

The catfish genome database cBARBEL: an informatic platform for genome biology of ictalurid catfish

Jianguo Lu¹, Eric Peatman¹, Qing Yang², Shaolin Wang¹, Zhiliang Hu³, James Reecy³, Huseyin Kucuktas¹ and Zhanjiang Liu^{1,*}

¹The Fish Molecular Genetics and Biotechnology Laboratory, Aquatic Genomics Unit, Department of Fisheries and Allied Aquacultures and Program of Cell and Molecular Biosciences, ²Department of Computer Science and Software Engineering, Auburn University, Auburn, AL 36849 and ³Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University, Ames, IA, USA

Received July 13, 2010; Accepted August 11, 2010

ABSTRACT

The catfish genome database, cBARBEL (abbreviated from catfish Breeder And Researcher Bioinformatics Entry Location) is an online open-access database for genome biology of ictalurid catfish (*Ictalurus spp.*). It serves as a comprehensive, integrative platform for all aspects of catfish genetics, genomics and related data resources. cBARBEL provides BLAST-based, fuzzy and specific search functions, visualization of catfish linkage, physical and integrated maps, a catfish EST contig viewer with SNP information overlay, and GBrowse-based organization of catfish genomic data based on sequence similarity with zebrafish chromosomes. Subsections of the database are tightly related, allowing a user with a sequence or search string of interest to navigate seamlessly from one area to another. As catfish genome sequencing proceeds and ongoing quantitative trait loci (QTL) projects bear fruit, cBARBEL will allow rapid data integration and dissemination within the catfish research community and to interested stakeholders. cBARBEL can be accessed at <http://catfishgenome.org>.

INTRODUCTION

Catfish (*Ictalurus spp.*) is an important aquaculture species in the United States, accounting for >60% of domestic aquaculture production. While channel catfish (*Ictalurus*

punctatus) accounts for the large majority of farm-raised catfish, increasing numbers of channel catfish female × blue catfish (*Ictalurus furcatus*) male hybrids are being cultured. Facing rising feed costs and stiff international competition, catfish producers require improvement in fish production and performance traits such as disease resistance, growth rate and feed conversion efficiency to maintain profitability. Efficient utilization of the natural diversity of trait phenotypes already present in different species, strains and hybrids of catfish for selection of superior broodstock, requires the identification of genetic underpinnings of trait differences. Toward this end and eventual marker-assisted selection (MAS), significant genome resources have been developed in catfish. These include over a half million expressed sequence tags (ESTs) (1–6), a large number of genome sequences generated from bacterial artificial chromosome (BAC) ends (7), genetic linkage maps (8–10), genome physical maps (11,12), tens of thousands of microsatellite markers (13,14), hundreds of thousands of single-nucleotide polymorphisms (SNPs) (15), over 10 000 full-length cDNAs (fcdNA; 16) and an alternative splicing database (17). Additionally, USDA NIFA funding has been secured to allow catfish whole-genome sequencing and the development of high-density SNP chips for catfish. With these and many more genome-oriented projects from catfish currently underway, the catfish research community needed a central repository for storing and integrating genomic data and a bioinformatic entry location for public access to currently inaccessible specialized data sets. To meet this need, we have created a catfish genome database, cBARBEL, the Catfish Breeder and

*To whom correspondence should be addressed. Email: zliu@acesag.auburn.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Researcher Bioinformatic Entry Location, a title that makes use of the distinctive whisker-like organs that give catfish their name.

cBARBEL represents one of the first comprehensive bioinformatic databases for an aquaculture species, although genome sequencing is planned or proceeding in close to a dozen different species from this fast-growing sector. cBARBEL provides wide-ranging query functions to facilitate user access to a host of catfish genome resources and integrates a variety of previously scattered data types. Here, we present an overview of cBARBEL search tools, platforms and functions connecting catfish EST, fl-cDNA, SNP, BAC-end sequence (BES), molecular marker, linkage map and physical map data. cBARBEL can be accessed at <http://catfishgenome.org/>.

RESULTS AND DISCUSSION

cBARBEL database schematic

The cBARBEL database schematic is shown in Figure 1, allowing visualization of potential data connections. The integration of existing and forthcoming catfish genome resources provided by the cBARBEL platform should speed research progress. Utilizing cBARBEL, for example, a user searching for a given gene in catfish is able to search with a similar sequence from a related species, such as zebrafish, identify matching catfish ESTs, identify the corresponding EST contig, identify SNP markers within that contig, visualize linkage map position of these markers and relate linkage and physical map locations through a comparative map system, all by navigating through a series of intuitive links. To do a similar search without cBARBEL would require interrogating a series of separate public and local databases, relating disparate nomenclatures, would not include visualization and could take well over 30 min for an individual query.

