A high-quality chromosome-level genome assembly of rohu carp, Labeo rohita, and discovery of SNP markers

This manuscript (permalink) was automatically generated from IGBB/rohu-genome@ac1ca5b on February 7, 2022.

Authors

Mark A. Arick II[1], Chuan-Yu Hsu[1], Zenaida Magbanua[1], Corrinne E. Grover[2], Emma R. Miller[2], Olga Pechanova[1], Adam Thrash[1], Ramey C. Youngblood[1], Lauren Ezzell[1], Samsul Alam[3], John Benzie[3], Matthew Hamilton[3], Attila Karsi[3], Mark L. Lawrence[3], Daniel G. Peterson[1]

Abstract

Introduction

Methods & Materials

Genome Sequencing & Flowcytometry

ALAM – Rohu blood collection

Flow cytometry

The genome size of *Labeo rohita* was estimated for five samples via flow cytometry using trout erythrocyte nuclei (TENs; https://www.biosure.com/tens.html) as a standard (genome size=6.5pg). For each sample, nuclei were stabilized in 200 ul of LB01-propidium iodide (PI) buffer as per [1], and two drops of TENs standard were used per 50ul of fish blood. Each sample was measured twice, totaling 10 runs overall. Only measurements with greater than 5,000 nuclei and a coefficient of variation (CV) of less than 3% were retained [1].

Illumina short-read sequencing

A total of 2 µg of extracted genomic DNA was used for DNA-Seq library preparation using Illumina TruSeq DNA PCR-free Library Prep Kit (Illumina, San Diego, CA, USA). The final DNA-Seq library with the insert size range of 350 bp to 450 bp was submitted to Novogene company (www.en.novogene.com) for total of 2 lanes of PE150 Illumina HiSeq X-Ten (Illumina, San Diego, CA, USA) sequencing.

Oxford Nanopore sequencing

For each Nanopore sequencing run, 2 to 2.5 µg of genomic DNA was used in the library preparation with Nanopore Genomic DNA Ligation Sequencing Kit SQK-LSK 109 (Oxford Nanopore Technologies, Oxford, UK). The final library (about 700 to 750 ng) was loaded on a Nanopore Flow Cell R9.4.1 (Oxford Nanopore Technologies, Oxford, UK) and sequenced on GridION sequencer (Oxford Nanopore

Technologies, Oxford, UK) for 48 hours. A total of 10 flow cell runs were conducted for the genome assembly.

Hi-C sequencing

One hundred µl of fish blood was subjected to the Hi-C library preparation using the Proximo Hi-C Animal Kit (Phase Genomics, Seattle, WA, USA). The final DNA-Seq library was submitted to Novogene (www.en.novogene.com) for 1 lane of PE150 Illumina HiSeq X-Ten (Illumina, San Diego, CA, USA) sequencing.

Bionano

Assembly & Annotation

Jellyfish (v2.2.10) [2] and GenomeScope (v1.0) [3] estimated the genome size using the Illumina paired reads digested into 50-mers.

Nanopore data was filtered to remove the control lambda-phage and sequences shorter than 1000 bases using the nanopack tool suite (v1.0.1) [4]. Trimmomatic (v0.32) [5] removed adapters, trimmed low quality bases, and filtered reads shorter than 85bp. The filtered nanopore data was assembled into contigs using wtdbg2 (v2.4) [6]. The contigs were polished using the two iterations of racon (v1.4.0) [7] with minimap2 (v2.17) ([8]) mapping the nanopore reads. The contigs were further polished using pilon (v1.23) [9] with bwa (v0.7.10) [10] mapping the Illumina paired reads. The contigs were scaffolded using Bionano Solve (Solve3.4.1_09262019) [[11]/) and SALSA (v2.3) [12]. That scaffolds larger than 10Mb were linked and oriented based on the Onychostoma macrolepis genome [13], the most similar chromosome assembly available on NCBI, using ragtag (v1.1.1) [14].

