Whole genome sequencing of Chinese clearhead icefish,

2 Protosalanx hyalocranius

4	Kai Liu ¹ [†] , Dongpo Xu ¹ [†] , Jia Li ² [†] , Chao Bian ² [†] , Jinrong Duan ¹ [†] , Yanfeng Zhou ¹ [†] ,
5	Minying Zhang ¹ , Xinxin You ² , Yang You ¹ , Jieming Chen ² , Hui Yu ² , Gangchun Xu ¹ ,
6	Di-an Fang ¹ , Jun Qiang ¹ , Shulun Jiang ¹ , Jie He ¹ , Junmin Xu ^{2,4,5} , Qiong Shi ^{2,4,5,6*} ,
7	Zhiyong Zhang ^{3*} , Pao Xu ^{1,5*}
8	
9	¹ Freshwater Fisheries Research Center, Chinese Academy of Fishery Sciences, Wuxi
10	214081, China
11	² Shenzhen Key Lab of Marine Genomics, Guangdong Provincial Key Lab of
12	Molecular Breeding in Marine Economic Animals, BGI, Shenzhen 518083, China
13	³ Institute of Oceanology & Marine Fisheries, Jiangsu 226007, China
14	⁴ BGI Zhenjiang Institute of Hydrobiology, Zhenjiang 212000, China
15	⁵ BGI Research Center for Aquatic Genomics, Chinese Academy of Fishery Sciences,
16	Shenzhen 518083, China
17	⁶ Laboratory of Aquatic Genomics, College of Ecology and Evolution, School of Life
18	Sciences, Sun Yat-Sen University, Guangzhou 510275, China
19	
20	† Equal contributors
21	*Correspondence: xup@ffrc.cn (PX); shiqiong@genomics.cn (QS);
22	13906292412@139.com (ZZ)
23	
24	Email addresses: liuk@ffrc.cn (KL); xudp@ffrc.cn(DX); lijia1@genomics.cn (JL);
25	bianchao@genomics.cn (CB); duanjr@ffrc.cn(JD); zhouyf@ffrc.cn(YZ);

zhangmy@ffrc.cn(MZ); youxinxin@genomics.cn (XY); youy@ffrc.cn (YY); chenjieming@genomics.cn (JC); yuhui@genomics.cn (HY); xugc@ffrc.cn(GX); fangda@ffrc.cn(DF); qiangj@ffrc.cn(JQ); 420219380@qq.com(SJ); hej@ffrc.cn(JH); б xujunmin@genomics.cn (JX); shiqiong@genomics.cn (QS); 13906292412@139.com (ZZ); xup@ffrc.cn (PX) Abstract Background: Chinese clearhead icefish, Protosalanx hyalocranius, is a representative icefish species with economic importance and special appearance. Due to its great economic values in China, the fish was introduced into Lake Dianchi and several other lakes from the Lake Taihu half a century ago. Similar to the Sinocyclocheilus cavefish, the clearhead icefish has certain cavefish-like traits, such as transparent body and nearly scaleless skin. Here, we provide the whole genome sequence of this surface-dwelling fish and generated a draft genome assembly, aiming at exploring molecular mechanisms for the biological interests. Findings: A total of 252.1 gigabases (Gb) of raw reads were sequenced. Subsequently, a novel draft genome assembly was generated, with the scaffold N50 reaching 1.163 Mb. The genome completeness was estimated to be 98.39% by using the CEGMA evaluation. Finally, we annotated 19,884 protein-coding genes and observed that repeat sequences account for 24.43% of the genome assembly. **Conclusion**: We report the first draft genome of the Chinese clearhead icefish. The genome assembly will provide a solid foundation for further molecular breeding and germplasm resource protection in Chinese clearhead icefish, as well as other icefishes. It is also a valuable genetic resource for revealing the molecular mechanisms for the cavefish-like characters. **Keywords:** Clearhead icefish; *Protosalanx hyalocranius*; Whole genome sequencing; Genome assembly; Gene prediction; Repetitive sequences

55 Data description

56 Background

Icefishes (Osmeriformes, Salangidae) are widely distributed in freshwater, coastal and estuarine habitats in East Asian countries [1-3]. Chinese clearhead icefish (Protosalanx hyalocranius, Figure 1), a diadromous fish, mainly inhabits in coastal areas and adjacent freshwaters [4-6]. As an economically important fish in China, the clearhead icefish was widely introduced into some lakes from the original Lake Taihu half a century ago, and it has developed a resident life history in these water areas [2, 7, 8]. Because of its transparent body and nearly scaleless skin, similar to the Sinocyclocheilus cavefishes [9], we are very interested in this surface-dwelling fish and are performing comparative genomics studies to explore the mechanisms for these biological phenotypes. However, with the rapid development of the Chinese economy in recent decades, population size of the clearhead icefish has been seriously declining because of overfishing, construction of water conservancy facilities and water pollution in the ecological systems [10]. To maintain its sustainable development in China, here we performed whole genome sequencing of Chinese clearhead icefish to support its biological and economic importance.

73 Sample and Sequencing

In this study, we applied Illumina whole genome sequencing (WGS) strategy to
sequence the genome of Chinese clearhead icefish (NCBI Taxonomy ID: 418454;
Fishbase ID: 12236). Genomic DNA was isolated from the muscle tissue of an
individual collected from the Lake Taihu of Jiangsu Province in China. We
constructed seven paired-end libraries with three short-insert libraries (250, 500 and
800 bp) and four long-insert libraries (2, 5, 10 and 20 kb) using the standard protocol
provided by Illumina (San Diego, USA). Subsequent paired-end sequencing was

performed by the Illumina HiSeq 2000 platform for each library. Finally, we obtained
252.1 Gb of raw reads for further analysis.

84 Genome size estimation and genome assembly

The SOAPfilter v2.2 software [11] with optimized parameters (-y -p -g 1 -o clean -M 2 -f 0) was utilized to remove low-quality raw reads (including reads with 10 or more Ns and low-quality bases) and PCR-replicates as well as adaptor sequences. In total, we obtained 169.0 Gb of clean reads. Subsequently, we estimated the genome size based on the 17-mer depth frequency distribution method [12]. We applied the following formula to calculate the genome size: G=k_num/k_depth=b_num/b_depth (k_num is the total number of K-mers from the sequencing data, k_depth is the expected coverage depth for k-mers, b_num is the total number of bases, b_depth is the expected coverage depth of bases; As one read with length L generates L-K+1 k-mers, k num/b num=(L-K+1)/L). In our current study, the K num was 10,500,000,000 and the K_depth was 20. Hence, we estimated that the genome size of Chinese clearhead icefish is 525 Mb. The filtered reads were assembled using SOAPdenovo2 v2.04.4 software [13] with optimized parameters (pregraph -K 79 -d 1; contig -M 1; scaff -F -b 1.5 -p 16) to generate contigs and original scaffolds. The gaps were filled using GapCloser v1.12 software [14] with default parameters and -p set to 25. Finally, we generated a draft genome assembly of 536 Mb, with the scaffold N50 reaching 1.163 Mb (Table 1). The completeness of our assembly was evaluated by using both CEGMA [15] and BUSCO [16]. The CEGMA program (Core Eukaryotic Genes Mapping Approach; version 2.4) assessment with 248 conserved Core Eukaryotic Genes (CEGs) was performed for evaluation of the gene space completeness. Our results revealed that the assembled genome had a CEGMA completeness score at 90.32% and 98.39%, which

- 107 was calculated from the complete gene set and the partial gene set, respectively.
- 108 Meanwhile, we used the representative metazoa gene set [17], which contains 843

single-copy genes that are widely present in metazoan, as a reference. The assessment
demonstrated that the BUSCO values is 89%, containing [D: 10%], F: 7.7%, M: 2.9%,
n: 843 (C: complete [D: duplicated], F: fragmented, M: missed, n: genes). These data
from CEGMA and BUSCO indicate that the assembled genome covered majority of
the gene space.

Repeat annotation

Firstly, a *de novo* repeat library was constructed by the RepeatModeller v1.05 [18] and LTR FINDER.x86 64-1.0.6 [10] with default parameters. Then, the assembled genome sequences were aligned against the RepBase v21.01 [19] and the de novo repeat libraries to recognize the known and novel transposable elements (TEs) using the RepeatMasker v4.06 [20]. Meantime, the Tandem Repeat Finder v4.07 [21] with parameters "Match=2, Mismatch=7, Delta=7, PM=80, PI=10, Minscore=50, and MaxPeriod=2000" was utilized for annotation of tandem repeats. Furthermore, the RepeatProteinMask software v4.0.6 [20] was used to predict TE relevant proteins in our genome assembly. Finally, we observed that the repeat sequences account for 24.43% of the assembled genome (Table 1), and the de novo annotation method predicted the most abundant repeat sequence among the four methods (Table 2).

128 Genome Annotation

129 In brief, we utilized two different methods to predict total gene set of the clearhead130 icefish.

131 1) *de novo* annotation. The AUGUSTUS v2.5 [22] and GENSCAN v1.0 [23] were
132 executed to *ab initio* predict genes within the assembled genome, with the repetitive
133 sequences masked as "N" in order to discard pseudo gene prediction. Those
134 low-quality genes with short length (<150 bp), premature termination or
135 frame-shifting were removed. Finally, we identified 23,132 and 21,379 pro-coding
136 genes by using the AUGUSTUS and GENSCAN software (Table 3).

