and sequencing of a microsatellite-enriched library (e.g., Malausa et al. 2011). The shotgun approach is to generate enough random sequences to isolate a satisfying number of microsatellite-containing sequences by chance. On the other hand, the enrichment approach is to construct a microsatellite-enriched library by conducting hybridization with repeat probes to genomic DNA fragments. After washing the non-hybridized DNA that presumably lacks repeat regions, the remaining DNA fragments are sequenced to isolate microsatellite-containing sequences. In each approach, the constructed libraries are appropriately pooled for multiplexed sequencing on an NGS platform.

Although the discovery of microsatellite-containing sequences is more efficient with the enrichment approach than with the shotgun approach (Malausa et al. 2011), the construction of a microsatellite-enriched library requires specialized bench work and is somewhat complicated in a routine laboratory setting (Gonzalez and Zardoya 2013). Consequently, many researchers prefer to use external services for the microsatellite-enrichment step, which requires more time to acquire candidate sequences for the development of microsatellite markers (Malausa et al. 2011; Gonzalez and Zardoya 2013). A solution to this problem is to improve the efficiency and speed of the conventional microsatellite-enrichment method in a routine laboratory setting.

In this study, we attempted to simplify the process of constructing a microsatellite-enriched library for multiplexed parallel sequencing. To simplify the enrichment method in order to reduce the time required, we applied an easy-to-use commercially available kit for the hybridization and wash steps in the target capture method. In the conventional enrichment method, after microsatellite-enriched library preparation, the library is ligated with NGS adapters and sequenced on an NGS platform. In contrast, in the improved enrichment method, after NGS shotgun library preparation, we captured microsatellite-containing DNA fragments directly from the NGS shotgun library by using the commercially available kit. Genomic DNA from each fish species was fragmented by digesting with enzyme. The prepared NGS shotgun library was hybridized with the biotinylated CA repeat probe, and then was subjected to wash step for exclusion of the non-hybridized DNA by using the commercially available kit. The enrichment libraries of 30 fish species were prepared for multiplexed parallel sequencing on a bench-top NGS platform (454 GS Junior, Roche). The three runs of multiplexed parallel sequencing were conducted to isolate the microsatellite-containing

sequences. Sequences were analyzed using the program to identify and select the microsatellite sequences and to design primer pairs. Furthermore, to validate the effectiveness of the present approach, we checked the ability of the designed primer sets to amplify and detect polymorphisms based on the microsatellite-containing sequences for ayu (*Plecoglossus altivelis*), a commercially important species in Japanese inland fisheries.

## Materials and methods

### DNA extraction and library preparation

Genomic DNA was extracted from 30 fish species (Table 1) with the DNeasy Blood and Tissue kit (Qiagen), the Genra Puregene kit (Qiagen), or the Wizard Genomic DNA Purification kit (Promega). For each fish species, approximately 0.5–1 µg of genomic DNA was fragmented by digesting with dsDNA Fragmentase (New England Biolabs) for 15–20 min at 37 °C. The fragmented DNAs were purified using the MinElute PCR Purification kit (Qiagen). The size distribution of fragmented DNAs was assessed on a 2100 Bioanalyzer by using the DNA High Sensitivity kit (Agilent Technologies). For each fragmented DNA, a shotgun library was prepared using the NEBNext Quick DNA Library Prep Master Mix Set for 454 (New England Biolabs). Each fragmented DNA was end repaired, A-tailed, and ligated to one of the 12 multiplex identifier (MID) oligonucleotide adaptors (Roche) for multiplexed sequencing (Table 1). For each shotgun library, small DNA fragments were subsequently removed using the AMPure purification system (Agencourt Bioscience).

### Target capture of microsatellite-containing DNA fragments

We performed target capture of microsatellite-containing DNA fragments by using the CA repeat probe and the SeqCap EZ hybridization and wash kit (Roche), according to the general guidelines provided in the NimbleGen SeqCap EZ Library LR User's Guide ver.1.0 (Roche) with slight modifications, as briefly described below. The prepared shotgun libraries were amplified by 12–15 cycles of pre-capture linker-mediated PCR (LM-PCR). The size distribution of pre-capture LM-PCR products was assessed on the 2100 Bioanalyzer by using a DNA 7500 kit (Agilent Technologies). The pre-capture LM-PCR products were quantified using the Nanodrop 2000 (Thermo Scientific) or the Qubit dsDNA HS assay kit (Invitrogen). One shotgun library was prepared from the pre-capture LM-PCR product for comparison. Approximately 0.5–1 µg of each pre-capture LM-PCR product was hybridized to 20 picomoles

**Table 1** Summary of the sequencing libraries, sequencing results, and post-sequencing selection of microsatellite loci for each species

