ASSEMBLY ASM291031v2 (GENBANK: GCA_002910315.2) IDENTIFIED AS ASSEMBLY OF THE NORTHERN DOLLY VARDEN (*Salvelinus malma malma*) GENOME, AND NOT THE ARCTIC CHAR (*S . alpinus*) GENOME

SV Shedko

Federal Scientific Center of the East Asia Terrestrial Biodiversity,

Far Eastern Branch of the Russian Academy of Sciences, Vladivostok, 690022

Russia

e-mail: shedko@biosoil.ru

To date, twelve complete genomes representing eleven species belonging to six genera have been sequenced in salmonids. For the genus *Salvelinus* , it was supposed to sequence the genome of Arctic char, one of the most variable species of vertebrate animals. Sequencing was carried out (Christensen et al., 2018) using the tissues of the female IW2-2015 obtained from the company engaged in industrial aquaculture of chars - Icy Waters Ltd. The company exploits two of its own broodstocks - NL and TR, originating from the chars from the Nauyuk Lake and the Tree River (Nunavut, Canada). Since the complete mitochondrial genome of the female IW2-2015 was absent in the published assembly ASM291031v2, we determined its type and complete sequence from the sequence read archives taken from Genbank. It was found that the female's mitogenome belongs to the BERING haplogroup, which is characteristic of Northern Dolly Varden *S. malma malma* . Analysis of other unlinked diagnostic loci encoded by nuclear DNA (ITS1, RAG1, SFO-12, SFO-18, SMM-21) also revealed distinctive characters of Northern Dolly Varden in female IW2-2015. It was concluded that the genomic assembly ASM291031v2 was obtained not from an individual of Arctic char *S. alpinus* , but from an individual of a related species - Northern Dolly Varden *S. malma malma* . The identical to the IW2-2015 female characteristics of diagnostic loci were found in other individuals from the broodstock TR. Apparently, the broodstock TR is entirely a strain derived from Northern Dolly Varden. The choice of a specific specimen for sequencing and assembling the genome of a species is no less important than the choice of a species,

since both of these choices determine the directions for subsequent data use. Assembly ASM291031v2, of course, is important for the study of the genomic architecture of the aquaculture strain TR. However, since it was obtained from a specimen originated from the marginal population isolated from the main range of the species and with some traces of introgressive hybridization, this assembly can hardly be considered as a description of a typical genome of Northern Dolly Varden.

Supplementary Data: https://github.com/svshedko/CHARgenome

Keywords: nuclear genome, mitochondrial genome, assembly, Salmonidae

Page 2

UDC 575.11 + 575.1]: 597.552.51

ASSEMBLY ASM291031v2 (GENBANK: GCA_002910315.2) IDENTIFIED HOW TO ASSEMBLY THE NORTHERN MALMA GENOMA (*Salvelinus malma malma*), NOT ARCTIC RING GENOMES (*S. alpinus*)

S.V. Shedko

Federal Research Center for Biodiversity of Terrestrial Biota East Asia

Far Eastern Branch of the Russian Academy of Sciences, Vladivostok, 690022

e-mail: shedko@biosoil.ru

ANNOTATION

In salmon fish, twelve complete genomes representing eleven species belonging to six genera. IN the genus *Salvelinus was* supposed to be sequenced by the genome of arctic char - one of the most variable vertebrate species. Sequencing done (Christensen et al., 2018) using female tissue IW2-2015 obtained from Icy Waters Ltd, an industrial aquaculture char company. Icy Waters Ltd operates two of its own NL broodstock in production and TR, originating from loaches from Lake. Nuayak and r. Three (Nunavut Province, Canada) respectively. Because in the published assembly ASM291031v2 the complete mitochondrial genome of the female IW2-2015 was absent, then its type and the complete sequence was established by us from the archives of readings taken from Genbank. As a result, it turned out that the mitogen of female IW2-2015 refers to haplogroup BERING, characteristic of northern malma *S. malma malma* . Analysis other unlinked diagnostic loci encoded by nuclear DNA (ITS1, RAG1, SFO-12, SFO-18, SMM-21), also revealed distinctive features in the female IW2-2015 features of northern malma. It is concluded that the genomic assembly of ASM291031v2 obtained not from an individual of the Arctic char *S. alpinus* , but from an individual of a close species - northern malma *S. malma malma* . Same as female IW2-2015, conditions characteristics of diagnostic loci were found in other individuals from

broodstock TR. TR broodstock appears to be the entire line,
derived from northern malm. Selecting a specific individual for sequencing and
assembling the genome of a species is just as important as choosing a species, since both
These choices determine the directions for subsequent use of the data.
Assembling ASM291031v2 is certainly important for research
genomic architecture of the aquaculture line TR. However, since she was
obtained from an individual originating from a marginal population isolated from
the main range of the species and with some traces of the introgressive
hybridization, this assembly can hardly be considered as a description of a typical
genome of northern malma.