2) Homology annotation. We aligned the protein sequences from six published genomes, including Danio rerio [24], Oryzias latipes [25], Takifugu rubripes [26], Tetraodon nigroviridis [27], Esox lucius [28] and Gasterosteus aculeatus [29], against our assembly to predict homology-based genes. The potential homology-based genes were searched by TblastN [30] with an e-value of 10⁻⁵. The TblastN results were then processed by SOLAR (Sorting Out Local Alignment Result [31]) to obtain the best hit of each alignment. Subsequently, GeneWise v2.2.0 [32] was performed to detect the possible gene structure for the best hit of each alignment. The low-quality genes were also removed as described in the above-mentioned de novo annotation.

3) Integration of annotation results. We employed the GLEAN [33] to generate a
non-redundant and comprehensive gene set. Finally, the best hit of each protein was
obtained through all protein sequences from the GLEAN results aligned to the
databases of the SwissProt and TrEMBL [34] (Uniprot release 2011.06) by BlastP
with an e-value of 10⁻⁵. Overall, we generated a final gene set with 19,884 genes for
the Chinese clearhead icefish (Table 3).

CEGMA was performed again to evaluate the coverage rate between KOG (EuKaryotic Orthologous Groups) genes predicted by CEGMA and the predicted total gene set. It demonstrates that the predicted gene set mapped 96.4% of the KOGs. Simultaneously, the BUSCO was implemented again to assess completeness of the predicted gene set. The BUSCO values were calculated as follows: C: 79% [D: 16%], F: 9.8%, M: 10%, n: 843 (C: complete [D: duplicated], F: fragmented, M: missed, n: genes). The assessment values from both CEGMA and BUSCO proved high accuracy of the annotation.

4) Function annotation. The predicted protein sequences of the clearhead icefish
were aligned against several public databases (Pfam [35], PRINTS [36], ProDom [37]
and SMART [38]) for detection of functional motifs and domains. Finally, we found
that 96.2% of the predicted total gene set had been annotated with at least one

functional assignment from other public databases (Swiss-Prot [39], Interpro [40],
TrEMBL [41] and KEGG [42]).

Genome evolution

We performed phylogenomic analyses with orthologues from representative species for each clade. We used the Ensembl BioMart (www.ensembl.org/biomart; Ensembl version 76) to extract orthologues for zebrafish [24], fugu [26], stickleback [29], medaka [25] and spotted gar [43]. This generated orthologue dataset from six species was filtered out to retain only one-to-one orthologues. Meanwhile, a new Asian arowana gene set stem from our recent work [44]. In order to extrapolate the Biomart orthologues to the arowana and clearhead icefish gene sets, we used zebrafish as the reference. We ran InParanoid [45] for the three species pairs (zebrafish-arowana and zebrafish-clearhead icefish) at default settings (i.e., a minimum BLASTP score of 40 bits, minimum 50% alignment span, minimum 25% alignment coverage, and minimum inparalog confidence level of 0.05). By comparing the three InParanoid outputs, we narrowed down the list of one-to-one orthologues, presented in all the seven species, to 454 genes. Multiple alignments were subsequently performed on proteins of each selected family using MUSCLE (version 3.8.31) [46] and protein alignments were converted to their corresponding CDS alignments using an in-house perl script (see supporting data). All the translated CDS sequences were linked into one "supergene" for each species. Non-degenerated sites extracted from the supergenes were subsequently joined into new sequence of each species to construct a phylogenetic tree (Figure 2) using MrBayes [47] (GTR+gamma model, Version 3.2). Our phylogenetic data demonstrate the phylogenetic position of the clearhead icefish (Figure 2).

190 Synteny blocks and genome duplication

Genomic homology between the clearhead icefish and Nile tilapia [48] was examined using i-ADHoRe 3.0 [49] using the following settings: alignment method gg2, gap size 30, tandem gap 30, cluster gap 35, q value of 0.85, prob cutoff 0.01, anchor points 5 and using multiple hypothesis correction FDR. The output of this was processed by the pipeline and incorporated in a relational database to which visualization programs can connect and on which additional statistical analysis can then be performed. For synteny detection, the cloud mode was enabled (cluster_type = cloud) and appropriate settings were selected as follows: cloud_gap_size 20, cloud cluster gap 20, cloud filter method binomial, prob cutoff 0.01, anchor points 5, multiple hypothesis correction FDR and level_2_only true. Finally, we identified 771 synteny blocks containing 7,057 genes between the clearhead icefish and Nile tilapia. Subsequently, Protein sequences of homologous gene pairs in the identified syntenic

regions were aligned using MUSCLE [46], and the protein alignments were then
converted to the CDS alignments. Finally, four-fold degenerative third-codon
transversion (4DTV) values were calculated on these CDS alignments and corrected
using the HKY model in the PAML package [50]. These data indicate that the
clearhead icefish also experienced the teleost-specific whole genome duplication
(WGD) (Figure 3).

211 Conclusion

We generated a draft genome assembly of the Chinese clearhead icefish. The novel
genome data were deposited in publicly accessible repositories to promote further
biological research, molecular breeding and resource protection of this representative
and valuable icefish.

217 Availability of supporting data

Supporting data and materials are available in the *GigaScience* GigaDB database [51],
with the raw genome sequences deposited in the SRA under the bioproject number
PRJNA328051.

222 Competing interests

223 The authors declare that they have no competing interests.

225 Funding

226 This study was supported by a grant from Natural Science Foundation of Jiangsu

227 Province (No.BK2012093), fish investigation in Taihu Lake (No.TH2016WT007),

228 National Infrastructure of Fishery Germplasm Resources (No.2016DKA30470), Basic

229 Research Funds from Freshwater Fisheries Research Center (No. 2013JBFM07),

230 Special Project on the Integration of Industry, Education and Research of Guangdong

231 Province (No. 2013B090800017), Shenzhen Special Program for Future Industrial

232 Development (N 192 o. JSGG20141020113728803), and Zhenjiang Leading Talent

233 Program for Innovation and Entrepreneurship.

234 Author's Contributions

235 KL, PX, QS, DX, JX, CB and ZZ conceived the project. MZ, XY, HY, JC, GX, DF,

236 JQ, SJ and JH collected the samples and extracted the genomic DNA. JL, CB and HY

performed the genome assembly and data analysis. JL, CB, QS, KL, XP, KL, YY andZZ wrote the paper.

References

Wang ZS, Cui Zhang FU: Biodiversity of Chinese Icefishes (Salangidae)
 and their conserving strategies. *Chinese Biodiversity* 2002, 10(4):416-424.
 Zhang J, Li M, Xu M, Takita T, Wei F: Molecular phylogeny of icefish
 Salangidae based on complete mtDNA cytochrome b sequences, with

	245		comments on estuarine fish evolution. Biological Journal of the Linnean
1	246		<i>Society</i> 2007, 91 (2):325-340.
3	247	3.	Wang Z, Lu C, Hu H, Xu C, Lei G: Dynamics of Icefish (Salangidae) Stocks
4	248		in Nanyi Lake, Eastern China: Degradation and Overfishing. Journal of
5 6	249		Freshwater Ecology 2004, 19 (2):271-278.
7	250	4.	Xia DQ, Cao Y, Ting ting WU, Yang H: Study on lineages of Protosalanx
8 9	251		chinensis, Neosalanx taihuensis and N.oligodontis in Taihu Lake with
10	252		RAPD technique . Journal of Fisheryences of China 2000, 7 (01):12-15.
11	253	5.	Xia DQ, Cao Y, Ting Ting WU, Yang H: Genetic Structures of Population
13	254		of Protosalanx Chinensis, Neosalanx Taihuensis and Neosalanx
14	255		Oligodontis in Lake Taihu . Journal of Fisheries of China 1999(03):254-260.
15 16	256	б.	Armani A, Castigliego L, Tinacci L, Gianfaldoni D, Guidi A: Molecular
17	257		characterization of icefish, (S alangidae family), using direct sequencing
18 19	258		of mitochondrial cytochrome b gene. Food Control 2011, 22(6):888-895.
20	259	7.	Wang Z, Lu C, Hu H, Zhou Y, Xu C, Lei G: Freshwater icefishes
21 22	260		(Salangidae) in the Yangtze River basin of China: Spatial distribution
23	261		patterns and environmental determinants. Environmental Biology of Fishes
24	262		2005, 73 (3):253-262.
25 26	263	8.	Ye S, Yang J, Liu H, Oshima Y: Use of elemental fingerprint analysis to
27	264		identify localities of collection for the large icefish protosalanx chinensis
28 29	265		in Taihu Lake, China. Journal of the Faculty of Agriculture, Kyushu
30	266		<i>University</i> 2011, 56 (1):41-45.
31 32	267	9.	Yang J, Chen X, Jie B, Fang D, Ying O, Jiang W, Hui Y, Chao B, Jiang L, He
33	268		S: The Sinocyclocheilus cavefish genome provides insights into cave
34 25	269		adaptation. Bmc Biology 2016, 14 (1):1-13.
36	270	10.	Xu J, Xie P, Zhang M, Zhou O, Zhang L, Wen Z, Cao T: Icefish (salangidae)
37	271		as an indicator of anthropogenic pollution in freshwater systems using
38 39	272		nitrogen isotope analysis . Bulletin of environmental contamination and
40	273		toxicology 2007, 79 (3):323-326.
41 42	274	11.	Kar HK, Narayan R, Gautam RK, Jain RK, Doda V, Sengupta D, Bhargava
43	275		NC: Mucocutaneous disorders in Hiv positive patients. Indian journal of
44 45	276		dermatology, venereology and leprology 1996, 62 (5):283-285.
46	277	12.	Liu B, Shi Y, Yuan J, Hu X, Zhang H, Li N, Li Z, Chen Y, Mu D, Fan W:
47	278		Estimation of genomic characteristics by analyzing k-mer frequency in de
40 49	279		novo genome projects . <i>Quantitative Biology</i> 2013, 35 (s 1–3):62-67.
50	280	13.	Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y et
51 52	281		al: SOAPdenovo2: an empirically improved memory-efficient short-read
53	282		de novo assembler. <i>Gigascience</i> 2012, 1 :18.
54 55	283	14.	Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J: SOAP2: an
56	284		improved ultrafast tool for short read alignment. Bioinformatics 2009,
57 58	285		25 (15):1966-1967.
59			
60			
ь⊥ 62			
63			