RepeatModeler (v2.0.1) [15] and RepeatMasker (v4.1.1) [**repeatmasker?**] were used to create a species-specific repeat database, and mask those repeats in the genome. All available RNA-seq libraries for *L. rohita* [1] were downloaded from NCBI and mapped to the masked genome using hisat2 (v2.1.0) [16]. These alignments were used in the mikado (v2.0rc2) [17] and braker2 (v2.1.5) [18] pipelines. Mikado uses the putative transcripts assembled from the RNA-seq alignments generated via stringtie (v2.1.2) [19], cufflinks (v2.2.1) [20], and trinity (v2.11.0) [21]; along with the junction site prediction from portcullis (v1.2.2) [22], the alignments of the putative transcripts with UniprotKB Swiss-Prot (v2021.03) [**uniprot?**], and the ORFs from prodigal (v2.6.3) [23] to select the best representative for each locus. Braker2 uses the RNA-seq alignments and the gene prediction from GeneMark-ES (v4.61) [24] to train a species-specific Augustus (v3.3.3) [25] model. Maker2 (v2.31.10) [26] predicts genes based on the new Augustus, GeneMark, and SNAP models; modifying the predictions based on the available RNA and protein evidence from the *Cyprinidae* family in the NCBI RefSeq database. Any predicted genes with an AED above 0.47 were removed from further analysis. The surviving genes were functionally annotated using InterProScan (v5.47-82.0) [27] and BLAST+ (v2.9.0) [28] alignments against the UniprotKB Swiss-Prot database. Busco (v5.2.2) [29] was used to verify the completeness of both the genome and annotations against the actinopterygii_odb10 database. Lastly, genes spanning large gaps or completely contained within another gene on the opposite strand were removed using a custom perl script (GITHUB).

Table 1: List of SRA accessions used in annotation pipeline. A table of all metadata available for these accessions can be found here .

BioProject	BioSample	Run	Instrument	sex	Tissue
PRJNA401304	SAMN07602342	SRR6003546	Illumina HiSeq 2000	female	Brain
PRJNA401304	SAMN07602341	SRR6003547	Illumina HiSeq 2000	female	Brain

BioProject	BioSample	Run	Instrument	sex	Tissue
PRJNA401304	SAMN07602344	SRR6003548	Illumina HiSeq 2000	female	Pituitary
PRJNA401304	SAMN07602343	SRR6003549	Illumina HiSeq 2000	female	Pituitary
PRJNA401304	SAMN07602346	SRR6003550	Illumina HiSeq 2000	female	Gonad
PRJNA401304	SAMN07602345	SRR6003551	Illumina HiSeq 2000	female	Gonad
PRJNA401304	SAMN07602348	SRR6003552	Illumina HiSeq 2000	female	Liver
PRJNA401304	SAMN07602347	SRR6003553	Illumina HiSeq 2000	female	Liver
PRJNA449818	SAMN08918388	SRR6987066	NextSeq 500	female	Pooled tissue
PRJNA449818	SAMN08918389	SRR6987067	NextSeq 500	male	Pooled tissue
PRJNA449818	SAMN08918390	SRR6987068	NextSeq 500	female	whole body
PRJNA528865	SAMN11246839	SRR8816555	Illumina HiSeq 2500	not applicable	Liver
PRJNA528865	SAMN11246841	SRR8816556	Illumina HiSeq 2500	not applicable	Liver
PRJNA528865	SAMN11246840	SRR8816557	Illumina HiSeq 2500	not applicable	Liver
PRJNA450719	SAMN08944450	SRR7027730	NextSeq 500	female	Pooled tissue
PRJNA450719	SAMN08944449	SRR7027731	NextSeq 500	male	Pooled tissue
PRJNA450719	SAMN08944451	SRR7027732	NextSeq 500	male	Whole body

ddRAD-Seq & SNP Discovery

WORLDFISH – Fin clipping collection

ZENAIDA – ddRAD-Seq method

CORRINNE – SNP Discovery, population analyses, interested in using the gender, which contigs associated with sex

Data Availability

The data used for the *L. rohita* genome and annotation are available at NCBI under the BioProject PRJNA650519. The assembled genome sequence and annotations are available at GenBank under accessions JACTAM000000000. The raw data is available at the SRA (Sequence Read Archive) under accessions SRR12580210 – SRR12580221.

