

A Second Generation Integrated Map of the Rainbow Trout (*Oncorhynchus mykiss*) Genome: Analysis of Conserved Synteny with Model Fish Genomes

Yniv Palti · Carine Genet · Guangtu Gao · Yuqin Hu ·
Frank M. You · Mekki Boussaha ·
Caird E. Rexroad III · Ming-Cheng Luo

Received: 9 August 2011 / Accepted: 18 October 2011 / Published online: 19 November 2011
© Springer Science+Business Media, LLC (outside the USA) 2011

Abstract DNA fingerprints and end sequences from bacterial artificial chromosomes (BACs) from two new libraries were generated to improve the first generation integrated physical and genetic map of the rainbow trout (*Oncorhynchus mykiss*) genome. The current version of the physical map is composed of 167,989 clones of which 158,670 are assembled into contigs and 9,319 are singletons. The number of contigs was reduced from 4,173 to 3,220. End sequencing of clones from the new libraries generated a total of 11,958 high quality sequence reads. The end sequences were used to develop 238 new microsatellites of which 42 were added to the genetic map. Conserved synteny between the rainbow trout genome and

model fish genomes was analyzed using 188,443 BAC end sequence (BES) reads. The fractions of BES reads with significant BLASTN hits against the zebrafish, medaka, and stickleback genomes were 8.8%, 9.7%, and 10.5%, respectively, while the fractions of significant BLASTX hits against the zebrafish, medaka, and stickleback protein databases were 6.2%, 5.8%, and 5.5%, respectively. The overall number of unique regions of conserved synteny identified through grouping of the rainbow trout BES into fingerprinting contigs was 2,259, 2,229, and 2,203 for stickleback, medaka, and zebrafish, respectively. These numbers are approximately three to five times greater than those we have previously identified using BAC paired ends. Clustering of the conserved synteny analysis results by linkage groups as derived from the integrated physical and genetic map revealed that despite the low sequence homology, large blocks of macrosynteny are conserved between chromosome arms of rainbow trout and the model fish species.

Electronic supplementary material The online version of this article (doi:10.1007/s10126-011-9418-z) contains supplementary material, which is available to authorized users.

Y. Palti (✉) · G. Gao · C. E. Rexroad III
National Center for Cool and Cold Water Aquaculture,
ARS-USDA,
11861 Leetown Road,
Kearneysville, WV 25430, USA
e-mail: yniv.palti@ars.usda.gov

C. Genet · M. Boussaha
INRA, UMR1313, Génétique Animale et Biologie Intégrative,
Domaine de Vilvert,
78352 Jouy en Josas Cedex, France

Y. Hu · F. M. You · M.-C. Luo
Department of Plant Sciences, University of California,
One Shields Ave.,
Davis, CA 95616, USA

Present Address:

F. M. You
Cereal Research Center, Agriculture and Agri-Food Canada,
Winnipeg, MB R3T 2M9, Canada

Keywords *Oncorhynchus mykiss* · Physical map · Genetic map · Comparative genomics · Conserved synteny

Introduction

Rainbow trout (*Oncorhynchus mykiss*) are the most-widely cultivated cold freshwater fish in the world and are considered by many to be the “aquatic lab-rat.” Interests in the utilization of rainbow trout as a model species for genome-related research activities focusing on carcinogenesis, toxicology, comparative immunology, disease ecology, physiology, transgenics, evolutionary genetics, and nutrition have been well documented (Thorgaard et al. 2002).

Coupling great interest in this species as a research model with the need for genetic improvement for aquaculture production efficiency and product quality justifies the continued development of genome resources facilitating selective breeding.

The rainbow trout genome is large and complex. Genome size estimates derived from determining the molecular weight of DNA per cell for rainbow trout and other salmonids vary from 2.4 to 3.0×10^9 bp (Ng et al. 2005; Young et al. 1998). A common ancestor of rainbow trout and most salmonids experienced a recent genome duplication event resulting in a semi-tetraploid state (i.e., after an autotetraploid event in the salmonids, their genome is undergoing reversion to a diploid state; Allendorf and Thorgaard 1984). This makes them an excellent model group for studying the differential evolution and loss of duplicated genes in the process of re-diploidization and speciation.

Current genomic resources available for rainbow trout research include multiple bacterial artificial chromosome (BAC) libraries and a BAC fingerprinting physical map (Katagiri et al. 2001; Palti et al. 2004, 2009), a database of ~200,000 BAC end sequences (BES) (Genet et al. 2011), doubled haploid (DH) clonal lines (Robison et al. 2001, 1999; Young et al. 1996; Quillet et al. 2007), multiple genetic maps based on clonal lines and outbred populations (Guyomard et al. 2006; Nichols et al. 2003b; Sakamoto et al. 2000; Young et al. 1998; Rexroad et al. 2008; Palti et al. 2011), large expressed sequence tag (EST) databases and a reference transcriptome (Rexroad et al. 2003; Govoroun et al. 2006; Salem et al. 2010), a micro-RNAs database (Salem et al. 2009), and high density DNA microarrays (Rise et al. 2004; Salem et al. 2008).

Qualitative trait loci (QTL) mapping experiments in rainbow trout are facilitated by their high fecundity, external fertilization, and ease of gamete handling and manipulation. Many QTL have been identified for production and life-history traits including resistance to the parasite *Ceratomyxa shasta* (Nichols et al. 2003a), resistance to IHNV (Barroso et al. 2008; Rodriguez et al. 2004) and to IPNV (Ozaki et al. 2001), whirling disease resistance (Baerwald et al. 2010), killer cell-like activity (Zimmerman et al. 2004), upper thermal tolerance (Perry et al. 2001, 2005), embryonic development rate (Nichols et al. 2007; Robison et al. 2001; Sundin et al. 2005; Miller et al. 2011), spawning time (O'Malley et al. 2003; Sakamoto et al. 1999), confinement stress response (Drew et al. 2007), early maturation (Haidle et al. 2008), osmoregulation (Le Bras et al. 2011), and smoltification (Nichols et al. 2008). The availability of a BAC physical map integrated with the genetic map can facilitate fine mapping of QTL, the selection of positional candidate genes, and the incorporation of marker-assisted selection (MAS) into rainbow trout breeding programs. However, a major shortcoming of QTL

studies is that they are limited to the variation present in a limited number of families. This can be overcome by whole genome association studies and other approaches, such as genomic selection, that capture the effects of most QTL that contribute to the population-wide variation in a trait (Cole et al. 2011; Wiggans et al. 2009). Hence, a high-quality reference genome sequence assembly is needed, and a robust integrated physical and genetic map can provide an excellent framework for a whole-genome sequence assembly.

Previously, we generated the first BAC-based physical map of the rainbow trout genome using DNA fingerprints of 154,439 clones from the $10 \times$ *Hind*III Swanson library (Palti et al. 2009) and integrated approximately 12% of the physical map with the genetic maps through anchoring of 238 BAC contigs to chromosomes (or linkage groups) (Palti et al. 2011). In addition, we generated 176,485 high-quality BAC end sequences from the same $10 \times$ *Hind*III library, which we used for developing a large number of DNA markers, producing a repeat elements database, masking of repeat sequences, and for identifying regions of conserved synteny with model fish species (Genet et al. 2011). Here, we report on the construction of a second generation integrated physical and genetic map of the rainbow trout genome using BAC clones from two new *Bam*HI and *Eco*RI libraries. In addition, we demonstrate through bioinformatic analysis that the BAC-based map can improve comparative genome analyses in species that lack a reference genome sequence and use it to identify regions of conserved synteny between rainbow trout and model fish species genomes. This new integrated map of the rainbow trout genome provides a framework for a robust composite genome map and for future reference genome sequence assemblies.

Methods

BAC Libraries and DNA Fingerprinting

High molecular weight genomic DNA from YY Swanson doubled haploid homozygous male rainbow trout was isolated from whole blood cells for the construction of a $10 \times$ BAC library by Amplicon Express, Pullman, WA, USA. The blood samples were contributed courtesy of Gary Thorgaard and Paul Wheeler (Washington State University). A standard library construction protocol (Palti et al. 2004) was followed. Briefly, genomic DNA was partially digested with *Bam*HI or *Eco*RI and fractionated on 1% agarose gel by pulsed field gel electrophoresis (PFGE). Fractions ranging from 120 to 400 kb were gel-isolated by electroelution and ligated to *Bam*HI or *Eco*RI cut pCCBAC1 (Epicentre). Ligated DNA was transformed by electroporation on a Cell-Porator (Invitrogen, Carlsbad, CA, USA) into DH10B

electro competent cells (Invitrogen). Transformed cells were recovered, and individual white colonies were picked and arrayed into 576-microtiter plates of 384 wells (288 plates *Bam*HI and 288 plates *Eco*RI) with Luria–Bertani medium containing 4.4% glycerol and 12.5 µg/ml chloroamphenicol. Plates were incubated for 18 h, replicated, and then frozen at -80°C . The replicated copy was used as a source plate for BAC DNA fingerprinting and end sequencing. For quality control, Amplicon Express randomly picked 100 clones and estimated their insert size by *Not*I digestion and PFGE analysis (CHEF mapper, BioRad, Hercules, CA, USA).