287 core genes in eukaryotic genomes. Bioinformatics 2007, 23(9):1061-1067. 288 16. Sim AFO, Waterhouse MR, Ioannidis P, Kriventseva VE, Zdobnov ME: 289 BUSCO: assessing genome assembly and annotation completeness with 590 single-copy orthologs. Bioinformatics 2015, 31(19):3210-3212. 791 17. Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simão FA, 792 Pozdnyakov IA, Ioannidis P, Zdobnov EM: OrtholD8 v8: update of the 711 Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simão FA, 792 Pozdnyakov IA, Ioannidis P, Zdobnov EM: OrthoDB v8: update of the 711 Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simão FA, 712 Pozdnyakov IA, Ioannidis P, Zdobnov EM: OrthoDB v8: update of the 713 Hierarchical catalog of orthologs and the underlying free software. Nucleic 714 Aziade M, Bouchard S, Gingras N, Charron L, Cardinal A, Roy MA, 715 Gauthier B, Tremblay G, Cote S, Fournier C et al: Long-term stability of 716 diagnosis and symptom dimensions in a systematic sample of patients 718 With onset of schizophrenia in childhood predictors of adult outcome. The 719 Postnegative distinction and childhood predictors of adult outcome. The 720 Iof(3):371-378.
288 16. Sim AFO, Waterhouse MR, Ioannidis P, Kriventseva VE, Zdobnov ME: 289 BUSCO: assessing genome assembly and annotation completeness with 290 single-copy orthologs. Bioinformatics 2015, 31(19):3210-3212. 291 17. Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simão FA, 292 Pozdnyakov IA, Ioannidis P, Zdobnov EM: OrthoDB V8: update of the 293 hierarchical catalog of orthologs and the underlying free software. Nucleic 294 Acids Research 2015, 43(Database issue):D250-D256. 18 Maziade M, Bouchard S, Gingras N, Charron L, Cardinal A, Roy MA, 295 18. Maziade M, Bouchard S, Gingras N, Charron L, Cardinal A, Roy MA, 296 Gauthier B, Tremblay G, Cote S, Fournier C et al: Long-term stability of 297 diagnosis and symptom dimensions in a systematic sample of patients 298 with onset of schizophrenia in childhood and early adolescence. II: 299 Boritish journal of psychiatry : the journal of mental science 1996, 201 169(3):371-378. 2302 19. Jurka, J., Kapitonov, V. V., Pavlicek, A., et al: Repbase Update, a database 203 of eukaryotic repetitive elements. Cytogenetic & Genome Research 2005, 204 10(1-4):462-467. <t< td=""></t<>
4 289 BUSCO: assessing genome assembly and annotation completeness with 6 290 single-copy orthologs. Bioinformatics 2015, 31(19):3210-3212. 7 291 17. Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simão FA, 9 292 Pozdnyakov IA, Ioannidis P, Zdobnov EM: OrtholDs Ns: update of the 11 10: hierarchical catalog of orthologs and the underlying free software. Nucleic 12 294 Acids Research 2015, 43(Database issue):D250-D256. 13 295 18. Maziade M, Bouchard S, Gingras N, Charron L, Cardinal A, Roy MA, 14 296 Gauthier B, Tremblay G, Cote S, Fournier C et al: Long-term stability of 15 297 Postnegative distinction and childhood predictors of adult outcome. The 17 298 with onset of schizophrenia in childhood predictors of adult outcome. The 18 299 Postnegative distinction and childhood predictors of adult outcome. The 18 201 169(3):371-378. 213 302 19. Jurka, J., Kapitonov, V. V., Pavlicek, A., et al: Repbase Update, a database 230 of eukaryotic repetitive elements. Cytogenetic & Genome Research 2005, 110(1-4):462-467. 230 20. Chen N:
290 single-copy orthologs. Bioinformatics 2015, 31(19):3210-3212. 7 291 17. Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simão FA, 9 292 Pordnyakov IA, Ioannidis P, Zdobnov EM: OrthoDB v8: update of the 11 293 hierarchical catalog of orthologs and the underlying free software. Nucleic 12 294 Acids Research 2015, 43(Database issue):D250-D256. 13 295 18. Maziade M, Bouchard S, Gingras N, Charron L, Cardinal A, Roy MA, 14 296 Gauthier B, Tremblay G, Cote S, Fournier C et al: Long-term stability of 15 297 diagnosis and symptom dimensions in a systematic sample of patients 16 297 miagnosis and symptom dimensions in a systematic sample of patients 17 298 with onset of schizophrenia in childhood and early adolescence. II: 18 298 postnegative distinction and childhood predictors of adult outcome. The 203 10 169(3):371-378. 219. Jurka, J., Kapitonov, V. V., Pavlicck, A., et al: Repbase Update, a database 303 of eukaryotic repetitive elements. Cytogenetic & Genome Research 2005, 210 Loten N:: Using RepeatMasker to Identify Repetitive Elements in Genomic 2305
7 291 17. Kriventseva EV, Tegenfeldt P, Petty TJ, Waterhouse RM, Simão FA, 7 292 Pozdnyakov IA, Ioannidis P, Zdobnov EM: OrthoDB v8: update of the 10 293 hierarchical catalog of orthologs and the underlying free software. Nucleic 12 294 Acids Research 2015, 43(Database issue):D250-D256. 12 294 Acids Research 2015, 43(Database issue):D250-D256. 12 294 Gauthier B, Tremblay G, Cote S, Fournier C et al: Long-term stability of 13 197 diagnosis and symptom dimensions in a systematic sample of patients 18 With onset of schizophrenia in childhood predictors of adult outcome. The 19 299 Postnegative distinction and childhood predictors of adult outcome. The 200 British journal of psychiatry : the journal of mental science 1996, 213 101 169(3):371-378. 220 Den N: Using RepeatMasker to Identify Repetitive Elements in Genomic 236 of eukaryotic repetitive elements. Cytogenetic & Genome Research 2005, 210 Lhen N: Using RepeatMasker to Identify Repetitive Elements in Genomic 236 Sequences[J]. 2004, Chapter 4(Unit 4):4.10.14.10.14. 230 21. Benson G, Tandem repeats finder: a program to analyze DNA se
9292Pozdnyakov IA, Ioannidis P, Zdobnov EM: OrthoDB v8: update of the10293hierarchical catalog of orthologs and the underlying free software. Nucleic11294Acids Research 2015, 43(Database issue):D250-D256.1229518.Maziade M, Bouchard S, Gingras N, Charron L, Cardinal A, Roy MA,12296Gauthier B, Tremblay G, Cote S, Foumier C et al: Long-term stability of15297diagnosis and symptom dimensions in a systematic sample of patients17298with onset of schizophrenia in childhood and early adolescence. II:18299Postnegative distinction and childhood predictors of adult outcome. The19300British journal of psychiatry : the journal of mental science 1996,16169(3):371-378.20219Jurka, J., Kapitonov, V. V., Pavlicek, A., et al: Repbase Update, a database203of eukaryotic repetitive elements. Cytogenetic & Genome Research 2005,204110(1-4):462-467.20520.Chen N: Using RepeatMasker to Identify Repetitive Elements in Genomic206Sequences[J]. 2004, Chapter 4(Unit 4):4.10.14.10.14.20721.Benson G, .Tandem repeats finder: a program to analyze DNA sequences.208Nucleic Acids Research 1999, 27(2):573-580.20922.Mario S, Oliver K, Irfan G, Alec H, Stephan W, Burkhard M: AUGUSTUS:201310ab initio prediction of alternative transcripts. Nucleic Acids Research 2006,21131223.Burge C., Karlin S., Prediction of complete gene structures in human213genom
10 293 hierarchical catalog of orthologs and the underlying free software. Nucleic 11 294 Acids Research 2015, 43(Database issue):D250-D256. 13 295 18. Maziade M, Bouchard S, Gingras N, Charron L, Cardinal A, Roy MA, 14 296 Gauthier B, Tremblay G, Cote S, Fournier C et al: Long-term stability of 15 297 diagnosis and symptom dimensions in a systematic sample of patients 17 298 with onset of schizophrenia in childhood and early adolescence. II: 18 299 Postnegative distinction and childhood predictors of adult outcome. The 200 British journal of psychiatry : the journal of mental science 1996, 213 101 169(3):371-378. 223 01 169(3):371-378. 233 02 19. Jurka, J, Kapitonov, V. V., Pavlicek, A., et al: Repbase Update, a database 243 303 of cukaryotic repetitive elements. Cytogenetic & Genome Research 2005, 211 110(1-4):462-467. 210 Chen N: Using RepeatMasker to Identify Repetitive Elements in Genomic 230 Sequences[J]. 2004, Chapter 4(Unit 4):4.10.1-4.10.14. 307 21. Benson G, . Tandem repeats finder: a program to analyze DNA sequences. </td
11294Acids Research 2015, 43(Database issue):D250-D256.1329518.Maziade M, Bouchard S, Gingras N, Charron L, Cardinal A, Roy MA,14296Gauthier B, Tremblay G, Cote S, Fournier C et al: Long-term stability of15297diagnosis and symptom dimensions in a systematic sample of patients17298with onset of schizophrenia in childhood and early adolescence. II:18299Postnegative distinction and childhood predictors of adult outcome. The19299Postnegative distinction and childhood predictors of adult outcome. The200British journal of psychiatry : the journal of mental science 1996,213301169(3):371-378.2330219.304110(1-4):462-467.230520.Chen N: Using RepeatMasker to Identify Repetitive Elements in Genomic2306Sequences[J]. 2004, Chapter 4(Unit 4):4.10.1-4.10.14.30721.Benson G, . Tandem repeats finder: a program to analyze DNA sequences.308Nucleic Acids Research 1999, 27(2):573-580.30922.Mario S, Oliver K, Irfan G, Alec H, Stephan W, Burkhard M: AUGUSTUS:310ab initio prediction of alternative transcripts. Nucleic Acids Research 2006,31134:435-439.31223.313genomic DNA. Journal of Molecular Biology 1997, 268(1):78-94.31424.315into the zebrafish Ensembl genebuild. Genome Research 2012,31622(10):2067-2078.31725.318Nagayasu Y, Doi K, Kasai Y: The medaka draft ge
 295 18. Maziade M, Bouchard S, Gingras N, Charron L, Cardinal A, Roy MA, 296 Gauthier B, Tremblay G, Cote S, Fournier C <i>et al</i>: Long-term stability of 297 diagnosis and symptom dimensions in a systematic sample of patients 298 with onset of schizophrenia in childhood and early adolescence. II: 299 Postnegative distinction and childhood predictors of adult outcome. <i>The</i> 291 British journal of psychiatry : the journal of mental science 1996, 202 169(3):371-378. 203 169(3):371-378. 204 169(3):371-478. 205 20. Chen N.: Using RepeatMasker to Identify Repetitive Elements in Genomic 206 sequences[J]. 2004, Chapter 4(Unit 4):4.10.1-4.10.14. 210 Benson G, . Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research 1999, 27(2):573-580. 210 Mario S, Oliver K, Irfan G, Alec H, Stephan W, Burkhard M: AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Research 2006, 311 34:435-439. 312 23. Burge C., Karlin S., Prediction of complete gene structures in human genomic DNA. Journal of Molecular Biology 1997, 268(1):78-94. 216 Collins JE, White S, Searle SMJ, Stemple DL: Incorporating RNA-seq data into the zebrafish Ensembl genebuild. Genome Research 2012, 22(10):2067-2078. 316 22(10):2067-2078. 317 25. Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y: The medaka draft genome and insights into vertebrate genome evolution. Nature 2007, 447(7145):714-719. 320 267. 2075585):1301-1310. 321 32 32 7. Jaillon O, Aury JM, Brunet F, Petit JL, Stangethomann N, Mauceli E, 322 327 Jaillon O, Aury JM, Brunet F, Petit JL, Stangethomann N, Mauceli E, 324 Bouneau L, Fischer C, Ozouf