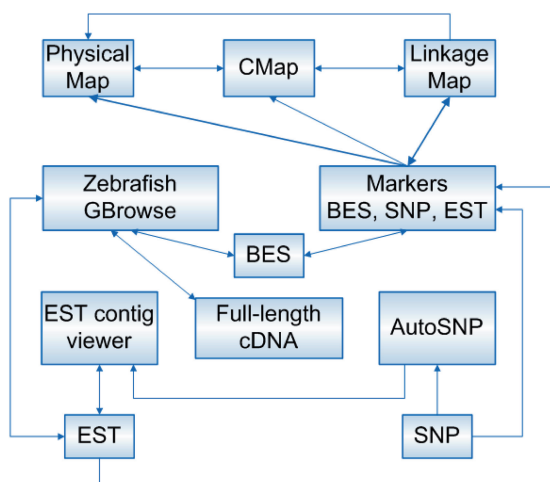


Figure 1. The cBARBEL database schematic showing database components and clickable, searchable connections between catfish genome resources. BES, BAC-end sequence; SNP, single-nucleotide polymorphism; EST, expressed sequence tag.

Implementation

Several software packages were used in the construction of cBARBEL database prerequisites including: (i) the operating system, Ubuntu 9.10 Linux system (<http://www.ubuntu.com/>); (ii) Apache web server version 2.0 (<http://www.apache.org/>); (iii) MySQL database management system, version 5.1 (<http://www.mysql.com/>); and (iv) Selective Perl modules and configuration of Bioperl and PHP (www.perl.org; www.bioperl.org; <http://php.net/>). The Generic Genome Browser (GBrowse) package (v1.7), a component of the Generic Model Organism Project (GMOD), was utilized for display of the catfish physical map, EST contig viewer and zebrafish comparative GBrowse. Another GMOD component, CMap (18), was used to display and align genetic and physical (FPC) maps.

General features of cBARBEL components and tools

cBARBEL currently is organized around three components: nucleotide sequence, genetic markers and maps. These components are brought together by search tools and multi-directional links. Features of each of these components are described briefly below.

Sequence similarity searches of catfish databases. BLAST searches can be carried out to query all or a subset of catfish sequences including catfish ESTs, catfish BES, full-length cDNA and catfish all (Figure 2). Search results provide further links to NCBI records and location of the hit on the catfish physical map (BES), EST contig viewer (EST) and zebrafish GBrowse chromosomes (All).

Specific search function of catfish databases. The specific search function can provide data access using a variety of search terms (Figure 2). Catfish ESTs can be queried using GenBank accession numbers, marker names for those ESTs containing a SNP or microsatellite marker, and EST contig ID. Search results provide sequence links as well as deeper connections to the EST contig viewer, zebrafish GBrowse comparative alignments and linkage map position, where appropriate. Similar searches can be carried out for BES and fl-cDNA. Those looking for a marker of interest can use fuzzy or specific search terms. For example, a search such as 'AUEST' returns all marker information for EST microsatellite markers generated at Auburn University. The resulting marker table contains the accession number of the relevant sequence, primer information, marker status (polymorphic, not polymorphic, untested), linkage map position, physical map position (BES) and EST-contig (EST).

Zebrafish GBrowse genomic viewer versus catfish genomic data sets. GBrowse is a GMOD tool that displays features of the genome aligned to a genomic sequence (19). GBrowse is easily customized to allow a variety of data tracks and third-party data types to be visualized. Catfish whole-genome sequencing is underway, but, in the interim, we have aligned a number of catfish genome data types to the genome sequence of zebrafish, the