Results & Discussion

Sequencing & Assembly

The C-value of *L. rohita* was previously reported as 1.99 pg (~1.95Gb) using Feulgen densitometry [30] or 1.5Gb using k-mer estimation [31]. However, the flow cytometry results (Table [2]) show a C-value of 0.99 pg (~0.97Gb) with a standard deviation of 0.02 across all measurements. The smaller C-value is also closer to the genome estimate produced by GenomeScope (0.97Gb) and the final genome assembly size of 0.95 Gb.

Table 2: Flow cytometry results for 5 *L. rohita* blood samples, measured twice. 1) Trout erythrocyte nuclei: Genome size = 5.19pg. 2) Genome estimate calculated as (average sample fluorescence/ average standard fluorescence * standard

genome size) in picograms DIPLOID (i.e., 2C).

Specimen Name	Number of Sample nuclei	Average sample fluorescence	Number of standard nuclei	Average standard fluorescence	Estimated Genome size [2]	HAPLOID
Fish 1 Sample 1	16020	27350	2065	69247	2.049857756	1.024928878
Fish 2 Sample 1	13082	25929	6570	66671	2.018441451	1.009220725 5
Fish 2 Sample 2	15402	25665	4354	67489	1.973674969	0.986837 4845
Fish 3 Sample 1	15124	25798	4442	68195	1.963364176	0.981682088
Fish 3 Sample 2	14923	25763	4823	68837	1.942414254	0.971207127
Fish 4 Sample 1	13320	26346	5913	69665	1.962760927	0.981380 4635
Fish 4 Sample 2	5624	26612	4097	68876	2.005288925	1.002644 4625
Fish 5 Sample 1	6771	25761	3080	68825	1.942602107	0.971301 0535
Fish 5 Sample 2	15926	26369	3352	68832	1.988248344	0.994124172
Standard only [1]	3	25258	3311	64331	NA	
				Average	1.982961434	0.991480 7172
				Standard Deviation	0.03607582	0.018037 90999

A total of 130.5 Gb of Nanopore long reads from 44.7 million read, and 261 Gb of Illumina short reads from 870 million pairs were produced, along with 382 million pairs (114 Gb) for the Hi-C library. The initial *de novo* assembly consisted of 4999 contigs with an N50 of 1.28 Mb. After the Bionano and HiC data was incorporated, the total number of sequences dropped to 2899 and the N50 increased to 29.9 Mb. The final assembly consisted of 25 chromosome length scaffolds and 2844 unplaced scaffolds, ranging in size from 1,479bp to 7.18 Mb. Table [3] contains a common assembly statistics for each step. The final genome size is 97.9% of the estimated genome size. The annotation pipeline produced 51,079 primary transcripts, 31,274 surviving the AED, gap, and overlapping filters. BUSCO analysis show the genome completely contains 98.1% of the 3640 orthologs in the actinopterygii_odb10 database with 37 (1%) duplicated; while, the filtered transcriptome contains 84.5% of the total orthologs complete with 74 duplicated. A complete comparison of the BUSCO analyses can be found in Table [4].