Key words : nuclear genome, mitochondrial genome, assembly, Salmonidae.

Page 3

The *char of the* genus *Salvelinus* is one of the evolutionary lines of the salmon fish of this family.
Salmonidae, common in the Northern Hemisphere, mainly in its
high latitudes. Many of the species of loaches have long attracted attention
researchers with their exceptional morphological diversity,
not only when comparing geographically isolated
populations, but also when considering them inside the same water system.
Loaches are one of the dominants in the poor species of northern ichthyocenoses. Leading in
mostly a passing way of life, char, however, is not able to tolerate
low winter temperature of waters with sea salinity. Therefore most
years they spend in fresh waters, while being available to the local
to the population as a source of valuable gourmet food.

Arctic char *S. alpinus* is a circumpolar species, in
where the distinctive features of loaches are most pronounced. Arctic
char can rightfully be considered one of the most variable vertebrate species
animals (Klemetsen et al., 2003; Klemetsen, 2010, 2013). Its range is closer to
to the north pole than the range of any type of freshwater or migratory fish.
Arctic char is important as an object of sport fishing,
fisheries and aquaculture (Johnston, 2006). Therefore, it is not surprising that among all
Salvelinus species first in line for sequencing and annotation of the complete genome
It turned out to be the Arctic char.

For sequencing the genome of arctic char, its assembly and annotation (Christensen et al., 2018) a small (20 cm) female (her identifier - IW2-2015) obtained from Icy Waters Ltd, engaged in industrial aquaculture char. Company is one from the largest manufacturers of products (meat-fillet, caviar) from artificially grown char for sale in North America. In his Icy Waters Ltd uses two brood stocks of char - NL and TR, originating from 15–25 individuals taken respectively from natural populations of loaches of Lake Nuayak and r. Three located on the arctic coast Canada to the hall. Coronations, Kitikmeot Region, Nunavut Province (Goel, 2004; Johnston, 2006; McGowan et al., 2009).

In familiarization with the assembly of the genome of Arctic char ASM291031v2 and the publication accompanying it (Christensen et al., 2018) took over attention to the following contradiction. The article on the second page states that char from lake Nuayak and r. Three have mitochondrial DNA (mtDNA), Haplogroup ARCTIC. However, at Genbank as an assembly

Page 4

non-nuclear genomic arctic char (RefSeq: GCF_002923155.1) mitogen from the haplogroup ACADIA (GenBank: AF154851.1 = NC_000861.1), obtained earlier in another study (Doiron et al., 2002) and for other char from northeast Quebec (Canada).

Our mapping attempt to clarify this situation readings of genomic DNA (Genbank: SRX3776048 – SRX3776052), as well as RNA sequencing (Genbank: SRX2635048, SRX2635052) on diagnostic haplotypes of mtDNA loach showed that the mitogen of the female IW2-2015 clearly belongs to the mtDNA haplogroup BERING, characteristic as it is known, not for arctic char, but for northern malma - *S. malma malma* (Brunner et al., 2001). Other involved diagnostic markers are also testified that the female IW2-2015 most likely is representative of northern Malma, and not arctic char. Results of this The analysis is outlined below.

MATERIAL AND METHOD

As the material, the assembly of the female genome IW2-2015 was used ASM291031v2 (GenBank: GCA_002910315.2), as well as archives of readings of her DNA (GenBank: SRX3776048 – SRX3776052) and RNA from liver and kidney tissues (GenBank: SRX2635048, SRX2635052).

In addition, archives of readings of genomic DNA were analyzed, obtained for two individuals from Icy Waters TR broodstock Ltd - SRX5282523, SRX5282524 and several combinations of tribal hybrids lines TR and NL: TR ♂ × NL ♀ - SRX5282528, SRX5282529; (TR ♂ × NL ♀) ♂ × NL ♀ - SRX5282525; TR ♂ × ((TR ♂ × NL ♀) ♂ × TR ♀) ♀ - SRX5282526.

Identification of ownership of individuals that served as a source genomic data was carried out using six markers: mtDNA and five nuclear DNA loci - ITS1, RAG1, SFO-12, SFO-18, SMM-21.