The five-color high-information-content fingerprinting (HICF) SNaPshot method for BAC DNA fingerprinting (Luo et al. 2003) was used as previously described (Palti et al. 2009). Outputs of size-calling files were automatically edited with the FPMiner program (<http://bioinformsoft.com/>) using the program's default setting. This software package was used to distinguish peaks corresponding to restriction fragments from peaks generated by background noise in the fingerprints' profile of each BAC and to remove vector restriction fragments from all profiles. The program also removed substandard profiles that could negatively affect contigs' assembly. The fragment size files generated by FPMiner were used in the FPC contig assembly.

Contigs Assembly

Contigs were assembled from fragments within size range of 70–1,000 bp using FPC program version 9.3 (Nelson and Soderlund 2009; Nelson et al. 2005; Soderlund et al. 2000). FPC parameters were adjusted for the HICF method as previously described (Luo et al. 2003; Nelson et al. 2005; Quiniou et al. 2007; Palti et al. 2009). The new BAC clone fingerprints were added to our previous assembly (Palti et al. 2009) using a tolerance of 0.5 bp and a Sulston score (Soderlund et al. 1997) of 1×10^{-30} . DQer function was recurrently used at the "Step" value of 2 and stopped at the Sulston score of 1×10^{-67} when no more than 1% of the contigs possessed 15% or more questionable (Q-) clones. The "Best of" function was set to 100 builds. The DQer function was used to improve the integrity of the assembly, as high occurrence of Q-clones in contigs is related to high occurrence of falsely assembled contigs.

BAC-end Sequencing

Aliquots of BAC DNA that were extracted for DNA fingerprinting were used as templates for the sequencing reactions. Sequencing reactions were carried out with T7 or Sp6 universal primers, using ABI kit version 3.1. Generated raw sequence files were subsequently processed using the PHRED software (Ewing and Green 1998), and Q20 values

were achieved by setting the sequence quality PHRED score cut-off value to 20. Vector and bacterial sequences were identified and removed using the SeqTrim V0.111-w0.19 program (Falgueras et al. 2010).

Identification of Microsatellites and Simple Sequence Repeats

Microsatellites and other SSR motifs were identified using tandem repeat finder software (Benson 1999). We examined ten classes of SSRs using a maximum period size of ten with default settings. BES containing microsatellites were subsequently masked using RepeatMasker with INRA RT rep1.0 custom library file: BES harboring SSRs with at least 50 bp flanking sequences were then selected, and forward and reverse primers were designed using Primer3 software (Rozen and Skaletsky 2000). The primer product size range was chosen between 150 and 450 nucleotides. The optimum size of primers was set to 20 nucleotides (range from 18 to 27 nucleotides) with an optimum melting temperature of 60.0°C (range from 57°C to 63°C).

Microsatellites Genotyping

The NCCCWA mapping panel of five families was genotyped with microsatellites as previously described (Rexroad et al. 2008; Palti et al. 2011). Microsatellite markers isolated from BAC end sequences were genotyped using the tailed protocol (Boutin-Ganache et al. 2001). Primers were obtained from a commercial source (Alpha DNA, Montreal, QC, Canada). Three oligonucleotide primers were used in each DNA amplification reaction (forward, 5' GAGTTTTCCCAGTCACGAC-primer sequence 3'; reverse, 5' GTTT-primer sequence 3'; fluorescent labeled primer with FAM, 5' GAGTTTTCCCAGTCACGAC 3'). Primers were optimized for amplification by varying annealing temperatures and MgCl_2 concentrations. PCR reactions (12 µl total volume) included 50 ng DNA, 1.5–2.5 mM MgCl_2 , 2 pmol of forward primer, 6 pmol of reverse primer, 1 pmol of fluorescent labeled primer, 200 µM dNTPs, 1× manufacturer's reaction buffer, and 0.5 U Taq polymerase (ABI, Foster City, CA, USA). Amplifications were conducted in an MJ Research DNA engine thermal cycler model PTC 200 (MJ Research, Waltham, MA, USA) as follows: an initial denaturation at 95°C for 10 min, 30 cycles consisting of 94°C for 60 s, annealing temperature for 45 s, 72°C extension for 45 s; followed by a final extension of 72°C for 10 min. PCR products were visualized on agarose gels after staining with ethidium bromide. Three microliters of each PCR product was added to 20 µl of water, and 1 µl of the diluted sample was added to 12.5 µl of loading mixture made up with 12 µl of HiDi formamide and 0.5 µl of

Genscan 400 ROX internal size standard. Samples were denatured at 95°C for 5 min and kept on ice until loading on an ABI 3730 DNA analyzer (ABI, Foster City, CA, USA). Output files were analyzed using GeneMapper version 3.7 (ABI, Foster City, CA, USA), formatted using Microsoft Excel and stored in a Microsoft Access database.

Linkage Analysis

The new microsatellites were placed on the rainbow trout genetic map using the genetic linkage mapping programs MULTIMAP (Matisse et al. 1994) and CRI-MAP (Lander and Green 1987). First, genotype data combined for both sexes were formatted into the standard LINKAGE (Lathrop et al. 1984) file format and checked for Mendelian inheritance using PEDCHECK (O’Connell and Weeks 1998). RECODE (O’Connell and Weeks 1995) was then used to convert the allele sizes into number-coded alleles. Genotype data and locus names were converted into CRI-MAP input format using an in-house Perl script. The resulting file was then added to the NCCCWA reference map data file (Palti et al. 2011) using another in-house Perl script. MULTIMAP was used to conduct two-point linkage analyses for identifying the closest markers with $\text{LOD} \geq 8.75$ and recombination fraction $r \leq 0.2$. Multipoint linkage analyses were conducted on individual linkage groups to assign LOD scores for the specific position of each marker within the linkage group. Framework maps were constructed at $\text{LOD} \geq 4$ for all linkage groups but OMY21, for which the framework map was created at $\text{LOD} \geq 3$. Markers were added to comprehensive maps by lowering the LOD threshold one integer at a time and starting with the previous order. Resulting maps are consensus maps, accounting for co-informative meiosis across the five NCCCWA reference families. Chromosome numbers were assigned to linkage groups using the rainbow trout integrated cytogenetic/linkage map (Phillips et al. 2006).

Assessment of Regions of Conserved Synteny with Other Fish Genomes

Sequence Homology Searches and Results Filtration

Masked BES reads with more than 100 base pairs of contiguous nonrepetitive sequences were analysed for sequence homology by BLASTN using ENSEMBL DNA databases for zebrafish (Danio_rerio.Zv9.61.dna_rm.toplevel.fa), stickleback (Gasterosteus_aculeatus.BROADS1.61.dna_rm.toplevel.fa), and medaka (Oryzias_latipes.MEDAKA1.61.dna_rm.toplevel.fa) and for gene content by BLASTX using the ENSEMBL nonredundant protein databases for zebrafish (Danio_rerio.Zv9.61.pep.all.fa),

stickleback (Gasterosteus_aculeatus.BROADS1.61.pep.all.fa), and medaka (Oryzias_latipes.MEDAKA1.61.pep.all.fa). BLASTN searches were carried out using an e value cut-off of $1e^{-5}$ with the following parameters in the command line (Altschul et al. 1990):

– m9 –r1 –q-1 –G4 –E2 –W9 –F “m D” –U

BLASTX searches were carried out with the following parameters: (Altschul et al. 1997)

– m8 –e 1.0e-5

For BLASTX, the Ensembl protein IDs were renamed by their corresponding Ensembl gene IDs as each gene may encode several peptides due to alternative splicing.

The BLAST search results were filtered to remove nonspecific sequences using the following filtration steps: (1) For each BES read with multiple BLAST hits, results were filtered to keep only the hits with the minimal e value score; (2) BES reads with multiple hits having the same minimal e value were filtered to keep the hits with the highest HSPs (high-scoring segment pairs; calculated as the product of % identity multiplied by alignment length); and (3) only BES reads with single “unique” hits following filtration steps 1 and 2 were kept for comparative synteny analyses.

Comparative Synteny Analysis

Regions of conserved synteny between rainbow trout and model fish species were identified using multiple BAC ends from a single fingerprinting (FP) contig with unique hits. Grouping of BES to FP contigs was accomplished in two steps: (1) a list of all the BACs from the FPC assembly and their contig assignments was extracted from the FPC output text file and (2) BES with significant Blast hits were linked to contigs containing the same BAC name. A region of microsynteny with the target genome was established if a single contig was linked through BES Blast hits to less than four chromosomes (1–3) from a model fish genome and at least two BAC ends from a single contig were mapped to the same chromosome with a space of 10 kb to 1 Mb between the left- and right-end hits. The upper limit of 1 Mb is approximately two times the average contig length, which is the same method we used previously to identify microsynteny using BAC pair ends (Genet et al. 2011). In addition, we defined regions of putative conserved synteny as those in which at least two BES hits were mapped to the same chromosome of the model species spanning more than 1 Mb. However, to simplify the discussion in the text, moving forward, we will use the term “regions of macrosynteny” to describe those segments of putative conserved synteny.

