

RESEARCH ARTICLE

Open Access

Whole genome comparative analysis of channel catfish (*Ictalurus punctatus*) with four model fish species

Yanliang Jiang¹, Xiaoyu Gao¹, Shikai Liu¹, Yu Zhang¹, Hong Liu¹, Fanyue Sun¹, Lisui Bao¹, Geoff Waldbieser² and Zhanjiang Liu^{1*}

Abstract

Background: Comparative mapping is a powerful tool to study evolution of genomes. It allows transfer of genome information from the well-studied model species to non-model species. Catfish is an economically important aquaculture species in United States. A large amount of genome resources have been developed from catfish including genetic linkage maps, physical maps, BAC end sequences (BES), integrated linkage and physical maps using BES-derived markers, physical map contig-specific sequences, and draft genome sequences. Application of such genome resources should allow comparative analysis at the genome scale with several other model fish species.

Results: In this study, we conducted whole genome comparative analysis between channel catfish and four model fish species with fully sequenced genomes, zebrafish, medaka, stickleback and *Tetraodon*. A total of 517 Mb draft genome sequences of catfish were anchored to its genetic linkage map, which accounted for 62% of the total draft genome sequences. Based on the location of homologous genes, homologous chromosomes were determined among catfish and the four model fish species. A large number of conserved syntenic blocks were identified. Analysis of the syntenic relationships between catfish and the four model fishes supported that the catfish genome is most similar to the genome of zebrafish.

Conclusion: The organization of the catfish genome is similar to that of the four teleost species, zebrafish, medaka, stickleback, and *Tetraodon* such that homologous chromosomes can be identified. Within each chromosome, extended syntenic blocks were evident, but the conserved syntenies at the chromosome level involve extensive inter-chromosomal and intra-chromosomal rearrangements. This whole genome comparative map should facilitate the whole genome assembly and annotation in catfish, and will be useful for genomic studies of various other fish species.

Keywords: Catfish, Genome, Comparative mapping, Linkage mapping, Conserved synteny

Background

With the advances of next generation sequencing technology, genomic resources are rapidly expanding, even for non-model species. Among teleost species, whole genomes of five model species have been fully sequenced and assembled, including zebrafish (*Danio rerio*) (<http://www.ensembl.org>), fugu (*Fugu rubripes*) [1], *Tetraodon*

(*Tetraodon nigroviridis*) [2], medaka (*Oryzias latipes*) [3,4] and three-spined stickleback (*Gasterosteus aculeatus*) [5]. Among aquaculture fish species, whole genome reference sequence has been only published for Atlantic cod [6], although genomes of many aquaculture species have been or are being sequenced. In recent years, great efforts on generating genomic resources have been made for economically important aquaculture species [7], such as Atlantic salmon [8-11], European sea bass [12-16], tilapia [17-22], rainbow trout [23-28], gilthead sea bream [29,30], and catfish (for reviews, see [31,32]). These genomic resources included expressed sequence tags (ESTs),

* Correspondence: liuzhan@auburn.edu

¹The Fish Molecular Genetics and Biotechnology Laboratory, Department of Fisheries and Allied Aquacultures, Program of Cell and Molecular Biosciences, Aquatic Genomics Unit, 203 Swingle Hall, Auburn University, Auburn, AL 36849, USA

Full list of author information is available at the end of the article

BAC end sequences, physical maps, genetic linkage maps, and radiation hybrid maps.

In the absence of whole genome sequences for most aquaculture species, comparative genomic analysis is useful. Comparative mapping allows identification of evolutionarily conserved chromosomal regions, i.e., conserved syntenies, which facilitate the understanding of genome organization, rearrangement, duplication and evolution [33-37]. Moreover, the conserved syntenies provide physical evidence for orthologies and genome annotation, which is particularly important when dealing with multi-gene families [38,39]. Comparative genome analysis can also enhance the efficiency for the identification of candidate genes controlling production traits of interest, when coupled with quantitative trait loci (QTL) mapping analysis [7].

Comparative mapping was initially demonstrated by Fujiyama et al. [40] for constructing the human-chimpanzee comparative map using chimpanzee BAC end sequences to hit against human genome sequences. Putative orthologues were identified between these two closely related species. Later on, comparative mapping was extensively performed among mammals, such as the construction of human-cattle [41], human-porcine [42], human-horse [43] and human-sheep [44] comparative maps. High percentage of BLAST hits and/or high level of genome colinearity made the comparative mapping successful [45]. However, whole genome comparative mapping in most teleost species is still limited due to lacking of genomic resources.

Channel catfish, *Ictalurus punctatus*, is the predominant aquaculture species in the United States. To gain understanding of the catfish genome, considerable efforts have been made toward the development of genomic resources, including genetic linkage maps [46-49], large-insert libraries [50,51], physical maps [52,53], BAC end sequences [45,54], a large number of Sanger sequenced ESTs from various tissues and developmental stages [55-59], full length cDNAs [60], RNA-Seq transcriptome assemblies [61,62], and whole genome shotgun sequence reads (unpublished). Such genomic resources provided a foundation for comparative analysis. For instance, Wang et al. [51] utilized catfish BAC end sequences to compare with zebrafish and *Tetraodon* genome, and identified conserved syntenic regions in the catfish genome. More recently, Liu et al. [45] conducted comparative mapping analysis by using a large number of BAC end sequences. Genetic linkage map containing type I gene-associated markers was also used for comparative analysis [46]. With next-generation sequencing data, Jiang et al. [63] conducted comparative analysis between an approximately 1 Mb DNA region in catfish genome with other model fish species. Recently, one catfish linkage group (chromosome) was compared with model fish [64]. These studies allowed for identification of

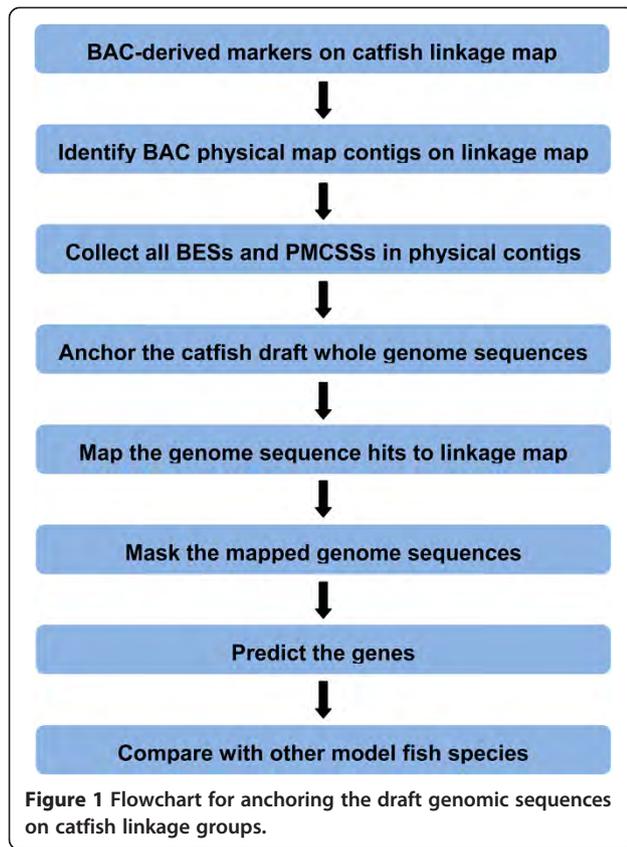
conserved syntenies in the catfish genome as compared with other sequenced fish genomes. In these studies, however, only a small number of gene markers or only one chromosome was used for comparative analysis.