The screenshot displays the cBARBEL web interface. At the top, the logo 'cBARBEL' is shown in blue, followed by the tagline 'Catfish Breeder And Researcher Bioinformatics Entry Location'. Below this is a navigation bar with tabs for 'Sequence Search', 'Maps', 'Expression', 'Tools', and 'Project Team'. Under 'Sequence Search', there are sub-links for 'Search', 'BES', 'EST', 'Full-length cDNA', 'Marker', 'Gbrowse', 'EST Contig Viewer', 'Alternative Splicing', and 'Genome Sequence'. The main content area is divided into two panels. The left panel, titled 'BLAST SEARCH', features a dropdown menu for 'Program' set to 'blastn' and a dropdown for 'Database' set to 'Catfish BES'. A list of database options is shown: 'Catfish BES', 'Catfish EST', 'Catfish FLcDNA', and 'All'. A text input field contains the sequence 'genomic 3', genomic survey sequence' followed by a DNA sequence: 'GTACTGIGAAATATTAGTGGAGTTCACCATTTTATTTTCAACACCACATTT' and 'AATAAATTATTTGTACATGA'. Below the sequence is a 'Search Sequence' button. The right panel, titled 'SPECIFIC SEARCH', has four search options: 'Search for EST Example' (with 'Acc. Number' dropdown set to 'FD150991'), 'Search for BES Example' (with 'Acc. Number' dropdown), 'Search for Full-length cDNA Example' (with 'Acc. Number' dropdown set to 'GU588469'), and 'Search for Marker Example' (with 'Marker Name' dropdown set to 'AUBES365'). Each search option has a 'Search' button. Green callout boxes with red text point to various elements: 'blastn; blastp; blastx' and 'tblastn; tblastx' point to the program dropdown; 'Genbank Acc. Number', 'Marker name', 'Clone ID', and 'Contig ID' point to the database dropdown; 'Genbank Acc. Number', 'Marker name', and 'Contig ID' point to the EST search dropdown; 'Genbank Acc. Number' and 'Full-length cDNA' point to the cDNA search dropdown; and 'Marker Name', 'Contig ID', 'Status', and 'View all' point to the marker search dropdown.

Figure 2. cBARBEL database sequence search functions. Searches can be conducted by BLAST against local databases of catfish ESTs, catfish BES, full-length cDNA or all. Searches can also be conducted using fuzzy or specific searches for ESTs, BES, flcDNA and genetic markers and associated subcategories. Examples for all query types are provided.

closest evolutionarily related species with an available sequenced genome (Figure 3). The alignments (based on tblastx similarity) help to organize catfish data on a genome scale, and harnessing synteny between the two species has proven useful in gene isolation and QTL fine-mapping studies. As catfish genome assembly proceeds, this comparative approach should also prove useful in scaffolding catfish supercontigs. By default, cBARBEL presents a view of catfish EST contigs, catfish singleton ESTs, catfish BES, and catfish fl-cDNA aligned to zebrafish chromosomal sequences (Figure 3). For each of these tracks, clicking on a feature provides a related link to the NCBI server or a link to a local copy of the sequence information. Users can locate a specific region of interest by entering a specific sequence range or by dragging the selection box to specify a chromosomal region. Additionally, cBARBEL BLAST and specific search outputs include zebrafish GBrowse links to allow connectivity to other database sectors.

Catfish EST contig viewer. The GBrowse package was also utilized to create a catfish EST contig viewer displaying the alignment of individual ESTs on the contig consensus sequence. A SNP track was also added allowing visualization of SNP density and position within the EST contig (Figure 4). For the EST track, clicking on an individual feature provides a related link to NCBI-based sequence information. Clicking on SNP entries allows the user to navigate to further SNP information in the AutoSNP program including SNP allele frequency, type, and position. For the contig track, a link is provided to local sequence information. As with the

zebrafish GBrowse view, cBARBEL BLAST and specific search outputs include links to the EST contig viewer.

Catfish physical map based on BAC contig. We previously reported the construction of a fingerprint contig (FPC) BAC-based physical map with 3307 contigs spanning the catfish genome (12), and are currently engaged in efforts to integrate the physical map with catfish linkage maps using BES-associated microsatellites. This map was previously displayed using a Java-based program, WebFPC, which did not allow efficient search or data integration options. For example, BES associated with BAC clones could not be searched or visualized with WebFPC. To remedy these issues, we adapted GBrowse to display both BAC contigs and BES information in a searchable, integrated format (Figure 5). BAC clones are displayed using their FPC position within a given contig. Custom scripts were developed to indicate the presence of BES within a clone using blue boxes and to allow BES sequence retrieval via NCBI link by clicking on the desired clone. As with other sections, cBARBEL BLAST and specific search outputs include physical map links where appropriate.

Catfish linkage map. A linkage map of catfish has been constructed based on genotyping of EST-based microsatellites, SNP markers and BES-based microsatellite markers on backcross hybrid (channel \times blue) families (10). Efforts are ongoing to increase marker density using BES-associated microsatellites and SNP markers. Marker information for the 29 linkage groups is displayed in table format with marker name and map position (cM).