Results

BAC Fingerprinting and Contigs' Assembly

Two 5× genome coverage BAC libraries (110,592 clones each) from the Swanson doubled haploid clonal line were prepared using *Bam*HI and *Eco*RI partial genomic digestion to complement the previously characterized and fingerprinted 10× *Hind*III library (Palti et al. 2004, 2009). The average insert size of each library was estimated from PFGE of 100 *Not*I digested clones to be 145 kb. The percent of clones that did not contain insert (blue colonies) was 0.65% and 1.0% for the *Bam*HI and *Eco*RI libraries, respectively, and the percent of wells with poor or no growth was 0.56% and 0.12% for the *Bam*HI and *Eco*RI libraries, respectively.

We used the five-color HICF SNaPshot method (Luo et al. 2003) to fingerprint 5,376 clones from the *Eco*RI library (RE) and 10,752 clones from the *Bam*HI library (RB). Following strict quality editing of the fingerprints, we have successfully added 13,550 clones from the two new libraries to our previous FPC assembly, and the number of contigs in the physical map assembly was reduced from 4,173 to 3,220. The current version of the physical map is composed of 167,989 clones of which 158,670 are assembled into contigs and 9,319 remained singletons (Table 1). The average number of BACs per contig is 49.3. The average number of fingerprinting fragments per BAC is 75.9, and the average insert size is estimated at 135 kb. Therefore, each fragment is estimated on average to represent approximately 1.78 kb of genome DNA. The total number of unique fingerprinting fragments (consensus bands) in contigs is 1,049,643, which corresponds to an estimated physical length of 1.87 Gb (~80% of the rainbow trout genome).

The average number of consensus bands (CB) per contig is 326, and the estimated contig size is 580 kb. The current version of the rainbow trout physical map can be browsed online via WebFPC:

<http://www.genome.clemson.edu/rainbowTrout2>

The previous version of the physical map is still available for end users at <http://www.genome.clemson.edu/activities/projects/rainbowTrout>.

For specific contig modifications with the previous version, end users can compare the location of specific clones on interest between the two map versions. The FPC assembly files of both versions are available upon request from the corresponding author.

BAC-end Sequencing Statistics

A total of 10,752 fingerprinted clones from the new RB and RE libraries (5,376 each) were used to generate BAC end sequence (BES) reads. Following trimming and filtration of low-quality sequence and reads with high similarity to bacterial and vector sequences, the total of high quality sequence reads longer than 100 bp was 11,958 (RB, 6,350; RE, 5,608). The number of clones with both ends sequenced was 4,016 and with only one end sequenced 3,926. The combined linear length of the BES reads was 6,834,308 bp, representing approximately 0.3% of the trout genome. The GC content was estimated to be 43%, which is very similar to what was previously found for rainbow trout (Genet et al. 2011). Repeat element analysis using the new INRA RT rebase1.0 library (Genet et al. 2011) masked almost 38% of the RB/RE BES database in base pairs. All processed BES were submitted to the GenBank public database with consecutive accession numbers of HR794783–HR806740.

Table 1 BAC fingerprinting and FPC assembly statistics; second generation physical map of the rainbow trout genome

Number of clones fingerprinted	208,224	~11.6× genome coverage
Used in FPC assembly	167,989	~9.3× genome coverage
Average insert size (kb)	135	
Average number of fragments per clone	75.9	
Estimated average fragment size (kb)	1.78	
Number of clones in contigs	158,670	~8.8× genome coverage
Number of contigs	3,220	
Average number of clones per contig	49.3	
Average contig size in consensus bands	326	
Estimated average contig size (kb)	580	
Number of Q contigs	1,290	40%
Number of Q clones	3,332	2.1%
Number of CB in contigs	1,049,643	
Estimate total length in contigs (Mb)	1,870	80% genome size

Development of Microsatellite Markers

A total of 557 microsatellites were identified in 513 BES reads. Approximately 52% of the microsatellites (267) were suitable for PCR primers design as they were flanked by sequences of at least 50 bp. We were able to define 238 primer pairs (~43%) from 211 distinct BES reads as more than one microsatellite can be detected in a single BES. Dinucleotides were the most abundant repeat motif (64.3%) followed by tetra-nucleotides (10.5%). The microsatellite markers and corresponding flanking sequences were submitted to the GenBank STS database with consecutive accession numbers of GF110823–GF110876 and GF111062–GF111272.

BES Microsatellite Genotypes

We selected 54 BES microsatellites (RB, 32; RE, 22) for PCR optimization and genotyping using two criteria: (1) the BAC of origin was in a fingerprinting contig of at least 50 clones and (2) the number of Q-clones in the contig was smaller than 5. Of the 54 markers genotyped, 40 markers (RB, 23; RE, 17) appeared to amplify single marker regions and were polymorphic. Seven markers were monomorphic, and five markers could not be resolved or unambiguously scored. Two markers generated duplicated patterns, of which one could be scored only for a single marker region and one produced a scorable duplicated pattern. The BACs of origin from which the 42 informative marker loci were isolated were mapped to 39 unique fingerprinting contigs. The 54 BES microsatellites selected for genotyping are listed in Supplemental file 1, with the corresponding PCR primers and conditions for each marker, number of alleles and size range, GenBank accession, primers sequences, genetic map chromosome (if mapped), and physical map contig.

The Genetic Map

The 42 informative microsatellites were added to our previous marker genotypes database to expand the most recent version of the NCCCWA genetic map (Palti et al. 2011). The 42 new markers were mapped to 24 of the 29 linkage groups. Two-point linkage analysis placed 1,459 loci in 29 linkage groups at $\text{LOD} \geq 3.0$. An additional six markers with two-point $\text{LOD} < 3.0$ were added to linkage groups manually (2.90, 2.89, 2.64, 2.12, 2.10, and 1.80). The specific best of two-point LOD score for each marker is provided in Supplemental file 2, Worksheet 1. The total combined sex averaged map distance was 3,360.1 cM (Kosambi). A sample map representing chromosome 2 is presented in Fig. 1, and linkage group diagrams representing the 24 chromosomes for which new markers were added are presented in Supplemental file 3. Multipoint

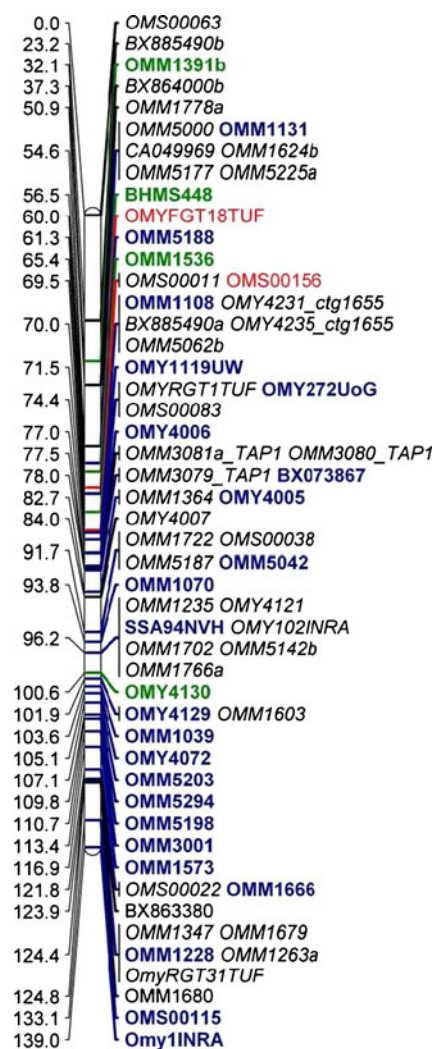


Fig. 1 Chromosome 2 from the new NCCCWA genetic map is shown as an example of the updated linkage map. Annotation of genes linked to the marker or BAC contig from the second generation physical map are connected to the marker name by underscore (e.g., OMM3080_TAP1 or OMY4231_ctg1655). Annotation of “or ?” means that the marker is duplicated and only one of two BAC contigs was identified for the marker. BAC contigs’ annotation is only shown for new microsatellites isolated from BAC end sequences of the RB and RE libraries. *Blue, green, red, black, and italicized font markers* were mapped to their specific location on the linkage group at LOD scores of 4, 3, 2, 1, and 0, respectively. Sex average distances between markers are shown in centimorgan

linkage analyses were conducted on individual linkage groups to assign LOD scores for the specific position of each marker within the linkage group. The number of markers included in a framework map created at $\text{LOD} \geq 4$ for the specific position of the marker in the linkage group was 477. The only chromosome that did not contain any framework markers at $\text{LOD} \geq 4$ was OMY21, for which a framework map was created at $\text{LOD} \geq 3$. Additional loci were added at $\text{LOD} \geq 3$ (84), ≥ 2 (84) ≥ 1 (56), and ≥ 0 (764). Supplemental file 2, Worksheet 1 contains this information

and can be used to recreate maps for all the 29 chromosomes using MapChart software (Voorrips 2002). The average resolution of the new genetic map was 2.29 cM with intermarker distances ranging from 1.55 to 3.23 cM for individual chromosomes (Supplemental file 2, Worksheet 2).