In alpinoids (arctic char and those close to or derived from it) forms) and malmoid (northern malma *S. malma malma* and related or derivatives of forms) of loaches according to the structure of the controlling region mtDNA previously identified six haplogroups - ATLANTIC, SIBERIA, ACADIA, ARCTIC, BERING (Bruner et al., 2001) and OKHOTSKIA (Shedko et al., 2007). At Arctic char of the Northwest Territories and the province of Nunavut haplogroup ARCTIC is widespread with absolute dominance in the samples haplotype ARC19 (Moore et al., 2015). In northern Malma in Alaska and North- In the Western Territories, as in the Asian part of the range, the haplogroup dominates

Page 5

BERING. The share of the BER12 haplotype from this haplogroup in the samples often reaches 50 and more percent, as, for example, in samples of Malma from the Commander Islands or Kamchatka [(Soshnina et al., 2016): haplotype under number KT962126]. Haplotypes ARC19 (Genbank: EU310899) and BER12 (Genbank: JX261984) are nine different from 499 nucleotide positions of the marker region of the control region mtDNA (eight of them are diagnostic for haplogroups in general). These two haplotypes acted as reference in express type diagnostics mtDNA haplogroups in genomic data. To assemble whole mitogenomes readings were mapped to the corresponding variants of complete mitogenomes: ARC19 - MF621741 (Genbank) and BER12 - KJ746618 (Genbank).

The first internal transcribed spacer (ITS1) is a composite

part of the tandem repeats of ribosomal DNA (rDNA) located in
alpinoid and malmoid loaches on one or several pairs of chromosomes
(Phillips et al., 1999). Alpinoid and malmoid (excluding southern
American malm *S. malma lordi*) loaches are characterized by different variants of ITS1,
characterized by unambiguous substitutions at positions 92 (C / T), 285 (T / A),
448 (C / A), 449 (T / A), 450 (G / C) and 453 (A / C) (Phillips et al., 1999). Coordinates
are given relative to the sequence AF059893 (Genbank) selected as
marker for alpinoid variant. To identify malmoid
option ITS1 in this capacity was the sequence AF059900 (Genbank).

In the genomes of salmon fish, the RAG1 gene (gene activating recombination 1)
present as a single copy. This gene in salmon has a relatively
low rate of evolution, but carries a clear phylogenetic signal
(Shedko et al., 2012a). Alpinoid and malmoid loaches are characterized by different
variants of the RAG1 gene (namely, a fragment of the second exon, 1524 bp long),
characterized by unambiguous substitutions at positions 285 (G / A) and 1113
(G / A) (PopSet: 306977732). Coordinates are given relative to the sequence.
gene RAG1 of northern malma *S. malma malma* GQ871481 (Genbank), selected in
as a reference.

Study of the composition of alleles in the microsatellite locus SFO-12 in different
species of char from their various populations showed (Angers, Bernatchez, 1997;
Wilson, Bernatchez, 1998) that Arctic char (emphasize that in the first of
These works were samples directly from Lake. Nuayak) peculiar
allele SFO-12 (225), and northern malmö *S. malma malma* and its satellite forms
(white char of the Kamchatka River, for example) SFO-12 allele (223). Difference of these alleles
determined by a different number of TG repeats on either side of the GG dimer:

Page 6

$(TG)_4$ GG $(TG)_8$ and $(TG)_5$ GG $(TG)_6$, respectively. Because of this, the identification of these
the two alleles is quite unambiguous, despite the minimal difference in
their length. Sequences of these alleles along with their flanking
plots were restored based on Fig. 1-2 works (Angers, Bernatchez,
1997). As a result, their length turned out to be equal to 221 (for the allele indicated in
cited work as 223) and 223 (for the allele under code 225). Their same lengths
obtained by analyzing genomic data.

Alleles of the microsatellite locus SFO-18 (155–157, on the one hand, and 161–164, on the other hand) differentiate malm and Arctic char from Alaska and east of the arctic coast of Canada (Angers, Bernatchez, 1996; Crane et al., 2014; Ditlecadet et al., 2006; Hart et al., 2015; Taylor, May-McMally, 2015). Alleles differ in the number of repetitions of the CA motive in the basic variants - 7 in malma versus 10 in Arctic char (Genbank: MN530967, MN530968). Populations
The Arctic char of Nunavut province is unusual in that in them approximately
Both alleles are found at equal frequency (Harris et al., 2016).

The SMM-21 microsatellite locus is considered another diagnostic locus.
marker for arctic char and northern malma. For arctic char and
close forms are characterized by alleles 105–111 bp long (Crane et al., 2014; Gordeeva et al., 2010; Senchukova, 2014; Hart et al., 2015; Taylor, May-McMally, 2015), and for northern malma - 115–158 bp (Crane et al., 2005, 2014; Salmenkova et al., 2009; Senchukova, 2014; Harris et al., 2015; Hart et al., 2015; Taylor, May-McMally, 2015). The differences in lengths are caused by the variation in the number of repeats of the TC motive (Genbank: AY327128, MN530969 – MN530972).