To obtain detailed comparative information at the genome level, whole genome comparative analysis is much needed. We report here the whole genome comparative analysis of catfish with four model fish species, zebrafish, medaka, *Tetraodon* and stickleback, utilizing all currently available catfish genomic resources. With the whole genome comparative mapping, homologous chromosomes were identified and a large number of conserved syntenies were identified.

Results

Identification of genomic sequences mapped on the catfish linkage map

Comparative sequence analysis of species with fully sequenced genomes is relatively straight-forward. However, it is difficult for most teleost fish such as catfish whose whole genome sequence assembly is not yet available. In such cases, one of the key steps for genome-scale comparative analysis is to identify whole genome sequences that are anchored on the linkage map. For catfish, the sequence-tagged markers existing on the linkage map are the BAC-derived microsatellite markers [48]. Therefore, as shown in Figure 1, this study starts with BAC-derived microsatellites on the linkage map, followed by several steps utilizing available genome resources including the integrated genetic linkage and physical map [48], BAC end sequences [45,54], physical map contig-specific sequences (PMCSSs) [65], and the anchored catfish genome scaffolds (unpublished data, Table 1). A total of 2,099 BAC end sequence-derived microsatellite markers were mapped on the catfish linkage map ([48], Table 2). Based on these marker-associated BACs, a total of 931 physical map contigs were linked with the linkage map. A total of 32,500 available BAC end sequences were identified from the BAC clones of the 931 physical map contigs. In addition, a total of 57,861,110 physical map contig-specific sequences were identified ([65], Table 2). Both BAC end sequences and physical map contig-specific sequences were used to anchor the draft catfish whole genome sequences to catfish genetic linkage map. Taken together, a total of 517 Mb (62%) draft catfish genome sequences were anchored to the linkage map (Table 2). As described in another study [65], PMCSSs are short specific sequences randomly distributed in each catfish physical map contig, which can serve as anchor points to map the draft catfish genome sequences to the linkage group. Without PMCSSs, only 26% of the draft genome sequences can be map onto linkage group. A number of the draft genome sequences were stacked together due to the relatively low resolution of the



catfish linkage map. In those cases, the gene orders and orientations were not resolved, though the relatively large-scale chromosome locations were known when examined at a “zoomed out” fashion.

Identification of homologous genes

The 517 Mb genome sequences retrieved from the draft catfish genome scaffolds were used for further comparative genome analysis. Genes located in these sequences were identified by BLASTX search against ENSEMBL protein database, including protein sequences from zebrafish, medaka, *Tetraodon*, and stickleback. Homologous genes in these species were identified as summarized in Table 2. The largest number of homologous genes (14,035) was found in zebrafish genome. Of the 14,035 homologous genes, 13,784 genes have chromosome information based on current zebrafish genome annotation in

Table 1 Summary of statistics of anchored scaffolds

No. of scaffolds	62,461
N50 of scaffolds (bp)	3,016,365
Total span (Mb)	832
No. of the anchored scaffolds	41,061
N50 of the anchored scaffolds (bp)	1,935,203
Total length of the anchored scaffolds (Mb)	517

Table 2 Summary of genome resources used to anchor genes to catfish linkage groups

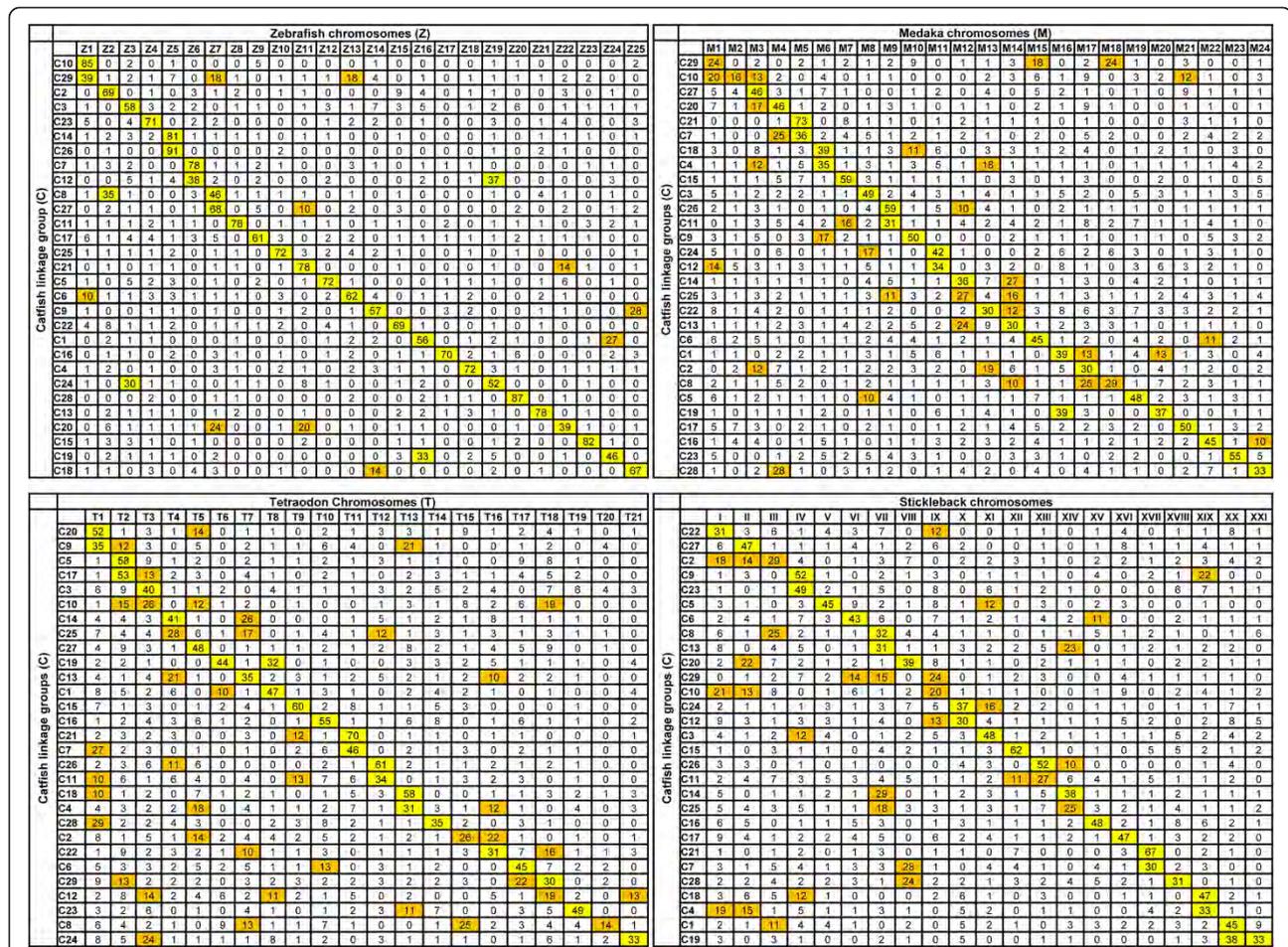
Category	Number
BAC-associated markers in linkage map	2,099
BAC contigs containing the BAC-associated markers	931
All BAC-end sequences (BES) from mapped BAC contigs	32,500
Physical map contig-specific reads (PMCSS)	57,861,110
Total length of mapped draft genome contigs (Mb)	517 (62%)
Unique medaka genes with mapped genome contig hits	9,949
Medaka gene hits mapped to chromosomes	9,036
Unique <i>Tetraodon</i> genes with mapped genome contig hits	9,920
<i>Tetraodon</i> gene hits mapped to chromosomes	7,181
Unique stickleback genes with mapped genome contig hits	10,430
Stickleback gene hits mapped to chromosomes	9,465
Unique zebrafish genes with mapped genome contig hits	14,035
Zebrafish gene hits mapped to chromosomes	13,784