A high frequency of duplicated microsatellite loci was reported in the previous NCCCWA genetic maps (Rexroad et al. 2008; Palti et al. 2011), but in many cases, only one locus was successfully ordered on the map. Overall, 89 duplicated markers were successfully mapped to two loci in this version of the genetic map (178 loci), which means that the total number of markers mapped was 1,376.

The Integrated Map

The newly added BES microsatellites combined with the previously mapped RT (*Hind*III) library BES microsatellites anchor physical map contigs to 28 of the 29 linkage groups with an average of 5.86 BES microsatellites per linkage group (Supplemental file 2, Worksheet 3). Overall, 27 physical map contigs were added to the integrated genome map through genetic mapping of 28 new microsatellite markers, which represents an estimated addition of at least 15,660 kb (or 0.65% of the rainbow trout genome) to the integrated map. The other 14 new BES microsatellites that were added to the genetic map in this paper represent 12 physical map contigs that were already anchored to the genetic map in the previous version of the integrated map (Palti et al. 2011).

In the current version of the integrated map we have identified 39 BAC contigs that are useful for estimating the relationship between the physical and genetic distances as they were anchored to the genetic linkage group through multiple markers. The ratio of physical to genetic linkage distances varied substantially among the 39 contigs, which is similar to other vertebrate genomes (Nievergelt et al. 2004; Quiniou et al. 2007). The 39 contigs represent segments from 23 of the 29 chromosomes (Supplemental file 2, Worksheet 4). The kilobase pair/centimorgan ratio ranged from 37 to 13,600 with an average of 2,088.

Homology with Other Fish Genomes

Of the 188,443 high quality BES reads that were generated from the RB, RE, and RT libraries, a total of 106,958 BES reads (~57%) had more than 100 base pairs of contiguous nonrepetitive sequences and were used for assessing conserved synteny by BLASTN and BLASTX similarity searches against the ENSEMBL genome and peptide databases of zebrafish, medaka, and stickleback. The fractions of all high quality trout BES reads (188,443) that had significant BLASTN hits against the zebrafish, medaka, and stickleback genome databases were 8.8%, 9.7%, and 10.5%, respectively, while the fractions of BES reads that

had significant BLASTX hits against the zebrafish, medaka, and stickleback protein databases were 6.2%, 5.8%, and 5.5%, respectively (Table 4).

Identification of Regions of Conserved Synteny with Other Fish Genomes

We used the physical map BAC fingerprinting contigs to group the BES Blast hits for identifying regions of conserved synteny between the genomes of rainbow trout and the model fish species. However, before we could conduct comparative analyses, we had to identify optimal thresholds for two critical parameters: (1) the number of BES Blast hits linking a single rainbow trout contig with a single model fish chromosome and (2) the number of chromosomes from a model fish genome that can be linked to a single rainbow trout contig. As shown in Fig. 2, we

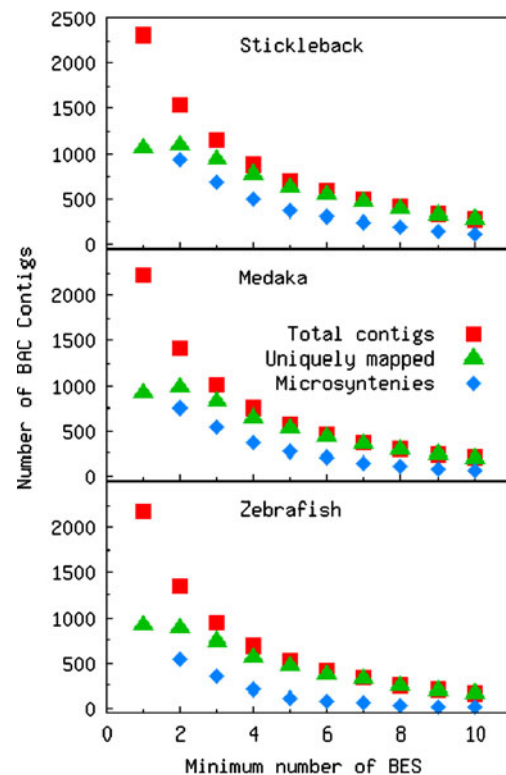


Fig. 2 Plot of the number of rainbow trout BAC contigs that were mapped to chromosomes of model fish genomes through BlastN hit matches against minimum threshold number hits of BAC end sequences (BES) from a single contig per chromosome. Total contigs: Total number of trout BAC contigs that were mapped to the model fish genome chromosomes corresponding to the minimum number of BES threshold on the X-axis. Each trout contig can be mapped to one or to multiple chromosomes from each of the three model fish genomes. Uniquely mapped: the number of BAC contigs that are mapped to only one chromosome at the corresponding minimum number of BES threshold. Microsynteny: the number of mapped BAC contigs (can be uniquely mapped or not) to which microsynteny was found in each model fish genome at the corresponding minimum number of BES threshold

determined that the highest number of uniquely mapped contigs and microsynteny were observed for all three model species at the threshold of two BlastN hits and that at the much more conservative threshold of five BES hits almost all the contigs were uniquely mapped (e.g., each mapped contig was only linked to one chromosome per species). To achieve maximum prediction of microsynteny in the current analyses, we decided to use the more permissive threshold of two BES hits, but higher thresholds can be easily set by re-sorting the data in Supplemental files 4 and 5. Using the threshold of two BES hits, we then determined that the majority of contigs that were linked to chromosomes were uniquely mapped to one chromosome (65.7%, 69.6%, and 70.9% for zebrafish, medaka, and stickleback, respectively) and that approximately 95% of them were linked to one, two, or three chromosomes in all three species (Fig. 3; Table 2). The evolutionary histories that separate rainbow trout from each of the three model species include a salmonid-specific genome duplication and partial re-diploidization in addition to numerous independent chromosomal rearrangements and localized duplications. Hence, we decided to account for multiple local synteny in our analyses and only discarded the ~5% of the contigs that were mapped to at least four chromosomes.

For BlastN searches, we identified 2,185, 2,228, and 2,303 BAC contigs with significant unique hits to the zebrafish, medaka, and stickleback genomes, respectively (Table 2). Out of those, we identified 1,272 (58%), 1,350 (61%), and 1,469 (64%) contigs mapping to less than four chromosomes with at least two unique hits per chromosome in zebrafish, medaka, and stickleback, respectively. For those contigs, we identified 1,756, 1,810, and 1,950 macro-

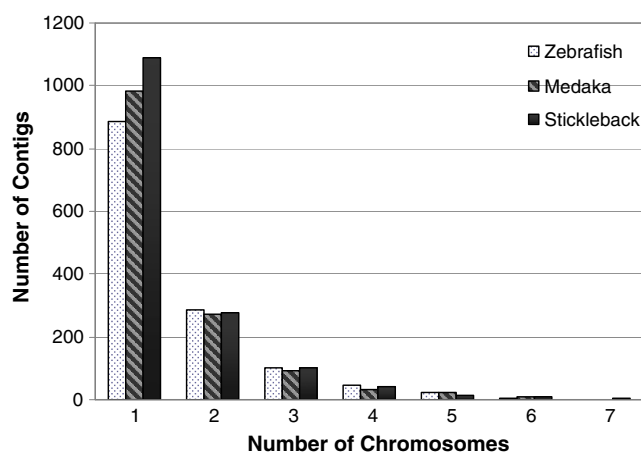


Fig. 3 Distribution of rainbow trout BAC contigs with conserved synteny to target model fish genomes through matching of at least two significant BlastN hits per chromosome, plotted against the number of chromosomes from the target genomes that matched a single rainbow trout contig

synteny (contig–chromosome pairs) with at least two BES BlastN hits in zebrafish, medaka, and stickleback, respectively. Of those, we identified 592 (34%), 807 (45%), and 992 (51%) regions of microsynteny between rainbow trout and zebrafish, medaka, and stickleback, respectively (Table 2; Supplemental file 4). Of the contigs mapping to less than four chromosomes with at least two unique hits per chromosome, 1,010 had macro-synteny in all three model species, of which 502 (49.7%) had hits to a single chromosome in all three species (Supplemental file 6, Worksheet 1). Anchored linkage groups and the respective genetic distance in centimorgans are also presented in Supplemental file 4 for the rainbow trout contigs that were integrated with the genetic map. For physical map contigs that were only anchored to the INRA genetic map (Guyomard et al. 2006; Palti et al. 2011), the linkage group number was converted to the corresponding NCCCWA linkage group in Supplemental file 4, but the distance in centimorgan is not presented (marked as N/A). A schematic diagram of a typical conserved synteny that can be drawn between stickleback chromosome 16 and rainbow trout linkage groups based on the BlastN analysis of the BES reads, and the available integrated physical map contigs is presented in Fig. 4. In Supplement file 4, we also show the boundaries and size of the conserved synteny regions using consensus bands (CB) and base pairs (bp) as the measurement units for the trout BAC contigs and the model species chromosomes, respectively.