Species	Order	Library	Sequence read archive accession no.	Number of sequences	Number of sequences at least 100 bp	Number of sequences with microsatellite motif (%)	Number of sequences in which primers were designed (%)	Number of sequences in which primers were designed (with CA motif) (%)	Number of sequences in which primers were designed (with perfect microsatellites) (%)	Number of sequences in which primers were designed (with compound microsatellites) (%)
<b>Sequencing run 1</b>										
<i>Carassius auratus</i>	Cypriniformes	Captured	DRA004663	11,405	9909	7205 (63)	2163 (19)	1969 (91)	1578 (73)	585 (27)
<i>Silurus biwaensis</i>	Siluriformes	Captured	DRA004684	15,107	12,200	8325 (55)	1971 (13)	1676 (85)	1323 (67)	648 (33)
<i>Silurus lithophilus</i>	Siluriformes	Captured	DRA004683	14,346	11,303	7512 (52)	1551 (11)	1294 (83)	1061 (68)	490 (32)
<i>Silurus asotus</i>	Siluriformes	Captured	DRA004669	13,615	11,017	7194 (53)	1325 (10)	1089 (82)	895 (68)	430 (32)
<i>Plecoglossus altivelis</i>	Osmeriformes	Captured	DRA004666	15,687	12,537	10,029 (64)	3405 (22)	2778 (82)	2274 (67)	1131 (33)
<i>Pseudoblenius cotoides</i>	Scorpaeniformes	Captured	DRA004685	13,240	11,288	9177 (69)	2551 (19)	2215 (87)	1598 (63)	953 (37)
<i>Cottus pollux</i> (large egg type)	Scorpaeniformes	Captured	DRA004680	12,392	10,686	8116 (65)	2221 (18)	1985 (89)	1542 (69)	679 (31)
<i>Cottus reinii</i>	Scorpaeniformes	Captured	DRA004686	13,687	11,687	9084 (66)	2496 (18)	2208 (88)	1740 (70)	756 (30)
<i>Lates japonicus</i>	Perciformes	Captured	DRA004681	11,730	10,719	8523 (73)	3622 (31)	3265 (90)	2537 (70)	1085 (30)
<i>Lates japonicus</i>	Perciformes	Shotgun	DRA004682	4987	4647	645 (13)	264 (5)	184 (70)	214 (81)	50 (19)
Total				1,26,196	1,05,993	75,810	21,569			
<b>Sequencing run 2</b>										
<i>Coilia nasus</i>	Clupeiformes	Captured	DRA004695	15,236	12,430	10,509 (69)	3405 (22)	2864 (84)	2159 (63)	1246 (37)
<i>Cynoglossus joyneri</i>	Pleuronectiformes	Captured	DRA004692	14,034	12,314	9394 (67)	3102 (22)	2533 (82)	2183 (70)	919 (30)
<i>Siganus javus</i>	Perciformes	Captured	DRA004687	11,893	10,799	8840 (74)	3721 (31)	3380 (91)	2550 (69)	1171 (31)
<i>Thunnus tonggol</i>	Perciformes	Captured	DRA004688	17,061	14,967	13,054 (77)	5241 (31)	4852 (93)	3558 (68)	1683 (32)
<i>Selar crumenophthalmus</i>	Perciformes	Captured	DRA004689	12,437	11,306	9589 (77)	3926 (32)	3381 (86)	2827 (72)	1099 (28)

Table 1 (continued)

Species	Order	Library	Sequence read archive accession no.	Number of sequences	Number of sequences at least 100 bp	Number of sequences with at least one microsatellite motif (%)	Number of sequences in which primers were designed (%)	Number of sequences in which primers were designed (with CA motif) (%)	Number of sequences in which primers were designed (with other motif) (%)	Number of sequences in which primers were designed (with perfect microsatellites) (%)	Number of sequences in which primers were designed (with compound microsatellites) (%)
<i>Sphyraena putnamae</i>	Perciformes	Captured	DRA004690	16,667	14,401	12,327 (74)	4442 (27)	4073 (92)	369 (8)	3022 (68)	1420 (32)
<i>Atule mate</i>	Perciformes	Captured	DRA004691	11,996	10,569	8676 (72)	3139 (26)	2827 (90)	312 (10)	2123 (68)	1016 (32)
<i>Centropyge flavissima</i>	Perciformes	Captured	DRA004693	18,165	15,941	14,369 (79)	6255 (34)	5847 (93)	408 (7)	4448 (71)	1807 (29)
<i>Acanthogobius flavimanus</i>	Perciformes	Captured	DRA004694	11,012	8888	5078 (46)	1029 (9)	895 (87)	134 (13)	615 (60)	414 (40)
<i>Gerres filamentosus</i>	Perciformes	Captured	DRA004696	9510	8750	7116 (75)	3285 (35)	3081 (94)	204 (6)	2453 (75)	832 (25)
<i>Scolopsis taenioptera</i>	Perciformes	Captured	DRA004697	12,595	11,007	9186 (73)	3103 (25)	2676 (86)	427 (14)	2050 (66)	1053 (34)
<i>Lepidion prologus mimicus</i>	Perciformes	Captured	DRA004698	15,019	13,305	10,881 (72)	3568 (24)	3385 (95)	183 (5)	2460 (69)	1108 (31)
Total				1,65,625	1,44,677	1,19,019	44,216				
Sequencing run 3											
<i>Konostirus punctatus</i>	Clupeiformes	Captured	DRA004705	17,374	14,310	12,931 (74)	5667 (33)	4929 (87)	738 (13)	3476 (61)	2191 (39)
<i>Hemigrammocyprius rasborella</i>	Cypriniformes	Captured	DRA004699	13,775	12,472	9,561 (69)	3855 (28)	3518 (91)	337 (9)	2820 (73)	1035 (27)
<i>Pseudorasbora parva</i>	Cypriniformes	Captured	DRA004707	20,117	16,253	12,443 (62)	3017 (15)	2797 (93)	220 (7)	2143 (71)	874 (29)
<i>Ablennes hians</i>	Belontiiformes	Captured	DRA004703	19,461	17,382	14,528 (75)	4450 (23)	4199 (94)	251 (6)	2965 (67)	1485 (33)
<i>Scolopsis monogramma</i>	Perciformes	Captured	DRA004700	33,196	25,002	23,864 (72)	1713 (5)	1551 (91)	162 (9)	1095 (64)	618 (36)
<i>Parapterois heterura</i>	Perciformes	Captured	DRA004706	13,101	11,943	10,194 (78)	2817 (22)	2516 (89)	301 (11)	1848 (66)	969 (34)
<i>Terapon jarbua</i>	Perciformes	Captured <sup>a</sup>	DRA004701	27,321	24,788	12,750 (47)	5789 (21)	4943 (85)	846 (15)	4118 (71)	1671 (39)
<i>Megalaspis cordyla</i>	Perciformes	Captured <sup>a</sup>	DRA004702	28,503	25,111	15,303 (54)	5559 (20)	3924 (71)	1 635 (29)	4061 (73)	1498 (27)