ENSEMBL (Table 2). A total of 9,949 homologous genes were identified in medaka genome. Of which, 9,036 genes were mapped on the chromosomes of medaka genome. Similar numbers of homologous genes were identified from *Tetraodon* and stickleback genome, with 7,181 and 9,465 being mapped to the chromosomes, respectively (Table 2).

Identification of homologous chromosomes

The identified homologous genes were used to determine the homologous chromosomes. All the genes identified for each of the catfish linkage group were used as queries to search against the protein sequences of the four model fish species. The homologous chromosomes for each catfish linkage group were determined based on the percentages of catfish genes that had significant hits with genes from the corresponding chromosomes of the species used for comparative analysis. For instance, 85% of the catfish genes in linkage group 10 had significant hits to genes located in zebrafish chromosome 1 (Figure 2), and therefore, catfish LG10 is the homologous chromosome of zebrafish chromosome 1. As shown in Figure 2, the relationships between catfish linkage groups and chromosomes of the four model fish species were established.

Eleven catfish linkage groups and zebrafish chromosomes had a one-to-one homologous relationship. These linkage groups included LG4, LG5, LG11, LG13, LG15, LG16, LG17, LG22, LG23, LG25 and LG28. Of all 29 catfish linkage groups, 17 linkage groups were found to be homologous to a single chromosome in zebrafish. Of the 17 catfish linkage groups, five are extremely highly conserved with over 81-91% of their genes shared between the catfish linkage groups and the zebrafish chromosomes. The catfish LG26 and zebrafish chromosome



29,305,847 to position 43,867,471 (Additional file 1). In catfish, as the whole genome assembly is not yet available, our analysis is limited to locate multiple genes within a single scaffold, followed by the analysis of the physical map and linkage map positions of the involved scaffolds. As shown in Additional file 1, many genes located in the same zebrafish genomic segments were also located in a single scaffold of the catfish draft genome sequence, and these scaffolds were mapped to similar locations on the linkage map.

Chromosome level conservation was the highest between catfish and zebrafish followed by stickleback, medaka, and *Tetraodon*. As shown in Figure 2, the one-to-one chromosome relationship was also observed between catfish and medaka, catfish and *Tetraodon*, and catfish and stickleback, but apparently at a lower level as compared with the situation between catfish and zebrafish. This was reflected at two levels. First, the percentage of homologous genes with a one-to-one relationship was much lower between catfish and medaka, catfish and *Tetraodon*, and catfish and stickleback as compared to catfish and zebrafish. Second, the level of chromosome rearrangements was much greater between catfish and medaka, catfish and *Tetraodon*, and catfish and stickleback as compared with catfish and zebrafish (Figure 2). The lowest level of chromosomal conservation was between catfish and *Tetraodon*.

Identification of conserved syntenic blocks between catfish and zebrafish

Conserved synteny, a block of genes in certain chromosome region in which the content and order are relatively conserved, is an interesting phenomenon in genome evolution [66]. Establishing the conserved syntenies by conducting comparative analysis between species is not only valuable for genome assembly and annotation, but also valuable for functional and evolutionary genomics studies such as gene duplication and chromosome rearrangement [4,67]. To gain a close insight into the conserved chromosomal segments, conserved syntenies were examined between catfish and zebrafish in this study. As shown in Table 3, a total of 1,943 conserved syntenic blocks were identified, spanning approximately 703 Mb. The average size of the conserved synteny is 362 Kb. A total of 10,876 homologous genes were involved, with an average of 6 genes per syntenic block. For each catfish linkage group, the number of conserved syntenic blocks ranged from 32 to 105, spanning genomic region from 6 Mb to 39 Mb, with the number of genes varying from 126 to 577 (Table 3). In most cases, the number of syntenic blocks, spanning lengths, and the number of genes located in these syntenic blocks were highly correlated. The more highly conserved chromosomes (linkage groups) tend to have greater number of syntenic blocks identified, with extended spanning lengths involved in a larger number of

genes (Table 3). Detailed information for the identified syntenic blocks is provided in Additional file 2.

Chromosomal level structural conservations

To gain detailed understanding of evolutionary relationship between catfish and zebrafish, a comparative map was constructed between the catfish linkage groups with their homologous chromosomes in zebrafish. Only gene sequences were used for this comparative analysis because gene sequences are more conserved than sequences in intergenic regions. The positions of physical map contigs were determined in the linkage group based on the locations of BES-associated markers. However, the positions and orders of genes within each physical contig cannot be determined because of the incompletely assembled genome sequences. In addition, a number of catfish genes were stacked because of the low resolution provided by the current genetic linkage map.

A comparative map of catfish linkage group 13 to the corresponding zebrafish chromosome 21 is shown in Figure 3. A total of 447 genes on catfish LG13 spanned a genetic distance 109 cM [48]. The homologous genes of the 447 genes on catfish LG13 are distributed almost across the entire zebrafish chromosome 21, from position 64 Kb to 44,353 Kb (Figure 3). The results indicated a high level of genome-level conservation existing between the catfish linkage group and the zebrafish chromosome. However, as shown in Figure 3, numerous chromosome rearrangements were involved during evolution.

Similar results were observed in the other 28 catfish linkage groups, with general large scale chromosome level of genome conservation, but with numerous chromosome breaks, shuffling and rearrangements, which are consistent with previous studies [46,51].

Discussion

Rapid development of genomic resources in fish species has provided the opportunity for comparative genome analysis, shedding lights on the structure, organization, function and evolution of vertebrate genomes. In this study, we conducted the whole genome comparative analysis of channel catfish, an important aquaculture species, with several model fish species. By comparing with other fully sequenced model fish species including zebrafish, medaka, stickleback, and *Tetraodon*, homologous chromosomes among these species were determined and a large number of conserved syntenies were identified, providing valuable information for whole genome assembly and annotation in catfish, and for comparative genome analysis of other teleost species.

Comparative map is a powerful tool in genomics studies, especially for non-model species, by transferring the genomic information from well-studied model species.