For BLASTX analyses, we identified 1,918, 1,900, and 1,848 BAC contigs with significant unique hits to the zebrafish, medaka, and stickleback genomes, respectively (Table 3). Further analysis revealed 1,029 (54%; zebrafish), 1,004 (53%; medaka), and 1,051 (57%; stickleback) contigs mapping to less than four chromosomes with at least two unique hits per chromosome. For those contigs, we identified 1,448, 1,397, and 1,370 macro-synteny with at least two BES BlastX hits in zebrafish, medaka, and stickleback, respectively. Of those, we identified 618 (43%), 710 (51%), and 745 (54%) regions of microsynteny between rainbow trout and zebrafish, medaka, and stickleback, respectively (Table 3; Supplemental file 5). Anchored linkage groups and the respective genetic distance in centimorgans are also presented in Supplemental file 5 for the rainbow trout contigs that were integrated with the genetic map. Of the contigs mapping to less than four chromosomes with at least two unique hits per chromosome, 810 had macro-synteny in all three model species, of which 398 (49.1%) had hits to a single chromosome in all three species (Supplemental file 6, Worksheet 2).

Our BlastN analyses identified 755 (zebrafish), 832 (medaka), and 889 (stickleback) Contig-Chr pairs with unique BAC end hits in macro-synteny that were not detected using BlastX analyses (Table 2), and conversely,

Table 2 Rainbow trout BAC end sequences (BES) BlastN statistics

	Zebrafish	Medaka	Stickleback
BES summary			
Total BES with BlastN hits	16,640 (8.8%) ^b	18,199 (9.7%)	19,843 (10.5%)
Unique BES using minimum <i>e</i> value criterion	15,604	17,067	18,812
Added BES using maximum HSP criterion	303	402	299
Total BES with significant BlastN hits selected (“unique”) ^a	15,907 (8.4%)	17,469 (9.3%)	19,111 (10.1%)
Contigs summary			
Contigs having BES with “unique” BlastN hits	2,185	2,228	2,303
Contig-Chr pairs linked by BES BlastN hits	5,393	6,127	5,573
Number of contigs linked to chromosomes by at least two BlastN hits	1,352	1,414	1,536
Contig-Chr pairs linked by at least two BlastN hits	2,135	2,111	2,264
Number of contigs mapped to less than four chromosomes	1,272	1,350	1,469
Of which number of Contig-Chr pairs with at least two BlastN hits	1,756	1,810	1,950
Of which in micro-synteny (span 10 kb–1 Mb)	592 (34%)	807 (45%)	992 (51%)
Of which were not detected by BlastX	755 (43%)	832 (46%)	889 (46%)

^a The bioinformatic filtering criteria used for identifying ‘unique’ BlastN hits are described in “Methods”

^b Percent of total rainbow trout BES reads

our BlastX analyses identified 447 (zebrafish), 419 (medaka), and 309 (stickleback) Contig-Chr pairs with unique BAC end hits in macrosynteny that were not detected using BlastN analyses (Table 3). Of the 502 contigs that were identified by BlastN to have a unique chromosome hit in each of the three model fish species, 232 were also identified by BlastX to have a unique chromosome hit in each of the three species.

The number of trout contigs that were mapped to the genome of a model fish species by at least two BES hits in the BlastN or BlastX analyses was 1,422, 1,458, and 1,576 for zebrafish, medaka, and stickleback, respectively. Given that 2,888 contigs had at least two BES reads containing at least 100 bp of unmasked nonrepetitive sequence, we determined that the probability of mapping a trout BAC contig with at least two unmasked BES reads to the

Fig. 4 A schematic diagram of a typical conserved synteny that can be drawn between stickleback chromosome 16 and rainbow trout linkage groups based on the BlastN analysis of the BES reads and the available integrated physical map contigs. The size of the rainbow trout linkage groups Omy03, 07, 17, and 22 is 200, 150, 140, and 88 cM, respectively. The size of stickleback chromosome 16 is 18 million bp. Size coordinates for the rainbow trout chromosomes are in centimorgan and for the stickleback chromosome in megabase pairs

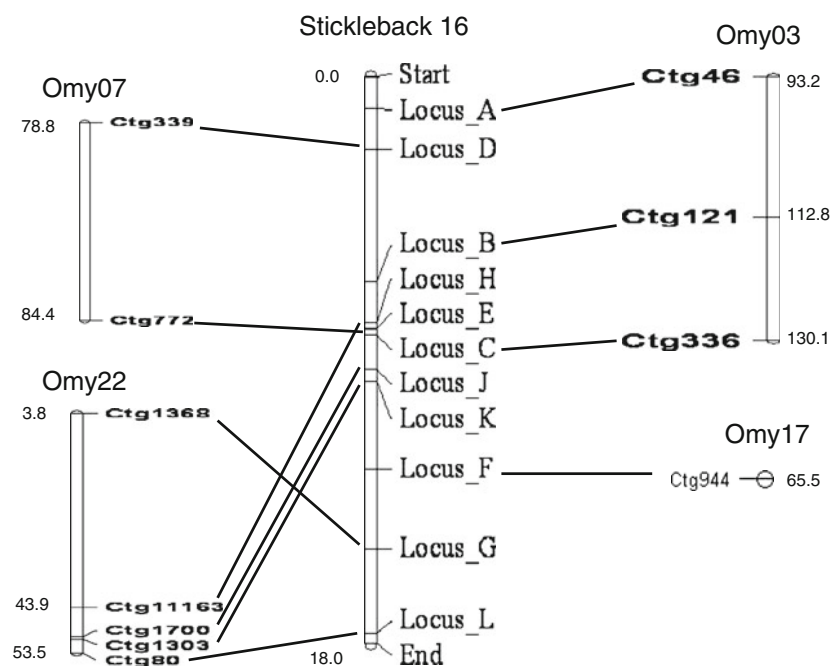


Table 3 Rainbow trout BAC end sequences BlastX statistics

	Zebrafish	Medaka	Stickleback
BES summary			
Total BES with BlastX hits	11,728 (6.2%) ^b	10,843 (5.8%)	10,325 (5.5%)
Total BES with significant BlastX hits selected (“unique”) ^a	11,185 (5.9%)	10,637 (5.6%)	10,053 (5.3%)
Contigs summary			
Contigs having BES with “unique” BlastX hits	1,918	1,900	1,848
Contig-Chr pairs linked by BES BlastX hits	4,463	4,448	3,589
No. of contigs linked to chromosomes by at least two BlastX hits	1,076	1,040	1,075
Contig-Chr pairs linked by at least two BlastX hits	1,668	1,560	1,473
No. of contigs mapped to less than four chromosomes	1,029	1,004	1,051
Of which number of Contig-Chr pairs with at least two BlastX hits	1,448	1,397	1,370
Of which in micro-synteny (span 10 Kb–1 Mb)	618 (43%)	710 (51%)	745 (54%)
Of which were not detected by BlastN	447 (31%)	419 (30%)	309 (23%)

^a The bioinformatic filtering criteria used for identifying ‘unique’ BlastX hits are described in the [Methods](#) section

^b Percent of total rainbow trout BES reads

genomes of zebrafish, medaka, and stickleback, is 49.2%, 50.5%, and 54.6%, respectively.

Discussion

The number of clones we fingerprinted from the RB and RE libraries for the second generation integrated map of the rainbow trout genome was modest compared to our previous effort in the first generation map, but the number of contigs in the physical map assembly was reduced significantly from 4,173 to 3,220 using similar FPC assembly stringency. The average contig size grew from 35 clones and 482 kb to 49 clones and 580 kb, respectively, demonstrating the utility of the new libraries that were produced using partial genomic digestion with *Bam*HI and *Eco*RI restriction enzymes to complement the 10× *Hind*III library. Thus, it is likely that fingerprinting of additional clones from the new libraries will further reduce the number of contigs in the physical map and contribute to a better assembly of a reference genome sequence from the minimal tiling path of the physical map.

The total number of high-quality and filtered BAC-end sequence (BES) reads longer than 100 bp was 11,958 (RB, 6,350; RE, 5,608). The GC content of the RB/RE libraries BES was estimated to be 43%, which is very similar to what was previously found for rainbow trout (Genet et al. 2011). Compared to other fishes, it is slightly lower than channel catfish (Xu et al. 2006) and stickleback (44%), but higher than zebrafish (36%) and medaka (40%) (<http://genome.ucsc.edu>). Repeat element analysis using the new INRA RT rebase1.0 library (Genet et al. 2011) masked almost 38% of the RB/RE BES database in base pairs. The

percentage of sequence from the rainbow trout *Hind*III BAC library that was masked by the same repeat database was more than 59% (Genet et al. 2011), which is substantially higher. These results likely indicate a bias toward repeat elements near the *Hind*III recognition site in the rainbow trout genome, but it is also possible that the database from the RB/RE libraries is too small to accurately predict the frequency and distribution of repeat elements in the trout genome.