Establishment of allele variants of six diagnostic loci in
genomic data was produced as follows.

The haplotype sequences of the control region ARC19 and
BER12, two variants of ITS1 (AF059893 and AF059900), alleles 223 and 221 of the locus SFO-12, as well as the RAG1 gene, were combined into a single file that performed in as a composite reference. For the SFO-18 and SMM-21 loci, the references were individual. Readings from one or another archive were aligned with
reference using the blastn program, available on the site
https://blast.ncbi.nlm.nih.gov. Alignment was performed in "megablast" mode (search for sequences with a high level of similarity) (Zhang et al., 2000) with limiting the maximum number of target sequences to 1000 pieces.
This method is easy to use and suitable for our case.
sensitivity and the necessary level of selectivity. Received Out

Page 7

blastn program, the SAM file was then analyzed using the Tablet program v.1.19.05.28 (Milne et al., 2013), and by the nature of the coating and the type of nucleotides in diagnostic positions established a variant of alleles of marker loci.

To clarify megablast alignment at the SMM-21 locus, aligned
the readings were extracted and mapped to the reference again, but with the help of
Bowtie 2 v2.3.2 (Langmead, Salzberg, 2012).

In those cases when it was required to use archive files as fully as possible
readings (to verify megablast alignment, coverage assessment, or
subsequent assembly of the target locus sequence), primary data
downloaded from the European Nucleotide Archive (www.ebi.ac.uk/ena) as FASTQ-
files containing broken reading directions. Then using
bbduk utility from BBMap filtering data for
the residual presence of adapters used in sequencing. Further
alignment was performed using the bbmap utility from the BBMap v.38.00 package
(Bushnell, 2014) with the keys "local" and "vslow" or Bowtie 2 in
"local" or "end-to-end". If necessary, aligned readings in the format
FASTQ or FASTA served as input to build the target
sequences using SPAdes v3.11.1 (Nurk et al., 2013) or
idba_ud v1.0.9 (Peng et al., 2012).

For resource-intensive operations and calculations, we used the opportunities
multiprocessor computing complex IRUS17 Center collective
use "Far Eastern Computing Resource" FEB RAS (g.
Vladivostok).

RESULTS AND DISCUSSION

Genomic data analysis results for the entire used set
marker loci are presented in the table.

As indicated in the introduction, authentic for the female IW2-2015 is complete
the mtDNA sequence in assembly ASM291031v2 is missing. Blastn analysis is not
also revealed in this assembly correctly assembled sections of ribosomal DNA.
The most similarities with the reference sequence ITS1 showed
plot NC_036862.1: 28683274-28682712 (Genbank). But the level of this similarity
was relatively small (95–96%). If added to this reference
its flanking sections 18S and 5.8S RNA, then the level of its similarity with
the leveled area (Genbank: NC_036862.1: 28683274-28682712) fell to 91%.
At the same time, ITS1 sequences obtained from archive analysis
readings of genomic DNA or RNA (all three ITS1 variants found together with

Page 8

adjacent sections of the 18S and 5.8S RNA genes were deposited at Genbank: MN530964 – MN530966), had a 100% similarity with those already known for char (table). From this we can understand that in the assembly of ASM291031v2, the collected rDNA sites are essentially absent.

Sequence variants of four other marker loci (RAG1, SFO-12, SFO-18 and SMM-21) installed as a result of assembly blastn analysis ASM291031v2, in each case identified as alleles characteristic of northern malmo.

Analysis of archives of readings of the genomic DNA of the female IW2-2015 taken from Genbank, revealed the following picture. Female IW2-2015 has mtDNA from haplogroup BERING (haplotype - BER12). RAG1, SFO-18 and SMM-21, in full agreement with the results of the analysis of the assembly ASM291031v2, belong to the variant of northern malm.

The vast majority belongs to the variant of northern malma readings for locus ITS1. As a minor component among the latter made alpinoid variant A2. It is worth emphasizing that, judging by the results analysis of reads obtained during RNA sequencing (Genbank: SRX2635048, SRX2635052), only one option is expressed from them - malmoid.

At the locus SFO-12, both variants of alleles (and alpinoid, and malmoid). In two experiments (Genbank: SRX3776048–SRX3776049) is dominated by the variant of malmoid char, in others (Genbank: SRX3776050 – SRX3776052) - their ratio is approximately equal.

Analysis of reading archives obtained for two individuals from broodstock TR (Genbank: SRX5282523 – SRX5282524), yielded results completely identical results of the analysis of the female IW2-2015. That is, judging by them, the female IW2-2015 certainly belongs to the TR line.