14296Gauthier B, Tremblay G, Cote S, Fournier C et al: Long-term stability of15297diagnosis and symptom dimensions in a systematic sample of patients17298with onset of schizophrenia in childhood and early adolescence. II:18299Postnegative distinction and childhood predictors of adult outcome. The203300British journal of psychiatry : the journal of mental science 1996,213301169(3):371-378.22330219.304Jurka, J., Kapitonov, V. V., Pavlicek, A., et al: Repbase Update, a database205of eukaryotic repetitive elements. Cytogenetic & Genome Research 2005,216100(1-4):462-467.27330520.20.Chen N.: Using RepeatMasker to Identify Repetitive Elements in Genomic293Sequences[J]. 2004, Chapter 4(Unit 4):4:10.1-4.10.14.20330721.21.Benson G, . Tandem repeats finder: a program to analyze DNA sequences.308Nucleic Acids Research 1999, 27(2):573-580.31330922.31434:435-439.315ab initio prediction of alternative transcripts. Nucleic Acids Research 2006,31134:435-439.31223.31424.24.Collins JE, White S, Searle SMJ, Stemple DL: Incorporating RNA-seq data31622(10):2067-2078.31725.25.Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T,318Nagayasu Y, Doi K, Kasai Y: The medaka draft genome and insights into3
15297diagnosis and symptom dimensions in a systematic sample of patients17298with onset of schizophrenia in childhood and early adolescence. II:19299Postnegative distinction and childhood predictors of adult outcome. The200British journal of psychiatry : the journal of mental science 1996,213169(3):371-378.22330219.304169(3):371-378.305219.305219.306repetitive elements. Cytogenetic & Genome Research 2005,216304307210.308Sequences[J]. 2004, Chapter 4(Unit 4):4.10.1-4.10.14.30922.301Sequences[J]. 2004, Chapter 4(Unit 4):4.10.1-4.10.14.30721.308Nucleic Acids Research 1999, 27(2):573-580.30922.30130930223.30330930431030531131232.313genomic DNA. Journal of Molecular Biology 1997, 268(1):78-94.31424.315into the zebrafish Ensembl genebuild. Genome Research 2012,31622(10):2067-2078.31725.318Nagayasu Y, Doi K, Kasai Y: The medaka draft genome and insights into319vertebrate genome evolution. Nature 2007, 447(7145):714-719.32026.331Kesteven GL: Whole-Genome Shotgun Assembly and Analysis of the3319Genome of Fugu rubripes. Science (New York, NY) 2002,331223.<
17298with onset of schizophrenia in childhood and early adolescence. II:18299Postnegative distinction and childhood predictors of adult outcome. The200British journal of psychiatry : the journal of mental science 1996,21301169(3):371-378.2230219.303of eukaryotic repetitive elements. Cytogenetic & Genome Research 2005,24303110(1-4):462-467.2730520.20.Chen N.: Using RepeatMasker to Identify Repetitive Elements in Genomic29306Sequences[J]. 2004, Chapter 4(Unit 4):4.10.1-4.10.14.30721.Benson G, . Tandem repeats finder: a program to analyze DNA sequences.308Nucleic Acids Research 1999, 27(2):573-580.30922.Mario S, Oliver K, Irfan G, Alec H, Stephan W, Burkhard M: AUGUSTUS:310ab initio prediction of alternative transcripts. Nucleic Acids Research 2006,31134:435-439.31223.313genomic DNA. Journal of Molecular Biology 1997, 268(1):78-94.31424.315into the zebrafish Ensembl genebuild. Genome Research 2012,31622(10):2067-2078.31725.318Nagayasu Y, Doi K, Kasai Y: The medaka draft genome and insights into319vertebrate genome evolution. Nature 2007, 447(7145):714-719.32026.33127.34239.35120:2067-2078.36320:207(5585):1301-1310.3643237920:207(5585):1301-13
18299Postnegative distinction and childhood predictors of adult outcome. The20300British journal of psychiatry : the journal of mental science 1996,21301169(3):371-378.2330219.303of eukaryotic repetitive elements. Cytogenetic & Genome Research 2005,24303of eukaryotic repetitive elements. Cytogenetic & Genome Research 2005,25304110(1-4):462-467.26306Sequences[J]. 2004, Chapter 4(Unit 4):4.10.1-4.10.14.2730721.308Nucleic Acids Research 1999, 27(2):573-580.30922.301Mario S, Oliver K, Irfan G, Alec H, Stephan W, Burkhard M: AUGUSTUS:303ab initio prediction of alternative transcripts. Nucleic Acids Research 2006,31134:435-439.31223.313genomic DNA. Journal of Molecular Biology 1997, 268(1):78-94.40314315into the zebrafish Ensembl genebuild. Genome Research 2012,31622(10):2067-2078.31725.318Nagayau Y, Doi K, Kasai Y: The medaka draft genome and insights into319vertebrate genome evolution. Nature 2007, 447(7145):714-719.32026.321Genome of Fugu rubripes. Science (New York, NY) 2002,322297(5585):1301-1310.32327.32327.324Fourier C, Ozoufcostaz C, Bernot A: Genome duplication in the325teleost fish Tetraodon nigroviridis reveals the early vertebrate
20300British journal of psychiatry : the journal of mental science 1996,21301169(3):371-378.2230219.Jurka, J., Kapitonov, V. V., Pavlicek, A., et al: Repbase Update, a database2330219.Jurka, J., Kapitonov, V. V., Pavlicek, A., et al: Repbase Update, a database24303of cukaryotic repetitive elements. Cytogenetic & Genome Research 2005,26304110(1-4):462-467.2730520.Chen N.: Using RepeatMasker to Identify Repetitive Elements in Genomic29306sequences[J]. 2004, Chapter 4(Unit 4):4.10.1-4.10.14.30721.Benson G, . Tandem repeats finder: a program to analyze DNA sequences.308Nucleic Acids Research 1999, 27(2):573-580.30922.Mario S, Oliver K, Irfan G, Alec H, Stephan W, Burkhard M: AUGUSTUS:310ab initio prediction of alternative transcripts. Nucleic Acids Research 2006,31134:435-439.31223.313genomic DNA. Journal of Molecular Biology 1997, 268(1):78-94.31424.315into the zebrafish Ensembl genebuild. Genome Research 2012,31622(10):2067-2078.31725.318Nagayasu Y, Doi K, Kasai Y: The medaka draft genome and insights into319vertebrate genome evolution. Nature 2007, 447(7145):714-719.32026.31421.315cance of Fugu rubripes. Science (New York, NY) 2002,316222317255555):1301-1310.318Sagayasu Y, Doi K, Kasa
21301169(3):371-378.2230219.Jurka, J., Kapitonov, V. V., Pavlicek, A. ,et al: Repbase Update, a database24303of eukaryotic repetitive elements. Cytogenetic & Genome Research 2005,26304110(1-4):462-467.2730520.Chen N.: Using RepeatMasker to Identify Repetitive Elements in Genomic29306Sequences[J]. 2004, Chapter 4(Unit 4):4.10.1-4.10.14.30721.Benson G, . Tandem repeats finder: a program to analyze DNA sequences.308Nucleic Acids Research 1999, 27(2):573-580.33330922.Mario S, Oliver K, Irfan G, Alec H, Stephan W, Burkhard M: AUGUSTUS:310ab initio prediction of alternative transcripts. Nucleic Acids Research 2006,31134:435-439.31223.Burge C, Karlin S., Prediction of complete gene structures in human313genomic DNA. Journal of Molecular Biology 1997, 268(1):78-94.403144124.421Collins JE, White S, Searle SMJ, Stemple DL: Incorporating RNA-seq data42331622(10):2067-2078.4331725.4431725.45318463184731948vertebrate genome evolution. Nature 2007, 447(7145):714-719.49320403214132142Collins JE, Whole-Genome Shotgun Assembly and Analysis of the43Genome of Fugu rubripes. Science (New York, NY) 2002,44319 </td
2230219.Jurka, J., Kapitonov, V. V., Pavlicek, A., et al: Repbase Update, a database24303of eukaryotic repetitive elements. Cytogenetic & Genome Research 2005,26304110(1-4):462-467.2730520.Chen N.: Using RepeatMasker to Identify Repetitive Elements in Genomic2830721.Benson G, . Tandem repeats finder: a program to analyze DNA sequences.308Nucleic Acids Research 1999, 27(2):573-580.30922.Mario S, Oliver K, Irfan G, Alec H, Stephan W, Burkhard M: AUGUSTUS:31ab initio prediction of alternative transcripts. Nucleic Acids Research 2006,31134:435-439.31223.Burge C., Karlin S., Prediction of complete gene structures in human39313genomic DNA. Journal of Molecular Biology 1997, 268(1):78-94.4031424.Collins JE, White S, Searle SMJ, Stemple DL: Incorporating RNA-seq data315into the zebrafish Ensembl genebuild. Genome Research 2012,31822(10):2067-2078.319vertebrate genome evolution. Nature 2007, 447(7145):714-719.32026.Kesteven GL: Whole-Genome Shotgun Assembly and Analysis of the32132327.32327.324324325324326327.327323328327.329326330339340343413434234344343453434634
24 25303 304of eukaryotic repetitive elements. Cytogenetic & Genome Research 2005, 304 30526 27 28305 30520.Chen N.: Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences[J]. 2004, Chapter 4(Unit 4):4.10.1-4.10.14.30 307 307 30821.Benson G, Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research 1999, 27(2):573-580.33 309 309 309 310 31122.Mario S, Oliver K, Irfan G, Alec H, Stephan W, Burkhard M: AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Research 2006, 311 34:435-439.37 312 313 313 313 314Burge C., Karlin S., Prediction of complete gene structures in human genomic DNA. Journal of Molecular Biology 1997, 268(1):78-94.40 41 41 31424.Collins JE, White S, Searle SMJ, Stemple DL: Incorporating RNA-seq data into the zebrafish Ensembl genebuild. Genome Research 2012, 22(10):2067-2078.44 45 46 317 31925.Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y: The medaka draft genome and insights into vertebrate genome evolution. Nature 2007, 447(7145):714-719.49 300 301 322 322 322 327(5585):1301-1310.323 327.323 323 327.27.Jaillon O, Aury JM, Brunet F, Petit JL, Stangethomann N, Mauceli E, Bouneau L, Fischer C, Ozoufcostaz C, Bernot A: Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate
226304110(1-4):462-467.27830520.Chen N.: Using RepeatMasker to Identify Repetitive Elements in Genomic298306Sequences[J]. 2004, Chapter 4(Unit 4):4.10.1-4.10.14.30721.Benson G., Tandem repeats finder: a program to analyze DNA sequences.308Nucleic Acids Research 1999, 27(2):573-580.30922.Mario S, Oliver K, Irfan G, Alec H, Stephan W, Burkhard M: AUGUSTUS:31ab initio prediction of alternative transcripts. Nucleic Acids Research 2006,3134:435-439.31genomic DNA. Journal of Molecular Biology 1997, 268(1):78-94.31424.Collins JE, White S, Searle SMJ, Stemple DL: Incorporating RNA-seq data315into the zebrafish Ensembl genebuild. Genome Research 2012,31622(10):2067-2078.31725.Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T,318Nagayasu Y, Doi K, Kasai Y: The medaka draft genome and insights into319vertebrate genome evolution. Nature 2007, 447(7145):714-719.32026.Kesteven GL: Whole-Genome Shotgun Assembly and Analysis of the321Genome of Fugu rubripes. Science (New York, NY) 2002,322297(5585):1301-1310.32327.Jaillon O, Aury JM, Brunet F, Petit JL, Stangethomann N, Mauceli E,32327.Jaillon O, Aury JM, Brunet F, Petit JL, Stangethomann N, Mauceli E,324Bouneau L, Fischer C, Ozoufcostaz C, Bernot A: Genome duplication in the325teleost fish Tetraodon nigroviridis reveals the early vertebrate
27 28 29305 30620.Chen N.: Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences[J]. 2004, Chapter 4(Unit 4):4.10.1-4.10.14.307 307 30821.Benson G, . Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research 1999, 27(2):573-580.309 309 309 309 309 30022.Mario S, Oliver K, Irfan G, Alec H, Stephan W, Burkhard M: AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Research 2006, 34:435-439.311 312 313 313 313 313 314 31424.Collins JE, White S, Searle SMJ, Stemple DL: Incorporating RNA-seq data into the zebrafish Ensembl genebuild. Genome Research 2012, 22(10):2067-2078.44 45 41 4125.Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y: The medaka draft genome and insights into vertebrate genome evolution. Nature 2007, 447(7145):714-719.49 320 321 322 322 322 32327.Jaillon O, Aury JM, Brunet F, Petit JL, Stangethomann N, Mauceli E, Bouneau L, Fischer C, Ozoufcostaz C, Bernot A: Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate
28306Sequences[J]. 2004, Chapter 4(Unit 4):4.10.1-4.10.14.30721.Benson G, . Tandem repeats finder: a program to analyze DNA sequences.308Nucleic Acids Research 1999, 27(2):573-580.30922.Mario S, Oliver K, Irfan G, Alec H, Stephan W, Burkhard M: AUGUSTUS:310ab initio prediction of alternative transcripts. Nucleic Acids Research 2006,31134:435-439.31223.Burge C., Karlin S., Prediction of complete gene structures in human39313genomic DNA. Journal of Molecular Biology 1997, 268(1):78-94.4031424.Collins JE, White S, Searle SMJ, Stemple DL: Incorporating RNA-seq data4131622(10):2067-2078.4431725.Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T,45319vertebrate genome evolution. Nature 2007, 447(7145):714-719.4932026.Kesteven GL: Whole-Genome Shotgun Assembly and Analysis of the50321Genome of Fugu rubripes. Science (New York, NY) 2002,52222297(5585):1301-1310.5332327.Jaillon O, Aury JM, Brunet F, Petit JL, Stangethomann N, Mauceli E,54324Bouneau L, Fischer C, Ozoufcostaz C, Bernot A: Genome duplication in the54325teleost fish Tetraodon nigroviridis reveals the early vertebrate
30721.Benson G, Tandem repeats finder: a program to analyze DNA sequences.318308Nucleic Acids Research 1999, 27(2):573-580.31930922.Mario S, Oliver K, Irfan G, Alec H, Stephan W, Burkhard M: AUGUSTUS:310ab initio prediction of alternative transcripts. Nucleic Acids Research 2006,31134:435-439.31223.Burge C., Karlin S., Prediction of complete gene structures in human319313genomic DNA. Journal of Molecular Biology 1997, 268(1):78-94.4031424.Collins JE, White S, Searle SMJ, Stemple DL: Incorporating RNA-seq data31622(10):2067-2078.4131725.4131725.42Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T,48319vertebrate genome evolution. Nature 2007, 447(7145):714-719.4932026.41321Genome of Fugu rubripes. Science (New York, NY) 2002,42322297(5585):1301-1310.4332327.4432445324463244732448325493234132442324433244431545324463244732548324493204932140322413244232543324443
31308Nucleic Acids Research 1999, 27(2):573-580.3330922.Mario S, Oliver K, Irfan G, Alec H, Stephan W, Burkhard M: AUGUSTUS:34310ab initio prediction of alternative transcripts. Nucleic Acids Research 2006,3631134:435-439.3731223.38genomic DNA. Journal of Molecular Biology 1997, 268(1):78-94.4031424.4124.42Collins JE, White S, Searle SMJ, Stemple DL: Incorporating RNA-seq data4331622(10):2067-2078.4431725.4531725.46318Nagayasu Y, Doi K, Kasai Y: The medaka draft genome and insights into47319vertebrate genome evolution. Nature 2007, 447(7145):714-719.4932026.41Steine of Fugu rubripes. Science (New York, NY) 2002,42297(5585):1301-1310.4332327.443244532546325473244831949320493204032141Genome of Fugu rubripes. Science (New York, NY) 2002,4232343324444545454647484949494041414243444546<
3330922.Mario S, Oliver K, Irfan G, Alec H, Stephan W, Burkhard M: AUGUSTUS:310ab initio prediction of alternative transcripts. Nucleic Acids Research 2006,31134:435-439.31223.Burge C., Karlin S., Prediction of complete gene structures in human39313genomic DNA. Journal of Molecular Biology 1997, 268(1):78-94.4031424.Collins JE, White S, Searle SMJ, Stemple DL: Incorporating RNA-seq data41315into the zebrafish Ensembl genebuild. Genome Research 2012,4222(10):2067-2078.4331725.44318Nagayasu Y, Doi K, Kasai Y: The medaka draft genome and insights into47319vertebrate genome evolution. Nature 2007, 447(7145):714-719.4832026.4932026.40Science (New York, NY) 2002,4132242297(5585):1301-1310.4332327.443244532546325
34 35310ab initio prediction of alternative transcripts. Nucleic Acids Research 2006,36 31134:435-439.37 31223.Burge C., Karlin S., Prediction of complete gene structures in human genomic DNA. Journal of Molecular Biology 1997, 268(1):78-94.40 4131424.Collins JE, White S, Searle SMJ, Stemple DL: Incorporating RNA-seq data into the zebrafish Ensembl genebuild. Genome Research 2012,41 42 43316 316 31722(10):2067-2078.44 45 47 48317 31925.48 49 41319 32026.49 41 42320 32026.41 43 44 45317 32026.42 43 44 45319 32026.44 45 46 47320 32026.45 46 47 48 48321 321 322 322 322 322327.46 47 48 48 49323 323 32327.47 49 49 40 40321 321 323 32327.48 49 49 49 49 49 49320 32226.49 40 41 410321 40 410Genome of Fugu rubripes. Science (New York, NY) 2002, 322 323 323 323 323 32341 42 431324 431324 43142 431 44 44324 44 44 44431 45 450324 450 4644 450 46 47324 47451 47 48 49324 47452 47 48324 47453
 311 34:435-439. 312 312 313 313 314 314 314 314 315 316 317 316 317 317 318 319 319 319 319 319 310 320 310 320 321 321 321 322 321 322 321 322 323 323 327 321 323 327 321 323 327 323 327 323 323 325 324 325 325 325 324 325 325 325 326 326 327 328 329 329 320 321 323 323 323 323 323 324 325 325 325 326 327 327 328 324 329 325 324 325 325 325 325 326 327 327 328 329 329 320 321 323 321 323 323 323 321 323 323 323 324 324 325 325 325 325 326 327 328 329 329 329 320 320 320 321 321 323 323 323 323 323 324 325 325 325 327 328 329 329 329 320 320 320 321 321 321 322 321 323 323 323 323 323 324 325 325 325
37 3831223.Burge C., Karlin S., Prediction of complete gene structures in human genomic DNA. Journal of Molecular Biology 1997, 268(1):78-94.40 4131424.Collins JE, White S, Searle SMJ, Stemple DL: Incorporating RNA-seq data into the zebrafish Ensembl genebuild. Genome Research 2012,43 44 45316 31722(10):2067-2078.44 4531725.Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y: The medaka draft genome and insights into vertebrate genome evolution. Nature 2007, 447(7145):714-719.49 40 41 42032026.Kesteven GL: Whole-Genome Shotgun Assembly and Analysis of the Genome of Fugu rubripes. Science (New York, NY) 2002, 222 227(5585):1301-1310.53 54 55 5243243010 O, Aury JM, Brunet F, Petit JL, Stangethomann N, Mauceli E, Bouneau L, Fischer C, Ozoufcostaz C, Bernot A: Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate
 313 313 313 313 314 314 314 315 316 317 316 317 316 317 317 318 318 319 319 319 320 320 320 321 321 321 322 321 322 321 322 321 323 323 323 323 323 323 325 324 325 324 325 325 325 324 316 325 325 324 325 325 325 326 326 327 327 328 324 327 329 325 324 325 324 325 325 325 326 327 327 327 328 324 325 324 325 324 325 325 325 326 327 327 327 328 325 328 329 329 329 320 321 321 321 323 323 323 324 325 324 325 325 325 326 327 327 328 328 329 329 329 329 329 320 320 320 321 321 321 321 323 323 323 324 325 327 328 329 329 329 329 320 320 320 320 321 321 321 322 321 323 321 323 321 323 321 323 321 321 321 321 321 321 321 321 321
 314 24. Collins JE, White S, Searle SMJ, Stemple DL: Incorporating RNA-seq data 315 into the zebrafish Ensembl genebuild. Genome Research 2012, 316 22(10):2067-2078. 317 25. Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, 318 Nagayasu Y, Doi K, Kasai Y: The medaka draft genome and insights into 319 vertebrate genome evolution. Nature 2007, 447(7145):714-719. 320 26. Kesteven GL: Whole-Genome Shotgun Assembly and Analysis of the 321 Genome of Fugu rubripes. Science (New York, NY) 2002, 322 297(5585):1301-1310. 323 27. Jaillon O, Aury JM, Brunet F, Petit JL, Stangethomann N, Mauceli E, Bouneau L, Fischer C, Ozoufcostaz C, Bernot A: Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate
 into the zebrafish Ensembl genebuild. Genome Research 2012, 22(10):2067-2078. 316 317 25. Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y: The medaka draft genome and insights into vertebrate genome evolution. Nature 2007, 447(7145):714-719. 320 26. Kesteven GL: Whole-Genome Shotgun Assembly and Analysis of the Genome of Fugu rubripes. Science (New York, NY) 2002, 322 297(5585):1301-1310. 323 323 27. Jaillon O, Aury JM, Brunet F, Petit JL, Stangethomann N, Mauceli E, Bouneau L, Fischer C, Ozoufcostaz C, Bernot A: Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate
 316 316 316 317 25. Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, 318 318 319 319 320 320 321 323 323 323 323 323 323 323 324 325 325 325 325 325 326 326 326 327 328 329 329 320 321 323 323 323 323 323 324 325 325 325 325 325 325 326 325 327 328 329 329 329 320 321 321 323 323 323 323 324 325 325 325 325 325 326 327 328 329 329 329 320 320 321 321 321 321 323 323 323 323 324 325 325 325 325 326 327 328 329 329 329 329 320 320 320 320 321 321 321 323 323 323 323 323 323 324 325 325 326 327 328 329 329 329 329 329 320 320 320 320 320 321 321 321 321 321 321 321 321<!--</td-->
 317 25. Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, 318 Nagayasu Y, Doi K, Kasai Y: The medaka draft genome and insights into 319 vertebrate genome evolution. <i>Nature</i> 2007, 447(7145):714-719. 320 26. Kesteven GL: Whole-Genome Shotgun Assembly and Analysis of the 321 Genome of Fugu rubripes. <i>Science</i> (<i>New York, NY</i>) 2002, 322 297(5585):1301-1310. 323 27. Jaillon O, Aury JM, Brunet F, Petit JL, Stangethomann N, Mauceli E, 324 Bouneau L, Fischer C, Ozoufcostaz C, Bernot A: Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate
46318Nagayasu Y, Doi K, Kasai Y: The medaka draft genome and insights into47319vertebrate genome evolution. Nature 2007, 447(7145):714-719.4932026.Kesteven GL: Whole-Genome Shotgun Assembly and Analysis of the50321Genome of Fugu rubripes. Science (New York, NY) 2002,52322297(5585):1301-1310.5332327.54324Bouneau L, Fischer C, Ozoufcostaz C, Bernot A: Genome duplication in the56325teleost fish Tetraodon nigroviridis reveals the early vertebrate
 ⁴⁷ ⁴⁸ ⁴⁷ ⁴⁸ ⁴⁹ ⁵⁰ ⁵⁰ ⁵¹ ⁵¹ ⁵² ⁵² ⁵² ⁵³ ⁵³ ⁵³ ⁵³ ⁵⁴ ⁵⁵ ⁵⁴ ⁵⁵ ⁵⁶ ⁵⁷ ⁵⁸ ⁵⁹ ⁵¹ ⁵¹ ⁵³ ⁵⁴ ⁵⁵ ⁵⁵ ⁵⁶ ⁵⁷ ⁵⁷ ⁵⁷ ⁵⁸ ⁵¹ ⁵¹ ⁵¹ ⁵² ⁵³ ⁵⁴ ⁵⁵ ⁵⁴ ⁵⁵ ⁵⁵ ⁵⁵ ⁵⁶ ⁵⁷ ⁵⁷ ⁵⁶ ⁵⁷ ⁵⁷ ⁵⁸ ⁵⁹ ⁵¹ ⁵¹ ⁵¹ ⁵¹ ⁵² ⁵³ ⁵⁴ ⁵⁵ ⁵⁴ ⁵⁵ ⁵⁵ ⁵¹ ⁵² ⁵³ ⁵⁴ ⁵⁵ ⁵⁴ ⁵⁵ ⁵⁵ ⁵⁵ ⁵⁶ ⁵⁷ ⁵⁷ ⁵⁸ ⁵¹ ⁵¹ ⁵¹ ⁵² ⁵³ ⁵⁴ ⁵⁵ ⁵⁴ ⁵⁵ ⁵⁵ ⁵⁵ ⁵⁵
 320 26. Kesteven GL: Whole-Genome Shotgun Assembly and Analysis of the 321 Genome of Fugu rubripes. Science (New York, NY) 2002, 322 297(5585):1301-1310. 323 27. Jaillon O, Aury JM, Brunet F, Petit JL, Stangethomann N, Mauceli E, Bouneau L, Fischer C, Ozoufcostaz C, Bernot A: Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate
50 51321Genome of Fugu rubripes. Science (New York, NY) 2002,52 52322297(5585):1301-1310.53 54 5532327.54 55324Bouneau L, Fischer C, Ozoufcostaz C, Bernot A: Genome duplication in the56 57325teleost fish Tetraodon nigroviridis reveals the early vertebrate
51322 297 (5585):1301-1310.5332327.5432327.55324Bouneau L, Fischer C, Ozoufcostaz C, Bernot A: Genome duplication in the56325teleost fish Tetraodon nigroviridis reveals the early vertebrate
53 54 5532327.Jaillon O, Aury JM, Brunet F, Petit JL, Stangethomann N, Mauceli E,55 55324Bouneau L, Fischer C, Ozoufcostaz C, Bernot A: Genome duplication in the56 57325teleost fish Tetraodon nigroviridis reveals the early vertebrate
54 55324Bouneau L, Fischer C, Ozoufcostaz C, Bernot A: Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate56 57325
56325teleost fish Tetraodon nigroviridis reveals the early vertebrate
⁵ / ₅₈ 326 proto-karyotype . <i>Nature</i> 2004, 431 (7011):946-957.
59
60 61
62
63