In this new version of the integrated genome map, we added 27 anchored BAC contigs to the 238 BAC contigs that were already anchored to chromosomes of the genetic map, and the integration of 12 previously anchored contigs was expanded by the mapping of 14 new BES microsatellites. The integrated map now covers approximately 11% of the genome across segments from all 29 chromosomes. This map provides a framework for a robust composite genome map. The availability of an integrated physical and genetic map is useful for detailed comparative genome analyses, fine mapping of QTL, positional cloning, selection of positional candidate genes for economically important traits, and the incorporation of MAS into rainbow trout breeding programs. A comprehensive integrated map also provides a minimal tiling path for genome sequencing and a framework for a whole genome sequence assembly. The integrated map and large number of SSR markers we developed for the rainbow trout genome will also facilitate comparative genomics studies with other salmonids. Many microsatellite markers can be used for genetic mapping across salmonid species, which is very useful for comparative genome mapping (Danzmann et al. 2006; Phillips et al. 2009) and can benefit research in salmonid species with less developed genome maps.

The estimated megabase pair/centimorgan ratio from 38 integrated contigs ranged from 37 kb to 13.6 Mb with an average of 2.142 Mb. This estimated average is much higher than what was found for other vertebrates with more advanced reference genome assemblies like the human genome (Nievergelt et al. 2004) and is likely an overestimate for rainbow trout as well. As the sex-average length of the genetic map is approximately 3,000 cM and genome size estimates based on DNA content placed the rainbow trout between 2.4 and 3.0 Gb, we predict that an average ratio of approximately 1 Mb/cM is also more likely for rainbow trout. As this ratio was already reduced from our previous estimate of approximately 3 Mb/cM (Palti et al. 2011), it is likely that with improved genome coverage and the addition of reference data points this average ratio estimate will become more accurate.

The overall genome-wide sequence homology between rainbow trout and the three model fish species genomes was low, most likely due to the large evolutionary distance between the salmonids and the model species (Davidson et al. 2011). The fractions of trout BES reads that had significant BlastN hits against the zebrafish, medaka, and stickleback genome databases were 8.8%, 9.7%, and 10.5%, respectively. In comparison, the fractions in similar studies that used BES reads in economically important nonmodel fishes were 76% for Gilthead sea bream matched with the stickleback genome (Kuhl et al. 2011), 60% for common carp with zebrafish (Xu et al. 2011), 52.4% for European sea bass with stickleback (Kuhl et al. 2010), and 17.3% for catfish with zebrafish (Liu et al. 2009).

However, despite the relatively low sequence homology, clustering of the conserved synteny analysis results by linkage groups as derived from the integrated map (Supplementary files 4 and 5) revealed that large blocks of macrosynteny are conserved between chromosome arms of rainbow trout and the model fish species. The conser-

vation of ancestral chromosome arms was recently described by Danzmann et al. (Danzmann et al. 2008), and an example is illustrated in Fig. 4 of this paper. Here Omy7 and Omy3 markers share homology to the upper half of stickleback Chr 16, while Omy22 shares extensive homology to the lower half of the chromosome. This is consistent with the ancestral chromosomes model as stickleback Chr 16, and all the rainbow trout chromosomes depicted (i.e., Omy7p=RT-12p, Omy3q=RT-31q, and Omy22=RT-5p&q in the Danzmann et al. map) are derived from the C ancestral lineage of teleost chromosomes (Danzmann et al. 2008). Close examination of the clusters in Supplementary files 4 and 5 reveals a mosaic pattern of conserved synteny in some of the rainbow trout linkage groups, suggesting that major rearrangements and inversions have occurred following the 4R whole genome duplication in the salmonid ancestral genome.

While the number of significant BES hits with BlastN was between 1.4 (zebrafish) to 1.9 (stickleback) times greater than those with BlastX hits, the difference in number of microsynteny identified was much smaller. The number of microsynteny regions identified with BlastN was actually slightly smaller than BlastX for zebrafish, and for stickleback, it was only 1.3 times greater than BlastX. The improved BlastX prediction of synteny in zebrafish compared to stickleback and medaka may be attributed in part to the better annotation of the zebrafish transcriptome and open reading frames (ORFs). In our previous evaluation of conserved synteny using paired BAC ends from the RT library, we observed a much lower percent of BlastX micro-synteny across species compared to BlastN. In the current analysis, we increased the maximum limit of microsynteny from 300 kb to 1 Mb as the rainbow trout genome units we used were BAC fingerprinting contigs rather than individual BACs. This size limit increase most likely contributed for the improvement in BlastX identification of

Table 4 Paired BAC ends from the RT library that were previously included in the synteny analysis and were also included in the current fingerprinting contigs-based analysis

	Zebrafish	Medaka	Stickleback
BlastN			
RT library paired BAC ends matched on the same chromosome	357	414	609
Of which paired BAC ends in FP contigs	283	322	478
Of which were in contigs mapped to more than 3 chromosomes	46	36	61
Number of paired BAC ends included in the current genome synteny analysis	237 (66%)	286 (69%)	417 (67%)
BlastX			
RT library paired BAC ends matched on the same chromosome	148	122	141
Of which paired BAC ends in FP contigs	116	97	110
Of which were in contigs mapped to more than 3 chromosomes	14	11	5
Number of paired BAC ends included in the current genome synteny analysis	102 (69%)	86 (70%)	105 (74%)

microsynteny because the points of matching on the chromosomes of the reference genomes using BlastX are determined by the ORF boundaries as opposed to the exact points of nucleotide sequence matches when using BlastN.

Our results demonstrated that use of both the BlastN and BlastX search algorithms for identifying regions of synteny significantly improved the efficiency of the synteny analysis. Our BlastN analyses identified 755 (zebrafish), 832 (medaka), and 889 (stickleback) Contig-Chr pairs with unique BAC end hits in macro-synteny that were not detected using BlastX analyses. This can be explained by several factors including incomplete annotations of the model fish genomes and the presence of pseudogenes and conserved noncoding sequences that were not included in the peptide databases. Conversely, our BlastX analyses identified 447 (zebrafish), 419 (medaka), and 309 (stickleback) Contig-Chr pairs with unique BAC end hits in macrosynteny that were not detected using BlastN analyses. This may be at least partially caused by non- or less-conserved peptides whose coding sequences are not under strong selection pressure and have evolved enough to escape detection as significant unique hits by BlastN. In addition, the lower rate of peptide sequence evolution compared to their underlying nucleotide sequences may result in the detection of conserved duplicated peptide sequences from the same gene family in the target genome as two “unique” hits where actually only one is truly an orthologous sequence.

The overall number of unique regions of synteny (Contig-Chr Pairs) we identified with the rainbow trout fingerprinting contigs by both BlastN and BlastX was 2,259, 2,229, and 2,203 for stickleback, medaka, and zebrafish, respectively. These numbers are approximately three to five times greater than the numbers of syntenies we identified using BAC paired ends (Genet et al. 2011) demonstrating the improved power of the comparative genomics analysis when it is based on the BAC FPC map. However, approximately 20% of the BAC paired ends with significant Blast hits were not included in the contigs-based analysis (Table 4). This fraction of missing BACs is expected as indeed approximately 20% of the BAC fingerprints failed our quality editing criteria and were filtered out of the FPC assembly. In addition, 4–13% of the BAC paired ends that were used in our previous synteny analysis (Genet et al. 2011) were filtered out of the current analysis because they were part of the contigs that matched more than three chromosomes on the reference genomes (Table 4). Our results show a likelihood of approximately 50% for a contig with at least two unmasked BES to be mapped by BlastN or BlastX to a chromosome in the genome of three model species. Microsynteny was found in approximately 40% (zebrafish) to 50% (stickleback) of the contigs that were mapped to chromosomes of the model species. Therefore, the likelihoods of identifying conserved microsynteny

between a rainbow trout physical map contig with at least two unmasked BAC end sequences and the genomes of zebrafish and stickleback are 20% and 25%, respectively.

The results of our comparative analyses clearly demonstrate that for species currently lacking a close relative with a reference genome sequence the use of the fingerprinting physical map for analysis of conserved synteny in addition to the BAC paired ends is highly desirable. However, once a close reference genome sequence is available and an *in silico* physical map can be constructed based on the BAC end sequences, the use of BAC paired ends for comparative genomics is preferred because they also provide information on the sequence orientation matching between the compared genomes (Kuhl et al. 2010, 2011; Dalrymple et al. 2007; Larkin et al. 2003).