Zebrafish Gbrowse

Showing 890 kbp from chr14, positions 29,390,000 to 30,280,000

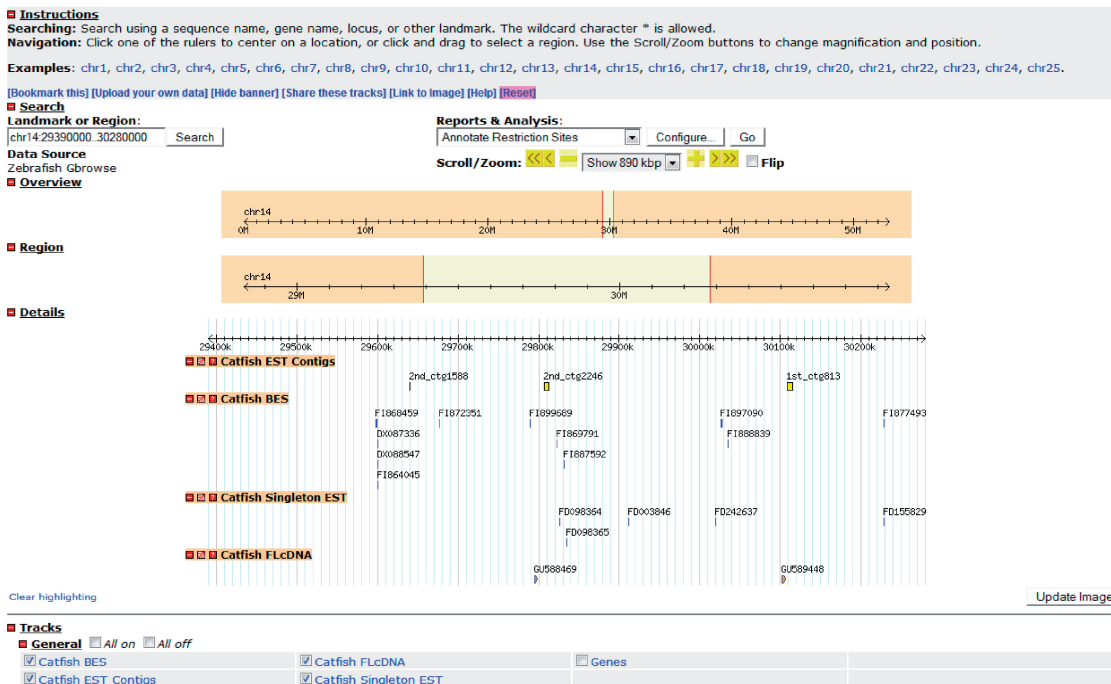


Figure 3. Mapping and organization of catfish genomic data based on homology with zebrafish chromosomal regions. In the absence of genome sequence for catfish, catfish EST contigs, BES, fclDNA were mapped based on tblastx results onto zebrafish chromosomes in the GBrowse package. Zoom features allow for examination of gene level to full chromosome level alignments. Mapped elements within each track are clickable to NCBI or local databases providing specific sequence information.