Table 3: Assembly statistics for each stage of the assembly

n	n:500	L50	min	N75	N50	N25	E-size	max	sum	name
4999	4999	202	1348	514919	1281850	2395030	1727184	7832582	9.43E+08	wtdbg2
3709	3706	15	1479	1.13E+07	2.65E+07	3.08E+07	2.20E+07	3.79E+07	9.46E+08	bionano
2899	2896	14	1479	2.64E+07	2.99E+07	3.43E+07	2.69E+07	4.45E+07	9.46E+08	hic
2872	2869	13	1479	2.88E+07	3.25E+07	3.61E+07	3.00E+07	4.53E+07	9.46E+08	ragtag

Table 4: BUSCO analysis for the genome and transcriptome, before and after AED filtering.

Type	Genome	Unfiltered Transcriptome	Filtered Transcriptome
Complete BUSCOs (C)	3571	3139	3078
Complete and single-copy BUSCOs (S)	3534	3064	3001
Complete and duplicated BUSCOs (D)	37	75	74
Fragmented BUSCOs (F)	23	192	170
Missing BUSCOs (M)	46	309	392
Total BUSCO groups searched	3640	3640	3640

```
==================================================
file name: Rohu.genome.fa
sequences:          2872
total length: 1128029156 bp (945637473 bp excl N/X-runs)
GC level:         36.05 %
bases masked:  465289941 bp (41.25 %)
==================================================
               number of      length   percentage
               elements*    occupied  of sequence
--------------------------------------------------
Retroelements       123482     53424977 bp    4.74 %
   SINEs:                0            0 bp    0.00 %
   Penelope              0            0 bp    0.00 %
   LINEs:            29838     13842585 bp    1.23 %
    CRE/SLACS            0            0 bp    0.00 %
     L2/CR1/Rex      15537      8265099 bp    0.73 %
     R1/LOA/Jockey     981       643697 bp    0.06 %
     R2/R4/NeSL        936       732043 bp    0.06 %
     RTE/Bov-B         271       181676 bp    0.02 %
     L1/CIN4          7978      1801119 bp    0.16 %
   LTR elements:     93644     39582392 bp    3.51 %
     BEL/Pao          1287      1138420 bp    0.10 %
     Ty1/Copia         221       110739 bp    0.01 %
     Gypsy/DIRS1     32714     19073624 bp    1.69 %
       Retroviral     2443      1815904 bp    0.16 %

DNA transposons      79815     22485920 bp    1.99 %
   hobo-Activator     2242       639067 bp    0.06 %
   Tc1-IS630-Pogo    49209     16661924 bp    1.48 %
   En-Spm                0            0 bp    0.00 %
   MuDR-IS905            0            0 bp    0.00 %
   PiggyBac              0            0 bp    0.00 %
   Tourist/Harbinger  1794       259934 bp    0.02 %
   Other (Mirage,      900       209002 bp    0.02 %
    P-element, Transib)

Rolling-circles        196        71055 bp    0.01 %

Unclassified:       1774260    361869982 bp   32.08 %

Total interspersed repeats:   437780879 bp   38.81 %


Small RNA:                0            0 bp    0.00 %

Satellites:               1          267 bp    0.00 %
Simple repeats:      504151     23839806 bp    2.11 %
Low complexity:       58354      3597934 bp    0.32 %
==================================================
```

```
* most repeats fragmented by insertions or deletions
  have been counted as one element


RepeatMasker version 4.1.1 , default mode

run with rmblastn version 2.10.0+
The query was compared to classified sequences in "Rohu-families.fa"
```

Comparative genomics

So, I might recommend doing some sort of comparative genomics here or throughout. This becomes relevant when suggesting this genome is an improvement. BUSCO comparison, number of genes predicted, orthogroup analysis, synteny, etc.

SNP discovery and population similarities among rohu fisheries

Labeo rohita is an important aquaculture fish in many areas of south Asia [I don't know how to put in this reference FAO (2019) Food and Agricultural Organization. Yearbook of Fishery and Aquaculture Statistics 2017].