	327	28.	Rondeau EB, Minkley DR, Leong JS, Messmer AM, Jantzen JR, von
1 2	328		Schalburg KR, Lemon C, Bird NH, Koop BF: The genome and linkage map
3	329		of the northern pike (Esox lucius): conserved synteny revealed between
4	330		the salmonid sister group and the Neoteleostei. PLoS ONE 2014,
5 6	331		9 (7):e102089.
7	332	29.	Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford
8 9	333		R, Pirun M, Zody MC, White S: The genomic basis of adaptive evolution in
10	334		threespine sticklebacks. Nature 2012, 484(7392):55-61.
11 12	335	30.	Pevsner J: Basic Local Alignment Search Tool (BLAST): John Wiley &
13	336		Sons, Inc.; 2005.
14 15	337	31.	Yu XJ, Zheng HK, Wang J, Wang W, Su B: Detecting lineage-specific
16	338		adaptive evolution of brain-expressed genes in human using rhesus
17	339		macaque as outgroup. Genomics 2006, 88(6):745-751.
18 19	340	32.	Birney E, Clamp M, Durbin R: GeneWise and Genomewise. Genome
20	341		Research 2004, 14(5):988-995.
21 22	342	33.	Elsik CG, Mackey AJ, Reese JT, Milshina NV, Roos DS, Weinstock GM:
23	343		Creating a honey bee consensus gene set. Genome Biology 2007,
24 25	344		8 (1):90-105.
26	345	34.	Bairoch A, Apweiler R: The SWISS-PROT protein sequence database and
27	346		its supplement TrEMBL in 2000. Nucleic Acids Res 2000, 28(1):45-48.
28 29	347	35.	Finn RD: Pfam: the protein families database. Nucleic Acids Research 2014,
30	348		42 (Database issue):D222-230.
31 32	349	36.	Attwood TK: The PRINTS database: A resource for identification of
33	350		protein families. Briefings in Bioinformatics 2002, 3(3):252-263.
34 35	351	37.	Bru C, Courcelle E, Beausse Y, Dalmar S, Kahn D: The ProDom database of
36	352		protein domain families: more emphasis on 3D. Nucleic Acids Research
37	353		2005, 33 (Database issue):212-215.
39	354	38.	Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting
40	355		CP, Bork P: SMART 4.0: towards genomic data integration. Nucleic Acids
4⊥ 42	356		Research 2004, 32(Databaseissue):D142-D144.
43	357	39.	Boeckmann B., Bairoch A., Apweiler R., Blatter M. C., Estreicher A et al:
44 45	358		The Swiss-Prot knowledgebase and its supplement TREMBL in 2003.
46	359		<i>Nucleic Acids Research</i> 2003, 31 (1):365-370.
47 48	360	40.	Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P,
49	361		Das U, Daugherty L, Duquenne L: InterPro: the integrative protein
50	362		signature database. Nucleic Acids Research 2009, 37(suppl 1):D211-D215.
51 52	363	41.	Hingamp P, Broek AEVD, Stoesser G, Baker W: The EMBL nucleotide
53	364		sequence database. Molecular Biotechnology 1999, 12(3):255-267.
54 55	365	42.	Kanehisa M, Goto S: KEGG: kyoto encyclopedia of genes and genomes.
56	366		Nucleic Acids Research 2000, 27 (1):29-34(26).
57 58	367	43.	Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J,
59	368		Amores A, Desvignes T, Batzel P, Catchen J et al: The spotted gar genome
60 61			
62			