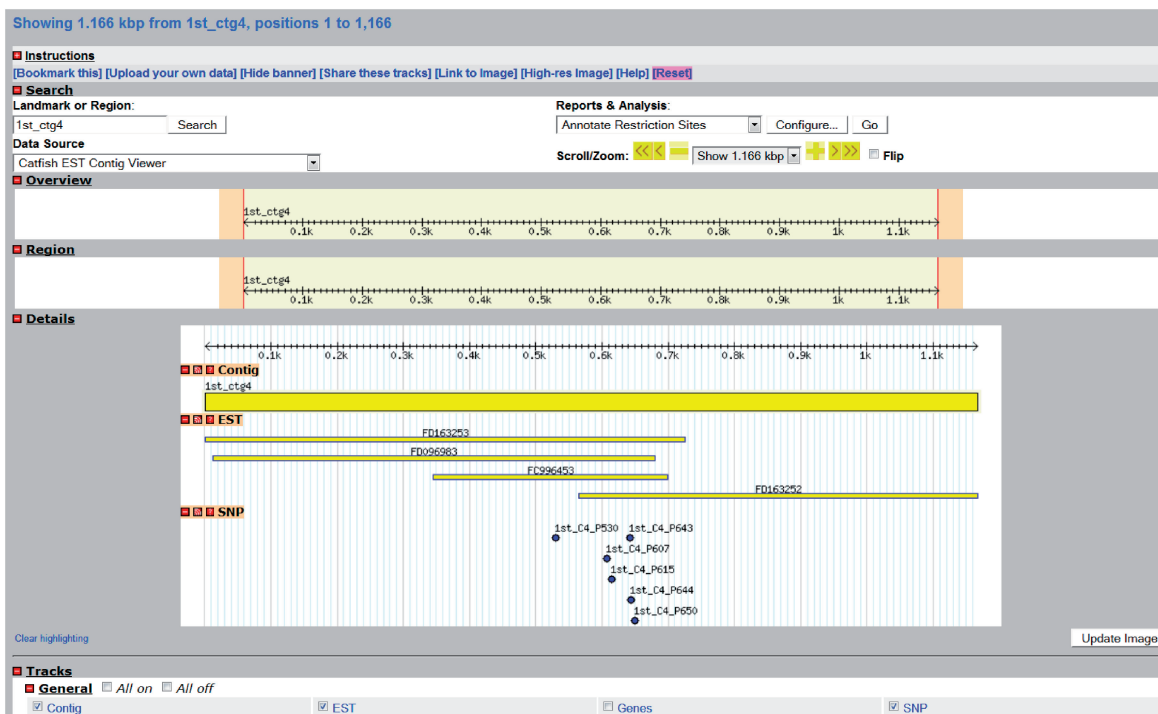


Figure 4. Catfish EST contig viewer within the GBrowse package. Tracks include Contig, EST and SNP and allow easy visualization of contig component sequences and physical position of SNPs along the consensus sequence. Each track is clickable to sequence information via NCBI (EST) or local database (contig) or AutoSNP (SNP) for detailed SNP allele frequency, type and position.

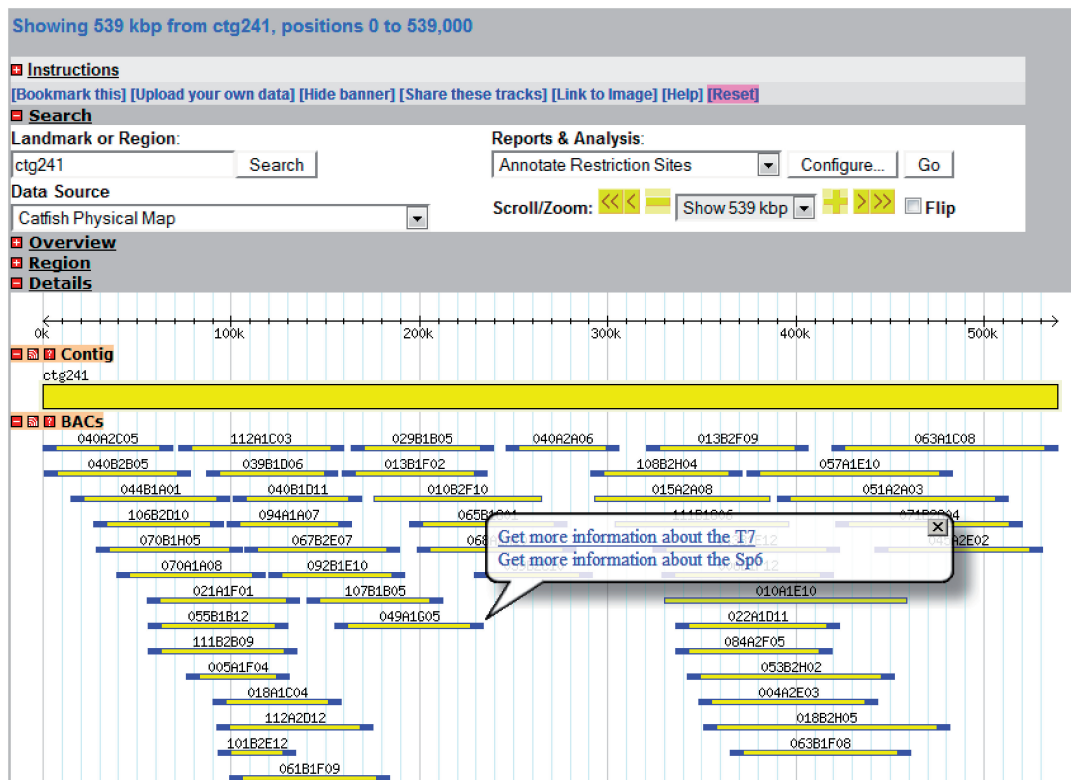


Figure 5. Catfish BAC-based physical map. Catfish BAC contigs are viewable and searchable in the GBrowse format. Associated BESs are indicated by blue clone ends and NCBI-based sequences can be retrieved by clicking a clone of interest.