Table 3 Summary of conserved syntenic blocks between catfish linkage groups and zebrafish chromosomes

Group	Total length spanned (Kb)	No. of syntenies	Ave. size of the synteny (Kb)	Max. size of the synteny (Kb)	N50 of the synteny (Kb)	No. of genes involved	Ave. gene number/synteny
LG1	39,522	105	361	1,997	1,276	577	6
LG2	11,105	54	206	942	340	202	5
LG3	26,165	93	281	1,475	519	439	5
LG4	21,282	65	327	2,091	854	335	6
LG5	36,292	80	454	3,189	964	503	8
LG6	36,722	102	360	2,835	810	570	7
LG7	21,884	48	456	2,341	984	319	8
LG8	26,601	92	289	3,269	631	434	5
LG9	27,314	66	414	3,702	932	431	8
LG10	15,589	41	380	3,247	709	277	9
LG11	26,953	72	374	4,995	1,121	475	8
LG12	22,483	79	285	2,731	521	389	6
LG13	29,613	78	380	2,363	707	472	8
LG14	13,468	43	313	2,631	749	231	6
LG15	27,758	78	356	2,466	632	453	8
LG16	36,037	105	343	1,803	684	535	5
LG17	35,452	92	385	1,651	658	464	6
LG18	19,451	68	286	1,536	526	333	6
LG19	20,872	36	580	3,201	1,604	269	9
LG20	31,788	73	435	3,800	844	456	9
LG21	23,200	56	414	3,730	835	315	6
LG22	28,266	69	410	2,233	842	431	7
LG23	11,486	47	244	977	465	221	5
LG24	24,625	46	535	2,529	1,075	363	10
LG25	23,276	79	295	1,539	759	407	6
LG26	18,055	46	393	3,541	803	256	7
LG27	15,045	35	430	2,465	1,103	204	7
LG28	26,602	63	422	2,382	1,010	389	8
LG29	5,860	32	183	705	345	126	5
Total	702,766	1,943				10,876	

Comparative map not only allows better understanding genome arrangement during evolution, but also benefits the discovery or confirmation of orthologies among species.

Without a well assembled whole genome sequences, comparative genome analysis can be achieved by using various other genomic resources containing information for genome level conservation. Markers with low levels of conservation in evolution have very limited value for comparative genome analysis, and most often only for very closely related species. For instance, when microsatellite markers on linkage maps were used for comparative analysis, a small number of microsatellites could be successfully mapped, indicating relative low levels of

conservation of microsatellites which were derived from non-coding regions of the genome [68]. Comparative analysis among different species using gene-derived markers on linkage maps was more effective because genes are well conserved through evolution [46]. Higher resolution of comparative maps can be achieved by using integrated physical and genetic linkage maps with BAC end sequences. For instance, Zhang et al. [64] conducted comparative analysis of one catfish chromosome (linkage group 8) with four model fish species utilizing catfish linkage map, physical map, BAC end sequences and draft genome sequences. In that work, 287 unique genes were identified, and a number of conserved syntenies were identified. Although that work demonstrated the

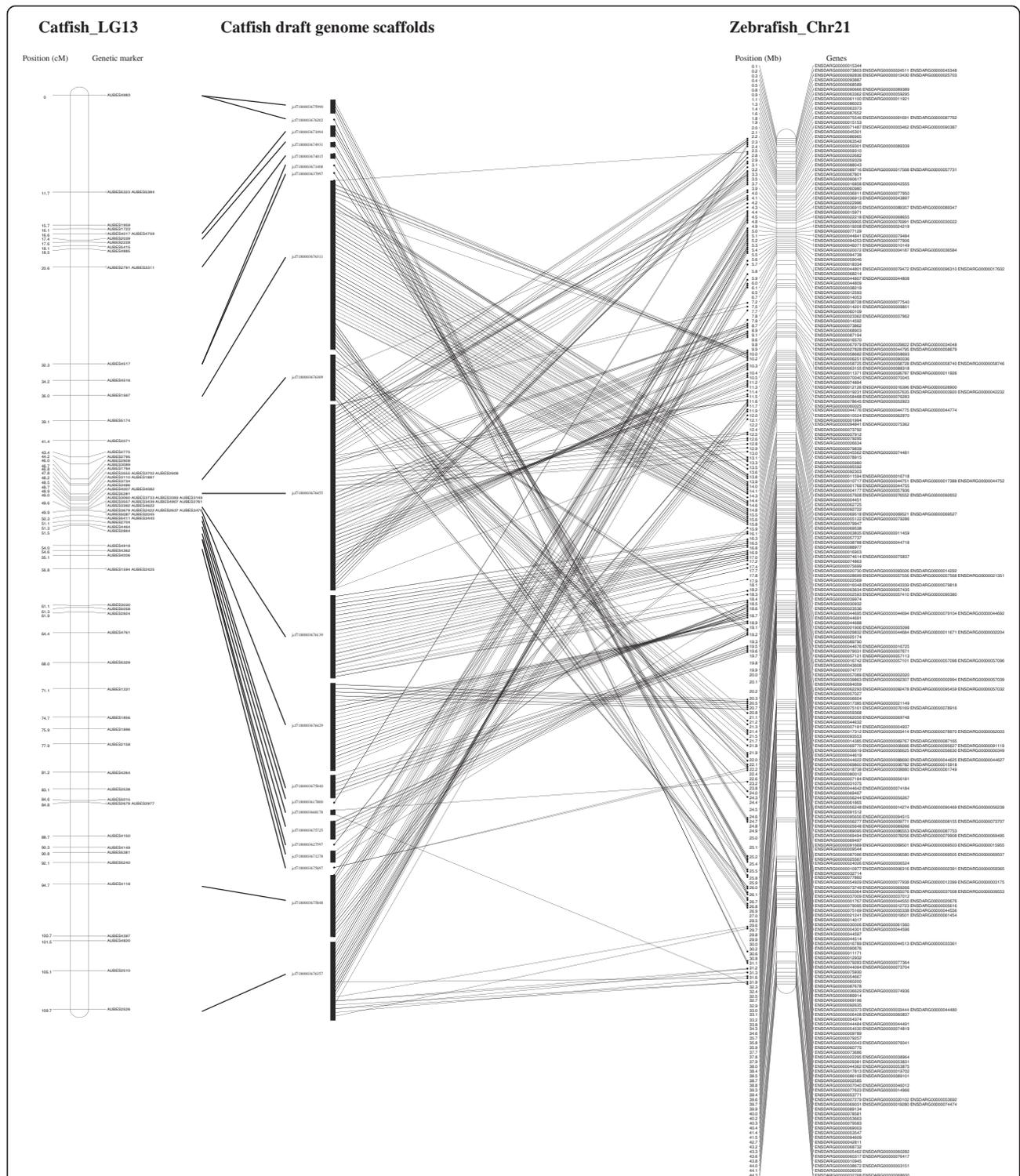


Figure 3 Comparative map of catfish LG13 and zebrafish chromosome 21. The catfish LG13 is presented on the left panel while the zebrafish chromosome 21 is on the right panel. The gene-associated catfish draft genome contigs (scaffolds) is presented in the middle panel. For catfish LG13, genetic linkage position is indicated in cM on the left of the bar, and the BAC-derived microsatellite markers are indicated on the right of the bar. For zebrafish chromosome 21, gene locations along the chromosome are indicated in Mb on the left while gene names are indicated on the right.

utilities of linkage maps when integrated with physical maps with BAC end sequences, the ability to establish whole genome comparative map was hindered by the lack of internal BAC sequences. In order to increase the power of comparative genome analysis, we recently developed one additional valuable genome resource, the physical map contig-specific sequences [65]. In this study, we used all the existing genome resources of catfish for whole genome comparative analysis. The ability to identify long conserved syntenic blocks was much enhanced. For instance, with the same linkage group, LG8, as used in Zhang et al. [64], we were able to identify 585 unique genes, more than the double of the 287 genes identified in the previous study [64], and the size of syntenic blocks were much increased.