369		illuminates vertebrate evolution and facilitates human-teleost
370		comparisons . <i>Nature genetics</i> 2016, 48 (4):427-437.
371	44.	Bian C, Hu Y, Ravi V, Kuznetsova IS, Shen X, Mu X, Sun Y, You X, Li J, Li
372		X et al: The Asian arowana (Scleropages formosus) genome provides new
373		insights into the evolution of an early lineage of teleosts. Scientific reports
374		2016, 6 :24501.
375	45.	Ostlund G, Schmitt T, Forslund K, Kostler T, Messina DN, Roopra S, Frings
376		O, Sonnhammer EL: InParanoid 7: new algorithms and tools for
377		eukaryotic orthology analysis. Nucleic acids research 2010, 38(Database
378		issue):D196-203.
379	46.	Edgar RC: MUSCLE: multiple sequence alignment with high accuracy
380		and high throughput. Nucleic acids research 2004, 32 (5):1792-1797.
381	47.	Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S,
382		Larget B, Liu L, Suchard MA, Huelsenbeck JP: MrBayes 3.2: efficient
383		Bayesian phylogenetic inference and model choice across a large model
384		space . Systematic biology 2012, 61 (3):539-542.
385	48.	Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, Simakov O, Ng
386		AY, Lim ZW, Bezault E <i>et al</i> : The genomic substrate for adaptive
387		radiation in African cichlid fish. <i>Nature</i> 2014, 513 (7518):375-381.
388	49.	Proost S, Fostier J, De Witte D, Dhoedt B, Demeester P, Van de Peer Y,
389		Vandepoele K: i-ADHoRe 3.0fast and sensitive detection of genomic
390		homology in extremely large data sets. Nucleic acids research 2012,
91		40 (2):e11.
392	50.	Yang Z: PAML: a program package for phylogenetic analysis by
693		maximum likelihood. Computer applications in the biosciences : CABIOS
394		1997, 13 (5):555-556.
395	51.	Kai Liu; Dongpo Xu; Jia Li; Chao Bian; Jinrong Duan; Yanfeng Zhou;
96		Minying Zhang; Xinxin You; Yang You; Jieming Chen; Hui Yu; Gangchun
397		Xu; Di-an Fang; Jun Qiang; Shulun Jiang; Jie He; Junmin Xu; Qiong Shi;
398		Zhiyong Zhang; Pao Xu (2016) Supporting data for "Whole genome
399		sequencing of Chinese clearhead icefish, Protosalanx hyalocranius''.
400		GigaScience Database. http://doi.org/10.5524/100262
401		
402		
403		
404	Tab	les
405	Tabl	e 1 . The statistics of genome assembly and annotation for <i>P. hyalocranius</i> .
	Gen	ome assembly
	Con	tig N50 size (kb) 17.2
	Scaf	fold N50 size (Mb) 1.163