The high fecundity and relative ease of cultivation has made rohu successful in breeding programs, leading to the generation of a fast-growing strain called Jayanti Historically, rohu breeding and research has relied on low-throughput information, such as cloning individual genes of interest (reviewed in [32]) or microsatellite-based selective breeding [33]; however, recent years have seen increased interest in applying high-throughput methods to rohu breeding, such as SNP panels [34] [35], genotyping-by-sequencing [36], transcriptomics [34] [37], and degradomics [38]. Recently, a draft genome was published for L. rohita along with pooled resequencing from three individuals in 10 populations [39]; however, genetic diversity and divergence among the populations was not reported.

Acknowledgements

ISU Office of Biotechnology Flow Cytometry Facility

ResearchIT Iowa State University

References

1. **The application of flow cytometry for estimating genome size and ploidy level in plants.**
Jaume Pellicer, Ilia J Leitch
Methods in molecular biology (Clifton, N.J.) (2014)
https://www.ncbi.nlm.nih.gov/pubmed/24415480
DOI: 10.1007/978-1-62703-767-9_14 · PMID: 24415480

2. **A fast, lock-free approach for efficient parallel counting of occurrences of k-mers**
Guillaume Marçais, Carl Kingsford
Bioinformatics (2011-03-15) https://doi.org/b7gkd6
DOI: 10.1093/bioinformatics/btr011 · PMID: 21217122 · PMCID: PMC3051319

3. **GenomeScope: fast reference-free genome profiling from short reads**
Gregory W Vurture, Fritz J Sedlazeck, Maria Nattestad, Charles J Underwood, Han Fang, James Gurtowski, Michael C Schatz
Bioinformatics (Oxford, England) (2017-07-15)
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5870704/
DOI: 10.1093/bioinformatics/btx153 · PMID: 28369201 · PMCID: PMC5870704

4. **NanoPack: visualizing and processing long-read sequencing data**
Wouter De Coster, Svenn D'Hert, Darrin T Schultz, Marc Cruts, Christine Van Broeckhoven
Bioinformatics (2018-08-01) https://doi.org/gf38cx
DOI: 10.1093/bioinformatics/bty149 · PMID: 29547981 · PMCID: PMC6061794

5. **Trimmomatic: a flexible trimmer for Illumina sequence data**
Anthony M Bolger, Marc Lohse, Bjoern Usadel
Bioinformatics (2014-08-01) https://doi.org/f6cj5w
DOI: 10.1093/bioinformatics/btu170 · PMID: 24695404 · PMCID: PMC4103590

6. **Fast and accurate long-read assembly with wtdbg2**
Jue Ruan, Heng Li
Nature Methods (2020-02) https://doi.org/ggd8j9
DOI: 10.1038/s41592-019-0669-3 · PMID: 31819265 · PMCID: PMC7004874

7. **Fast and accurate de novo genome assembly from long uncorrected reads**
Robert Vaser, Ivan Sović, Niranjan Nagarajan, Mile Šikić
Genome Research (2017-05) https://doi.org/gfkmzr
DOI: 10.1101/gr.214270.116 · PMID: 28100585 · PMCID: PMC5411768

8. **Minimap2: pairwise alignment for nucleotide sequences**
Heng Li
Bioinformatics (2018-09-15) https://doi.org/gdhbqt
DOI: 10.1093/bioinformatics/bty191 · PMID: 29750242 · PMCID: PMC6137996

9. **Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement**
Bruce J Walker, Thomas Abeel, Terrance Shea, Margaret Priest, Amr Abouelliel, Sharadha Sakthikumar, Christina A Cuomo, Qiandong Zeng, Jennifer Wortman, Sarah K Young, Ashlee M Earl
PLoS ONE (2014-11-19) https://doi.org/gfkmz5
DOI: 10.1371/journal.pone.0112963 · PMID: 25409509 · PMCID: PMC4237348