Clicking on marker name allows navigation to a table providing more detailed information including primer sequences, marker status and relevant EST contig (EST) or physical map (BES) locations. As additional markers are genotyped and new maps constructed, this information is rapidly updated.

Catfish CMap-map integration. Integration of catfish linkage and physical maps is ongoing, based largely on the mapping of BES microsatellites from physical map contigs onto the catfish linkage map. While details of integration can be gathered in table format in other cBARBEL sections, we used CMap to provide visualization of LG-level map integration (Figure 6). Links are provided to each of the 29 linkage groups arrayed alongside corresponding physical map contigs using predetermined settings. A link is also provided to allow users to access numerous different display settings including the choice to view graphic representations of linkage groups alone. As with the linkage maps, data is continually updated as additional markers are genotyped.

Other tools and features. Under Tools, cBARBEL provides links to simple daily-use informatics tools that database users may need while accessing cBARBEL, including the external NCBI BLAST server, SMART domain search, ClustalW-based multiple sequence alignments and basic nucleotide to amino acid translation. These tools are provided in-frame so that the user does not need to leave the cBARBEL site for small-scale sequence analysis. Publications relevant to catfish

genomics and genetics are also linked. Additionally, the recently created Teleost Alternative Splicing Database with alternative splicing data for catfish ESTs (17) is linked to cBARBEL.

cBARBEL is intended to serve as a central database for catfish researchers. In addition to the data components and tools described above, the database frontpage is updated with relevant links, news and meeting information of importance to the catfish research community.

Data availability

All of the data in cBARBEL is freely available. Users can contact the Auburn cBARBEL project team or email the corresponding author to request a specific subset of the data or to make suggestions about future database content and features.

FUTURE PLANS

cBARBEL is continuously updated to include newly generated data. These new data types are incorporated and linked to the existing data when appropriate. New data and features we anticipate adding to cBARBEL in the near-term include:

- Catfish next-generation sequence data (454 and Illumina) utilizing GBrowse v2.0 from genome and transcriptome sequencing.