Conclusions

The second generation integrated map of the rainbow trout genome that we constructed and described in this paper provides a framework for a robust composite genome map and a minimal tiling path for a draft genome sequence assembly. The availability of an integrated physical and genetic map enables detailed comparative genome analyses with other salmonids, fine mapping of QTL, positional cloning, selection of positional candidate genes for economically important traits, and the incorporation of marker assisted selection (MAS) into rainbow trout breeding programs. The comparative genome analyses reported here provide a survey of conserved synteny between rainbow trout and three model fish species. Sequence homology between the rainbow trout BAC end sequences and the three model fish species genomes was relatively low, but clustering of the conserved synteny analysis results by linkage groups as derived from the integrated physical and genetic map revealed that large blocks of macrosynteny are conserved between chromosome arms of rainbow trout and the model fish species.

Acknowledgments This project was supported by National Research Initiative Competitive Grant number 2007-35616-17875 from the USDA National Institute of Food and Agriculture and by internal base funds provided by the Agricultural Research Service project number 1930-31000-009. Brian Smith, Kristy Shewbridge, and Roseanna Long carried out the microsatellites genotyping described in this manuscript. We thank Gary Thorgaard and Paul Wheeler (Washington State University) for the Swanson doubled haploid samples that were used to construct the new BAC libraries and Eric Peatman (Auburn University) for reviewing an earlier draft of the manuscript. We also thank anonymous reviewer 2 for his insightful comments on the blocks of conserved macro-synteny that can be identified through the chromosomes of the integrated map. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture. USDA is an equal opportunity provider and employer.

References

- Allendorf FW, Thorgaard GH (1984) Tetraploidy and the evolution of salmonid fishes. In: Turner BJ (ed) Evolutionary genetics of fishes. Plenum, New York
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Baerwald MR, Petersen JL, Hedrick RP, Schisler GJ, May B (2010) A major effect quantitative trait locus for whirling disease resistance identified in rainbow trout (*Oncorhynchus mykiss*). *Heredity* 106:920–926
- Barroso RM, Wheeler PA, Lapatra SE, Drew RE, Thorgaard GH (2008) QTL for IHNV resistance and growth identified in a rainbow (*Oncorhynchus mykiss*)xYellowstone cutthroat (*Oncorhynchus*). *Aquaculture* 277:156–163
- Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27:573–580
- Boutin-Ganache I, Raposo M, Raymond M, Deschepper CF (2001) M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods. *Biotechniques* 31(24–6):28
- Cole J, Wiggans G, Ma L, Sonstegard T, Lawlor T, Crooker B, Van Tassell C, Yang J, Wang S, Matukumalli L, Da Y (2011) Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. *BMC Genomics* 12:408
- Dalrymple B, Kirkness E, Nefedov M, McWilliam S, Ratnakumar A, Barris W, Zhao S, Shetty J, Maddox J, O'grady M, Nicholas F, Crawford A, Smith T, De Jong P, Mcewan J, Oddy VH, Cockett N, The International Sheep Genomics C (2007) Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. *Genome Biol* 8:R152
- Danzmann RG, Cairnes M, Davidson WS, Ferguson MM, Gharbi K, Guyomard R, Holm LE, Leder E, Okamoto N, Ozaki A, Rexroad CE, Sakamoto T, Taggart JB, Woram RA (2006) A comparative analysis of the rainbow trout genome with 2 other species of fish (Arctic char and Atlantic salmon) within the tetraploid derivative Salmonidae family (subfamily: Salmoninae). *Genome* 48:1037–1051
- Danzmann R, Davidson E, Ferguson M, Gharbi K, Koop B, Hoyheim B, Lien S, Lubieniecki K, Moghadam H, Park J, Phillips R, Davidson W (2008) Distribution of ancestral proto-Actinopterygian chromosome arms within the genomes of 4R-derivative salmonid fishes (Rainbow trout and Atlantic salmon). *BMC Genomics* 9:557
- Davidson W, Koop B, Jones S, Iturra P, Vidal R, Maass A, Jonassen I, Lien S, Omholt S (2011) Sequencing the genome of the Atlantic salmon (*Salmo salar*). *Genome Biol* 11:403
- Drew RE, Schwabl H, Wheeler PA, Thorgaard GH (2007) Detection of QTL influencing cortisol levels in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 272:S183–S194
- Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 8:186–194
- Falgueras J, Lara A, Fernandez-Pozo N, Canton F, Perez-Trabado G, Claros MG (2010) SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read. *BMC Bioinforma* 11:38
- Genet C, Dehais P, Palti Y, Gao G, Gavory F, Wincker P, Quillet E, Boussaha M (2011) Analysis of BAC-end sequences in rainbow trout: content characterization and assessment of synteny between trout and other fish genomes. *BMC Genomics* 12:314
- Govoroun M, Le Gac F, Guiguen Y (2006) Generation of a large scale repertoire of expressed sequence tags (ESTs) from normalised rainbow trout cDNA libraries. *BMC Genomics* 7:196
- Guyomard R, Mauger S, Tabet-Canale K, Martineau S, Genet C, Krieg F, Quillet E (2006) A type I and type II microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) with presumptive coverage of all chromosome arms. *BMC Genomics* 7:302
- Haidle L, Janssen J, Gharbi K, Moghadam H, Ferguson M, Danzmann R (2008) Determination of quantitative trait loci (QTL) for early maturation in rainbow trout (*Oncorhynchus mykiss*). *Mar Biotechnol* 10:579–592
- Katagiri T, Asakawa S, Minagawa S, Shimizu N, Hirono I, Aoki T (2001) Construction and characterization of BAC libraries for three fish species; rainbow trout, carp and tilapia. *Anim Genet* 32:200–204
- Kuhl H, Beck A, Wozniak G, Canario A, Volckaert F, Reinhardt R (2010) The European sea bass *Dicentrarchus labrax* genome puzzle: comparative BAC-mapping and low coverage shotgun sequencing. *BMC Genomics* 11:68
- Kuhl H, Sarropoulou E, Tine M, Kotoulas G, Magoulas A, Reinhardt R (2011) A comparative BAC map for the gilthead sea bream (*Sparus aurata* L.). *J Biomed Biotechnol* 2011:1–7
- Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in humans. *Proc Natl Acad Sci U S A* 84:2363–2367
- Larkin DM, Wind A, Rebeiz M, Schweitzer PA, Bachman S, Green C, Wright CL, Campos EJ, Benson LD, Edwards J, Liu L, Osoegawa K, Womack JE, De Jong PJ, Lewin HA (2003) A cattle-human comparative map built with cattle BAC-ends and human genome sequence. *Genome Res* 13:1966–1972
- Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci U S A* 81:3443–3446
- Le Bras Y, Dechamp N, Krieg F, Filangi O, Guyomard R, Boussaha M, Bovenhuis H, Pottinger T, Prunet P, Le Roy P, Quillet E (2011) Detection of QTL with effects on osmoregulation capacities in the rainbow trout (*Oncorhynchus mykiss*). *BMC Genet* 12:46
- Liu H, Jiang Y, Wang S, Ninwichian P, Somridhivej B, Xu P, Abernathy J, Kucuktas H, Liu Z (2009) Comparative analysis of catfish BAC end sequences with the zebrafish genome. *BMC Genomics* 10:592
- Luo M-C, Thomas C, You FM, Hsiao J, Ouyang S, Buell CR, Malandro M, Mcguire PE, Anderson OD, Dvorak J (2003) High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. *Genomics* 82:378–389
- Matise TC, Perlin M, Chakravarti A (1994) Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map. *Nat Genet* 6:384–390
- Miller MR, Brunelli JP, Wheeler PA, Liu S, Rexroad Iii CE, Palti Y, Doe CQ & Thorgaard GH (2011) A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Molecular Ecology*. doi:10.1111/j.1365-294X.2011.05305.x
- Nelson W, Soderlund C (2009) Integrating sequence with FPC fingerprint maps. *Nucl Acids Res* 37:e36
- Nelson WM, Bharti AK, Butler E, Wei F, Fuks G, Kim H, Wing RA, Messing J, Soderlund C (2005) Whole-genome validation of high-information-content fingerprinting. *Plant Physiol* 139:27–38
- Ng SH, Artieri CG, Bosdet IE, Chiu R, Danzmann RG, Davidson WS, Ferguson MM, Fjell CD, Hoyheim B, Jones SJ, De Jong PJ, Koop BF, Krzywinski MI, Lubieniecki K, Marra MA, Mitchell LA, Mathewson C, Osoegawa K, Parisotto SE, Phillips RB, Rise ML, Von Schalburg KR, Schein JE, Shin H, Siddiqui A, Thorsen J, Wye N, Yang G, Zhu B (2005) A physical map of the genome of Atlantic salmon, *Salmo salar*. *Genomics* 86:396–404