10. **Fast and accurate short read alignment with Burrows-Wheeler transform**

H Li, R Durbin
Bioinformatics (2009-07-15) https://doi.org/dqt59j
DOI: 10.1093/bioinformatics/btp324 · PMID: 19451168 · PMCID: PMC2705234

11. **Bionano Access Software**
Bionano Genomics
https://bionanogenomics.com/support-page/bionano-access-software/

12. **Integrating Hi-C links with assembly graphs for chromosome-scale assembly**
Jay Ghurye, Arang Rhie, Brian P Walenz, Anthony Schmitt, Siddarth Selvaraj, Mihai Pop, Adam M Phillippy, Sergey Koren
PLOS Computational Biology (2019-08-21) https://doi.org/gf74qw
DOI: 10.1371/journal.pcbi.1007273 · PMID: 31433799 · PMCID: PMC6719893

13. **Chromosome-level genome assembly of a cyprinid fish Onychostoma macrolepis by integration of nanopore sequencing, Bionano and Hi-C technology**
Lina Sun, Tian Gao, Feilong Wang, Zuliang Qin, Longxia Yan, Wenjing Tao, Minghui Li, Canbiao Jin, Li Ma, Thomas D Kocher, Deshou Wang
Molecular Ecology Resources (2020-09) https://doi.org/gmx33w
DOI: 10.1111/1755-0998.13190 · PMID: 32419357

14. **RaGOO: fast and accurate reference-guided scaffolding of draft genomes**
Michael Alonge, Sebastian Soyk, Srividya Ramakrishnan, Xingang Wang, Sara Goodwin, Fritz J Sedlazeck, Zachary B Lippman, Michael C Schatz
Genome Biology (2019-12) https://doi.org/ggctnf
DOI: 10.1186/s13059-019-1829-6 · PMID: 31661016 · PMCID: PMC6816165

15. **RepeatModeler2 for automated genomic discovery of transposable element families**
Jullien M Flynn, Robert Hubley, Clément Goubert, Jeb Rosen, Andrew G Clark, Cédric Feschotte, Arian F Smit
Proceedings of the National Academy of Sciences (2020-04-28) https://doi.org/ggsnv2
DOI: 10.1073/pnas.1921046117 · PMID: 32300014 · PMCID: PMC7196820

16. **Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype**
Daehwan Kim, Joseph M Paggi, Chanhee Park, Christopher Bennett, Steven L Salzberg
Nature Biotechnology (2019-08) https://doi.org/gf5395
DOI: 10.1038/s41587-019-0201-4 · PMID: 31375807 · PMCID: PMC7605509

17. **Leveraging multiple transcriptome assembly methods for improved gene structure annotation**
Luca Venturini, Shabhonam Caim, Gemy George Kaithakottil, Daniel Lee Mapleson, David Swarbreck
GigaScience (2018-08-01) https://doi.org/gmx7k4
DOI: 10.1093/gigascience/giy093 · PMID: 30052957 · PMCID: PMC6105091

18. **BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database**
Tomáš Brůna, Katharina J Hoff, Alexandre Lomsadze, Mario Stanke, Mark Borodovsky
NAR Genomics and Bioinformatics (2021-01-06) https://doi.org/gmx7k6
DOI: 10.1093/nargab/lqaa108 · PMID: 33575650 · PMCID: PMC7787252

19. **Transcriptome assembly from long-read RNA-seq alignments with StringTie2**
Sam Kovaka, Aleksey V Zimin, Geo M Pertea, Roham Razaghi, Steven L Salzberg, Mihaela Pertea
Genome Biology (2019-12) https://doi.org/ghrk2k
DOI: 10.1186/s13059-019-1910-1 · PMID: 31842956 · PMCID: PMC6912988

20. **Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks**
Cole Trapnell, Adam Roberts, Loyal Goff, Geo Pertea, Daehwan Kim, David R Kelley, Harold Pimentel, Steven L Salzberg, John L Rinn, Lior Pachter
Nature Protocols (2012-03) https://doi.org/f4pbzd
DOI: 10.1038/nprot.2012.016 · PMID: 22383036 · PMCID: PMC3334321