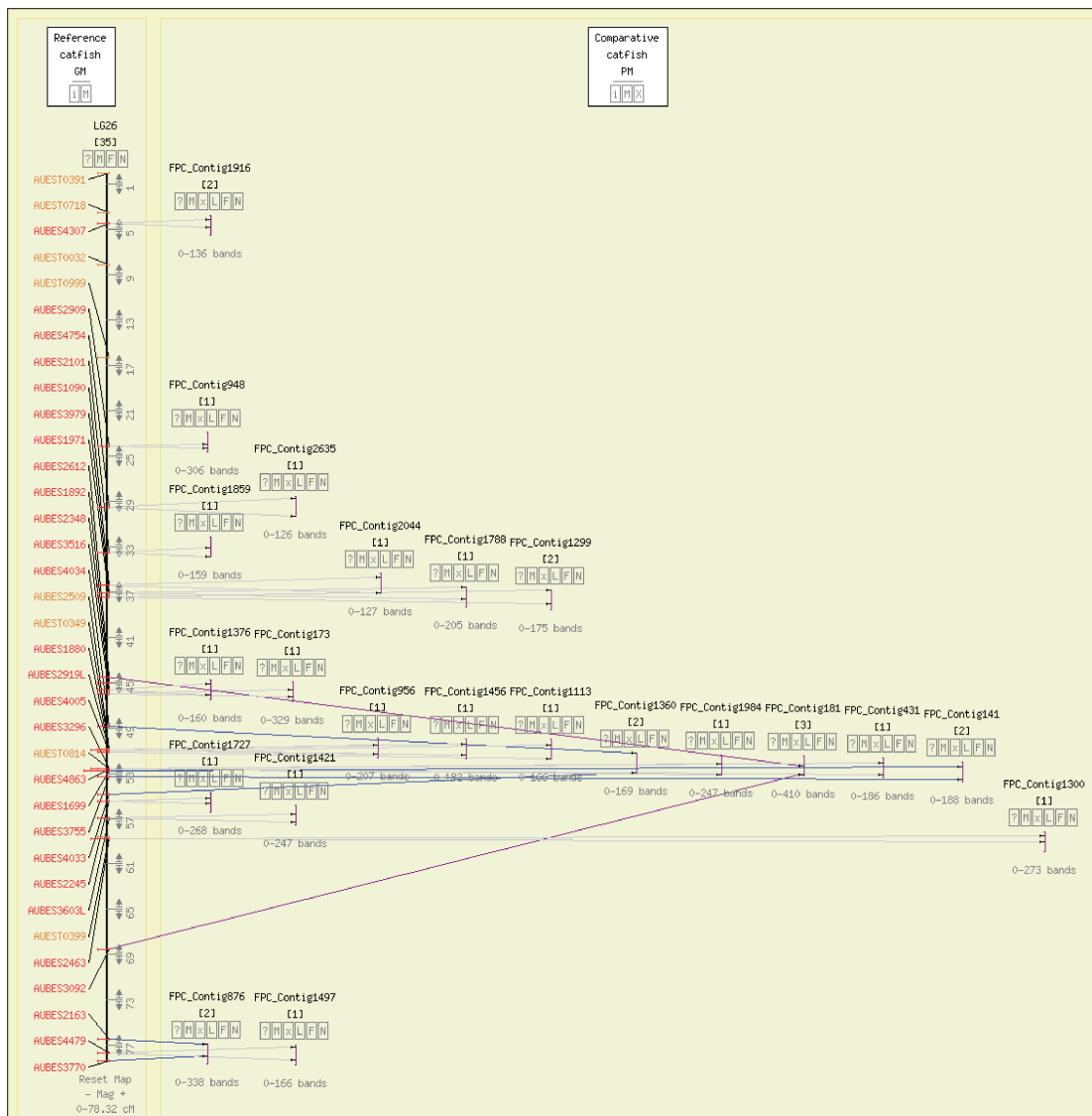


Figure 6. CMap-based visualization of early integration of catfish linkage and physical maps through mapping of BES microsatellites. Catfish linkage group 26 (LG26) is provided as an example on the left, and BES-associated contigs are shown on the right. CMap allows numerous options for viewing and comparing the catfish linkage and physical maps.

- Additional annotation and batch sequence retrieval tools (BLAST, gene ontology annotation and primer design).
- Expression data and links—microarray platform information and links to NCBI GEO-archived catfish data.

ACKNOWLEDGEMENTS

We wish to thank members of the Catfish Genome Consortium for their efforts in generating catfish genome resources, especially members of the USDA-ARS Catfish Genetics Research Unit, Drs. Geoff Waldbieser and Sylvie Quiniou. We also thank NRSP-8's Animal Genome Research Bioinformatics program (<http://www.animalgenome.org/>) for site hosting and assistance.

FUNDING

The USDA's National Institute of Food and Agriculture; a scholarship from the Chinese Scholarship Council (to J.L.); and a scholarship from Alabama EPSCoR Graduate Research Scholar's Program (to J.L.). Funding for open access charge: Discretionary Lab funds.

Conflict of interest statement. None declared.

REFERENCES

1. Cao,D., Kocabas,A., Ju,Z., Karsi,A., Li,P., Patterson,A. and Liu,Z. (2001) Transcriptome of channel catfish (*Ictalurus punctatus*): initial analysis of genes and expression profiles of the head kidney. *Anim. Genet.*, **32**, 169–188.
2. Ju,Z.L., Karsi,A., Kocabas,A., Patterson,A., Li,P., Cao,D.F., Dunham,R. and Liu,Z.J. (2000) Transcriptome analysis of