- Nichols KM, Bartholomew J, Thorgaard GH (2003a) Mapping multiple genetic loci associated with *Ceratomyxa shasta* resistance in *Oncorhynchus mykiss*. *Dis Aquat Organ* 56:145–154
- Nichols KM, Young WP, Danzmann RG, Robison BD, Rexroad C, Noakes M, Phillips RB, Bentzen P, Spies I, Knudsen K (2003b) A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*). *Anim Genet* 34:102–115
- Nichols KM, Broman KW, Sundin K, Young JM, Wheeler PA, Thorgaard GH (2007) Quantitative trait loci×maternal cytoplasmic environment interaction for development rate in *Oncorhynchus mykiss*. *Genetics* 175:335–347
- Nichols KM, Edo AF, Wheeler PA, Thorgaard GH (2008) The genetic basis of smoltification-related traits in *Oncorhynchus mykiss*. *Genetics* 179:1559–1575
- Nievergelt CM, Smith DW, Kohlenberg JB, Schork NJ (2004) Large-scale integration of human genetic and physical maps. *Genome Res* 14:1199–1205
- O'connell JR, Weeks DE (1995) The VITESSE algorithm for rapid exact multilocus linkage analysis via genotype set-recoding and fuzzy inheritance. *Nat Genet* 11:402–408
- O'connell JR, Weeks DE (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 63:259–266
- O'malley KG, Sakamoto T, Danzmann RG, Ferguson MM (2003) Quantitative trait loci for spawning date and body weight in rainbow trout: testing for conserved effects across ancestrally duplicated chromosomes. *J Hered* 94:273–284
- Ozaki A, Sakamoto T, Khoo S, Nakamura K, Coimbra MR, Akutsu T, Okamoto N (2001) Quantitative trait loci (QTLs) associated with resistance/susceptibility to infectious pancreatic necrosis virus (IPNV) in rainbow trout (*Oncorhynchus mykiss*). *Mol Genet Genomics* 265:23–31
- Palti Y, Gahr SA, Hansen JD, Rexroad CE (2004) Characterization of a new BAC library for rainbow trout: evidence for multi-locus duplication. *Anim Genet* 35:130–133
- Palti Y, Luo M-C, Hu Y, Genet C, You F, Vallejo R, Thorgaard G, Wheeler P, Rexroad C (2009) A first generation BAC-based physical map of the rainbow trout genome. *BMC Genomics* 10:462
- Palti Y, Genet C, Luo M-C, Charlet A, Gao G, Hu Y, Castano-Sanchez C, Tabet-Canale K, Krieg F, Yao J, Vallejo R, Rexroad C (2011) A first generation integrated map of the rainbow trout genome. *BMC Genomics* 12:180
- Perry GM, Danzmann RG, Ferguson MM, Gibson JP (2001) Quantitative trait loci for upper thermal tolerance in outbred strains of rainbow trout (*Oncorhynchus mykiss*). *Heredity* 86:333–341
- Perry GM, Ferguson MM, Sakamoto T, Danzmann RG (2005) Sex-linked quantitative trait loci for thermotolerance and length in the rainbow trout. *J Hered* 96:97–107
- Phillips RB, Nichols KM, Dekoning JJ, Morasch MR, Keatley KA, Rexroad C 3rd, Gahr SA, Danzmann RG, Drew RE, Thorgaard GH (2006) Assignment of rainbow trout linkage groups to specific chromosomes. *Genetics* 174:1661–1670
- Phillips R, Keatley K, Morasch M, Ventura A, Lubieniecki K, Koop B, Danzmann R, Davidson W (2009) Assignment of Atlantic salmon (*Salmo salar*) linkage groups to specific chromosomes: conservation of large syntenic blocks corresponding to whole chromosome arms in rainbow trout (*Oncorhynchus mykiss*). *BMC Genet* 10:46
- Quillet E, Dorson M, Le Guillou S, Benmansour A, Boudinot P (2007) Wide range of susceptibility to rhabdoviruses in homozygous clones of rainbow trout. *Fish Shellfish Immunol* 22:510–519
- Quiniou S, Waldbieser G, Duke M (2007) A first generation BAC-based physical map of the channel catfish genome. *BMC Genomics* 8:40
- Rexroad CE, Lee Y, Keele JW, Karamycheva S, Brown G, Koop B, Gahr SA, Palti Y, Quackenbush J (2003) Sequence analysis of a rainbow trout cDNA library and creation of a gene index. *Cytogenet Genome Res* 102:347–354
- Rexroad CE 3rd, Palti Y, Gahr SA, Vallejo RL (2008) A second generation genetic map for rainbow trout (*Oncorhynchus mykiss*). *BMC Genet* 9:74
- Rise ML, Von Schalburg KR, Brown GD, Mawer MA, Devlin RH, Kuipers N, Busby M, Beetz-Sargent M, Alberto R, Gibbs AR, Hunt P, Shukin R, Zeznik JA, Nelson C, Jones SRM, Smailus DE, Jones SJM, Schein JE, Marra MA, Butterfield YSN, Stott JM, Ng SHS, Davidson WS, Koop BF (2004) Development and application of a salmonid EST database and cDNA microarray: data mining and interspecific hybridization characteristics. *Genome Res* 14:478–490
- Robison BD, Wheeler PA, Thorgaard GH (1999) Variation in development rate among clonal lines of rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 173:131–141
- Robison BD, Wheeler PA, Sundin K, Sikka P, Thorgaard GH (2001) Composite interval mapping reveals a major locus influencing embryonic development rate in rainbow trout (*Oncorhynchus mykiss*). *J Hered* 92:16–22
- Rodriguez MF, Lapatra S, Williams S, Famula T, May B (2004) Genetic markers associated with resistance to infectious hematopoietic necrosis in rainbow and steelhead trout. *Aquaculture* 241:93–115
- Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 132:365–386
- Sakamoto T, Danzmann RG, Okamoto N, Ferguson MM, Ihssen PE (1999) Linkage analysis of quantitative trait loci associated with spawning time in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 173:33–43
- Sakamoto T, Danzmann RG, Gharbi K, Howard P, Ozaki A, Khoo SK, Woram RA, Okamoto N, Ferguson MM, Holm L-E, Guyomard R, Hoyheim B (2000) A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. *Genetics* 155:1331–1345
- Salem M, Kenney B, Rexroad CE, Yao J (2008) Development of a 37 k high-density oligonucleotide microarray: a new tool for functional genome research in rainbow trout. *J Fish Biol* 72:2187–2206
- Salem M, Xiao C, Womack J, Rexroad CE, Yao J (2009) MicroRNA repertoire for functional genome research in rainbow trout (*Oncorhynchus mykiss*). *Marine Biotech* 12:410–429
- Salem M, Rexroad C, Wang J, Thorgaard G, Yao J (2010) Characterization of the rainbow trout transcriptome using Sanger and 454-pyrosequencing approaches. *BMC Genomics* 11:564
- Soderlund C, Longden I, Mott R (1997) FPC: a system for building contigs from restriction fingerprinted clones. *Comput Appl Biosci* 13:523–535
- Soderlund C, Humphray S, Dunham A, French L (2000) Contigs built with fingerprints, markers, and FPC V4.7. *Genome Res* 10:1772–1787
- Sundin K, Brown KH, Drew RE, Nichols KM, Wheeler PA, Thorgaard GH (2005) Genetic analysis of a development rate QTL in backcrosses of clonal rainbow trout, *Oncorhynchus mykiss*. *Aquaculture* 247:75–83
- Thorgaard GH, Bailey GS, Williams D, Buhler DR, Kaattari SL, Ristow SS, Hansen JD, Winton JR, Bartholomew JL, Nagler JJ, Walsh PJ, Vijayan MM, Devlin RH, Hardy RW, Overturf KE, Young WP, Robison BD, Rexroad CE, Palti Y (2002) Status and opportunities for genomics research with rainbow trout. *Comp Biochem Physiol B Biochem Mol Biol* 133:609–646
- Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered* 93:77–78

- Wiggans GR, Sonstegard TS, Vanraden PM, Matukumalli LK, Schnabel RD, Taylor JF, Schenkel FS, Van Tassell CP (2009) Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. *J Dairy Sci* 92:3431–3436
- Xu P, Wang S, Liu L, Peatman E, Somridhivej B, Thimmapuram J, Gong G, Liu Z (2006) Channel catfish BAC-end sequences for marker development and assessment of syntenic conservation with other fish species. *Anim Genet* 37:321–326
- Xu P, Li J, Li Y, Cui R, Wang J, Wang J, Zhang Y, Zhao Z, Sun X (2011) Genomic insight into the common carp (*Cyprinus carpio*) genome by sequencing analysis of BAC-end sequences. *BMC Genomics* 12:188
- Young WP, Wheeler PA, Fields RD, Thorgaard GH (1996) DNA fingerprinting confirms isogenicity of androgenetically derived rainbow trout lines. *J Hered* 87:77–80
- Young WP, Wheeler PA, Coryell VH, Keim P, Thorgaard GH (1998) A detailed linkage map of rainbow trout produced using doubled haploids. *Genetics* 148:839–850
- Zimmerman A, Evenhuis J, Thorgaard G, Ristow S (2004) A single major chromosomal region controls natural killer cell-like activity in rainbow trout. *Immunogenetics* 55:825–835