From the read archives SRX3776048 and SRX5282523 two were assembled variants of complete mitochondrial genomes (Genbank: MN530959, MN530961), characterizing the line TR. These mitogenomes differed from each other in two nucleotide positions, and from mitogenomes of malmoid loam of Kamchatka (Genbank: KJ746618, KU674351, KU674352, KT266871, KT266870) - 30–39 each positions.

In TR ♂ × NL ♀ Hybrids (Genbank: SRX5282528 – SRX5282529) Found htlotype mtDNA from haplogroup ARCTIC, and alternative nuclear loci

allele variants in approximately equal proportions. Characteristics of one of

Page 9

nine

return crosses (Genbank: SRX5282526) were close to
characteristics of the TR line (by the way, the complete mitogen of this individual was completely identical to the mitogen recovered from the archive SRX5282523), and the other (Genbank: SRX5282525) - to the characteristics of an active char. Note that in in the latter case, as with TR ♂ × NL ♀ hybrids, a special variant of ITS1 was found - alpinoid variant A3. From the read archives SRX5282525 and SRX5282528 two variants of complete mitogenomes from the haplogroup ARCTIC were restored (Genbank: MN530963, MN530962), distinguished by two nucleotide positions. Variants of these mitogenomes, as well as alleles of nuclear loci, obtained for archive SRX5282525, apparently, can be considered as characteristics close to the characteristics (if not identical to them) of individuals from broodstock NL.

According to the results of the analysis, it is clear that the characteristics tribal lines TR and NL repeat the characteristics of northern Malm and Arctic char, respectively. In other words, these lines, throughout appearances come from individuals belonging to two different species of char. IN In this regard, it can be argued that the published assembly of the IW2- female genome 2015 (ASM291031v2, Genbank: GCA_002910315.2) is a genome assembly northern malma *S. malma malma* .

Three River in the hall. Coronations, where the founding individuals of the TR line come from, one of the well-known [especially among lovers of sport of alpine fishing from for the opportunity to mark a large trophy weighing up to 14 kg (Moshenko et al., 1984)] Habitat for malma in the Kitikmeot region of Nunavut province (Alfonso et al., 2018). For Malma, this region is the limit of its eastward distribution. along the arctic coast of North America. Note that here passes the eastern border of the ranges of many other species of the Pacific marine or anadromous fish (Alfonso et al., 2018): Pacific herring *Clupea pallasii* , chum salmon *Oncorhynchus keta* , sockeye salmon *O. nerka* , *Chinook* salmon *O. tshawytscha* , Asian toothy smelt *Osmerus dentex* , Pacific navaga *Eleginus gracilis* , star Flounders *Platichthys stellatus* , Far Eastern catfish *Anarhichas orientalis* . So

Thus, it should be borne in mind that the assembly ASM291031v2 characterizes the genome individuals originating from a marginal population isolated from the main range species ending in the river basin. Mackenzie

Marginal populations are often small in number and may therefore be susceptible to hybridization with populations of related species. It is very likely that the presence of the alpinoid variant A2 ITS1, as well as the SFO-12 allele (223) in

Page 10

genomes of individuals from the TR line are traces of hybridization northern malma and arctic char in the Kitikmeot region. Wide the presence of the SFO-18 allele (156) in Arctic char from this region (Harris et al., 2016) may also be a consequence of such hybridization. Therefore the female genome IW2-2015 cannot be considered as a description of a typical northern genome malma. On the other hand, this genome is not indicative of arctic char, since the ancestral lines of malmoid and alpinoid loaches dispersed quite a long time ago - 3.03 (1.18–5.00) million years ago (Shedko et al., 2012b). At the same time, the data obtained in (Christensen et al., 2018), certainly important for research on genomic architecture aquacultural TR and NL lines of these char species.

We emphasize that the completeness and quality of the existing assembly ASM291031v2 so far, unfortunately, cannot be considered high. This is evidenced by first, that the work (Christensen et al., 2018) did not contain mitochondrial genome and, secondly, that the assembly ASM291031v2 lacks the correct ribosomal DNA sequences.

The presence of an adequate reference genome is a basic condition for all genomic research in a particular group of organisms (Elmer, 2016). Here, of course, it's important not only to choose a species as a typical representative of this group, but also the choice of a specific individual, representing the view itself. Ideally, this individual should come from a population, typical for this species, and located in the central part of its range. Hopefully, in future attempts at sequencing the loach genomes (Malma, Arctic char or others) this circumstance will be given Special attention.

THANKS

The results are obtained using the equipment
"Far Eastern Computing Resource" IAPU FEB RAS
(https://www.cc.dvo.ru).