The strategy developed in this study allowed whole genome comparative analysis of catfish conducted without a well-assembled whole genome sequence. However, lacking of a continuous reference sequences, the order and orientation of catfish draft genome contigs/scaffolds within the same catfish physical contig cannot be determined at present. The reason for this inability is the low resolution of the linkage map. Because of the low resolution of the linkage map, many genes are mapped directly or indirectly to the same map location, forming stacks of sequence contigs and scaffolds with gene orders and orientations undetermined. Therefore, it is imperative to develop linkage maps with many more markers or using high density marker such as SNPs in the future, and more importantly with high resolution by using large resource families.

Comparative analysis was conducted based on the similarities between gene sequences in catfish and the homologous genes in the genomes of other model fish species. Only genes were used for analysis in this study because gene sequences are more conserved than intergenic sequences. The largest number of homologous genes was found in zebrafish among four model fish species, with 14,035 genes, followed by stickleback with 10,430 genes, medaka with 9,949 genes and *Tetraodon* with 9,920 genes. This difference of the homologous genes identified in different model fish species may be resulted from two reasons: First, the quality of the reference genome. For instance, of all annotated 32,574 zebrafish genes, only 1,225 (3.7%) are unmapped onto zebrafish chromosomes or mitochondria, while in *Tetraodon*, there are 6,487 (31.5%) of all 20,562 genes cannot be mapped (Ensembl database). Second, the phylogenetic relationship between catfish and these model fish species determined that zebrafish is the most closely species to catfish [63,69-71].

High levels of chromosomal conservations were observed between catfish and the four fish species. However, due to the difference of chromosome numbers

among those fish species, e.g. catfish has 29 chromosomes, while zebrafish has 25 chromosomes, *Tetraodon* has 21 chromosomes, medaka has 24 chromosomes and stickleback has 21 chromosomes, chromosome breakage or fusion would have occurred during evolution. For instance, 33% of genes identified on catfish LG19 were found to be homologous in zebrafish chr.16, and 46% of genes were found to be homologous in zebrafish chr.24 (Figure 2), suggesting catfish LG19 have been created by fusion of chromosomal segments similar to zebrafish chr.16 and chr.24, or inversely the two zebrafish chromosomes have been created by split of the chromosome similar to catfish LG19. Similar cases can be observed between the comparison of catfish and medaka, catfish and *Tetraodon*, catfish and stickleback, indicated that chromosomal fusions or splits occurred frequently during the teleost evolution.

Sarropoulou et al. [7] conducted a comparative study in which the syntenic relationship between six non-model fish species genomes were established by using ESTs and microsatellites sequences. Our study here extended that study by adopting a much larger numbers of genes. For instance, catfish LG15 was identified to be the homologous chromosome of medaka chr.7 (C7), *Tetraodon* chr.9 (T9), and stickleback Grp. XII (SXII), respectively, which indicated that C7, T9 and SXII were homologous chromosomes to one another. This was consistent with the results of Sarropoulou et al. [7]. Similarly, catfish LG9 (C9) corresponded to *Tetraodon* T1, T2 and T13 in this study, while T1 was reported to be homologous chromosome to medaka M10 [7], which also had the highest percentage homologous gene hits to catfish LG9 in our study.

Because zebrafish is the most closely related model fish to catfish, detailed comparative analyses were conducted between them. Catfish has 29 pairs of chromosomes while zebrafish has 25 pairs of chromosomes. Therefore, some zebrafish chromosomes are expected to be homologous to greater than one chromosome in catfish. This was found with several chromosomes (Figure 2). However, the opposite situation was also found with one catfish chromosome being homologous to several zebrafish chromosomes (Figure 2), suggesting extensive chromosome rearrangements during evolution.

A large number of conserved syntenic blocks between catfish and zebrafish were established. Analysis of the conserved syntenies should greatly benefit genome annotation in catfish. This is particularly true when dealing with large gene families and duplicated genes. As reported by Liu et al. [39], identities of genes involved in large gene families such as the ABC transporter gene families sometimes cannot be resolved by phylogenetic analysis alone. Syntenic analysis is essential to provide orthologous information for the identification of such

genes. The inferred orthologies are important not only for the identification and annotation of genes, but also for functional inference based on orthologies [64]. Apparently, catfish genome is well conserved at the chromosomal level with those in other model fish species. However, local chromosome shuffling and rearrangements are extensive (Figure 3). Our whole genome comparative analysis with four teleost species also indicated extensive inter-chromosomal rearrangements during evolution, consistent with the hypothesis that inter-chromosomal rearrangements were increased after whole genome duplication in the teleost lineage [67,72].

Conclusions

Whole genome comparative analysis of channel catfish was conducted by utilizing currently available catfish genomic resources including genetic linkage map, physical map, BAC end sequences, physical map contig-specific sequences, and the draft whole genome sequences. Homologous genes and homologous chromosomes of catfish as compared with four fully sequenced fish species were identified based on sequence similarities and arrangements of homologous genes along the chromosomes. Detailed comparative analysis between catfish and zebrafish allowed for the establishment of a large number of conserved synteny, with some being extended in large sizes. The whole genome comparative analysis should facilitate whole genome sequence assembly and annotation, as well as providing insight into genome evolution.

Methods

Anchorage of draft genomic sequences on catfish linkage groups

Various currently available catfish genome resources were utilized in this study, including the genetic linkage map [48], BAC-derived microsatellite markers [48], BAC-based physical map [53], BAC end sequences [45,54], and draft genome sequences (Unpublished data). As shown in Figure 1, the steps to anchor draft genomic contigs on catfish linkage groups are: 1) Starting with all BAC-derived microsatellite markers on the catfish linkage map; 2) Using the markers to identify BAC end sequences from which they were derived; 3) Using the BAC end sequences to determine the physical map contigs mapped on the linkage groups; 4) Collecting all BAC end sequences and the physical map contig-specific sequences from all mapped physical contigs; 5) Using the assembly of BAC end sequences and physical map contig-specific sequences to search for corresponding whole genome draft sequence contigs and scaffolds using BLAST. All identified draft genome sequences and their respective chromosome (linkage group) locations were used for further comparative analysis.

Identification of homologous genes

Before the gene identification, RepeatMasker (Version 3.2.7, <http://www.repeatmasker.org/>) was used to mask repetitive-elements within the draft genome sequences. The repeat-masked sequences were then used as query for BLASTX searches against the ENSEMBL protein database of other fully sequenced model fish species, including zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), stickleback (*Gasterosteus aculeatus*) and *Tetraodon* (*Tetraodon nigroviridis*), with an E-value cutoff of $1e-10$. Gene annotation information was retrieved using BioMart (www.ensembl.org/biomart/martview) with ENSEMBL protein ID.