21. **Full-length transcriptome assembly from RNA-Seq data without a reference genome**
Manfred G Grabherr, Brian J Haas, Moran Yassour, Joshua Z Levin, Dawn A Thompson, Ido Amit, Xian Adiconis, Lin Fan, Raktima Raychowdhury, Qiandong Zeng, … Aviv Regev
Nature Biotechnology (2011-07) https://doi.org/b2bctj
DOI: 10.1038/nbt.1883 · PMID: 21572440 · PMCID: PMC3571712

22. **Efficient and accurate detection of splice junctions from RNA-seq with Portcullis**
Daniel Mapleson, Luca Venturini, Gemy Kaithakottil, David Swarbreck
GigaScience (2018-12-01) https://doi.org/gkzg5r
DOI: 10.1093/gigascience/giy131 · PMID: 30418570 · PMCID: PMC6302956

23. **Prodigal: prokaryotic gene recognition and translation initiation site identification**
Doug Hyatt, Gwo-Liang Chen, Philip F LoCascio, Miriam L Land, Frank W Larimer, Loren J Hauser
BMC Bioinformatics (2010-12) https://doi.org/cktxnm
DOI: 10.1186/1471-2105-11-119 · PMID: 20211023 · PMCID: PMC2848648

24. **Eukaryotic Gene Prediction Using GeneMark.hmm-E and GeneMark-ES**
Mark Borodovsky, Alex Lomsadze
Current Protocols in Bioinformatics (2011-09) https://doi.org/b75r2g
DOI: 10.1002/0471250953.bi0406s35 · PMID: 21901742 · PMCID: PMC3204378

25. **Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources**
Mario Stanke, Oliver Schöffmann, Burkhard Morgenstern, Stephan Waack
BMC Bioinformatics (2006-12) https://doi.org/cv8xsn
DOI: 10.1186/1471-2105-7-62 · PMID: 16469098 · PMCID: PMC1409804

26. **MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects**
Carson Holt, Mark Yandell
BMC Bioinformatics (2011-12) https://doi.org/fz39nj
DOI: 10.1186/1471-2105-12-491 · PMID: 22192575 · PMCID: PMC3280279

27. **InterProScan 5: genome-scale protein function classification**
P Jones, D Binns, H-Y Chang, M Fraser, W Li, C McAnulla, H McWilliam, J Maslen, A Mitchell, G Nuka, … S Hunter
Bioinformatics (2014-05-01) https://doi.org/f53532
DOI: 10.1093/bioinformatics/btu031 · PMID: 24451626 · PMCID: PMC3998142

28. **BLAST+: architecture and applications**
Christiam Camacho, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, Thomas L Madden
BMC Bioinformatics (2009-12) https://doi.org/cnjxgz
DOI: 10.1186/1471-2105-10-421 · PMID: 20003500 · PMCID: PMC2803857

29. **BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes**
Mosè Manni, Matthew R Berkeley, Mathieu Seppey, Felipe A Simão, Evgeny M Zdobnov

Molecular Biology and Evolution (2021-09-27) https://doi.org/gmgv52
DOI: 10.1093/molbev/msab199 · PMID: 34320186 · PMCID: PMC8476166

30. http://zotero.org/users/local/gVY292am/items/GYK6YE2I

31. **De novo Assembly and Genome-Wide SNP Discovery in Rohu Carp, Labeo rohita**
Paramananda Das, Lakshman Sahoo, Sofia P Das, Amrita Bit, Chaitanya G Joshi, Basdeo Kushwaha, Dinesh Kumar, Tejas M Shah, Ankit T Hinsu, Namrata Patel, … Joykrushna Jena
Frontiers in genetics (2020-04-21) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7186481/
DOI: 10.3389/fgene.2020.00386 · PMID: 32373166 · PMCID: PMC7186481