**Table 1** (continued)

Species	Order	Library	Sequence read archive accession no.	Number of sequences at least 100 bp	Number of sequences with at least one microsatellite motif (%)	Number of sequences in which primers were designed (with CA motif) (%)	Number of sequences in which primers were designed (with other motif) (%)	Number of sequences in which primers were designed (with perfect microsatellites) (%)	Number of sequences in which primers were designed (with compound microsatellites) (%)
<i>Sebastes trivittatus</i>	Perciformes	Captured <sup>a</sup>	DRA004704	31,965	13,803 (43)	6606 (21)	5553 (84)	4831 (73)	1775 (27)
Total				2,04,813	1,25,377	39,473			

The sequencing results of the mixed libraries did not greatly differ from those of the captured libraries

<sup>a</sup>Concentration of captured and shotgun libraries was mixed in a 1:1 ratio

of biotinylated probe [B-ATAGAATAT(CA)<sub>16</sub>] at 55 °C for 1 h. During the hybridization, COT human DNA was not added to the hybridization component. The hybridization mixture was incubated with streptavidin-coated Dynabeads M-270 (Invitrogen), and then non-captured material (the unbound target DNA presumably lacking microsatellites and the unbound probe) was washed away. After washing, each captured library was amplified by 15 cycles of post-capture LM-PCR. The size distribution of the post-capture LM-PCR products—the microsatellite-captured library—was assessed on a 2100 Bioanalyzer by using the DNA 7500 kit (Agilent Technologies). The microsatellite-captured libraries were quantified using the Qubit dsDNA HS assay kit (Invitrogen) or the KAPA Library quantification kit (Kapa Biosystems). Finally, 30 microsatellite-captured libraries and one shotgun library were constructed and quantified.

**Pooling libraries, emulsion PCR, and pyrosequencing**

For each of the three runs of multiplexed parallel sequencing, a different library pool (made from 9 to 12 libraries) was sequenced (Table 1), which were distinguished by 12 MID tags. Each library pool was quantified using the KAPA Library quantification kit (Kapa Biosystems) and was then separately sequenced using the GS Junior System (Roche). Emulsion PCR with 0.2 DNA copies per amplification bead, breaking, and pyrosequencing were performed according to the manufacturer’s protocols for the Lib-L kit (Roche). After each multiplexed sequencing run, the output SFF (Standard Flowgram Format) files were separated according to the sequences of MID tags by using the ‘sfffile’ program (Roche), and the resulting SFF files were converted into FASTA format files by using the ‘sffinfo’ program (Roche).

**Data analysis and primer design**

Sequences were received in the form of FASTA files and analyzed using the pipeline QDD version 2 (Meglécz et al. 2010) to identify and select microsatellite sequences and to design primer pairs. Sequences longer than 100 bp and containing at least five repeats of perfect microsatellites, composed of one single motif of 2- to 6-bp length with no interruption, were selected for further analysis. Sequence similarities were identified by an “all against all” BLAST (Altschul et al. 1997) analysis, using an e-value of 1E-40 and with microsatellite sequences soft-masked. Sequences exceeded 95% pairwise similarity in flanking regions were grouped into contigs and a 2/3 majority rule consensus sequence was created from each contig. Sequences with significant BLAST hits to other sequences and an overall similarity in flanking region of less than 95% were

excluded to avoid potential duplicated loci and mobile elements. Primer sequences were then designed based on all of the unique sequences and the consensus sequences. Primer pairs were designed with Primer3 (Rozen and Skaletsky 1999), implemented in QDD with the following criteria: (1) PCR product lengths were set between 90 and 450 bp, several primer pairs designed *in silico* for each sequence with an interval of 50 bp; (2) an optimal primer length of 24 bp (range 20–30 bp); (3) an optimal primer pair annealing temperature of 63 °C (range 60–66 °C); and (4) 50% GC content (range 20–80%).

### Validation and characterization of microsatellite markers in ayu

To validate the effectiveness of the present approach, we screened for potential primer sets for ayu (*P. altivelis*). In ayu, 3405 unique sequences and consensus sequences fulfilled the requirements for primer design (Table 1). During primer selection, QDD grouped the primers into seven different primer designs (A–G). We chose only the most restrictive, design A (334 sequences), which fulfills the following conditions: (1) no repeats of a single base in the flanking and primer regions; (2) no other target microsatellites in the flanking region; (3) no nanosatellites, that is 3–4 tandem repetition of a 2- to 6-bp motif, in the flanking and primer regions; and (4) not allowing compound microsatellites. The final step was to select only sequences whose microsatellites comprise a motif repeated more than ten times. After all this exclusions, 88 sequences were retained for primer synthesis. To facilitate multiplex PCR, we selected 44 of 88 primer sets based on expected PCR product size and the forward primer of each locus was synthesized with one of four universal tag sequences (Table 2; see also Blacket et al. 2012). The four universal tag primers can be combined with four fluorophores to co-amplify multiple loci via multiplex PCR.