CONFLICT OF INTEREST

The author states that he has no conflict of interest.

FINANCING

The study was conducted as part of the research topic No. AAAA-A17-117062710083-0. Part of the work (mapping and assembly of sequences) funded by the Priority Scientific research for the integrated development of the Far Eastern Branch RAS "(grant No. 18-4-040).

APPENDICES

All significant nucleotide sequences recovered from read archives deposited with Genbank under access numbers MN530959–MN530972.

The main source data used to build the table, and other materials can also be found in the repository, available at https://github.com/svshedko/CHARgenome.

LIST OF REFERENCES

Gordeeva N.V., Osinov A.G., Alekseev S.S. et al. Genetic differentiation Arctic char *Salvelinus alpinus* complex Transbaikalia, identified by microsatellite markers // Vopr. ichthyology. 2010.V. 50. No. 3. P. 293–304.

Senchukova A.L. Genetic differentiation of char (genus *Salvelinus*) Kronotskaya river system (Kamchatka). The dissertation ... cand. biol. sciences. M .: Moscow State University. 2014.106 p.

Salmenkova E.A., Omelchenko V.T., Afanasyev K.I. et al. Genetic divergence of white loam populations of *Salvelinus albus* , northern and southern forms of Malma *S. malma* (Salmonidae) at microsatellite DNA loci // Vopr. ichthyology. 2009.V. 49. No. 6. P. 752–762.

Alfonso NR, Coad BW, Sawatzky CD, Reist JD . Distributional records for marine fishes of Arctic Canada // Can. Data Rep. Fish Aquat. Sci. 2018.No 1287. 295 p.

Angers B., Bernatchez L. Usefulness of heterologous microsatellites obtained from brook charr, *Salvelinus fontinalis* Mitchill, in other *Salvelinus* species // Mol. Ecol. 1996. V. 5. No. 2. P. 317–319.

Angers B., Bernatchez L. Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information // Mol. Biol. Evol. 1997. V. 14. No. 3. P. 230–238.

Brunner PC, Douglas MR, Osinov A. et al. Holarctic phylogeography of Arctic Charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences // Evolution. 2001. V. 55. No. 3. P. 573.

Bushnell B. BBMap: a fast, accurate, splice-aware aligner // Joint Genome Institute, department of energy. 2014. https://www.osti.gov/biblio/1241166-bbmap-fast-acc.

Page 12

Christensen KA, Rondeau EB, Minkley DR et al . The arctic charr (*Salvelinus alpinus*) genome and transcriptome assembly // PLoS One. 2018. V. 13. No. 9. P. 1–30.

P. Crane, Viavant T., J Wenburg . Patterns of Dolly overwintering Varden *Salvelinus malma* in the Sagavanirktok River in the Alaskan North Slope inferred using mixed-stock analysis // Alaska Fish. Tech. Rep. 2005. No. 84. P. 1–22.

Crane PA, Lisac MJ, Lewis CJ, Wenburg JK Genetic baseline development for Dolly Varden in southwestern Alaska // Alaska Fisheries Data Series Number 2014-2. 2014.40 p.

Ditlecadet D., Dufresne F., Le François NR, Blier PU Applying microsatellites in two commercial strains of Arctic charr (*Salvelinus alpinus*): Potential for a selective breeding program // Aquaculture. 2006. V. 257. No. 1-4. P. 37–43.

Doiron S., Bernatchez L., Blier PU A Comparative mitogenomic analysis of the potential adaptive value of Arctic charr mtDNA introgression in Brook charr populations (*Salvelinus fontinalis* Mitchill) // Mol. Biol. Evol. 2002. V. 19. No. 11. P. 1902-1909.

Elmer KR Genomic tools for new insights to variation, adaptation, and evolution in the salmonid fishes: a perspective for charr // Hydrobiologia. 2016. V. 783. No. 1. P. 191–208.

Goel AK Developing broodstock of Arctic charr (*Salvelinus alpinus* L.). M.Sc. Thesis. Simon Fraser Univ. 2004.186 p.

Harris LN, Bajno R., Gallagher CP et al. Life-history characteristics and landscape attributes as drivers of genetic variation, gene flow, and fine-scale population structure in Northern Dolly Varden (*Salvelinus malma malma*) in Canada // Can. J. Fish Aquat. Sci. 2015. V. 72. No. 10. P. 1477-1493.

Harris LN, Moore J.-S., Bajno R., Tallman RF Genetic stock structure of anadromous Arctic char in Canada's central Arctic: Potential implications for the management of Canada's largest Arctic char commercial fishery // North Am. J. Fish. Manag. 2016. V. 36. No. 6. P. 1473-1488.