Identification of homologous chromosomes

The genome locations of homologous genes from zebrafish, medaka, stickleback and *Tetraodon* were obtained by using BioMart with respective ENSEMBL gene IDs. The homologous chromosomes corresponding to each catfish linkage group (chromosome) from zebrafish, medaka, *Tetraodon* and stickleback were determined as the chromosome with a majority of homologous genes. In cases where significant fractions (more than 10%) of genes were located on several chromosomes, all these chromosomes were determined to contain homologous genomic segments.

Identification of conserved synteny

Conserved synteny were defined as preserved co-localization of genes on chromosomes from different species. Conserved synteny were identified based on the genetic locations of BAC-derived microsatellite markers and their associated genes on the linkage map and the genome positions of homologous genes from other model fish species. In this study, conserved synteny were established when at least two adjacent genes on the model fish chromosome were found within a single contig of the draft catfish genome or scaffold.

Comparative maps between each catfish linkage group with the homologous chromosome from zebrafish were conducted by using MapChart [73]. The genes within a physical map contig were located on catfish linkage group based on the position of BAC-derived microsatellite markers [48]. The comparative maps were drawn based the gene position on catfish linkage map and the position of their homologous genes from zebrafish chromosomes.

Additional files

Additional file 1: Homologous chromosome relationships with homologous genes between catfish and zebrafish. The conserved syntenic blocks are highlighted.

Additional file 2: The detail information of conserved syntenic blocks identified between catfish and zebrafish. The conserved syntenic blocks are highlighted.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

YJ conducted the major part of the project including data analysis and manuscript preparation. XG, SL, YZ, HL, FS and LB assisted in data analysis, GW involved in generation of catfish genome sequence resources used in this study. ZL designed and supervised the project, and finalized manuscript. All authors read and approved the manuscript.

Acknowledgments

This project was supported by Agriculture and Food Research Initiative Competitive Grants with grant no. 2010-65205-20356 and 2012-67015-19410 from the USDA National Institute of Food and Agriculture (NIFA).

Author details

¹The Fish Molecular Genetics and Biotechnology Laboratory, Department of Fisheries and Allied Aquacultures, Program of Cell and Molecular Biosciences, Aquatic Genomics Unit, 203 Swingle Hall, Auburn University, Auburn, AL 36849, USA. ²USDA-ARS Catfish Genetics Research Unit, 141 Experiment Station Road, Stoneville, MS 38776, USA.

Received: 19 May 2013 Accepted: 28 October 2013

Published: 11 November 2013

References

1. Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, et al: **Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes***. *Science* 2002, **297**(5585):1301–1310.
2. Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, et al: **Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype**. *Nature* 2004, **431**(7011):946–957.
3. Ahsan B, Kobayashi D, Yamada T, Kasahara M, Sasaki S, Saito TL, Nagayasu Y, Doi K, Nakatani Y, Qu W, et al: **UTGB/medaka: genomic resource database for medaka biology**. *Nucleic Acids Res* 2008, **36**(Database issue):D747–D752.
4. Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y, et al: **The medaka draft genome and insights into vertebrate genome evolution**. *Nature* 2007, **447**(7145):714–719.
5. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, et al: **The genomic basis of adaptive evolution in three spine sticklebacks**. *Nature* 2012, **484**(7392):55–61.
6. Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrom M, Gregers TF, Rounge TB, Paulsen J, Solbakken MH, Sharma A, et al: **The genome sequence of Atlantic cod reveals a unique immune system**. *Nature* 2011, **477**(7363):207–210.
7. Sarrpoulou E, Nousdili D, Magoulas A, Kotoulas G: **Linking the genomes of nonmodel teleosts through comparative genomics**. *Mar Biotechnol (NY)* 2008, **10**(3):227–233.
8. Lien S, Gidskehaug L, Moen T, Hayes BJ, Berg PR, Davidson WS, Omholt SW, Kent MP: **A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns**. *BMC Genomics* 2011, **12**:615.
9. Moen T, Hayes B, Baranski M, Berg PR, Kjolglum S, Koop BF, Davidson WS, Omholt SW, Lien S: **A linkage map of the Atlantic salmon (*Salmo salar*) based on EST-derived SNP markers**. *BMC Genomics* 2008, **9**:223.
10. Ng SH, Artieri CG, Bosdet IE, Chiu R, Danzmann RG, Davidson WS, Ferguson MM, Fjell CD, Hoyheim B, Jones SJ, et al: **A physical map of the genome of Atlantic salmon, *Salmo salar***. *Genomics* 2005, **86**(4):396–404.
11. Rise ML, von Schalburg KR, Brown GD, Mawer MA, Devlin RH, Kuipers N, Busby M, Beetz-Sargent M, Alberto R, Gibbs AR, et al: **Development and application of a salmonid EST database and cDNA microarray: data mining and interspecific hybridization characteristics**. *Genome Res* 2004, **14**(3):478–490.