Initial screening of the 44 designed primer sets was performed on four ayu individuals (collected from Hidaka River, Wakayama, Japan). For this screening, single-locus PCR reactions were carried out in 7 µl of final reaction mixture, containing 1×GoTaq Green Master mix (Promega), 0.86 µM of each primer, and approximately 50–250 ng of template DNA. The thermal cycling profile was 95 °C for 5 min; 40 cycles of 94 °C for 15 s, 59 °C for 15 s, and 72 °C for 30 s; and a final extension at 60 °C for 7 min. The PCR products were separated by electrophoresis on a 1.5% agarose gel, and the 37 primer sets that were successfully amplified were subsequently tested for ability to detect polymorphisms by using two forms of ayu, the amphidromous and landlocked Lake Biwa forms, between which substantial genetic differences have been revealed by

previous genetic studies using microsatellite DNA markers (Takagi et al. 1999; Takeshima et al. 2009, 2016).

Forty-eight individuals of each form (the amphidromous form collected from the Hidaka River and the landlocked form collected from the Ane River in the Lake Biwa system, Shiga, Japan) were used to test for polymorphism of the 37 primer sets. These primer sets were amplified using five multiplex PCR reactions (Table 2). PCR reactions were carried out in 7 µl of final reaction mixture (Sets I and II) or 4 µl (Sets III–V), containing 1×Type-it Microsatellite PCR kit (Qiagen), approximately 50–250 ng of template DNA, and the forward primer, reverse primer, and universal tail primer in a 1:2:1 ratio (Blacket et al. 2012). The final concentration of the forward primers was optimized for each marker (Table 2). Each universal tail primer was fluorescently labeled with 6-FAM, VIC, NED, or PET (Applied Biosystems). The thermal cycling profile was 95 °C for 5 min; 40 cycles of 94 °C for 15 s, 59 °C for 15 s, and 72 °C for 30 s; and a final extension at 60 °C for 30 min. Microsatellite products were analyzed on a 3130XL Genetic Analyzer (Applied Biosystems) with GeneScan 500 LIZ Size Standard (Applied Biosystems). Genotyping was performed with the software GeneMapper version 3.7 (Applied Biosystems).

For each form of ayu, genetic variability parameters, including the number of alleles per locus ( $N_A$ ), observed heterozygosity ( $H_O$ ), and expected heterozygosity ( $H_E$ ), were calculated at each locus by using GENETIX 4.05 (Belkhir et al. 2004). Departure from Hardy–Weinberg equilibrium (HWE) within each form by locus and across all loci, as well as linkage disequilibrium (LD) among loci, were tested by using GENEPOP 3.4 (Raymond and Rousset 1995). Tests for HWE and LD employed a Markov chain method to estimate without bias the exact  $P$ -values proposed by Guo and Thompson (1992), with the following chain parameters: 10,000 dememorization steps, 100 batches, and 5000 iterations per batch. To avoid type I error, a sequential Bonferroni correction (Rice 1989) was applied to  $P$ -values from multiple tests. In addition, the presence of null alleles at each locus was checked for each form by using the software MicroChecker ver. 2.2.3 (van Oosterhout et al. 2004). Genetic differentiation between the two forms of ayu was examined by the fixation index ( $F_{ST}$ ) for each locus (Weir and Cockerham 1984) by using Arlequin ver 3.0 (Excoffier et al. 2005), with statistical significance estimated from 16,000 permutations.

### Results and discussion

The results of the three runs of multiplexed parallel sequencing are summarized in Table 1. From the runs, we obtained 126,196, 165,625, and 204,813 sequences in