Hart LM, Bond MH, May-McNally SL et al. Use of otolith microchemistry and stable isotopes to investigate the ecology and anadromous migrations of Northern Dolly Varden from the Egegik River, Bristol Bay, Alaska // Environ. Biol. Fishes. 2015. V. 98. No. 6. P. 1633-1643.

Johnston G. Arctic charr aquaculture. Oxford, UK: Blackwell Publishing. 2006.274 p.

Klemetsen A. The charr problem revisited: Exceptional phenotypic plasticity promotes ecological speciation in postglacial lakes // Freshw. Rev. 2010. V. 3. No. 1. P. 49–74.

Page 13

A Klemetsen . The most variable vertebrate on Earth // J. Ichthyol. 2013. V. 53. No. 10. P. 781–791.

Klemetsen A., Amundsen PA, Dempson JB 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. V. 12. No. 1. P. 1–59.

Langmead B., Salzberg SL Fast gapped-read alignment with Bowtie 2 // Nat. Methods 2012. V. 9. No. 4. P. 357–359.

McGowan CR, Lucas JJ, Keddie E. et al. Application of salmonid genomic research in a mid-sized commercial broodstock facility: The Icy Waters Arctic charr experience // Bull. Aquacul. Assoc. Canada 2009. V. 107-3. P. 31–43.

Milne I., Stephen G., Bayer M. et al . Using Tablet for visual exploration of second-generation sequencing data. // Brief. Bioinform 2013. V. 14. No. 2. P. 193–202.

Moore J.-S., Bajno R., Reist JD, Taylor EB Post-glacial recolonization of the North

American Arctic by Arctic char (*Salvelinus alpinus*): genetic evidence of multiple northern refugia and hybridization between glacial lineages // J. Biogeogr. 2015. V. 42. No. 11. P. 2089–2100.

Moshenko RW, Peet RF, Dahlke LW, Dowler DH The Arctic charr sport fishery at the Tree River, Northwest Territories, Canada // Biology of the Arctic charr. Proceedings of the international symposium on Arctic charr (Eds Johnson L., Burns BL). Winnipeg: University of Manitoba Press. 1984. P. 1964-1978.

Nurk S., Bankevich A., Antipov D. et al. Assembling genomes and mini-metagenomes from highly chimeric reads // Lecture Notes in Computer Science (Eds. Deng M., Jiang R., Sun FZX) Berlin: Springer. 2013. P. 158–170.

Peng Y., Leung HCM, Yiu SM, Chin FYL IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth // Bioinformatics. 2012. V. 28. No. 11. P. 1420-1428.

RB Phillips, Gudex LI, KM Westrich, DeCicco A . Combined phylogenetic analysis of ribosomal ITS1 sequences and new chromosome data supports three subgroups of Dolly Varden char (*Salvelinus malma*) // Can. J. Fish. Aquat. Sci. 1999. V. 56. No. 8. P. 1504-1511.

Shedko SV, Ginatulina LK, Miroshnichenko IL, Nemkova GA Phylogeography of mitochondrial DNA in South Asian Dolly Varden char *Salvelinus curilus* Pallas, 1814 (Salmoniformes, Salmonidae): Mediated gene introgression? // Russ. J. Genet. 2007. V. 43. No. 2. P. 165–176.

Page 14

Shedko SV, Miroshnichenko IL, Nemkova GA Phylogeny of salmonids (Salmoniformes: Salmonidae) and its molecular dating: Analysis of nuclear RAG1 gene // Russ. J. Genet. 2012a. V. 48. No. 5. P. 575-579.

Shedko SV, Miroshnichenko IL, Nemkova GA Data from: Phylogeny of salmonids (Salmoniformes: Salmonidae) and its molecular dating: analysis of nuclear RAG1 gene. Dryad Digital Repository. 2012b. https://doi.org/10.5061/dryad.ph602

Soshina VA, Pavlov SD, Zelenina DA Genetic diversity of charrs of the Commander Islands based on the analysis of mitochondrial DNA // Russ. J. Genet. 2016.V. 52. No. 11. P. 1208–1213.

Taylor EB, May-McNally SL Genetic analysis of Dolly Varden (*Salvelinus malma*) across its North American range: evidence for a contact zone in southcentral Alaska // Can. J. Fish. Aquat. Sci. 2015. V. 72. No. 7. P. 1048-1057.

Wilson CC, Bernatchez L. The ghost of hybrids past: fixation of arctic charr (*Salvelinus alpinus*) mitochondrial DNA in an introgressed population of lake trout (*S. namaycush*) // Mol. Ecol. 1998. V. 7. No. 1. P. 127–132.