32. **Status of genetic and genomic approaches for delineating biological information and improving aquaculture production of farmed rohu, Labeo rohita (Ham, 1822)**
Kiran D Rasal, Jitendra Kumar Sundaray
Reviews in Aquaculture (2020-11) https://doi.org/gnj5hk
DOI: 10.1111/raq.12444

33. **A molecular tool for parentage analysis in Indian major carp, Labeo rohita (Hamilton, 1822)**
L Sahoo, PK Meher, K Das Mahapatra, JN Saha, P Jayasankar, P Das
Aquaculture International (2017-06) https://doi.org/gbgqt5
DOI: 10.1007/s10499-016-0104-z

34. **A linkage map of transcribed single nucleotide polymorphisms in rohu (Labeo rohita) and QTL associated with resistance to Aeromonas hydrophila**
Nicholas Robinson, Matthew Baranski, Kanta Mahapatra, Jatindra Saha, Sweta Das, Jashobanta Mishra, Paramananda Das, Matthew Kent, Mariann Arnyasi, Pramoda Sahoo
BMC Genomics (2014) https://doi.org/gb3g4v
DOI: 10.1186/1471-2164-15-541 · PMID: 24984705 · PMCID: PMC4226992

35. **Status and future perspectives of single nucleotide polymorphisms (SNPs) markers in farmed fishes: Way ahead using next generation sequencing**
Kiran Dashrath Rasal, Vemlawada Chakrapani, Amrendra Kumar Pandey, Avinash Rambhau Rasal, Jitendra K Sundaray, Arun Ninawe, Pallipuram Jayasankar
Gene Reports (2017-03) https://doi.org/gjg4d7
DOI: 10.1016/j.genrep.2016.12.004

36. **Single Nucleotide Polymorphisms (SNPs) Reveal Sibship Among Founders of a Bangladeshi Rohu (Labeo rohita) Breeding Population**
Matthew Gray Hamilton, Wagdy Mekkawy, Andrzej Kilian, John AH Benzie
Frontiers in Genetics (2019-06-19) https://doi.org/gnj5hm
DOI: 10.3389/fgene.2019.00597 · PMID: 31275363 · PMCID: PMC6593075

37. **Molecular characterization, computational analysis and expression profiling of Dmrt1 gene in Indian major carp, Labeo rohita (Hamilton 1822)**
L Sahoo, S Sahoo, M Mohanty, M Sankar, S Dixit, P Das, KD Rasal, MA Rather, JK Sundaray
Animal Biotechnology (2021-08-02) https://doi.org/gnj5hj
DOI: 10.1080/10495398.2019.1707683 · PMID: 31880491

38. **Revealing liver specific microRNAs linked with carbohydrate metabolism of farmed carp, Labeo rohita (Hamilton, 1822)**
Kiran D Rasal, Mir Asif Iquebal, Amrendra Pandey, Parmeswari Behera, Sarika Jaiswal, Manohar Vasam, Sangita Dixit, Mustafa Raza, Lakshman Sahoo, Samiran Nandi, … Jitendra Kumar Sundaray

Genomics (2020-01) https://doi.org/gnj5hh
DOI: 10.1016/j.ygeno.2019.07.010 · PMID: 31325488

39. **De novo Assembly and Genome-Wide SNP Discovery in Rohu Carp, Labeo rohita**
Paramananda Das, Lakshman Sahoo, Sofia P Das, Amrita Bit, Chaitanya G Joshi, Basdeo Kushwaha, Dinesh Kumar, Tejas M Shah, Ankit T Hinsu, Namrata Patel, … Joykrushna Jena
Frontiers in Genetics (2020-04-21) https://doi.org/gnj5hn
DOI: 10.3389/fgene.2020.00386 · PMID: 32373166 · PMCID: PMC7186481