**Table 2** Primer sequences and characteristics of 37 microsatellite loci for ayu, *Plecoglossus altivelis*

Multiplex set	Locus	Primer sequence 5'–3'	Tail sequence (fluorescence)	Concentration of forward primer (μM)	Repeat motif	Amphidromous form					
						Accession no.	<i>n</i>	<i>N<sub>A</sub></i>	<i>R</i> (bp)	<i>H<sub>E</sub></i>	<i>H<sub>O</sub></i>
Set I	<i>Plal-001</i>	F:TCGCACATGCATACACATACAATC	A (6-FAM)	0.05	(AC) <sub>10</sub>	LC151564	48	20	102–140	0.916	0.917
		R:ATGGAGGAGCAGCACTGATGT									
	<i>Plal-002</i>	F:GTACGCACAGAGATACAGACG	B (VIC)	0.02	(AC) <sub>10</sub>	LC151565	47	7	94–108	0.534	0.489
		R:CCATCATGACTCAGCAGTGACCT									
	<i>Plal-003</i>	F:CTTGGATCAGGCCATGTTG	C (NED)	0.05	(AC) <sub>13</sub>	LC151566	48	12	94–124	0.848	0.750
		R:AGTGCACGCCATTCACATAAG									
	<i>Plal-004</i>	F:TTAAGACTGAGGATTCAGTCCAGC	D (PET)	0.05	(AC) <sub>11</sub>	LC151567	48	13	100–128	0.799	0.708
		R:TGTAGTCAATCCATCTCGTAGCAC									
	<i>Plal-005</i>	F:CACTAGTGTGGAAGAGTCTCTGG	A	0.15	(AC) <sub>11</sub>	LC151568	47	16	202–236	0.866	0.894
		R:CCTAACTGTATGGCACATGTTGG									
	<i>Plal-006</i>	F:CAGCTTCAGGTCTGTTGATGTCAG	B	0.15	(AC) <sub>15</sub>	LC151569	48	24	202–262	0.933	0.688*
		R:CACTATATGATTCACACCCGTAACCA									
	<i>Plal-007</i>	F:CCACTCACACCAAGTATGCAACAC	C	0.05	(AC) <sub>13</sub>	LC151570	47	9	208–224	0.776	0.723
		R:GAAGGCAGATAAGATGGAGACTGC									
	<i>Plal-008</i>	F:TCTGATGCACTGGCAAGAAGAC	D	0.15	(AC) <sub>11</sub>	LC151571	47	7	206–220	0.536	0.575
		R:CCTAAGACCTGTCATGGTAGAGCA									
	<i>Plal-009</i>	F:CAATAGGATGCCAGACATGATGAA	A	0.2	(AC) <sub>10</sub>	LC151572	47	5	276–284	0.556	0.532
		R:ACGTCTCCACGAGACTCTGTGACT									
	<i>Plal-010</i>	F:GACACTACAGGATTCGCCAGCAT	B	0.075	(AC) <sub>11</sub>	LC151573	48	5	322–330	0.405	0.375
R:GGAGCAGTGGAGAGTGAATCAGAG											
<i>Plal-011</i>	F:GGCAATGCATGGATTCCTAA	C	0.2	(AC) <sub>12</sub>	LC151574	47	15	352–388	0.844	0.809	
	R:TAGCAATGCCTTGGTGCCTATAAT										
Set II	<i>Plal-013</i>	F:GGTCTGGACACTGAGACACTAGCA	A	0.05	(AC) <sub>11</sub>	LC151575	47	6	108–118	0.745	0.638
		R:CCTCAATTGTCAAGATTGTCTCC									
	<i>Plal-015</i>	F:ATCCAGACCTCGACATTCTACTGC	C	0.2	(ACCG) <sub>10</sub>	LC151576	43	20	128–236	0.936	0.465*
		R:TCAGGACAGCACAAACCGTACC									
	<i>Plal-016</i>	F:CTATCTGACGTCTGTGTAGCCAGC	D	0.025	(AC) <sub>10</sub>	LC151577	48	8	124–144	0.619	0.521
		R:CTCGTAGAACGGTGTGGTGTATT									
	<i>Plal-017</i>	F:CTTGTGAAGTTGAGGAAAGTGGACA	A	0.15	(AC) <sub>14</sub>	LC151578	47	11	240–260	0.794	0.745
		R:TTCTCCTGCAATCAAGGTTACACA									
	<i>Plal-018</i>	F:ITCACACTTCCTAGTCTCCTCCAAC	B	0.1	(AC) <sub>12</sub>	LC151579	47	6	256–266	0.630	0.660
		R:GCATCTCAGACACTCGTTCATCAT									
	<i>Plal-019</i>	F:CATGAATACTGCTCAGATGGCTCA	C	0.1	(AC) <sub>10</sub>	LC151580	48	14	258–296	0.696	0.688
		R:CCTGAGAACAGGAAGTGAAGGAG									

Table 2 (continued)

Multiplex set	Locus	Primer sequence 5'–3'	Tail sequence (fluorescence)	Concentration of forward primer ( $\mu\text{M}$ )	Repeat motif	Amphidromous form					
						Accession no.	$n$	$N_A$	$R$ (bp)	$H_E$	$H_O$
Set III	<i>Plal-020</i>	F: GTTCTGTCTCCATCTGGCAGG R: GCAGATGAGGTCATTGTTCAGCTT	D	0.15	(AC) <sub>10</sub>	LC151581	45	18	252–288	0.884	0.689
	<i>Plal-021</i>	F: ACTGGCTGAGGTGGACAGAGAC R: GAGACTGTACGCATGCTGAGTGAT	A	0.0375	(AC) <sub>14</sub>	LC151582	48	13	98–130	0.777	0.750
	<i>Plal-024</i>	F: GTTGCCCTGCAGAACATATGGA R: ACTTGCTCACAGAAAGCACAGCATA	D	0.025	(AC) <sub>11</sub>	LC151583	48	5	116–124	0.445	0.458
	<i>Plal-025</i>	F: TGGATCAGTAGAGATCATGTTAGCG R: GTGTCAGTCTTGAAGGCAGCATTA	A	0.01	(AC) <sub>12</sub>	LC151584	46	15	216–254	0.832	0.609*
	<i>Plal-027</i>	F: CAAGGATTTAGCGAGATAACCG R: AATCAGTGGTCTCAAAGCAGGTACT	C	0.01	(AC) <sub>10</sub>	LC151585	47	6	240–258	0.717	0.723
	<i>Plal-028</i>	F: GGTGTTATGTCCGAGCGTACTTG R: GGCTCTTGTCTCACAGGAATGAAT	D	0.01	(AC) <sub>10</sub>	LC151586	47	10	238–256	0.765	0.766
	<i>Plal-029</i>	F: CCTCCACCAATACCTGCTTATCAA R: GACGTATCACTCTGTTACATCACAGG	A	0.02	(AC) <sub>11</sub>	LC151587	48	7	110–122	0.665	0.479
	<i>Plal-030</i>	F: AGGTCTGTGAGACAGAAAGGCTCTC R: CTGATAACAGCTGATCACTGGCTG	B	0.045	(AC) <sub>10</sub>	LC151588	48	3	110–118	0.137	0.146
Set IV	<i>Plal-031</i>	F: GTTGGCATGCATACACTCCTCAC R: CTGTGCTTGTATATCTTGGATGGC	C	0.03	(AC) <sub>11</sub>	LC151589	48	3	120–124	0.118	0.125
	<i>Plal-033</i>	F: GCGTGACTAAGCCTCAGATCTCTT R: TAGTGTCTTTCATCTCGAGTAGA	A	0.45	(AC) <sub>10</sub>	LC151590	46	6	190–206	0.430	0.435
	<i>Plal-034</i>	F: TAGCAGTCAGCAGTGGCATTAGTC R: AGGCGTCTATTGTGAAGACAGACC	B	0.01	(AC) <sub>10</sub>	LC151591	46	8	188–206	0.539	0.587
	<i>Plal-035</i>	F: ATGTCCACATCCAGAGAGCTACC R: ATGACTTGCCTGATGACAGAATTG	C	0.125	(AC) <sub>10</sub>	LC151592	26	6	192–206	0.539	0.077*
	<i>Plal-036</i>	F: CCACTGTACGGCTGCTTCTTCT R: CGAAGTATTGCTGCTGAATTGTTG	D	0.1	(AC) <sub>13</sub>	LC151593	46	10	194–214	0.809	0.848
	<i>Plal-037</i>	F: CTGTACGAGAAAGCGTCAAGTGT R: ATCCAGTGTCTGTTGATGATGCT	A	0.0175	(AC) <sub>12</sub>	LC151594	48	10	116–136	0.630	0.625
	<i>Plal-038</i>	F: TGCACACTGGTTGGCCTAATTACT R: AGGACAACCCAGACTAGACCAGCC	B	0.0065	(AC) <sub>10</sub>	LC151595	47	4	126–132	0.573	0.575
	<i>Plal-039</i>	F: TCTTATCAATAAGTGCCTGGTGTGAG R: GTCAGCTAATGGAGTGTTCCTGGT	C	0.015	(AC) <sub>13</sub>	LC151596	47	9	128–148	0.791	0.894
<i>Plal-040</i>	F: ATCATGATCTCTGGACACTCAGCA	D	0.025	(AC) <sub>10</sub>	LC151597	47	4	134–140	0.159	0.170	
Set V											