12. Chistiakov DA, Hellemans B, Haley CS, Law AS, Tsigonopoulos CS, Kotoulas G, Bertotto D, Libertini A, Volckaert FA: **A microsatellite linkage map of the European sea bass *Dicentrarchus labrax* L.** *Genetics* 2005, **170**(4):1821–1826.
13. Chistiakov DA, Tsigonopoulos CS, Lagnel J, Guo YM, Hellemans B, Haley CS, Volckaert FA, Kotoulas G: **A combined AFLP and microsatellite linkage map and pilot comparative genomic analysis of European sea bass *Dicentrarchus labrax* L.** *Anim Genet* 2008, **39**(6):623–634.
14. Guyon R, Senger F, Rakotomanga M, Sadequi N, Volckaert FA, Hitte C, Galibert F: **A radiation hybrid map of the European sea bass (*Dicentrarchus labrax*) based on 1581 markers: Synteny analysis with model fish genomes**. *Genomics* 2010, **96**(4):228–238.
15. Kuhl H, Beck A, Wozniak G, Canario AV, Volckaert FA, Reinhardt R: **The European sea bass *Dicentrarchus labrax* genome puzzle: comparative BAC-mapping and low coverage shotgun sequencing**. *BMC Genomics* 2010, **11**:68.
16. Whitaker HA, McAndrew BJ, Taggart JB: **Construction and characterization of a BAC library for the European sea bass *Dicentrarchus labrax***. *Anim Genet* 2006, **37**(5):526.
17. Guyon R, Rakotomanga M, Azzouzi N, Coutanceau JP, Bonillo C, D'Cotta H, Pepey E, Soler L, Rodier-Goud M, D'Hont A, et al: **A high-resolution map of the Nile tilapia genome: a resource for studying cichlids and other percomorphs**. *BMC Genomics* 2012, **13**:222.
18. Katagiri T, Kidd C, Tomasino E, Davis JT, Wishon C, Stern JE, Carleton KL, Howe AE, Kocher TD: **A BAC-based physical map of the Nile tilapia genome**. *BMC Genomics* 2005, **6**:89.
19. Lee BY, Howe AE, Conte MA, D'Cotta H, Pepey E, Baroiller JF, di Palma F, Carleton KL, Kocher TD: **An EST resource for tilapia based on 17 normalized libraries and assembly of 116,899 sequence tags**. *BMC Genomics* 2010, **11**:278.
20. Lee BY, Lee WJ, Streebman JT, Carleton KL, Howe AE, Hulata G, Slettan A, Stern JE, Terai Y, Kocher TD: **A second-generation genetic linkage map of tilapia (*Oreochromis spp.*)**. *Genetics* 2005, **170**(1):237–244.
21. Soler L, Conte MA, Katagiri T, Howe AE, Lee BY, Amemiya C, Stuart A, Dossat C, Poulain J, Johnson J, et al: **Comparative physical maps derived from BAC end sequences of tilapia (*Oreochromis niloticus*)**. *BMC Genomics* 2010, **11**:636.
22. Van Bers NE, Crooijmans RP, Groenen MA, Dibbits BW, Komen J: **SNP marker detection and genotyping in tilapia**. *Mol Ecol Resour* 2012, **12**(5):932–941.
23. Genet C, Dehais P, Palti Y, Gao G, Gavory F, Wincker P, Quillet E, Boussaha M: **Analysis of BAC-end sequences in rainbow trout: content characterization and assessment of synteny between trout and other fish genomes**. *BMC Genomics* 2011, **12**:314.
24. Guyomard R, Mauger S, Tabet-Canale K, Martineau S, Genet C, Krieg F, Quillet E: **A type I and type II microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) with presumptive coverage of all chromosome arms**. *BMC Genomics* 2006, **7**:302.
25. Palti Y, Genet C, Gao G, Hu Y, You FM, Boussaha M, Rexroad CE III, Luo MC: **A second generation integrated map of the rainbow trout (*Oncorhynchus mykiss*) genome: analysis of conserved synteny with model fish genomes**. *Mar Biotechnol (NY)* 2012, **14**(3):343–357.
26. Palti Y, Genet C, Luo MC, Charlet A, Gao G, Hu Y, Castano-Sanchez C, Tabet-Canale K, Krieg F, Yao J, et al: **A first generation integrated map of the rainbow trout genome**. *BMC Genomics* 2011, **12**:180.
27. Palti Y, Luo MC, Hu Y, Genet C, You FM, Vallejo RL, Thorgaard GH, Wheeler PA, Rexroad CE III: **A first generation BAC-based physical map of the rainbow trout genome**. *BMC Genomics* 2009, **10**:462.
28. Rexroad CE III, Palti Y, Gahr SA, Vallejo RL: **A second generation genetic map for rainbow trout (*Oncorhynchus mykiss*)**. *BMC Genet* 2008, **9**:74.
29. Franch R, Louro B, Tsalavouta M, Chatziplis D, Tsigonopoulos CS, Sarrpoulou E, Antonello J, Magoulas A, Mylonas CC, Babbucci M, et al: **A genetic linkage map of the hermaphrodite teleost fish *Sparus aurata* L.** *Genetics* 2006, **174**(2):851–861.
30. Senger F, Priat C, Hitte C, Sarrpoulou E, Franch R, Geisler R, Bargelloni L, Power D, Galibert F: **The first radiation hybrid map of a perch-like fish: the gilthead sea bream (*Sparus aurata* L.)**. *Genomics* 2006, **87**(6):793–800.
31. Liu Z: **Development of genomic resources in support of sequencing, assembly, and annotation of the catfish genome**. *Comp Biochem Physiol Part D Genomics Proteomics* 2011, **6**(1):11–17.
32. Liu Z: **A review of catfish genomics: progress and perspectives**. *Comp Funct Genomics* 2003, **4**(2):259–265.
33. Woods IG, Kelly PD, Chu F, Ngo-Hazelett P, Yan Y-L, et al: **A Comparative Map of the Zebrafish Genome**. *Genome Res* 2000, **10**:1903–1914.
34. Taylor JS, Van De Peer Y, Braasch I, Meyer A: **Comparative genomics provides evidence for an ancient genome duplication event in fish**. *Philos Trans R Soc Lond B Biol Sci* 2001, **356**:1661–1679. 10.1098/rstb.2001.0975.