Assembl	ed genome size	(MD)	536				
Genome	coverage (X)		315				
The long	gest scaffold (bp)	5,398,3	389			
Gap leng	gth (Mb)		122				
Genome	annotation						
Protein-o	coding gene nur	nber	19,884				
Annotate	ed functional ge	ne number	: 19,125	(96.2%)			
Unannot	ated functional	gene numb	per 759 (3.	8%)			
Repeat c	ontent		24.43%	,)			
Table 2.	Detailed classif	ication of a	repeat seque	nces in the	assembled	genome.	
Table 2.	Detailed classif	ication of Rep	repeat sequer peat Size(bp)	nces in the	assembled % of Gene	genome.	
Table 2.TypeProteinM	Detailed classif	ication of r Rep 992	repeat sequer peat Size(bp) 5152	nces in the	assembled % of Geno 1.85	genome.	
Table 2. Type ProteinM RepeatM	Detailed classif Aask Iasker	ication of r Rep 992 594	repeat sequer peat Size(bp) 5152 8136	nces in the	assembled % of Geno 1.85 1.11	genome.	
Table 2. Type ProteinM RepeatM Tandem	Detailed classif Aask Iasker Repeat Finder	ication of 1 Rep 992 594 665	repeat sequen beat Size(bp) 5152 8136 95756	nces in the	assembled % of Geno 1.85 1.11 12.41	genome.	
Table 2.TypeProteinMRepeatMTandemDe novo	Detailed classif Aask Iasker Repeat Finder	ication of 1 Rep 992 594 665 937	repeat sequer peat Size(bp) 5152 8136 95756 26009	nces in the	assembled % of Geno 1.85 1.11 12.41 17.47	genome.	
Table 2.TypeProteinMRepeatMTandemDe novoTotal	Detailed classif Aask Iasker Repeat Finder	ication of 1 Rep 992 594 665 937 131	repeat sequer beat Size(bp) 5152 8136 95756 26009 090229	nces in the	assembled % of Geno 1.85 1.11 12.41 17.47 24.43	genome.	
Table 2.TypeProteinMRepeatMTandemDe novoTotal	Detailed classif Aask Aasker Repeat Finder	ication of 1 Rep 992 594 665 937 131	repeat sequer peat Size(bp) 5152 8136 95756 26009 090229	nces in the	assembled % of Geno 1.85 1.11 12.41 17.47 24.43	genome.	
Table 2.TypeProteinMRepeatMTandemDe novoTotal	Detailed classif Aask Aasker Repeat Finder	ication of 1 Rep 992 594 665 937 131	repeat sequer peat Size(bp) 5152 8136 95756 26009 090229	nces in the	assembled % of Geno 1.85 1.11 12.41 17.47 24.43	genome.	
Table 2.TypeProteinMRepeatMTandemDe novoTotal	Detailed classif Aask Aasker Repeat Finder	ication of 1 Rep 992 594 665 937 131 Number	repeat sequer peat Size(bp) 5152 8136 95756 26009 090229 Average	Average	assembled % of Geno 1.85 1.11 12.41 17.47 24.43 Average	genome.	Average
Table 2. Type ProteinM RepeatM Tandem De novo Total	Detailed classif Aask Aasker Repeat Finder	ication of 1 Rep 992 594 665 937 131 Number	repeat sequer eat Size(bp) 5152 8136 95756 26009 090229 Average Transcript	Average CDS	assembled % of Geno 1.85 1.11 12.41 17.47 24.43 Average Exons	genome. ome Average Exons	Average
Table 2.TypeProteinMRepeatMTandemDe novoTotal	Detailed classif Aask Aasker Repeat Finder	ication of 1 Rep 992 594 665 937 131 Number	repeat sequer peat Size(bp) 5152 8136 95756 26009 090229 Average Transcript Length	Average CDS Length	assembled % of Geno 1.85 1.11 12.41 17.47 24.43 Average Exons Per	genome. Dme Average Exons Length	Average Intron Length
Table 2. Type ProteinM RepeatM Tandem De novo Total	Detailed classif Aask Aasker Repeat Finder	ication of 1 Rep 992 594 665 937 131 Number	repeat sequer peat Size(bp) 5152 8136 95756 26009 090229 Average Transcript Length (bp)	Average CDS Length (bp)	assembled % of Geno 1.85 1.11 12.41 17.47 24.43 Average Exons Per Gene	genome. Dme Average Exons Length (bp)	Average Intron Length (bp)
Table 2.TypeProteinMRepeatMTandemDe novoTotal	Detailed classif Aask Aasker Repeat Finder	ication of 1 Rep 992 594 665 937 131 Number 23.132	Average Transcript Length (bp) 4.897.24	Average CDS Length (bp) 1.264 61	assembled % of Geno 1.85 1.11 12.41 17.47 24.43 Average Exons Per Gene 5.78	genome. Dme Average Exons Length (bp) 218.81	Average Intron Length (bp) 760.04
Table 2.TypeProteinMRepeatMTandemDe novoTotal	Detailed classif Aask Aasker Repeat Finder	ication of 1 Rep 992 594 665 937 131 Number 23,132 21,379	repeat sequer peat Size(bp) 5152 8136 95756 26009 090229 Average Transcript Length (bp) 4,897.24 17.213 49	Average CDS Length (bp) 1,264.61 1.973 56	assembled % of Geno 1.85 1.11 12.41 17.47 24.43 Average Exons Per Gene 5.78 10.22	genome. Dme Average Exons Length (bp) 218.81 193.05	Average Intron Length (bp) 760.04 1.652.4
Table 2.TypeProteinMRepeatMTandemDe novoTotalMethodDe novoDe novo	Detailed classif Aask Aasker Repeat Finder AUGUSTUS GeneScan Danio rerio	ication of 1 Rep 992 594 665 937 131 Number 23,132 21,379 25 390	repeat sequer beat Size(bp) 5152 8136 95756 26009 090229 Average Transcript Length (bp) 4,897.24 17,213.49 7 156 92	Average CDS Length (bp) 1,264.61 1,973.56 1 312 32	assembled % of Geno 1.85 1.11 12.41 17.47 24.43 Average Exons Per Gene 5.78 10.22 6.17	genome. Dme Average Exons Length (bp) 218.81 193.05 212.62	Average Intron Length (bp) 760.04 1,652.4
Table 2.TypeProteinMRepeatMTandemDe novoTotalMethodDe novoHomolog	Detailed classif Aask Aasker Repeat Finder AUGUSTUS GeneScan Danio rerio Orvzias	ication of n Rep 992 594 665 937 131 Number 23,132 21,379 25,390 25,319	repeat sequer peat Size(bp) 5152 8136 95756 26009 090229 Average Transcript Length (bp) 4,897.24 17,213.49 7,156.92 6,411.36	Average CDS Length (bp) 1,264.61 1,973.56 1,312.32 1 194 58	assembled % of Geno 1.85 1.11 12.41 17.47 24.43 Average Exons Per Gene 5.78 10.22 6.17 5.89	genome. Dme Average Exons Length (bp) 218.81 193.05 212.62 202.73	Average Intron Length (bp) 760.04 1,652.4 1,129.99
Table 2.TypeProteinMRepeatMTandemDe novoTotalWethodDe novoHomolog	Detailed classif Mask Masker Repeat Finder Masker Augustus GeneScan Danio rerio Oryzias Latines	ication of 1 Rep 992 594 665 937 131 Number 23,132 21,379 25,390 25,319	repeat sequer peat Size(bp) 5152 8136 95756 26009 090229 Average Transcript Length (bp) 4,897.24 17,213.49 7,156.92 6,411.36	Average CDS Length (bp) 1,264.61 1,973.56 1,312.32 1,194.58	assembled % of Gene 1.85 1.11 12.41 17.47 24.43 Average Exons Per Gene 5.78 10.22 6.17 5.89	genome. Dme Average Exons Length (bp) 218.81 193.05 212.62 202.73	Average Intron Length (bp) 760.04 1,652.41 1,129.99 1,066.29
Table 2.TypeProteinMRepeatMTandemDe novoTotalMethodDe novoHomolog	Detailed classif	ication of 1 Rep 992 594 665 937 131 Number 23,132 21,379 25,390 25,319	repeat sequer peat Size(bp) 5152 8136 95756 26009 090229 Average Transcript Length (bp) 4,897.24 17,213.49 7,156.92 6,411.36 7 990 91	Average CDS Length (bp) 1,264.61 1,973.56 1,312.32 1,194.58 1,759.17	assembled % of Geno 1.85 1.11 12.41 17.47 24.43 Average Exons Per Gene 5.78 10.22 6.17 5.89	genome. Dme Average Exons Length (bp) 218.81 193.05 212.62 202.73	Average Intron Length (bp) 760.04 1,652.41 1,129.99 1,066.29
Table 2.TypeProteinMRepeatMTandemDe novoTotalMethodDe novoHomolog	Detailed classif	ication of 1 Rep 992 594 665 937 131 Number 23,132 21,379 25,390 25,319 16,563	repeat sequer peat Size(bp) 5152 8136 95756 26009 090229 Average Transcript Length (bp) 4,897.24 17,213.49 7,156.92 6,411.36 7,990.91	Average CDS Length (bp) 1,264.61 1,973.56 1,312.32 1,194.58 1,759.17	assembled % of Gene 1.85 1.11 12.41 17.47 24.43 Average Exons Per Gene 5.78 10.22 6.17 5.89 11.59	genome. Dme Average Exons Length (bp) 218.81 193.05 212.62 202.73 151.75	Average Intron Length (bp) 760.04 1,652.4 1,129.99 1,066.29 588.32

Estimated genome size (Mb) 525





Figure 3. Distribution of 4DTV distances between the clearhead icefish and tilapia. The
horizontal axis stands for the 4DTV distance corrected using the HKY model. The

432 vertical axis represents the percentage of collinear gene pairs.