**Table 2** (continued)

Multiplex set	Locus	Primer sequence 5'–3'	Tail sequence (fluorescence)	Concentration of forward primer (μM)	Repeat motif	Accession no.	Amphidromous form					
							<i>n</i>	<i>N<sub>A</sub></i>	<i>R</i> (bp)	<i>H<sub>E</sub></i>	<i>H<sub>O</sub></i>	
Set I	<i>Plal-041</i>	R:CTGTGTCCTTCACCTGACATGG										
		F:ATGACTGCTCTTGTTCACCTGGAGA	A	0.35	(AC) <sub>15</sub>	LC151598	47	19	160–218	0.888	0.723	
		R:TACACGGATGATGAGTGTCTGAG										
	<i>Plal-042</i>	F:TTCTGGTAACAGTTGGCAGCAITTA	B	0.05	(AC) <sub>10</sub>		LC151599	47	5	182–192	0.652	0.596
		R:CGAACTGAAGAGGCAGAACAGAITT										
		F:CTACATGGCGGTGACAGGAAG	D	0.075	(ACAG) <sub>10</sub>		LC151600	47	15	176–232	0.911	0.872
	Set II	<i>Plal-011</i>	R:AACAGAGGTAGCGTTAGAGATGCC									
			<i>Plal-001</i>	48	11	100–128	0.842		0.813			<b>0.028</b>
			<i>Plal-002</i>	47	6	94–108	0.589		0.511*			<b>0.042</b>
<i>Plal-003</i>			48	11	100–120	0.831		0.792			0.001	
<i>Plal-004</i>			48	12	106–140	0.802		0.729			–0.004	
<i>Plal-005</i>			47	13	202–230	0.833		0.766			<b>0.068</b>	
<i>Plal-006</i>			46	22	202–270	0.898		0.826			<b>0.022</b>	
<i>Plal-007</i>			47	7	208–224	0.783		0.681			0.006	
<i>Plal-008</i>			46	6	204–218	0.583		0.587			0.003	
<i>Plal-009</i>			46	3	276–282	0.305		0.304			<b>0.109</b>	
<i>Plal-010</i>			46	3	324–330	0.043		0.044			<b>0.172</b>	
Set III	<i>Plal-028</i>											
		<i>Plal-013</i>	48	7	108–120	0.642		0.667			<b>0.042</b>	
		<i>Plal-015</i>	43	24	124–240	0.948		0.419*			0.011	
		<i>Plal-016</i>	48	4	126–132	0.483		0.396			<b>0.248</b>	
		<i>Plal-017</i>	47	10	242–264	0.773		0.830			<b>0.056</b>	
		<i>Plal-018</i>	46	6	254–266	0.670		0.739			<b>0.146</b>	
		<i>Plal-019</i>	46	6	262–278	0.584		0.544			0.012	
		<i>Plal-020</i>	47	15	252–288	0.814		0.745			<b>0.028</b>	
Set III	<i>Plal-028</i>											
		<i>Plal-021</i>	47	6	94–114	0.652		0.617			<b>0.072</b>	
		<i>Plal-024</i>	47	4	116–122	0.197		0.213			<b>0.046</b>	
		<i>Plal-025</i>	46	9	216–240	0.776		0.565*			0.004	
		<i>Plal-027</i>	47	6	240–258	0.683		0.617			<b>0.026</b>	
		<i>Plal-028</i>	47	6	238–252	0.616		0.617			<b>0.045</b>	

Table 2 (continued)

Multiplex Set	Locus	Landlocked Lake Biwa form			$F_{ST}$ value			
		$n$	$N_A$	$R$ (bp)	$H_E$	$H_O$	$F_{ST}$ value	
Set IV	<i>Plal-029</i>	47	6	110–120	0.622	0.319*	0.001	
	<i>Plal-030</i>	48	2	114–116	0.021	0.021	<b>0.033</b>	
	<i>Plal-031</i>	48	3	120–124	0.061	0.063	0.005	
	<i>Plal-033</i>	46	5	186–204	0.238	0.196	<b>0.064</b>	
	<i>Plal-034</i>	47	8	188–202	0.716	0.702	<b>0.055</b>	
	<i>Plal-035</i>	37	10	192–236	0.732	0.243*	0.023	
	<i>Plal-036</i>	48	9	200–216	0.604	0.604	<b>0.059</b>	
	Set V	<i>Plal-037</i>	48	10	116–134	0.613	0.542	<b>0.022</b>
		<i>Plal-038</i>	48	3	126–130	0.565	0.500	<b>0.034</b>
		<i>Plal-039</i>	48	8	128–142	0.643	0.729	<b>0.115</b>
<i>Plal-040</i>		48	3	138–142	0.137	0.146	0.013	
<i>Plal-041</i>		46	18	170–206	0.899	0.848	<b>0.019</b>	
<i>Plal-042</i>		48	8	180–198	0.565	0.500	0.016	
<i>Plal-044</i>		48	16	168–236	0.904	0.958	0.003	