Zhang Z., Schwartz S., Wagner L., Miller WA Greedy algorithm for aligning DNA sequences // J. Comput. Biol. 2000. V. 7. No. 1–2. P. 203–214.

Page 15

Table . Variants of alleles of marker loci characteristic of arctic char and northern Malma and females IW2-2015 (in curly brackets - the ratio of reading options in the case of polymorphism), two individuals from the broodstock TR, as well as four hybrid individuals obtained in several variants of crossing tribal lines TR and NL.

Table . Alleles of diagnostic loci characteristic for Arctic char and Northern Dolly Varden, revealed in female IW2-2015 (in curly brackets is the ratio of alternative reads in the case of polymorphism), two individuals from broodstock TR, as well as four variants of hybrids.

Taxon / Sample / data (numbers Access Genbank)	mtDNA	ITS1 [1]	RAG1	SFO-12	SFO-18	SMM-21
S. alpinus	ARCTIC	alpinus	alpinus	223	162–164	105–111 (alpinus)
S. malma malma	Bering	malma	malma	221	156	115–158 (malma)
IW2-2015						
GCA_002910315.2	-	-	malma [2]	221 [3]	156 [4]	malma [5]
SRX3776048	BER12 A2 / malma {~ 1: 55}		malma	221/223 {~ 13: 1}	156	malma
SRX3776049	BER12 A2 / malma {~ 1: 34}		malma	221/223 {~ 6: 1}	156	malma
SRX3776050	BER12 A2 / malma {~ 1: 55}		malma	221/223 {~ 1: 1}	156	malma
SRX3776051	BER12 A2 / malma {~ 1: 235}		malma	221/223 {~ 1: 2}	156	malma
SRX3776052	Ber12	malma	malma	221/223 {~ 1: 1}	156	malma
SRX2635048, SRX2635052	Ber12	malma	malma	-	-	-
TR						
SRX5282523	BER12 A2 / malma {~ 1: 30}		malma	221/223 {~ 1: 1}	156	malma
SRX5282524	BER12 A2 / malma {~ 1: 30}		malma	221/223 {~ 1.5: 1}	156	malma
Hybrids						
SRX5282528 TR ♂ × NL ♀	ARC19 A3 / malma {~ 1: 2}		alpinus [6] / malma {~ 1: 3}	221/223 {~ 2: 1}	156/162 {~ 2: 1}	alpinus / malma {~ 2: 1}
SRX5282529 TR ♂ × NL ♀	ARC19 A3 / malma {~ 1: 2}		alpinus [6] / malma {~ 1: 1}	221/223 {~ 2: 1}	156/162 {~ 1.4: 1}	alpinus / malma {~ 4: 1}
SRX5282525 (TR ♂ × NL ♀) ♂ × NL ♀	ARC19 A3 / malma {~ 5: 1}		alpinus [6]	223	156/162 {~ 2: 1}	alpinus
SRX5282526 TR ♂ × ((TR ♂ × NL ♀) ♂ × TR ♀) ♀	BER12 A2 / malma {~ 1: 33}		malma	221/223 {~ 1.5: 1}	156/162 {~ 2: 1}	alpinus / malma {~ 3: 1}

Notes . [1] A2 - alpinoid variant (Genbank: AF059899); A3 - alpinoid variant (Genbank: AF059897), found in arctic char from Lake Nuayak (Phillips et al., 1999), is unique to char by nucleotide substitution C> T at position 298. [2] NC_036842.1: 69870375–69871898 (Genbank), LG4q.1: 29. [3] NC_036863.1: 41094706-41094926 (Genbank), LG23. [4] NC_036875.1: 27971989–27972144 (Genbank), LG36. [5] NC_036838.1: 46248006–46248134 (Genbank), LG1. [6] RAG1 rare for alpinoid char (Genbank: GQ871484), so far it has been found only at Boganidskoy paley from Lake. Elgygytgyn in Chukotka (Shedko et al., 2012a).

The Notes . [1] A2 - alpinoid variant (Genbank: AF059899); A3 - alpinoid variant (Genbank: AF059897), found in Arctic char from Nuayak Lake (Phillips et al., 1999), is distinguished by a unique for alpinoid chars C> T nucleotide substitution in position 298. [2] NC_036842.1: 69870375–69871898 (Genbank), LG4q.1: 29. [3] NC_036863.1: 41094706–41094926 (Genbank), LG23. [4] NC_036875.1: 27971989–27972144 (Genbank), LG36. [5] NC_036838.1: 46248006–46248134 (Genbank), LG1. [6] A rare for alpinoid chars variant of RAG1 (Genbank: GQ871484), so far it has been found only in Boganida char from Lake El'gygytgyn in Chukotka (Shedko et al., 2012a).