- channel catfish (*Ictalurus punctatus*): genes and expression profile from the brain. *Gene*, **261**, 373–382.
3. Karsi, A., Cao, D., Li, P., Patterson, A., Kocabas, A., Feng, J., Ju, Z., Mickett, K.D. and Liu, Z. (2002) Transcriptome analysis of channel catfish (*Ictalurus punctatus*): initial analysis of gene expression and micro satellite-containing cDNAs in the skin. *Gene*, **285**, 157–168.
 4. Kocabas, A.M., Li, P., Cao, D.F., Karsi, A., He, C.B., Patterson, A., Ju, Z.L., Dunham, R.A. and Liu, Z.J. (2002) Expression profile of the channel catfish spleen: Analysis of genes involved in immune functions. *Mar. Biotechnol.*, **4**, 526–536.
 5. Wang, S.L., Peatman, E., Abernathy, J., Waldbieser, G., Lindquist, E., Richardson, P., Lucas, S., Wang, M., Li, P., Thimmapuram, J. *et al.* (2010) Assembly of 500,000 inter-specific catfish expressed sequence tags and large scale gene-associated marker development for whole genome association studies. *Genome Biol.*, **11**, R8.
 6. Li, P., Peatman, E., Wang, S.L., Feng, J.N., He, C.B., Baoprasertkul, P., Xu, P., Kucuktas, H., Nandi, S., Somridhivej, B. *et al.* (2007) Towards the ictalurid catfish transcriptome: generation and analysis of 31,215 catfish ESTs. *BMC Genomics*, **8**, 177.
 7. Liu, H., Jiang, Y.L., Wang, S.L., Ninwichian, P., Somridhivej, B., Xu, P., Abernathy, J., Kucuktas, H. and Liu, Z.J. (2009) Comparative analysis of catfish BAC end sequences with the zebrafish genome. *BMC Genomics*, **10**, 592.
 8. Liu, Z.J., Karsi, A., Li, P., Cao, D.F. and Dunham, R. (2003) An AFLP-based genetic linkage map of channel catfish (*Ictalurus punctatus*) constructed by using an interspecific hybrid resource family. *Genetics*, **165**, 687–694.
 9. Waldbieser, G.C., Bosworth, B.G., Nonneman, D.J. and Wolters, W.R. (2001) A microsatellite-based genetic linkage map for channel catfish, *Ictalurus punctatus*. *Genetics*, **158**, 727–734.
 10. Kucuktas, H., Wang, S.L., Li, P., He, C.B., Xu, P., Sha, Z.X., Liu, H., Jiang, Y.L., Baoprasertkul, P., Somridhivej, B. *et al.* (2009) Construction of genetic linkage maps and comparative genome analysis of catfish using gene-associated markers. *Genetics*, **181**, 1649–1660.
 11. Quiniou, S.M.A., Waldbieser, G.C. and Duke, M.V. (2007) A first generation BAC-based physical map of the channel catfish genome. *BMC Genomics*, **8**, 40.
 12. Xu, P., Wang, S.L., Liu, L., Thorsen, J., Kucuktas, H. and Liu, Z.J. (2007) A BAC-based physical map of the channel catfish genome. *Genomics*, **90**, 380–388.
 13. Serapion, J., Kucuktas, H., Feng, J.A. and Liu, Z.J. (2004) Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*). *Mar. Biotechnol.*, **6**, 364–377.
 14. Somridhivej, B., Wang, S.L., Sha, Z.X., Liu, H., Quilang, J., Xu, P., Li, P., Hue, Z.L. and Liu, Z.J. (2008) Characterization, polymorphism assessment, and database construction for microsatellites from BAC end sequences of channel catfish (*Ictalurus punctatus*): a resource for integration of linkage and physical maps. *Aquaculture*, **275**, 76–80.
 15. Wang, S.L., Sha, Z.X., Sonstegard, T.S., Liu, H., Xu, P., Somridhivej, B., Peatman, E., Kucuktas, H. and Liu, Z.J. (2008) Quality assessment parameters for EST-derived SNPs from catfish. *BMC Genomics*, **9**, 450.
 16. Chen, F., Lee, Y., Jiang, Y.L., Wang, S.L., Peatman, E., Abernathy, J., Liu, H., Liu, S.K., Kucuktas, H., Ke, C.H. *et al.* (2010) Identification and characterization of full-length cDNAs in channel catfish (*Ictalurus punctatus*) and blue catfish (*Ictalurus furcatus*). *PLoS ONE*, **12**, e11546.
 17. Lu, J.G., Peatman, E., Wang, W.Q., Yang, Q., Abernathy, J., Wang, S.L., Kucuktas, H. and Liu, Z.J. (2010) Alternative splicing in teleost fish genomes: same-species and cross-species analysis and comparisons. *Mol. Genet. Genomics*, **283**, 531–539.
 18. Youens-Clark, K., Faga, B., Yap, I.V., Stein, L. and Ware, D. (2009) CMap 1.01: a comparative mapping application for the Internet. *Bioinformatics*, **25**, 3040–3042.
 19. Stein, L.D., Mungall, C., Shu, S.Q., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J.E., Harris, T.W., Arva, A. *et al.* (2002) The Generic Genome Browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.