Four tail sequences (Blacket et al. 2012) were the followings: Tail A, GCCTCCCTCGGGCCA; Tail B, GCCTTGCCAGCCCCG; Tail C, CAGGACCAGGCTACCGTG; Tail D, CGGAGAGCC GAGAGGTTG

Bold for the  $F_{ST}$  value indicates the significant differences using 16,000 permutations ( $P < 0.05$ )

$n$  number of individuals genotyped,  $N_A$  number of alleles,  $R$  range of observed alleles,  $H_O$  observed heterozygosity,  $H_E$  expected heterozygosity

\*Significant deviation from Hardy–Weinberg equilibrium ( $P < 0.05$ )

total, respectively. The average sequence length for each sequencing run was 298, 280, and 320 bp, respectively. Among the 30 microsatellite-captured libraries, we found high proportions of microsatellite-containing sequences, ranging from 43 to 79%. Thus, sufficient numbers of primer sets for developing microsatellite markers (from 1029 to 6606) were effectively designed for each species. We also found high proportions of sequences in which primers were designed with CA motif, ranging from 71 to 95%, among the 30 microsatellite-captured libraries.

To validate the effectiveness of our approach for microsatellite isolation, we screened potential primer sets designed for ayu. The results of screening with 37 primer sets in the two forms of ayu are summarized in Table 2. The genetic variability parameters  $N_A$ ,  $H_E$ , and  $H_O$  in the two forms of ayu ranged from 2 to 24, from 0.021 to 0.948, and from 0.021 to 0.958, respectively. Significant deviations from HWE were observed in only 9 of the 74 probability tests (Table 2) after sequential Bonferroni correction ( $P < 0.05$ ). One possible cause for the observed heterozygote deficiency is the existence of null alleles. In microsatellite studies, primer-site mutations result in non-amplification of some alleles (null alleles). This possibility was further investigated by MicroChecker analyses, which detected signs of a null allele at *Plal-015*, *Plal-025*, *Plal-029*, and *Plal-035* in both forms and at *Plal-006*, *Plal-020*, and *Plal-041* in the amphidromous form only. The consistency of results from HWE and MicroChecker suggested there was a high probability of null alleles at *Plal-015*, *Plal-025*, *Plal-029*, and *Plal-035*.

For both forms of ayu, analyses of LD for each locus showed only 3 significant pairwise comparisons between loci (of 1330 comparisons) after sequential Bonferroni correction ( $P < 0.05$ ), without specific interlocus relationships, suggesting the overall independence of the loci examined.  $F_{ST}$  values between the two forms of ayu for each locus ranged from  $-0.004$  to  $0.248$ , and the values were significantly different from zero at 23 of the 37 loci ( $P < 0.05$ ). Allelic variability at these 23 microsatellite markers in ayu will be useful for studying population structure.

In this study we used a commercially available kit for target capture and performed multiplexed parallel sequencing on a bench-top NGS platform. Our improved approach effectively determined massive numbers of microsatellite-containing sequences, and the sequences provided enough information for designing potential primer sets of microsatellite markers. Based on the obtained sequences, we succeeded in developing 23 useful microsatellite DNA markers for analyzing the population structure of ayu. These results prove the effectiveness of our approach for microsatellite marker development in a routine laboratory setting. The present approach could be effectively applied to develop microsatellite markers for other fishes or other organisms.

In the present study, the time required for microsatellite capture and sequencing was approximately 1 week, which is substantially less than that of the conventional approach (usually more than 2 weeks; Malausa et al. 2011). In addition, preparing the microsatellite-captured library was a simple process.

To evaluate the benefit of microsatellite capture, we compared the results from the microsatellite-captured library and the shotgun library of *Lates japonicus* (Table 1). The proportion of sequences of primers designed in the microsatellite-captured library (3622/11,730 sequences: 31%) was 6.2 times that in the shotgun library (264/4987: 5%), indicating that microsatellite capture is substantially more effective than the shotgun method for the development of microsatellite markers.

Finally, we briefly discuss the use of this new approach for future research. In the present study, we used only the CA repeat probe to capture microsatellite DNA from target fishes because CA is the most common dinucleotide motif in the vertebrate genome (Chistiakov et al. 2006). If our approach is applied to other organisms (e.g., invertebrates and plants), various repeat motifs should be used as capture probes. Although Roche announced that the company will stop supporting the 454 GS Junior platform by mid-2016, our approach using the SeqCap EZ hybridization and wash kit (Roche) can readily be adopted with Illumina's the MiSeq platform using the 300 bp paired-end sequencing format. Therefore, our approach may be useful for further development of novel microsatellite markers for conservation genetics and population genetics research.