35. Guyomard R, Boussaha M, Krieg F, Hervet C, Quillet E: **A synthetic rainbow trout linkage map provides new insights into the salmonid whole genome duplication and the conservation of synteny among teleosts.** *BMC Genet* 2012, **13**:15.
36. Parveen N, Masood A, Iftikhar N, Minhas BF, Minhas R, et al: **Comparative genomics using teleost fish helps to systematically identify target gene bodies of functionally defined human enhancers.** *BMC Genomics* 2013, **14**:122. doi:10.1186/1471-2164-14-122.
37. Zheng X, Kuang Y, Zhang X, Lu C, Cao D, et al: **A genetic linkage map and comparative genome analysis of common carp (*Cyprinus carpio* L.) using microsatellites and SNPs.** *Mol Genet Genomics* 2011, **286**:261–277. doi:10.1007/s00438-011-0644-x.
38. Garrett M, Fullaondo A, Troxler L, Micklem G, Gubb D: **Identification and analysis of serpin-family genes by homology and synteny across the 12 sequenced *Drosophilid* genomes.** *BMC Genomics* 2009, **10**:489.
39. Liu S, Li Q, Liu Z: **Genome-Wide Identification, Characterization and Phylogenetic Analysis of 50 Catfish ATP-Binding Cassette (ABC) Transporter Genes.** *PLoS ONE* 2013, **8**(5):e63895.
40. Fujiyama A, Watanabe H, Toyoda A, Taylor TD, Itoh T, Tsai SF, Park HS, Yaspo ML, Lehrach H, Chen Z, et al: **Construction and analysis of a human-chimpanzee comparative clone map.** *Science* 2002, **295**(5552):131–134.
41. Larkin DM, Everts-van der Wind A, Rebeiz M, Schweitzer PA, Bachman S, Green C, Wright CL, Campos EJ, Benson LD, Edwards J, et al: **A cattle-human comparative map built with cattle BAC-ends and human genome sequence.** *Genome Res* 2003, **13**(8):1966–1972.
42. Meyers SN, Rogatcheva MB, Larkin DM, Yerle M, Milan D, Hawken RJ, Schook LB, Beever JE: **Piggy-BACing the human genome II. A high-resolution, physically anchored, comparative map of the porcine autosomes.** *Genomics* 2005, **86**(6):739–752.
43. Leeb T, Vogl C, Zhu B, de Jong PJ, Binns MM, Chowdhary BP, Scharfe M, Jarek M, Nordsiek G, Schrader F, et al: **A human-horse comparative map based on equine BAC end sequences.** *Genomics* 2006, **87**(6):772–776.
44. Dalrymple BP, Kirkness EF, Nefedov M, McWilliam S, Ratnakumar A, Barris W, Zhao S, Shetty J, Maddox JF, O'Grady M, et al: **Using comparative genomics to reorder the human genome sequence into a virtual sheep genome.** *Genome Biol* 2007, **8**(7):R152.
45. Liu H, Jiang Y, Wang S, Ninwichian P, Somridhvej B, Xu P, Abernathy J, Kucuktas H, Liu Z: **Comparative analysis of catfish BAC end sequences with the zebrafish genome.** *BMC Genomics* 2009, **10**:592.
46. Kucuktas H, Wang S, Li P, He C, Xu P, Sha Z, Liu H, Jiang Y, Baoprasertkul P, Somridhvej B, et al: **Construction of genetic linkage maps and comparative genome analysis of catfish using gene-associated markers.** *Genetics* 2009, **181**(4):1649–1660.
47. Liu Z, Karsi A, Li P, Cao D, Dunham R: **An AFLP-based genetic linkage map of channel catfish (*Ictalurus punctatus*) constructed by using an interspecific hybrid resource family.** *Genetics* 2003, **165**(2):687–694.
48. Ninwichian P, Peatman E, Liu H, Kucuktas H, Somridhvej B, Liu S, Li P, Jiang Y, Sha Z, Kaltenboeck L, et al: **Second-generation genetic linkage map of catfish and its integration with the BAC-based physical map.** *G3 (Bethesda)* 2012, **2**(10):1233–1241.
49. Waldbieser GC, Bosworth BG, Nonneman DJ, Wolters WR: **A microsatellite-based genetic linkage map for channel catfish, *Ictalurus punctatus*.** *Genetics* 2001, **158**(2):727–734.
50. Quiniou SM, Katagiri T, Miller NW, Wilson M, Wolters WR, Waldbieser GC: **Construction and characterization of a BAC library from a gynogenetic channel catfish *Ictalurus punctatus*.** *Genet Sel Evol* 2003, **35**(6):673–683.
51. Wang S, Xu P, Thorsen J, Zhu B, de Jong PJ, Waldbieser G, Kucuktas H, Liu Z: **Characterization of a BAC library from channel catfish *Ictalurus punctatus*: indications of high levels of chromosomal reshuffling among teleost genomes.** *Mar Biotechnol (NY)* 2007, **9**(6):701–711.
52. Quiniou SM, Waldbieser GC, Duke MV: **A first generation BAC-based physical map of the channel catfish genome.** *BMC Genomics* 2007, **8**:40.
53. Xu P, Wang S, Liu L, Thorsen J, Kucuktas H, Liu Z: **A BAC-based physical map of the channel catfish genome.** *Genomics* 2007, **90**(3):380–388.
54. Xu P, Wang S, Liu L, Peatman E, Somridhvej B, Thimmapuram J, Gong G, Liu Z: **Channel catfish BAC-end sequences for marker development and assessment of syntenic conservation with other fish species.** *Anim Genet* 2006, **37**(4):321–326.
55. Cao D, Kocabas A, Ju Z, Karsi A, Li P, Patterson A, Liu Z: **Transcriptome of channel catfish (*Ictalurus punctatus*): initial analysis of genes and expression profiles of the head kidney.** *Anim Genet* 2001, **32**(4):169–188.
56. Ju Z, Karsi A, Kocabas A, Patterson A, Li P, Cao D, Dunham R, Liu Z: **Transcriptome analysis of channel catfish (*Ictalurus punctatus*): genes and expression profile from the brain.** *Gene* 2000, **261**(2):373–382.
57. Karsi A, Cao D, Li P, Patterson A, Kocabas A, Feng J, Ju Z, Mickett KD, Liu Z: **Transcriptome analysis of channel catfish (*Ictalurus punctatus*): initial analysis of gene expression and microsatellite-containing cDNAs in the skin.** *Gene* 2002, **285**(1–2):157–168.
58. Li P, Peatman E, Wang S, Feng J, He C, Baoprasertkul P, Xu P, Kucuktas H, Nandi S, Somridhvej B, et al: **Towards the ictalurid catfish transcriptome: generation and analysis of 31,215 catfish ESTs.** *BMC Genomics* 2007, **8**:177.
59. Wang S, Peatman E, Abernathy J, Waldbieser G, Lindquist E, Richardson P, Lucas S, Wang M, Li P, Thimmapuram J, et al: **Assembly of 500,000 inter-specific catfish expressed sequence tags and large scale gene-associated marker development for whole genome association studies.** *Genome Biol* 2010, **11**(1):R8.
60. Chen F, Lee Y, Jiang Y, Wang S, Peatman E, Abernathy J, Liu H, Liu S, Kucuktas H, Ke C, et al: **Identification and characterization of full-length cDNAs in channel catfish (*Ictalurus punctatus*) and blue catfish (*Ictalurus furcatus*).** *PLoS One* 2010, **5**(7):e11546.
61. Liu S, Zhang Y, Zhou Z, Waldbieser G, Sun F, Lu J, Zhang J, Jiang Y, Zhang H, Wang X, et al: **Efficient assembly and annotation of the transcriptome of catfish by RNA-Seq analysis of a doubled haploid homozygote.** *BMC Genomics* 2012, **13**:595.
62. Liu S, Zhou Z, Lu J, Sun F, Wang S, Liu H, Jiang Y, Kucuktas H, Kaltenboeck L, Peatman E, et al: **Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array.** *BMC Genomics* 2011, **12**:53.
63. Jiang Y, Lu J, Peatman E, Kucuktas H, Liu S, Wang S, Sun F, Liu Z: **A pilot study for channel catfish whole genome sequencing and *de novo* assembly.** *BMC Genomics* 2011, **12**:629.
64. Zhang Y, Liu S, Lu J, Jiang Y, Gao X, Ninwichian P, Li C, Waldbieser G, Liu Z: **Comparative genomic analysis of catfish linkage group 8 reveals two homologous chromosomes in zebrafish with extensive inter-chromosomal rearrangements.** *BMC Genomics* 2013, **14**:387.
65. Jiang Y, Ninwichian P, Liu S, Zhang J, Kucuktas H, Sun F, Kaltenboeck L, Sun L, Bao L, Liu Z: **Generation of physical map contig-specific sequences useful for whole genome sequence assembly and scaffolding.** *PLoS ONE* 2013, **8**(13):e78872. doi:10.1371.
66. Lehmann J, Stadler PF, Prohaska SJ: **SynBlast: assisting the analysis of conserved synteny information.** *BMC Bioinformatics* 2008, **9**:351.
67. Ravi V, Venkatesh B: **Rapidly evolving fish genomes and teleost diversity.** *Curr Opin Genet Dev* 2008, **18**(6):544–550.
68. Sanetra M, Henning F, Fukamachi S, Meyer A: **A microsatellite-based genetic linkage map of the cichlid fish, *Astatotilapia burtoni* (Teleostei): a comparison of genomic architectures among rapidly speciating cichlids.** *Genetics* 2009, **182**(1):387–397.
69. Froschauer A, Braasch I, Volff J-N: **Fish genomes, comparative genomics and vertebrate evolution.** *Curr Genomics* 2006, **7**(1):43–57.
70. Steinke D, Salzburger W, Meyer A: **Novel relationships among ten fish model species revealed based on a phylogenomic analysis using ESTs.** *J Mol Evol* 2006, **62**(6):772–784.
71. Sarropoulou E, Fernandes JMO: **Comparative genomics in teleost species: Knowledge transfer by linking the genomes of model and non-model fish species.** *Comp Biochem Physiol D Genom Proteomics* 2011, **6**:92–102.
72. Otto SP: **The evolutionary consequences of polyploidy.** *Cell* 2007, **131**(3):452–462.
73. Voorrips RE: **MapChart: software for the graphical presentation of linkage maps and QTLs.** *J Hered* 2002, **93**(1):77–78.

doi:10.1186/1471-2164-14-780

Cite this article as: Jiang et al.: Whole genome comparative analysis of channel catfish (*Ictalurus punctatus*) with four model fish species. *BMC Genomics* 2013 **14**:780.