**Acknowledgements** We thank S. Awata, S. Chiba, M. Konishi, S. Matsuzaki, K. Nakayama, T. Natsumeda, D. Tahara, and K. Watanabe for providing fish samples. We also thank H. Itoh, S. Ueda, Y. Yamasaki, M. Yamashiro, and Y. Kodani for their assistance in experiments and data analyses. This study was supported in part by grants from the Japan Ministry of Agriculture, Forestry, and Fisheries, by KAKENHI (Grants-in-Aid for Scientific Research) from the Japan Society for the Promotion of Science, and by the "Coastal Area Capability Enhancement in Southeast Asia" project of the Research Institute for Humanity and Nature, Japan.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, duplication, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

## References

- Abdelkrim J, Robertson BC, Stanton JAL, Gemmill NJ (2009) Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *Biotechniques* 46:185–192. doi:10.2144/000113084

- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. doi:[10.1093/nar/25.17.3389](https://doi.org/10.1093/nar/25.17.3389)
- Belkhir K, Borsa P, Chikhi N et al. (2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Université de Montpellier II, Montpellier
- Blacket MJ, Robin C, Good RT, Lee SF, Miller AD (2012) Universal primers for fluorescent labelling of PCR fragments: an efficient and cost-effective approach to genotyping by fluorescence. *Mol Ecol Resour* 12:456–463. doi:[10.1111/j.1755-0998.2011.03104.x](https://doi.org/10.1111/j.1755-0998.2011.03104.x)
- Chistiakov DA, Hellemans B, Volckaert FAM (2006) Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. *Aquaculture* 255:1–29. doi:[10.1016/j.aquaculture.2005.11.031](https://doi.org/10.1016/j.aquaculture.2005.11.031)
- Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol Bioinform* 1:47–50
- Frankham R, Ballou JD, Briscoe DA (2010) Introduction to conservation genetics, 2nd edn. Cambridge University Press, Cambridge
- Gardner MG, Fitch AJ, Bertozzi T, Lowe AJ (2011) Rise of the machines—recommendations for ecologists when using next generation sequencing for microsatellite development. *Mol Ecol Resources* 11:1093–1101. doi:[10.1111/j.1755-0998.2011.03037.x](https://doi.org/10.1111/j.1755-0998.2011.03037.x)
- Gonzalez EG, Zardoya R (2013) Microsatellite DNA capture from enriched libraries. *Methods Mol Biol* 1006:67–87. doi:[10.1007/978-1-62703-389-3\\_5](https://doi.org/10.1007/978-1-62703-389-3_5)
- Guichoux E, Lagache L, Wagner S, Chaumeil P, LÉGer P, Lepais O, Lepoittevin C, Malausa T, Revardel E, Salin F, Petit RJ (2011) Current trends in microsatellite genotyping. *Mol Ecol Resources* 11:591–611. doi:[10.1111/j.1755-0998.2011.03014.x](https://doi.org/10.1111/j.1755-0998.2011.03014.x)
- Guo SW, Thompson EA (1992) Performing the exact test of Hardy–Weinberg proportions for multiple alleles. *Biometrics* 48:361–372. doi:[10.2307/2532296](https://doi.org/10.2307/2532296)
- Malausa T, Gilles A, Meglécz E et al (2011) High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Mol Ecol Resour* 11:638–644. doi:[10.1111/j.1755-0998.2011.02992.x](https://doi.org/10.1111/j.1755-0998.2011.02992.x)
- Megléc E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, Martin JF (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics* 26:403–404. doi:[10.1093/bioinformatics/btp670](https://doi.org/10.1093/bioinformatics/btp670)
- Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact test and ecumenicism. *J Heredity* 86:248–249
- Rice WR (1989) Analyzing tables of statistical tests. *Evol Int J Org Evol* 43:223–225. doi:[10.2307/2409177](https://doi.org/10.2307/2409177)
- Rozen S, Skaletsky H (1999) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 132:365–386. doi:[10.1385/1-59259-192-2:365](https://doi.org/10.1385/1-59259-192-2:365)
- Schoebel CN, Brodbeck S, Buehler D, Cornejo C, Gajurel J, Hartikainen H, Keller D, Leys M, Ríčanová S, Segelbacher G, Werth S, Csencsics D (2013) Lessons learned from microsatellite development for non-model organisms using 454 pyrosequencing. *J Evol Biol* 26:600–611. doi:[10.1111/jeb.12077](https://doi.org/10.1111/jeb.12077)
- Takagi M, Shoji E, Taniguchi N (1999) Microsatellite DNA polymorphism to reveal genetic divergence in ayu, *Plecoglossus altivelis*. *Fish Sci* 65:507–512. doi:[10.2331/fishsci.65.507](https://doi.org/10.2331/fishsci.65.507)
- Takeshima H, Iguchi K, Nishida M (2009) Ayu (*Plecoglossus altivelis*) at a contact zone between amphidromous and landlocked forms: genetic analyses of populations in the Yodo River system. *Zool Sci* 26:536–542. doi:[10.2108/zsj.26.536](https://doi.org/10.2108/zsj.26.536)
- Takeshima H, Iguchi K, Hashiguchi Y, Nishida M (2016) Using dense locality sampling resolves the subtle genetic population structure of the dispersive fish species *Plecoglossus altivelis*. *Mol Ecol* 25:3048–3064. doi:[10.1111/mec.13650](https://doi.org/10.1111/mec.13650)
- van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* 4:535–538. doi:[10.1111/j.1471-8286.2004.00684.x](https://doi.org/10.1111/j.1471-8286.2004.00684.x)
- Wei N, Bemmels JB, Dick CW (2014) The effects of read length, quality and quantity on microsatellite discovery and primer development: from Illumina to PacBio. *Mol Ecol Resour* 14:953–965. doi:[10.1111/1755-0998.12245](https://doi.org/10.1111/1755-0998.12245)
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evol Int J Org Evol* 38:1358–1370. doi:[10.2307/2408641](https://doi.org/10.2307/2408641)